Lie theory and the wave equation in space-time. I. The Lorentz group

E. G. Kalnins

Mathematics Department, University of Waikato, Hamilton, New Zealand

W. Miller, Jr.

School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455
(Received 30 January 1976)

In this article we begin a study of the relationship between separation of variables and the conformal symmetry group of the wave equation $\psi_{tt} - \Delta_3\psi = 0$ in space-time. In this first article we make a detailed study of separation of variables for the Laplace operator on the one and two sheeted hyperboloids in Minkowski space. We then restrict ourselves to homogeneous solutions of the wave equation and the Lorentz subgroup SO(3,1) of the conformal group SO(4,2). We study the various separable bases by using the methods of integral geometry as developed by Gel'fand and Graev. In most cases we give the spectral analysis for these bases, and a number of new bases are developed in detail. Many of the special function identities derived appear to be new. This preliminary study is of importance when we subsequently study models of the Hilbert space structure for solutions of the wave equation and the Klein–Gordon equation $\psi_{tt} - \Delta_3\psi = \lambda\psi$.

INTRODUCTION

In this article we continue[1–3] the investigation of the connection between separation of variables for the principal equations of mathematical physics and the associated symmetry groups of such equations. The object of our study in this current series of articles is the wave equation in space–time,

$$\psi_{tt} - \Delta_3\psi = 0. \qquad (*)$$

The motivation for such a study stems from the inherent physical importance of this equation as well as its intrinsic mathematical interest. In this article we initiate the study with a detailed investigation of separation of variables for the Laplace operator on the one and two sheeted hyperboloids $[X, X] = -1$ and $[X, X] = 1$, respectively. Here $X = (t, x, y, z) = (x_0, x_1, x_2, x_3)$ and $[X, X] = t^2 - x^2 - y^2 - z^2$ is the usual Lorentz space–time scalar product. In doing this we are also concerned with the corresponding problem on the cone $[X, X] = 0$. Here we are dealing only with the Lorentz subgroup of the SO(4, 2) symmetry group of (*). A detailed study of these manifolds is however of importance when the full symmetry group is utilized to study separation of variables for (*). This will be shown in a subsequent article where we introduce a Hilbert space structure for solutions of (*) and discuss various equivalent representations of this structure. The problem of separation of variable for the Laplace operator on the upper sheet of the two sheeted hyperboloid has been investigated by Olevski[4] who found 34 coordinate systems. In this article we perform harmonic analysis on the space $L^2(H_+)$ of square integrable functions on the upper sheet H_+ of the two sheeted hyperboloid for the majority of coordinate systems given by Olevski. In a number of cases we give the harmonic analysis for coordinate systems which are representative of a particular subclass of coordinates. We specifically exclude the treatment of elliptic coordinates which requires the solution of multiparameter eigenvalue problems. This analysis is performed using the methods of integral geometry as developed by Gel'fand, Graev, and Vilenkin. The resulting spectral problems are then reduced to spectral problems on the cone $[X, X] = 0$. The contents of this article are arranged as follows: In Sec. 1 we present all the mathematical preliminaries and notations necessary for subsequent sections. These include the formulas for harmonic analysis on $L^2(H_+)$ and the space $L^2(H_s)$ of square integrable functions on the single sheet hyperboloid H_s. In Sec. 2 we give explicitly the 34 coordinate systems, due to Olevski, together with the pair of operators which specify each system, expressed in terms of the generators of the Lorentz group. In Sec. 3 we compute the spectral decompositions corresponding to the various coordinate systems. In the cases where this is already known the result is merely listed. Section 4 is devoted to a presentation of the appropriate basis functions on $L^2(H_+)$ and some comments on overlap functions. Finally, in Sec. 5 we compute various expansions on $L^2(H_s)$.

1. HARMONIC ANALYSIS AND THE LORENTZ GROUP

The homogeneous Lorentz group SO(3, 1) consists of those proper real linear transformations which leave $[X, X]$ invariant. The Lie algebra of SO(3, 1) is six-dimensional, and is generated by the rotation generators

$$M_1 = y\partial_z - z\partial_y,$$
$$M_2 = x\partial_z - z\partial_x, \qquad (1.1)$$
$$M_3 = x\partial_y - y\partial_x,$$

and the Lorentz transformation generators

$$K_1 = t\partial_x + x\partial_t,$$
$$K_2 = t\partial_y + y\partial_t, \tag{1.2}$$
$$K_3 = t\partial_z + z\partial_t.$$

The communtation relations are

$$[M_i, M_j] = \varepsilon_{ijk}M_k,$$
$$[M_i, K_j] = \varepsilon_{ijk}K_k, \tag{1.3}$$
$$[K_i, K_j] = -\varepsilon_{ijk}K_k.$$

The group SO(3, 1) has two Casimir operators $A = \mathbf{M}^2 - \mathbf{K}^2$ and $A' = \mathbf{M}\cdot\mathbf{K}$. The irreducible representations of the identity component of SO(3, 1) are labelled by two numbers. (j_0, σ), where j_0 is an integer or half-integer and σ is in general complex. If $\psi_{j_0\sigma}$ transforms according to the irreducible representation (j_0, σ) then

$$A\psi_{j_0\sigma} = -[j_0^2 + \sigma(\sigma + 2)]\psi_{j_0\sigma},$$
$$A'\psi_{j_0\sigma} = -j_0(\sigma + 1)\psi_{j_0\sigma}. \tag{1.4}$$

If the irreducible representation is in addition unitary then σ must have the one of the forms:

1. $\sigma = -1 + is$, $s \in \mathbf{R}$. This is the principal series.

2. $-1 \leqslant \sigma \leqslant 0$, $\sigma \in \mathbf{R}$ and $j_0 = 0, \pm 1, \cdots$.

This is the complementary series.

Further details concerning the representation theory of the Lorentz group can be found in Naimark[5] and Gel'fand *et al.*[6] We now give the basic formulas necessary for the harmonic analysis of functions defined on the spaces $L^2(H_+)$ and $L^2(H_s)$ mentioned in the introduction. These formulas are due to Gel'fand *et al.*[7]

A. Harmonic analysis on $L^2(H_+)$

The space $L^2(H_+)$ consists of functions $f(X)$ defined on the upper sheet of the hyperboloid $[X, X] = 1$, $t \geqslant 1$, satisfying

$$\int |f(X)|^2\, dX < \infty, \tag{1.5}$$

where $dX = dxdydz/(1 + x^2 + y^2 + z^2)^{1/2}$. The harmonic analysis of a function $f(X) \in L^2(H_+)$ requires only the unitary irreducible representations $\sigma = -1 + is$ $(0 < s < \infty)$, $j_0 = 0$ corresponding to the principal series.

It is readily verified from the coordinate representation of the generators that $A' = 0$. The formulas which yield the harmonic analysis of $f(X)$ are

$$f(X) = \frac{1}{(4\pi)^3}\int_0^\infty s^2\, ds \int_\Gamma F(Y; s)\,[X, Y]^{-is-1}\, dw, \tag{1.6}$$

where Γ is any contour on the $[Y, Y] = 0$ cone which intersects each generator of the cone once and dw is the differential form defined by $dY = dPdw$ where $P(Y) = 1$ is the equation of Γ and $dY = dy_1dy_2dy_3/y_0$ is the invariant measure on the cone. Here

$$F(Y, s) = \int f(X)\,[X, Y]^{is-1}\, dX. \tag{1.7}$$

The function $f(X)$ is then decomposed into components which transform according to the unitary irreducible representations $\sigma = -1 + is$,

$$0 < s < \infty, \quad j_0 = 0.$$

B. Harmonic analysis on $L^2(H_s)$

The space $L^2(H_s)$ consists of functions $f(X)$ defined on the single sheet hyperboloid $[X, X] = -1$ and satisfying

$$f(X) = f(-X), \tag{1.8a}$$

$$\int |f(X)|^2\, dX < \infty, \tag{1.8b}$$

where $dX = dxdydz/(x^2 + y^2 + z^2 - 1)^{1/2}$. The harmonic analysis of a function $f(X) \in L^2(H_s)$ requires the unitary irreducible representations

(i) $\sigma = -1 + is$, $(0 < s < \infty)$, $j_0 = 0$, and

(ii) $\sigma = -1$, $j_0 = 2n$, $n = 1, 2, \cdots$.

The reason we choose the condition (1.8a) is that the harmonic analysis of a function satisfying this symmetry condition has been studied in detail by Gel'fand *et al.*[7] An example of an expansion which does not exhibit the property (1.8a) has been given by Zmuidzinas.[8] We should also mention the work of Limic *et al.*[9] who have examined the general problem of the expansion of square integrable functions defined on the transitivity surfaces of SO(p, q) in the canonical reduction. The expansion formulas for $L^2(H_s)$ are

$$f(X) = \frac{1}{2(4\pi)^3}\int_0^\infty s^2\, ds \int_\Gamma F(Y; s)\,|[X, Y]|^{-is-1}\, dw + \frac{4}{\pi^2}\sum_{n=1}^\infty n\int_\Gamma F(Y, B; 2n)e^{2in\theta}\,\delta([X, Y])\, dw,$$

with w and Γ as in (1.6). This expansion can be inverted via the formulas

$$F(Y; s) = \int f(X)\,|[X, Y]|^{is-1}\, dX,$$
$$F(Y, B; 2n) = \int f(X)e^{-2in\theta}\,\delta([X, Y])\, dX. \tag{1.10}$$

In both these formulas B is a four vector satisfying

$$[B, B] = -1, \quad [B, Y] = [Y, Y] = 0.$$

The first component of B is zero. The angle θ is given by the relation $\cos\theta = [X, B]$. For further details concerning these formulas we refer the reader to Gel'fand *et al.*[7]

2. SEPARABLE COORDINATE SYSTEMS FOR $\Delta\psi = -\sigma(\sigma+2)$ ON THE UPPER SHEET OF $[X, X]=1$

We now study the differential equation

$$(\mathbf{M}^2 - \mathbf{K}^2)\,\psi(X) = -\sigma(\sigma+2)\,\psi(X), \qquad (+)$$

where X ranges over the upper sheet of $[X, X] = 1$. As follows from (1.6), an arbitrary $f \in L^2(H_+)$ can be decomposed as an integral over solutions of (+). Furthermore, (+) arises when one looks for solutions of (∗) which are homogeneous in x, y, z, t.

Equation (+) has been studied by Olevski[4] who showed that it admitted exactly 34 separable orthogonal coordinate systems. Here we give Olevski's results in a somewhat more explicit form. We also give for the first time a characterization of each separable system in terms of a pair of commuting second-order elements L_1, L_2 in the enveloping algebra of the Lie algebra of SO(3, 1). The corresponding separated solutions are eigenfunctions of L_1 and L_2 and the eigenvalues are the separation constants. (Smorodinski and Tugov[10] have computed L_1 and L_2 earlier but only in the form of differential operators.) For each set of separable coordinates $\{\rho, \nu, \eta\}$ we list the metric

$$ds^2 = H_1^2\,d\rho^2 + H_2^2\,d\nu^2 + H_3^2\,d\eta^2, \quad H_i = H_i(\rho, \nu, \eta).$$

In terms of the metric coefficients, (+) becomes

$$\frac{1}{H_1H_2H_3}\left[\partial_\rho\left(\frac{H_2H_3}{H_1}\partial_\rho\psi\right) + \partial_\nu\left(\frac{H_1H_3}{H_2}\partial_\nu\psi\right) + \partial_\eta\left(\frac{H_1H_2}{H_3}\partial_\eta\psi\right)\right] = -\sigma(\sigma+2)\psi.$$

Setting $\psi = A_1(\rho)A_2(\nu)A_3(\eta)$ in this equation, one can easily derive the ordinary differential equations satisfied by the separated solutions A_i.

1. $ds^2 = d\rho^2 + \cosh^2\rho\,d\nu^2 + \sinh^2\rho\,d\eta^2,$
$t = \cosh\rho\cosh\nu, \quad x = \sinh\rho\cos\eta, \qquad (2.1)$
$y = \sinh\rho\sin\eta, \quad z = \cosh\rho\sinh\nu$
$-\infty < \rho < \infty, \quad -\infty < \nu < \infty, \quad 0 \leqslant \eta < 2\pi.$

The operators are

$$L_1 = K_3^2, \quad L_2 = M_3^2.$$

2. $ds^2 = d\rho^2 + e^{-2\rho}(d\nu^2 + d\eta^2),$
$t = \tfrac{1}{2}[e^\rho + (1+\nu^2+\eta^2)\,e^{-\rho}], \quad x = e^{-\rho}\nu, \qquad (2.2)$
$y = e^{-\rho}\eta, \quad z = \tfrac{1}{2}[e^\rho + (-1+\nu^2+\eta^2)\,e^{-\rho}],$
$-\infty < \rho, \nu, \eta < \infty.$

The operators are

$$L_1 = (K_1 + M_2)^2, \quad L_2 = (K_2 - M_1)^2.$$

3. $ds^2 = d\rho^2 + \sinh^2\rho(\mathrm{sn}^2(\nu, k) - \mathrm{sn}^2(\eta, k))(d\nu^2 - d\eta^2),$
$t = \cosh\rho, \quad x = (1/k)\sinh\rho\,\mathrm{dn}(\nu, k)\,\mathrm{dn}(\eta, k),$
$y = (ik/k')\sinh\rho\,\mathrm{cn}(\nu, k)\,\mathrm{cn}(\nu, k),$
$z = k\sinh\rho\,\mathrm{sn}(\nu, k)\,\mathrm{sn}(\eta, k), \qquad (2.3)$
$0 < k < 1, \quad k' = (1-k^2)^{1/2},$
$-\infty < \rho < \infty, \quad \nu \in [-2K, 2K],$
$\eta \in [-K, -K + 2iK'].$

The operators are

$$L_1 = M_1^2 + M_2^2 + M_3^2, \quad L_2 = M_1^2 + k^2M_2^2.$$

4. $ds^2 = d\rho^2 + \cosh^2\rho\,(\mathrm{sn}^2(\nu, k) - \mathrm{sn}^2(\eta, k))\,(d\nu^2 - d\eta^2),$
$t = ik\cosh\rho\,\mathrm{sn}(\nu, k)\,\mathrm{sn}(\eta, k),$
$x = (k/k')\cosh\rho\,\mathrm{cn}(\nu, k)\,\mathrm{cn}(\eta, k), \qquad (2.4)$
$y = (i/k')\cosh\rho\,\mathrm{dn}(\nu, k),\,\mathrm{dn}(\eta, k), \quad z = \sinh\rho,$
$-\infty < \rho < \infty, \quad \nu \in [K, K + 2iK],$
$\eta \in [iK', iK' + 2K].$

The operators are

$$L_1 = K_1^2 + K_2^2 - M_3^2, \quad L_2 = k^2M_3^2 + k'^2K_2^2.$$

5. Differential form as for system 4
$t = (ik/k')\cosh\rho\,\mathrm{cn}(\nu, k)\,\mathrm{cn}(\eta, k),$
$x = \varepsilon ik\cosh\rho\,\mathrm{sn}(\nu, k)\,\mathrm{sn}(\eta, k),$
$y = \varepsilon'(i/k')\cosh\rho\,\mathrm{dn}(\nu, k)\,\mathrm{dn}(\eta, k), \qquad (2.5)$
$z = \sinh\rho, \quad \varepsilon, \varepsilon' = \pm,$
$-\infty < \rho < \infty, \quad \nu \in [iK', iK' + 2K], \quad \eta \in [0, 2iK'].$

The operators are

$$L_1 = K_1^2 + K_2^2 - M_3^2, \quad L_2 = K_2^2 - k^2M_3^2.$$

6. $ds^2 = d\rho^2 + \tfrac{1}{4}(\nu - \eta)\cosh^2\rho\left[\frac{d\nu^2}{(\nu-a)(\nu-b)\nu} - \frac{d\eta^2}{(\eta-a)(\eta-b)\eta}\right].$

The coordinates are given by the equations

$$(t + iy)^2 = \frac{2(\nu-a)(\eta-a)}{a(a-b)}\cosh^2\rho, \qquad (2.6)$$
$$x = \sqrt{-\nu\eta/ab}\cosh\rho, \quad z = \sinh\rho,$$

where $a = b^* = \alpha + i\beta$, $\alpha, \beta \in \mathbf{R}$, $-\infty < \nu < 0$, $0 < \eta < \infty$. The operators are

$$L_1 = K_1^2 + K_2^2 - M_3^2,$$
$$L_2 = \beta(M_3K_2 + K_2M_3) + \alpha K_1^2.$$

7. $ds^2 = d\rho^2 + \cosh^2\rho\left(\frac{1}{\cos^2\nu} - \frac{1}{\cosh^2\eta}\right)(d\nu^2 + d\eta^2),$
$t = \tfrac{1}{2}\cosh\rho\left(\frac{\cosh\eta}{\cos\nu} + \frac{\cos\nu}{\cosh\eta}\right),$
$x = \cosh\rho\tanh\eta\tan\nu, \qquad (2.7)$
$y = \cosh\rho\left[\frac{1}{\cosh\eta\cos\nu} - \frac{1}{2}\left(\frac{\cosh\eta}{\cos\nu} + \frac{\cos\nu}{\cosh\eta}\right)\right],$
$z = \sinh\rho,$
$-\infty < \rho < \infty, \quad -\infty < \eta < \infty, \quad 0 \leqslant \nu < 2\pi.$

The operators are

$$L_1 = K_1^2 + K_2^2 - M_3^2,$$
$$L_2 = K_1^2 + K_2^2 + M_3^2 - M_3K_2 - K_2M_3.$$

8. $ds^2 = d\rho^2 + \cosh^2\rho\left(\frac{1}{\sinh^2\eta} + \frac{1}{\sin^2\nu}\right)(d\eta^2 + d\nu^2),$

$$t = \cosh\rho\left[\frac{1}{\sinh\eta\sin\nu} + \frac{1}{2}\left(\frac{\sinh\eta}{\sin\nu} - \frac{\sin\nu}{\sinh\eta}\right)\right],$$
$$x = \cosh\rho\coth\eta\cot\nu,$$
$$y = \tfrac{1}{2}\cosh\rho\left(\frac{\sin\nu}{\sinh\eta} - \frac{\sinh\eta}{\sin\nu}\right), \tag{2.8}$$
$$z = \sinh\rho, \qquad -\infty < \rho < \infty,$$
$$-\infty < \eta < \infty, \qquad 0 \leqslant \nu < 2\pi.$$

The operators are

$$L_1 = K_1^2 + K_2^2 - M_3^2,$$
$$L_2 = -K_1^2 + K_2^2 + M_3^2 - M_3K_2 - K_2M_3.$$

9. $ds^2 = d\rho^2 + \cosh^2\rho\left(\frac{1}{\nu^2} + \frac{1}{\eta^2}\right)(d\nu^2 + d\eta^2),$

$$t = \cosh\rho\frac{[(\nu^2+\eta^2)^2+4]}{8\nu\eta}, x = \tfrac{1}{2}\cosh\rho\left(\frac{\eta}{\nu} - \frac{\nu}{\eta}\right),$$
$$y = \cosh\rho\frac{[-(\nu^2+\eta^2)^2+4]}{8\nu\eta}, \qquad z = \sinh\rho, \tag{2.9}$$
$$-\infty < \rho < \infty, \quad -\infty < \nu < \infty, \quad -\infty < \eta < \infty.$$

The operators are

$$L_1 = K_1^2 + K_2^2 - M_3^2,$$
$$L_2 = K_1K_2 + K_2K_1 - K_1M_3 - M_3K_1.$$

10. $ds^2 = d\rho^2 + \sinh^2\rho\,(d\nu^2 + \sin^2\nu\,d\eta^2)$

$$t = \cosh\rho, \qquad x = \sinh\rho\sin\nu\cos\eta, \tag{2.10}$$
$$y = \sinh\rho\sin\nu\sin\eta, \qquad z = \sinh\rho\cos\nu,$$
$$-\infty < \rho < \infty, \qquad 0 \leqslant \nu < \pi, \qquad 0 \leqslant \eta < 2\pi.$$

The operators are

$$L_1 = M_1^2 + M_2^2 + M_3^2, \qquad L_2 = M_3^2.$$

11. $ds^2 = d\rho^2 + \cosh^2\rho\,(d\nu^2 + \sinh^2\nu\,d\eta^2),$

$$t = \cosh\rho\cosh\nu, \quad x = \cosh\rho\sinh\nu\cos\eta, \tag{2.11}$$
$$y = \cosh\rho\sinh\nu\sin\eta, \qquad z = \sinh\rho,$$
$$-\infty < \rho < \infty, \qquad -\infty < \nu < \infty, \qquad 0 \leqslant \eta < 2\pi.$$

The operators are

$$L_1 = K_1^2 + K_2^2 - M_3^2, \qquad L_2 = M_3^2.$$

12. $ds^2 = d\rho^2 + \cosh^2\rho\,(d\nu^2 + \cosh^2\nu\,d\eta^2),$

$$t = \cosh\rho\cosh\nu\cosh\eta,$$
$$x = \cosh\rho\cosh\nu\sinh\eta, \tag{2.12}$$
$$y = \cosh\rho\sinh\nu, \qquad z = \sinh\rho,$$
$$-\infty < \nu, \rho, \eta < \infty.$$

The operators are

$$L_1 = K_1^2 + K_2^2 - M_3^2, \qquad L_2 = K_1^2.$$

13. $ds^2 = d\rho^2 + \cosh^2\rho(d\nu^2 + e^{-2\nu}\,d\eta^2),$

$$t = \tfrac{1}{2}\cosh\rho\,[e^{\nu} + (1+\eta^2)\,e^{-\nu}],$$
$$x = \eta e^{-\nu}\cosh\rho, \tag{2.13}$$
$$y = \tfrac{1}{2}\cosh\rho\,[e^{\nu} + (-1+\eta^2)\,e^{-\nu}], \qquad z = \sinh\rho,$$
$$-\infty < \nu, \rho, \eta < \infty.$$

The operators are

$$L_1 = K_1^2 + K_2^2 - M_3^2, \qquad L_2 = (K_1 + M_3)^2.$$

14. $ds^2 = d\rho^2 + e^{-2\rho}(d\nu^2 + \nu^2\,d\eta^2),$

$$t = \tfrac{1}{2}[e^{\rho} + (1+\nu^2)e^{-\rho}], \qquad x = e^{-\rho}\,\nu\cos\eta,$$
$$y = e^{-\rho}\,\nu\sin\eta, \quad z = \tfrac{1}{2}[e^{\rho} + (-1+\nu^2)\,e^{-\rho}], \tag{2.14}$$
$$-\infty < \rho < \infty, \quad -\infty < \nu < \infty, \quad 0 \leqslant \eta < 2\pi.$$

The operators are

$$L_1 = (K_1 + M_2)^2 + (K_2 - M_1)^2, \qquad L_2 = M_3^2.$$

15. $ds^2 = d\rho^2 + e^{-2\rho}(\cosh 2\nu - \cos 2\eta)\,(d\nu^2 + d\eta^2),$

$$t = \tfrac{1}{2}[e^{\rho} + (1 + \cosh^2\nu - \sin^2\eta)e^{-\rho}],$$
$$x = e^{-\rho}\cosh\nu\cos\eta, \qquad y = e^{-\rho}\sinh\nu\sin\eta, \tag{2.15}$$
$$z = \tfrac{1}{2}[e^{\rho} + (-1 + \cosh^2\nu - \sin^2\eta)\,e^{-\rho}],$$
$$-\infty < \rho < \infty, \quad -\infty < \nu < \infty, \quad 0 \leqslant \eta < 2\pi.$$

The operators are

$$L_1 = (K_1 + M_2)^2 + (K_2 - M_1)^2,$$
$$L_2 = M_3^2 + (K_1 + M_2)^2.$$

16. $ds^2 = d\rho^2 + e^{-2\rho}(\eta^2 + \nu^2)\,(d\eta^2 + d\nu^2),$

$$t = \tfrac{1}{2}[e^{\rho} + (1 + \tfrac{1}{4}(\eta^2+\nu^2)^2)\,e^{-\rho}],$$
$$x_1 = \tfrac{1}{2}e^{-\rho}(\eta^2 - \nu^2), \qquad y = e^{-\rho}\,\eta\nu, \tag{2.16}$$
$$z = \tfrac{1}{2}[e^{\rho} + (-1 + \tfrac{1}{4}(\eta^2+\nu^2)^2)\,e^{-\rho}],$$
$$-\infty < \rho < \infty, \quad -\infty < \eta < \infty, \quad -\infty < \nu < \infty.$$

The operators are

$$L_1 = (K_1 + M_2)^2 + (K_2 - M_1)^2,$$
$$L_2 = M_3(K_1 + M_2) + (K_1 + M_2)\,M_3.$$

17. $ds^2 = (\mathrm{sn}^2(\rho,k) - \mathrm{sn}^2(\nu,k))(d\rho^2 - d\nu^2)$
$$+ (k^2/k'^2)\,\mathrm{cn}^2(\rho,k)\,\mathrm{cn}^2(\nu,k)\,d\eta^2,$$

$$t = k\,\mathrm{sn}(\rho,k)\,\mathrm{sn}(\nu,k),$$
$$x = (k/k')\,\mathrm{cn}(\rho,k)\,\mathrm{cn}(\nu,k)\cos\eta,$$
$$y = (k/k')\,\mathrm{cn}(\rho,k)\,\mathrm{cn}(\nu,k)\sin\eta, \tag{2.17}$$
$$z = (i/k')\,\mathrm{dn}(\rho,k)\,\mathrm{dn}(\nu,k),$$
$$\rho \in [K, K + 2iK'], \quad \nu \in [iK', iK' + 2K],$$
$$0 \leqslant \eta < 2\pi.$$

The operators are

$$L_1 = M_3^2, \qquad L_2 = K_1^2 + K_2^2 + k^2K_3^2 - k'^2M_3^2.$$

18. $ds^2 = (\mathrm{sn}^2(\rho,k) - \mathrm{sn}^2(\nu,k))\,(d\rho^2 - d\nu^2)$
$$- (1/k'^2)\,\mathrm{dn}^2(\rho,k)\,\mathrm{dn}^2(\nu,k)\,d\eta^2,$$

$$t = k\,\mathrm{sn}(\rho,k)\,\mathrm{sn}(\nu,k),$$
$$x = (i/k')\,\mathrm{dn}(\rho,k)\,\mathrm{dn}(\nu,k)\cos\eta,$$
$$y = (i/k')\,\mathrm{dn}(\rho,k)\,\mathrm{dn}(\nu,k)\sin\eta, \tag{2.18}$$
$$z = (k/k')\,\mathrm{cn}(\rho,k)\,\mathrm{cn}(\nu,k),$$
$$\rho \in [K, K + 2iK'], \quad \nu \in [iK', iK' + 2K],$$
$$0 \leqslant \eta < 2\pi.$$

The operators are

$$L_1 = M_3^2, \qquad L_2 = K_3^2 + k^2(K_1^2 + K_2^2) + k'^2\,M_3^2.$$

19. $ds^2 = (\mathrm{sn}^2(\rho, k) - \mathrm{sn}^2(\nu, k))\,(d\rho^2 - d\nu^2) + k^2\,\mathrm{sn}^2(\rho, k)\,\mathrm{sn}^2(\nu, k)\,d\eta^2,$

$$\begin{aligned} t &= k\,\mathrm{sn}(\rho, k)\,\mathrm{sn}(\nu, k)\cosh\eta,\\ x &= k\,\mathrm{sn}(\rho, k)\,\mathrm{sn}(\nu, \mathrm{k})\sinh\eta,\\ y &= (k/k')\,\mathrm{cn}(\rho, k)\,\mathrm{cn}(\nu, k),\\ z &= (i/k')\,\mathrm{dn}(\rho, k)\,\mathrm{dn}(\nu, k),\\ \rho &\in [K, K + 2iK'], \qquad \nu \in [iK', iK' + 2K],\\ -\infty &< \eta < \infty. \end{aligned} \tag{2.19}$$

The operators are

$$L_1 = K_1^2,$$
$$L_2 = K_2^2 - M_3^2 + k^2(K_3^2 - M_2^2) - (1 + k^2)\,K_1^2.$$

20. Differential form the same as in system 17 with

$$\begin{aligned} t &= (ik/k')\,\mathrm{cn}(\rho, k)\,\mathrm{cn}(\nu, k)\cosh\eta,\\ x &= (ik/k')\,\mathrm{cn}(\rho, \mathrm{k})\,\mathrm{cn}(\nu, k)\sinh\eta,\\ y &= (i/k')\,\mathrm{dn}(\rho, k)\,\mathrm{dn}(\nu, k),\\ z &= ik\,\mathrm{sn}(\rho, k)\,\mathrm{sn}(\nu, k),\\ \rho &\in [iK', iK' + 2K], \qquad \nu \in [-iK', iK'],\\ -\infty &< \eta < \infty. \end{aligned} \tag{2.20}$$

The operators are

$$L_1 = K_1^2, \qquad L_2 = K_3^2 - M_2^2 + k^2(K_1^2 - M_1^2).$$

21. Differential form the same as in system 18 with

$$\begin{aligned} t &= (ik/k')\,\mathrm{cn}(\rho, k)\,\mathrm{cn}(\nu, k),\\ x &= (i/k')\,\mathrm{dn}(\rho, k)\,\mathrm{dn}(\nu, k)\cos\eta,\\ y &= (i/k')\,\mathrm{dn}(\rho, k)\,\mathrm{dn}(\nu, k)\sin\eta,\\ z &= ik\,\mathrm{sn}(\rho, k)\,\mathrm{sn}(\nu, k),\\ \rho &\in [iK', iK' + 2K], \qquad \nu \in [0, 2iK'],\ 0 \leqslant \eta < 2\pi. \end{aligned} \tag{2.21}$$

The operators are

$$L_1 = M_3^2, \qquad L_2 = K_3^2 + M_3^2 - k^2(M_2^2 + M_1^2).$$

22. $ds^2 = \tfrac{1}{4}(\rho - \nu)\left[\dfrac{d\rho^2}{(\rho - a)(\rho - b)\rho} - \dfrac{d\nu^2}{(\nu - a)(\nu - b)\nu}\right] - \rho\nu d\eta^2.$

The coordinates are given by the equations

$$\begin{aligned} (t + iz)^2 &= 2(\rho - a)(\nu - a)/a(a - b),\\ x &= (\sqrt{-\rho\nu/ab})\cos\eta,\\ y &= (\sqrt{-\rho\nu/ab})\sin\eta, \end{aligned} \tag{2.22}$$

where

$$\begin{aligned} a = b^* &= \alpha + i\beta, \qquad \alpha, \beta \in \mathbf{R}, \qquad -\infty < \nu < 0,\\ 0 &< \rho < \infty, \qquad 0 \leqslant \eta < 2\pi. \end{aligned}$$

The operators are

$$L_1 = M_3^2,$$
$$L_2 = \alpha(K_1^2 + K_2^2 - M_1^2 - M_2^2) - \beta(K_1M_2 + M_2K_1 + K_2M_1 + M_1K_2) - 2\alpha M_3^2.$$

23. $ds^2 = \left(\dfrac{1}{\cos^2\rho} - \dfrac{1}{\cosh^2\nu}\right)(d\rho^2 + d\nu^2) + \dfrac{1}{\cos^2\rho\cosh^2\nu}\,d\eta^2,$

$$\begin{aligned} t &= \frac{1}{2}\left(\frac{\cosh\nu}{\cos\rho} + \frac{\cos\rho}{\cosh\nu}\right) + \frac{\eta^2}{2\cosh\nu\cos\rho},\\ x &= \frac{\eta}{\cosh\nu\cos\rho}, \qquad y = \tanh\nu\tan\rho,\\ z &= \left[\frac{1}{\cosh\nu\cos\rho} - \frac{1}{2}\left(\frac{\cosh\nu}{\cos\rho} + \frac{\cos\rho}{\cosh\nu}\right)\right] - \frac{\eta^2}{2\cosh\nu\cos\rho},\\ 0 &\leqslant \rho < 2\pi, \quad -\infty < \nu < \infty, \quad -\infty < \eta < \infty. \end{aligned} \tag{2.23}$$

The operators are

$$L_1 = (K_1 + M_2)^2,$$
$$L_2 = 2K_1^2 + K_2^2 + K_3^2 + M_1^2 - K_1M_2 - M_2K_1 - K_2M_1 - M_1K_2.$$

24. $ds^2 = \left(\dfrac{1}{\sinh^2\nu} + \dfrac{1}{\sin^2\rho}\right)(d\rho^2 + d\nu^2) + \dfrac{1}{\sinh^2\nu\,\sin^2\rho}\,d\eta^2,$

$$\begin{aligned} t &= \left[\frac{1}{\sinh\nu\,\sin\rho} + \frac{1}{2}\left(\frac{\sinh\nu}{\sin\rho} - \frac{\sin\rho}{\sinh\nu}\right)\right] + \frac{\eta^2}{2\sinh\nu\,\sin\rho},\\ x &= \frac{\eta}{\sinh\nu\,\sin\rho}, \qquad y = \cot\rho\,\coth\nu,\\ z &= \frac{1}{2}\left(\frac{\sinh\nu}{\sin\rho} - \frac{\sin\rho}{\sinh\nu}\right) - \frac{\eta^2}{\sinh\nu\,\sin\rho},\\ 0 &\leqslant \rho < 2\pi, \quad -\infty < \nu < \infty, \quad -\infty < \eta < \infty. \end{aligned} \tag{2.24}$$

The operators are

$$L_1 = (K_1 + M_2)^2,$$
$$L_2 = 2M_2^2 + M_1^2 + K_2^2 - K_3^2 - K_2M_1 - M_1K_2 - K_1M_2 - M_2K_1.$$

25. $ds^2 = \left(\dfrac{1}{\cos^2\rho} - \dfrac{1}{\cosh^2\nu}\right)(d\rho^2 + d\nu^2) + \tan^2\rho\,\tanh^2\nu\,d\eta^2, \qquad t = \dfrac{1}{2}\left(\dfrac{\cosh\nu}{\cos\rho} + \dfrac{\cos\rho}{\cosh\nu}\right),$

$$\begin{aligned} x &= \tan\rho\,\tanh\nu\cos\eta, \qquad y = \tan\rho\,\tanh\nu\sin\eta,\\ z &= \frac{1}{\cosh\nu\cos\rho} - \frac{1}{2}\left(\frac{\cosh\nu}{\cos\rho} + \frac{\cos\rho}{\cosh\nu}\right),\\ 0 &\leqslant \rho < 2\pi, \qquad -\infty < \nu < \infty, \qquad 0 \leqslant \eta < 2\pi. \end{aligned} \tag{2.25}$$

The operators are

$$L_1 = M_3^2, \quad L_2 = M_2^2 + M_1^2 + K_1^2 + K_2^2 + K_3^2 - M_2K_1 - K_1M_2 - M_1K_2 - K_2M_1.$$

26. $ds^2 = \left(\dfrac{1}{\sinh^2\nu} + \dfrac{1}{\sin^2\rho}\right)(d\nu^2 + d\rho^2) + \cot^2\rho\,\coth^2\nu\,d\eta^2,$

$$\begin{aligned} t &= \frac{1}{\sinh\nu\,\sin\rho} + \frac{1}{2}\left(\frac{\sinh\nu}{\sin\rho} - \frac{\sin\rho}{\sinh\nu}\right),\\ x &= \cot\rho\,\coth\nu\cos\eta,\\ y &= \cot\rho\,\coth\nu\sin\eta, \qquad z = \frac{1}{2}\left(\frac{\sinh\nu}{\sin\rho} - \frac{\sin\rho}{\sinh\nu}\right),\\ 0 &\leqslant \rho < 2\pi, \quad -\infty < \nu < \infty, \quad 0 \leqslant \eta < 2\pi. \end{aligned} \tag{2.26}$$

The operators are

$$L_1 = M_3^2,$$

$$L_2 = M_2^2 + M_1^2 + K_1^2 + K_2^2 - K_3^2 + K_1M_2 + M_2K_1 + M_1K_2 + K_2M_1.$$

27. $$ds^2 = \left(\frac{1}{\rho^2} + \frac{1}{\nu^2}\right)(d\rho^2 + d\nu^2) + \frac{1}{\rho^2\nu^2}\,d\eta^2,$$

$$t = \frac{(\rho^2+\nu^2)^2+4}{8\rho\nu} + \frac{\eta^2}{2\rho\nu}, \qquad x = \frac{\eta}{\rho\nu}, \tag{2.27}$$

$$y = \frac{1}{2}\left(\frac{\nu}{\rho} - \frac{\rho}{\nu}\right), \qquad z = \frac{-(\rho^2+\nu^2)^2+4}{8\rho\nu} - \frac{\eta^2}{2\rho\nu}.$$

The operators are

$$L_1 = (K_1 + M_2)^2,$$

$$L_2 = M_3(K_1 + M_2) + (K_1 + M_2)M_3 + K_3(K_2 - M_1) + (K_2 - M_1)K_3.$$

28. $$ds = \frac{1}{4}\left[\frac{(\rho-\eta)(\rho-\nu)}{P(\rho)}\,d\rho^2 + \frac{(\nu-\rho)(\nu-\eta)}{P(\nu)}\,d\nu^2 + \frac{(\nu-\rho)(\nu-\eta)}{P(\eta)}\,d\eta^2\right],$$

$$P(z) = (z-a)(z-b)(z-1)z,$$

$$t^2 = \frac{\rho\nu\eta}{ab}, \qquad x^2 = \frac{(\rho-1)(\nu-1)(\eta-1)}{(a-1)(b-1)},$$

$$y^2 = -\frac{(\rho-b)(\nu-b)(\eta-b)}{(a-b)(b-1)b}, \tag{2.28}$$

$$z^2 = \frac{(\rho-a)(\nu-a)(\eta-a)}{(a-b)(a-1)a},$$

$$0 < 1 < \eta < b < \nu < a < \rho.$$

The operators are

$$L_1 = ab\,K_1^2 + a\,K_2^2 + b\,K_3^2,$$

$$L_2 = (a+b)K_1^2 + (a+1)K_2^2 + (b+1)K_3^2 - a\,M_3^2 - b\,M_2^2 - M_1^2.$$

29. Differential form as in system 28 with

$$t^2 = -\frac{(\rho-1)(\nu-1)(\eta-1)}{(a-1)(b-1)}, \qquad x^2 = -\frac{\rho\nu\eta}{ab},$$

$$y^2 = -\frac{(\rho-b)(\nu-b)(\eta-b)}{(a-b)(b-1)}, \tag{2.29}$$

$$z^2 = \frac{(\rho-a)(\nu-a)(\eta-a)}{(a-b)(a-1)a},$$

$$\eta < 0 < 1 < b < \nu < a < \rho.$$

The operators are

$$L_1 = ab\,K_1^2 - a\,M_3^2 - b\,M_2^2,$$

$$L_2 = (a+b)K_1^2 - (a+1)M_3^2 - (b+1)M_2^2 + a\,K_2^2 + b\,K_3^2 - M_1^2.$$

30. Differential form as in system 28 with

$$a = b^* = \alpha + i\beta, \qquad \alpha, \beta \in \mathbf{R},$$

$$(x+it)^2 = \frac{2(\rho-a)(\nu-a)(\eta-a)}{(a-b)(b-1)b}, \tag{2.30}$$

$$y^2 = \frac{(\rho-1)(\nu-1)(\eta-1)}{(a-1)(b-1)}, \qquad z^2 = -\frac{\rho\nu\eta}{ab},$$

$$\eta < 0 < \nu < 1 < \rho.$$

The operators are

$$L_1 = -(\alpha^2+\beta^2)M_1^2 + \alpha(K_3^2 - M_2^2) - \beta(K_3M_2 + M_2K_3),$$

$$L_2 = -2\alpha M_1^2 + (\alpha+1)(K_3^2 - M_2^2) + \alpha(K_2^2 - M_3^2) + \beta(K_2M_3 + M_3K_2 - M_2K_3 - K_3M_2).$$

31. Differential form as in system 28 with

$$P(z) = (z-a)(z-1)z^2,$$

$$(t+x)^2 = \rho\nu\eta/a,$$

$$(t^2 - x^2) = (1/a^2)[a(\rho\nu + \rho\eta + \nu\eta) - (a+1)\rho\nu\eta],$$

$$y^2 = -(\rho-1)(\nu-1)(\eta-1)/(a-1), \tag{2.31}$$

$$z^2 = (\rho-a)(\nu-a)(\eta-a)/a^2(a-1),$$

$$0 < \eta < 1 < \nu < a < \rho.$$

The operators are

$$L_1 = (K_3 + M_2)^2 - a(K_2 + M_3)^2 + a\,K_1^2,$$

$$L_2 = (a+1)K_1^2 + K_3^2 - M_2^2 + a(M_3^2 - K_2^2) + (K_2 + M_3)^2 + (K_3 + M_2)^2.$$

32. Differential form as in system 28 with

$$(t+x)^2 = -\rho\nu\eta/a,$$

$$(t^2 - x^2) = (1/a^2)[a(\rho\nu + \nu\eta + \rho\eta) - (a+1)\rho\nu\eta],$$

$$y^2 = -(\rho-1)(\nu-1)(\eta-1)/(a-1), \tag{2.32}$$

$$z^2 = (\rho-a)(\nu-a)(\eta-a)/a^2(a-1),$$

$$-\eta < 0 < 1 < \nu < a < \rho.$$

The operators are

$$L_1 = -(K_3 + M_2)^2 + a(K_2 + M_3)^2 + a\,K_1^2,$$

$$L_2 = (a+1)\,K_1^2 + M_2^2 - K_3^2 + a(K_2^2 - M_3^2) - (K_2 + M_3)^2 - (K_3 + M_2)^2.$$

33. Differential form as in system 28 with

$$P(z) = (z-a)(z+1)\,z^2,$$

$$(t+x)^2 = -\nu\rho\eta/a,$$

$$(t^2 - x^2) = -(1/a^2)[a(\rho\eta + \rho\nu + \eta\nu) - (a-1)\nu\rho\eta],$$

$$y^2 = (\rho-a)(\nu-a)(\eta-a)/a^2(a+1), \tag{2.33}$$

$$z^2 = -(\rho+1)(\nu+1)(\eta+1)/(a+1),$$

$$\eta < -1 < 0 < \nu < a < \rho.$$

The operators are

$$L_1 = aK_1^2 - (K_2 + M_3)^2 + a(K_3 + M_2)^2,$$

$$L_2 = (a-1)\,K_1^2 - (K_2 + M_3)^2 + (K_3 + M_2)^2 + M_3^2 - K_2^2 + a(M_2^2 - K_3^2).$$

34. Differential form as in system 28 with

$$P(z) = (z-1)z^3, \qquad (t-x)^2 = -\nu\rho\eta,$$

$$2y(x-t) = \nu\rho + \nu\eta + \rho\eta - \nu\rho\eta,$$

$$x^2 + y^2 - t^2 = -\nu\rho\eta + \nu\rho + \nu\eta + \rho\eta - \nu - \rho - \eta$$

$$z^2 = (\nu - 1)(\rho - 1)(\eta - 1),$$
$$\eta < 0 < \nu < 1 < \rho. \tag{2.34}$$

The operators are

$$L_1 = (M_2 - K_3)^2 - K_1(K_2 - M_3) - (K_2 - M_3)K_2,$$
$$L_2 = M_2^2 - K_3^2 - M_1^2 - (M_2 - K_3)^2 - M_1(M_2 - K_3) - (M_2 - K_3)M_1.$$

3. THE SPECTRAL ANALYSIS OF SEPARABLE BASES ON $L_s^2(C)$

Following Vilenkin[21] we construct a Hilbert space $L_s^2(C)$ of homogeneous functions $f(Y)$ on the forward light cone C: $[Y\ Y] = 0$, $Y = (y_0, y_1, y_2, y_3)$, $y_0 > 0$. In particular we require that f be homogeneous of degree $\sigma = is - 1$,

$$f(\rho Y) = \rho^{is-1} f(Y), \qquad \rho > 0. \tag{3.1}$$

Let Γ be a contour on the cone C which cuts each generator exactly once. If $Y(\nu, \eta)$ is a parametrization for Γ then every Y on C can be expressed uniquely in the form

$$Y = \rho Y(\nu, \eta), \qquad \rho > 0.$$

Now the measure on C which is invariant under the identity component of SO(3, 1) is well known to be $dY = dy_1 dy_2 dy_3/y_0$. We define a measure dw on Γ by $dY = \rho d\rho dw$, i.e.,

$$dw = |\det A|\, \rho^{-2} y_0^{-1}\, d\nu d\eta,$$
$$A = \begin{bmatrix} y_1 & \partial_\nu y_1 & \partial_\eta y_1 \\ y_2 & \partial_\nu y_2 & \partial_\eta y_2 \\ y_3 & \partial_\nu y_3 & \partial_\eta y_3 \end{bmatrix}, \quad Y = (y_0, y_1, y_2, y_3). \tag{3.2}$$

Then $L_s^2(C)$ is the space of measurable functions $f(Y)$, in (3.1), on C such that

$$\int_\Gamma |f(Y)|^2\, dw < \infty.$$

The inner product on this space is

$$\langle f_1, f_2 \rangle = \int_\Gamma f_1(Y) \bar{f}_2(Y) dw, \qquad f_i \in L_s^2(C). \tag{3.3}$$

[It is easy to verify from (3.1) and (3.2) that the value of the inner product is independent of the contour Γ.]

Note that the function

$$h(X, Y) = [X, Y]^{is-1} \tag{3.4}$$

belongs to $L_s(C)$ for each $X \in H_+$. Furthermore, the function $F(X)$ defined by

$$F(X) = \langle f, h(X, \cdot\ ) \rangle = \int_\Gamma f(Y)\, [X, Y]^{-is-1}\, dw \tag{3.5}$$

is a solution of equation (+) for each $f \in L_s^2(C)$.

The action of the identity component of SO(3, 1) on the functions $F(X)$ as defined by operators (1.1) and (1.2) induces via (3.5) a corresponding action of SO(3, 1) on $L_s^2(C)$ given by

$$M_1 = y_2\partial_{y_3} - y_3\partial_{y_2}, \quad M_2 = y_1\partial_{y_3} - y_3\partial_{y_1},$$
$$M_3 = y_1\partial_{y_2} - y_2\partial_{y_1}, \quad K_1 = y_0\partial_{y_1} + y_1\partial_{y_0}, \tag{3.6}$$
$$K_2 = y_0\partial_{y_2} + y_2\partial_{y_0}, \quad K_3 = y_0\partial_{y_3} + y_3\partial_{y_0}.$$

The associated group action is unitary and irreducible. More explicity, if Γ_0 is the sphere $Y_0 = (1, \xi_1, \xi_2, \xi_3)$, $\xi_1^2 + \xi_2^2 + \xi_3^2 = 1$ so that $Y = \rho Y_0$, then the action of SO(3, 1) on elements of $L_s^2(C)$ restricted to Γ_0 is

$$M_1 = -\xi_3\partial_{\xi_2}, \quad M_2 = -\xi_3\partial_{\xi_1}, \quad M_3 = \xi_1\partial_{\xi_2} - \xi_2\partial_{\xi_1},$$
$$K_1 = (is - 1)\xi_1 + (1 - \xi_1^2)\partial_{\xi_1} - \xi_1\xi_2\partial_{\xi_2},$$
$$K_2 = (is - 1)\xi_2 - \xi_1\xi_2\partial_{\xi_1} + (1 - \xi_2^2)\partial_{\xi_2}, \tag{3.7}$$
$$K_3 = (is - 1)\xi_3 - \xi_1\xi_3\partial_{\xi_1} - \xi_2\xi_3\partial_{\xi_2}.$$

(We are taking ξ_1 and ξ_2 as the independent variables on Γ_0.) If Γ is another contour related to Γ_0 by Γ: $Y = \zeta Y_0$ then the operators L_0, in (3.7), are replaced by operators $L = L_0 + (1 - is)\ \zeta^{-1}\ (\tilde{L}_0\ \zeta)$, where $\tilde{L}_0$ is the purely differential part of L_0.

The 34 commuting pairs of operators L_1, L_2 defined in the preceding section can in an obvious manner be defined on $L_s^2(C)$ as commuting pairs of self-adjoint operators and the spectral resolution can be determined explicitly. For each of these systems we list a convenient contour Γ and in the most tractable cases a basis for $L_s^2(C)$ consisting of simultaneous eigenfunctions of L_1 and L_2. In the case of new results we give a full development of their derivation.

1. The contour Γ is given by

$$Y = (\cosh\nu, \cos\eta, \sin\eta, \sinh\nu),$$
$$-\infty < \nu < \infty, \qquad 0 \leqslant \eta < 2\pi.$$

The basis functions are

$$f^{(1)}_{\tau m} = (1/2\pi)\, e^{i\tau\nu}\, e^{im\eta}, \qquad -\infty < \tau < \infty,$$
$$m = 0, \pm 1, \pm 2, \cdots, \tag{3.8}$$
$$\langle f^{(1)}_{\tau m}, f^{(1)}_{\tau' m'} \rangle = \delta(\tau - \tau')\, \delta_{mm'}.$$

The eigenvalues of L_1 and L_2 are $-\tau^2$ and $-m^2$, respectively.

2. $Y = (\frac{1}{2}(1 + \nu^2 + \eta^2), \nu, \eta, \frac{1}{2}(-1 + \nu^2 + \eta^2))$

$$-\infty < \nu < \infty, \qquad -\infty < \eta < \infty.$$

The basis functions are

$$f^{(2)}_{\tau\kappa} = (1/2\pi)\, e^{i\tau\nu}\, e^{i\kappa\eta},$$
$$-\infty < \tau < \infty, \qquad -\infty < \kappa < \infty, \tag{3.9}$$
$$\langle f^{(2)}_{\tau\kappa}, f^{(2)}_{\tau'\kappa'} \rangle = \delta(\tau - \tau')\, \delta(\kappa - \kappa').$$

The eigenvalues of L_1 and L_2 are $-\tau^2$ and $-\kappa^2$, respectively.

3. $Y = (1, (1/k')\ \mathrm{dn}(\nu, k)\ \mathrm{dn}(\eta, k), (ik/k')\ \mathrm{cn}(\nu, k) \times \mathrm{cn}(\eta, k), k\ \mathrm{sn}(\nu, k)\ \mathrm{sn}(\eta, k))$, (3.10)

$$\nu \in [-2K, 2K], \qquad \eta \in [-K, -K + 2iK'].$$

The basis functions are

$$f^{(3)pq}_{lm} = E^{pq}_{lm}(\nu) E^{pq}_{lm}(\eta), \qquad l = 0, 1, 2, \cdots,$$

a product of Lamé polynomials.[12,13] Here p and q are the eigenvalues of the rotation $e^{i\pi M_3}$ and the reflection $P \times e^{i\pi M_1}$, respectively, where P is the parity operator and m is the number of zeros of the Lamé polynomials in the interval $[0, K]$. For l even, $0 \leqslant m \leqslant \frac{1}{2}l + 1$ if $p = q = +1$, and $0 \leqslant m \leqslant \frac{1}{2}l$ otherwise. If l is odd, $0 \leqslant m \leqslant$

$\frac{1}{2}(l-1)$ for $p=q=-1$, and $0 \leqslant m \leqslant \frac{1}{2}(l+1)$ otherwise. For further details on this see Ref. 13. The eigenvalues of L_1 and L_2 are $l(l+1)$ and λ_{jm}^{pq} respectively.

4. $Y_\pm = (ik\ \mathrm{sn}(\nu,k)\ \mathrm{sn}(\eta,k), (k/k')\ \mathrm{cn}(\nu,k)\ \mathrm{cn}(\eta,k), (i/k')\ \mathrm{dn}(\nu,k)\ \mathrm{dn}(\eta,k), \pm 1),$

$\nu \in [K, iK+2iK'], \qquad \eta \in [iK', iK'+2K].$

The orthonormal basis functions are

$$f_{sjm}^{(4)pq}(\nu,\eta) = C\varepsilon E_{jm}^{pq}(\nu)\, E_{jm}^{pq}(\eta), \quad \varepsilon = \pm, \tag{3.11}$$

where $C_+ = \begin{pmatrix}1\\0\end{pmatrix}, \qquad C_- = \begin{pmatrix}0\\1\end{pmatrix},$

and $j = -\frac{1}{2} + iq$, $0 < q < \infty$. These basis functions are a product of periodic Lamé functions.

Here $m = 0, 1, 2, \cdots$ and p and q are eigenvalues of the operators $P\, e^{i\pi M_2}$ and $P\, e^{i\pi M_1}$ respectively. The integer m denotes the number of zeros N_{pq} of the basis functions $E_{jm}^{pq}(z)$ in the interval $[0, 2K]$ according to the table:

N_{pq}	p	q
$2m$	1	1
$2m+1$	-1	1
$2m+2$	1	-1
$2m+1$	-1	-1

The eigenvalues of L_1 and L_2 are $j(j+1)$ and λ_{jm}^{pq} respectively.

5. $Y_\pm = \left(\frac{ik}{k'}\ \mathrm{cn}(\nu,k)\ \mathrm{cn}(\eta,k), \varepsilon\, ik\ \mathrm{sn}(\nu,k)\ \mathrm{sn}(\eta,k), \varepsilon' \frac{i}{k'}\ \mathrm{dn}(\nu,k)\ \mathrm{dn}(\eta,k), \pm 1\right),$

$\varepsilon, \varepsilon' = \pm, \qquad \nu \in [iK', iK'+2K], \qquad \eta \in [0, 2iK'].$

The basis functions are

$$f_{\pm jm}^{(5)\varepsilon\varepsilon'} = C_{\pm\varepsilon\varepsilon'}\, F_j^m(\nu,k)\, F_j^m(\eta,k), \tag{3.12}$$

where

$$C_{\pm\varepsilon\varepsilon'} C^{+}_{\pm\varepsilon\bar{\varepsilon}'} = \delta_{\varepsilon\bar{\varepsilon}}\delta_{\varepsilon'\bar{\varepsilon}'}, \quad j = -\tfrac{1}{2} + iq, \quad 0 < q < \infty.$$

The basis functions are products of Lamé Wangerin or finite Lamé functions.[14] The label $m = 0, 1, \cdots$ specifies the number of zeros of these functions in the interval $[iK', iK'+2K]$. The eigenvalues of L_1 and L_2 are $j(j+1)$ and λ_{jm} respectively. There is a Dirac $\delta(j-j')$ normalization on the parameter j and a Kronecker normalization on the remaining parameters.

6. $Y_\pm = \left(\mathrm{Re}\left\{\frac{2(\nu-a)(\eta-a)}{a(a-b)}\right\}^{1/2}, \left(\frac{-\nu\eta}{ab}\right)^{1/2}, \mathrm{Im}\left\{\frac{2(\nu-a)(\eta-a)}{a(a-b)}\right\} \pm 1\right),$

$-\infty < \nu < 0 < \eta < \infty.$

7. $Y_\pm = \left(\frac{1}{2}\left(\frac{\cosh\eta}{\cos\nu} + \frac{\cos\nu}{\cosh\eta}\right), \tanh\eta\tan\nu, \frac{1}{\cosh\eta\cos\nu} - \frac{1}{2}\left(\frac{\cosh\eta}{\cos\nu} + \frac{\cos\nu}{\cosh\eta}\right), \pm 1\right),$

$-\infty < \eta < \infty, \qquad 0 \leqslant \nu < 2\pi,$

8. $Y_\pm = \left(\frac{1}{\sinh\eta\sin\nu} + \frac{1}{2}\left(\frac{\sinh\eta}{\sin\nu} - \frac{\sin\nu}{\sinh\eta}\right), \coth\eta\cot\nu, \frac{1}{2}\left(\frac{\sin\nu}{\sinh\eta} - \frac{\sinh\eta}{\sin\nu}\right), \pm 1\right),$

$-\infty < \nu < \infty, \qquad 0 \leqslant \eta < 2\pi.$

9. $Y_\pm = \left(\frac{[(\nu^2+\eta^2)^2+4]}{8\nu\eta}, \frac{1}{2}\left\{\frac{\eta}{\nu} - \frac{\nu}{n}\right\}, \frac{[-(\nu^2+\eta^2)^2+4]}{8\nu\eta}, \pm 1\right),$

$-\infty < \nu < \infty, \qquad -\infty < \eta < \infty.$

These last four coordinate systems have been treated to some extent in Ref. 13 and we refer the reader to that work for further details.

10. $Y = (1, \sin\nu\cos\eta, \sin\nu\sin\eta, \cos\nu),$

$0 \leqslant \nu \leqslant \pi, \qquad 0 \leqslant \eta < 2\pi.$

The orthonormal basis functions are

$$f_{lm}^{(10)} = \left(\frac{2l+1}{4\pi}\,\frac{(l-|m|)!}{(l+|m|)!}\right)^{1/2} P_l^{|m|}(\cos\nu)\, e^{im\eta}, \tag{3.13}$$

$l = 0, 1, 2, \ldots, m = -l, -l+1, \ldots, l$. The eigenvalues of L_1 and L_2 are $l(l+1)$ and $-m^2$, respectively.

11. $Y = (\cosh\nu, \sinh\nu\cos\eta, \sinh\nu\sin\eta, \pm 1),$

$-\infty < \nu < \infty, \qquad 0 \leqslant \eta < 2\pi.$

The basis functions are

$$f_{\pm jm}^{(11)} = \frac{\Gamma(j+1-m)}{\Gamma(j+1)} P_j^m(\cosh\nu) e^{im\eta}, \tag{3.14}$$

where $j = -\frac{1}{2} + iq, \qquad 0 < q < \infty,$

which are normalized according to

$$\langle f_{\pm jm}^{(11)}, f_{\pm j'm'}^{(11)} \rangle = \frac{2\pi}{q\tanh\pi q}\,\delta(q-q')\,\delta_{mm'}.$$

The eigenvalues of L_1 and L_2 are $-j(j+1)$ and $-m^2$, respectively.

12. $Y = (\cosh\nu\cosh\eta, \cosh\nu\sinh\eta, \sinh\nu, \pm 1),$

$-\infty < \nu, \eta < \infty.$

The basis functions are

$$f_{\pm j\tau}^{(12)} = C_\varepsilon \frac{\Gamma(j+1+i\tau)\,\Gamma(-j+i\tau)}{\Gamma(j+1)} \times P_{-1/2+i\tau}^{-j-1/2}(\varepsilon\tanh\nu)\, e^{i\tau\eta}, \tag{3.15}$$

where

$$\varepsilon = \pm, \qquad C_+ = \begin{pmatrix}1\\0\end{pmatrix}, \qquad C_- = \begin{pmatrix}0\\1\end{pmatrix},$$

$j = -\frac{1}{2} + iq,$

$0 < q < \infty, \qquad$ and $\qquad -\infty < \tau < \infty.$

These basis functions are normalized according to

$$\langle f_{\pm j\tau}^{(12)\varepsilon}, f_{\pm j'\tau'}^{(12)\varepsilon'} \rangle = \frac{\delta(q-q')\,\delta(\tau-\tau')}{q\tanh\pi q}.$$

The eigenvalues of L_1 and L_2 are $-j(j+1)$ and $-\tau^2$, respectively.

13. $Y = (\frac{1}{2}[e^{\nu} + (1+\eta^2)e^{-\nu}],\ \eta e^{-\nu},$
$\frac{1}{2}[e^{\nu} + (-1+\eta^2)e^{-\nu}],\ \pm 1),$
$-\infty < \nu < \infty, \qquad -\infty < \eta < \infty.$

The basis functions are

$$f^{(13)}_{j\tau} = \frac{1}{\Gamma(j+1)} e^{-\nu/2} K_{j+1/2}(\tau e^{-\nu})\, e^{i\tau\eta}, \tag{3.16}$$

where

$$j = -\tfrac{1}{2} + iq, \quad 0 < q < \infty, \quad \text{and} \quad -\infty < \tau < \infty.$$

These functions are normalized according to

$$(f^{(13)}_{j\tau}, f^{(13)}_{j'\tau'}) = \frac{\delta(q-q')\,\delta(\tau-\tau')}{q \tanh \pi q}.$$

The eigenvalues of L_1 and L_2 are $-j(j+1)$ and $-\tau^2$, respectively.

14. $Y = (\frac{1}{2}(1+\nu^2),\ \nu\cos\eta,\ \nu\sin\eta,\ \frac{1}{2}(-1+\nu^2)),$
$0 \leqslant \nu < \infty, \qquad 0 \leqslant \eta < 2\pi.$

The basis functions are

$$f^{(14)}_{\chi m} = \left(\frac{\chi}{2\pi}\right)^{1/2} J_m(\chi\nu)\, e^{im\eta}, \tag{3.17}$$

$m = 0, \pm 1, \cdots, 0 < \chi < \infty$. The eigenvalues of L_1 and L_2 are $-\chi^2$ and $-m^2$, respectively,

$$\langle f^{(14)}_{\chi m}, f^{(14)}_{\chi' m'} \rangle = \delta(\chi - \chi')\,\delta_{mm'}.$$

15. $Y = (\frac{1}{2}(1 + \cosh^2\nu - \sin^2\eta),\ \cosh\nu\cos\eta,$
$\sinh\nu\sin\eta,\ \frac{1}{2}(-1 + \cosh^2\nu - \sin^2\eta)),$
$-\infty < \nu < \infty, \qquad 0 \leqslant \eta < 2\pi.$

The orthogonal basis functions are

$$f^{(15)}_{\chi n} = \begin{cases} Ce_n(\nu, \chi^2/4)\, ce_n(\eta, \chi^2/4), \\ Se_n(\nu, \chi^2/4)\, se_n(\eta, \chi^2/4), \end{cases} \tag{3.17}$$

products of Mathieu functions. Here $n = 0, 1, 2, \cdots$ is the number of zeros of the periodic Mathieu functions in the interval $0 \leqslant \eta \leqslant \frac{1}{2}\pi$.

The eigenvalues of L_1 and L_2 are $-\chi^2$ and a_n (even), b_n (odd), respectively, where even and oddness refer to the periodic Mathieu functions under the interchange $\eta \to -\eta$.

16. $Y = (\frac{1}{2}[1 + \frac{1}{4}(\eta^2+\nu^2)^2],\ \frac{1}{2}(\eta^2-\nu^2),\ \eta\nu,$
$\frac{1}{2}(-1 + \frac{1}{4}(\eta^2+\nu^2)^2)),$
$-\infty < \nu, \eta < \infty.$

The basis functions are

$$f^{(16)\varepsilon}_{\chi\lambda} = C_\varepsilon \frac{1}{\sqrt{2}\cosh\pi\lambda}\,[D_{-i\lambda-1/2}(\varepsilon\sigma\eta)\, D_{i\lambda-1/2}(\sigma\nu) + D_{-i\lambda-1/2}(-\varepsilon\sigma\eta)\, D_{i\lambda-1/2}(-\sigma\nu)], \tag{3.18}$$

where

$$\varepsilon = \pm,\ \sigma = e^{i\pi/4}\sqrt{2\chi}, \text{ and } C_+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ C_- = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$
$$-\infty < \lambda < \infty.$$

The eigenvalues of L_1 and L_2 are $-\chi^2$ and 2λ,

$$\langle f^{(16)\varepsilon}_{\chi\lambda}, f^{(16)\varepsilon'}_{\chi'\lambda'} \rangle = \delta(\chi-\chi')\,\delta(\lambda-\lambda')\,\delta_{\varepsilon\varepsilon'}.$$

17. For this coordinate system a suitable choice of contour on the cone is

$$Y = (k\,\mathrm{sn}\nu,\ (ik/k')\,\mathrm{cn}\nu\cos\eta,\ (ik/k')\,\mathrm{cn}\nu\sin\eta,\ (1/k')\,\mathrm{dn}\nu),$$

where

$$\nu \in [K, K + 2iK'], \qquad 0 \leqslant \eta < 2\pi.$$

The basis functions have the form

$$\Psi = \Phi(\nu) \begin{cases} \cos m\eta \\ \sin m\eta \end{cases}$$

where $m = 0, 1, 2, \cdots$ and Φ satisfies

$$\frac{1}{\mathrm{cn}(\nu,k)}\frac{d}{d\nu}\left(\mathrm{cn}(\nu,k)\frac{d\Phi}{d\nu}\right) - \left[\frac{m^2k'^2}{\mathrm{cn}^2(\nu,k)} + (1+s^2)\,\mathrm{cn}^2(\nu,k) - \lambda - (1+s^2)\right]\Phi = 0, \tag{3.19}$$

where λ is an eigenvalue of L_2, the eigenvalue of L_1 being $-m^2$. If we write $\Phi(\nu) = [\mathrm{cn}(\nu,k)]^m\,\Xi(\nu)$ then Ξ satisfies the equation

$$\frac{d^2\Xi}{d\nu^2} - \frac{(2m+1)\,\mathrm{sn}(\nu,k)\,\mathrm{dn}(\nu,k)}{\mathrm{cn}(\nu,k)}\frac{d\Xi}{d\nu} + [-k^2(1+s^2+m(m+2))\,\mathrm{cn}^2(\nu,k) + H_A]\,\Xi = 0, \tag{3.20}$$

where $H_A = k^2\lambda + k^2(1+s^2) + (k^2-k'^2)m(m+1)$.

There are then four types of solution to this equation and the imposition of periodic boundary conditions requires λ to assume a distinct set of discrete values λ_n. We now develop the solutions.

(1) $\Xi = \sum_{r=0}^{\infty} A_r\,[\mathrm{cn}(\nu,k)]^{2r}$,

the boundary condition is

$$\Xi(K + iK') = \Xi(K + 2iK') = 0.$$

The recurrence relations for the coefficients A_r are

$$\begin{aligned} &k^2H_AA_0 + 4k'^2(m+1)A_1 = 0, \\ &k^2[4(r-1)(r+m-1) + (1+s^2+m(m+2))]A_{r-1} \\ &\quad + [(k'^2-k^2)\,2r(2r+2m+1) - H_A]\,A_r \\ &\quad - \frac{4k'^2}{k^2}(r+1)(r+m+1)A_{r+1} = 0, \end{aligned} \tag{3.21}$$

where $H_A = \lambda + (1+s^2)$. We write this solution as

$$L^{++}_{smn}(\nu,k) = \sum_{r=0}^{\infty} A_r[\mathrm{cn}(\nu,k)]^{m+2r}, \qquad n = 0, 2, 4, \cdots.$$

(2) $\Xi = \mathrm{dn}(\nu,k)\sum_{r=0}^{\infty} B_r[\mathrm{cn}(\nu,k)]^{2r}$, the boundary condition is $\Xi'(K+iK') = \Xi(K+2iK') = 0$. The recurrence relations for the coefficients B_r are

$$k^2H_BB_0 + 4k'^2(m+1)B_1 = 0,$$

$$k^2[(2r-1)(2r+2m+1)+(1+s^2+m(m+2))]B_{r-1}+[(k'^2-k^2)\times 2r(2r+2m+1)-4rk^2-H_B]B_r-\frac{4k'^2}{k^2}(r+1)\times(r+m+1)B_{r+1}=0,\quad (3.22)$$

$$H_B=H_A+2k^2(m+1).$$

The solutions are written

$$L^{-+}_{smn}(\nu,k)=\mathrm{dn}(\nu,k)\sum_{r=0}^{\infty}B_r[\mathrm{cn}(\nu,k)]^{m+2r},\quad n=1,3,\cdots.$$

(3) $\Xi=\mathrm{sn}(\nu,k)\,\mathrm{dn}(\nu,k)\sum_{r=0}^{\infty}C_r[\mathrm{cn}(\nu,k)]^{2r}$, the boundary condition is $\Xi(K+iK')=\Xi'(K+2iK')=0$. The recurrence relations for C_r are

$$k^2H_CC_0+4k'^2(m+1)C_1=0,$$

$$k^2[(2r-1)(2r+2m+3)+(1+s^2+m(m+2))]C_{r-1}+[(k'^2-k^2)2r(2r+2m+3)-H_C]C_r-\frac{4k'^2}{k^2}(r+1)(r+m+1)C_{r+1}=0,\quad (3.23)$$

$$H_C=H_A+2(k^2-k'^2)(m+1).$$

The solutions are written

$$L^{+-}_{smn}(\nu,k)=\mathrm{sn}(\nu,k)\,\mathrm{dn}(\nu,k)\sum_{r=0}^{\infty}C_r[\mathrm{cn}(\nu,k)]^{m+2r},\quad n=2,4,\cdots.$$

(4) $\Xi=\mathrm{sn}(\nu,k)\sum_{r=0}^{\infty}D_r[\mathrm{cn}(\nu,k)]^{2r}$, the boundary condition is

$$\Xi'(K+iK')=\Xi'(K+2iK')=0.$$

The recurrence relations for the D_r are

$$k^2H_DD_0+4k'^2(m+1)D_1=0,$$

$$k^2[4(r-1)(r+m)-2(2m+1)+(1+s^2+m(m+2))]D_{r-1}+[(k'^2-k^2)2r(2r+2m+1)+4rk'^2-H_D]D_r-\frac{4k'^2}{k^2}(r+1)(r+m+1)D_{r+1}=0,\quad (3.24)$$

$$H_D=H_A-k'^2(4m+3).$$

The solutions are written

$$L^{--}_{smn}(\nu,k)=\mathrm{sn}(\nu,k)\sum_{r=0}^{\infty}D_r[\mathrm{cn}(\nu,k)]^{m+2r},\quad n=1,3,5,\cdots.$$

The general solution can be written as

$$f^{(17)\varepsilon\varepsilon'}_{smn}=L^{\varepsilon\varepsilon'}_{smn}(\nu,k)\begin{Bmatrix}\cos m\eta\\ \sin m\eta\end{Bmatrix},\quad (3.25)$$

where ε and ε' are the eigenvalues of the operators $Pe^{i\pi M_3}$ and $e^{i\pi M_3}$. The number n is the number of zeros of the basis functions Ξ in the interval $[K-iK', K+iK']$. We will call these solutions associated periodic Lamé functions of the first kind.

18. $Y=(k\,\mathrm{sn}(\nu,k),(1/k')\,\mathrm{dn}(\nu,k)\cos\eta,(1/k')\,\mathrm{dn}(\nu,k)\times\sin\eta,\varepsilon(ik/k')\,\mathrm{cn}(\nu,k))$, where

$$\varepsilon=\pm1,\quad \nu\in[K,K+iK'],\quad 0\leqslant\eta<2\pi.$$

The basis functions have the form

$$\Psi=\Phi(\nu)\begin{Bmatrix}\cos m\eta\\ \sin m\eta\end{Bmatrix},$$

where $m=0,1,2,\cdots$ and Φ satisfies

$$\frac{1}{\mathrm{dn}(\nu,k)}\frac{d}{d\nu}\left(\mathrm{dn}(\nu,k)\frac{d\Phi}{d\nu}\right)+\left[\frac{m^2k'^2}{\mathrm{dn}^2(\nu,k)}-(1+s^2)\,\mathrm{dn}^2(\nu,k)+\lambda+(1+s^2)\right]\Phi=0.\quad (3.26)$$

The solutions to an equation similar to the above, have been investigated in Ref. 15. The development of solutions to the above equation proceeds in direct analogy with the procedure used in Ref. 15 to obtain finite solutions. In the case of interest here however we have the solutions expressed as an infinite series. We denote the solutions of (3.20),

$$\Phi=[\mathrm{dn}(\nu,k)]^m\,K^{P\varepsilon}_{snm}(\mathrm{dn}(\nu,k)),$$

in analogy with the solutions in Ref. 15. The superscript $P=A$, B, C, or D indicates the form of the solution as an expansion in Jacobi elliptic functions, viz. $P=A$ corresponds to the function

$$K^{A\varepsilon}_{snm}(\mathrm{dn}(\nu,k))=\sum_{r=0}^{\infty}A_r[\mathrm{dn}(\nu,k)]^{2r}.$$

The recurrence relations for the expansion coefficients are those in Ref. 15 with $2F(2F+2)$ replaced by $(1+s^2)$ and k by k'. Similarly we have that $P=B$ gives

$$K^{B\varepsilon}_{snm}(\mathrm{dn}(\nu,k))=\mathrm{cn}(\nu,k)\sum_{r=0}^{\infty}B_r[\mathrm{dn}(\nu,k)]^{2r},$$

$P=C$ gives

$$K^{C\varepsilon}_{snm}(\mathrm{dn}(\nu,k))=\mathrm{sn}(\nu,k)\sum_{r=0}^{\infty}C_r[\mathrm{dn}(\nu,k)]^{2r},$$

and $P=D$ gives

$$K^{D\varepsilon}_{snm}(\mathrm{dn}(\nu,k))=\mathrm{sn}(\nu,k)\,\mathrm{cn}(\nu,k)\sum_{r=0}^{\infty}D_r[\mathrm{dn}(\nu,k)]^{2r}.$$

In each case the spectrum of L_2 is discrete and is labelled by the positive integer n. The basis functions are then of the form

$$f^{(18)}_{sn}=C_\varepsilon[\mathrm{dn}(\nu,k)]^m\,K^{P\varepsilon}_{snm}(\mathrm{dn}(\nu,k))\begin{bmatrix}\cos m\eta\\ \sin m\eta\end{bmatrix}\quad (3.27)$$

where

$$C_+=\begin{pmatrix}1\\0\end{pmatrix}\quad\text{and}\quad C_-=\begin{pmatrix}0\\1\end{pmatrix}.$$

The $K^{P\varepsilon}_{snm}$ functions we have introduced here will be called associated periodic Lamé functions of the second kind.

21. $Y=((k/k')\,\mathrm{cn}(\nu,k),(k/k')\,\mathrm{dn}(\nu,k)\cos\eta,(k/k')\,\mathrm{dn}(\nu,k)\sin\eta,ik\,\mathrm{sn}(\nu,k))$,

where $\nu\in[-iK',iK']$, $0\leqslant\eta<2\pi$. The basis functions then have the form

$$\Psi = \Phi(\nu)\begin{cases}\cos m\eta \\ \sin m\eta\end{cases}, \qquad \text{with} \quad m = 0, 1, 2, \cdots$$

where Φ satisfies the same equation as in system 18. The corresponding boundary value problem is singular at both ends $\nu = \pm iK'$. The spectrum is however discrete and suitable boundary conditions are as follows.

(1) $\mathrm{dn}(\nu, k)\ \Phi(\nu)$ bounded at $\nu = \pm iK'$, $\Phi(0) = 0$. Solutions satisfying these conditions can be developed in the form

$$\Phi(\nu) = \sum_{r=0}^{\infty} A_r[\mathrm{dn}(\nu, k)]^{-1-is-2r}.$$

Such solutions are denoted by $W_{sm}^{2n}(\nu, k)$ and correspond to an eigenvalue λ_{2n} of L_2. This solution has $2n$ zeros in the interval $[-iK', iK']$. The recurrence relations for the coefficients A_r are

$$[H + (1 + k'^2)(i - s)s]A_0 + 4(1 + is)A_1 = 0,$$
$$k'^2[m^2 - (2r - 1 + is)^2]A_{r-1} + [\lambda_{2n} + 1 + s^2 + (2r + 1 + is)(2r + is)]A_r - 4(r + 1)(r + 1 + is)A_{r+1} = 0.$$

(2) $\mathrm{dn}(\nu, k)\ \Phi(\nu)$ bounded at $\nu = \pm iK'$, $\Phi'(0) = 0$. Solutions satisfying these conditions can be developed in the form

$$\Phi(\nu) = \mathrm{sn}(\nu, k)\sum_{r=0}^{\infty} B_r[\mathrm{dn}(\nu, k)]^{-2-is-2r}.$$

Such solutions are denoted by $W_{sm}^{2n+1}(\nu, k)$ and correspond to an eigenvalue λ_{2n+1} of L_2. Each such solution has $2n + 1$ zeros in the interval $[-iK', iK']$. The recurrence relations for the coefficients B_r are

$$[H' + (1 + k'^2)(i - s)s]B_0 + 4(1 + is)B_1 = 0,$$
$$k'^2[m^2 - (2r - 1 + is)^2]B_{r-1} + [H' - 4k'^2r + (1 + k'^2)(is + 2r + 2)(is + 2r + 1)]B_r - 4(r + 1)(r + 1 + is)B_{r+1} = 0,$$
$$H' = -2k'^2(1 + is) + (1 + s^2) + \lambda_{2n+1}.$$

The complete set of basis function is

$$f_{nm}^{(21)} = W_{sm}^{n}(\nu, k)\, e^{im\eta}, \qquad m = 0, \pm 1, \cdots. \tag{3.28}$$

22. For this coordinate system suitable coordinates on the cone are given by the relations

$$(t + iz)^2 = \frac{2(\nu - a)}{a(a - b)}, \qquad x = \sqrt{\frac{-\nu}{ab}}\cos\eta,$$
$$y = \sqrt{\frac{-\nu}{ab}}\sin\eta,$$

with $-\infty < \nu < 0$.

23. $Y = (\frac{1}{2}[\cosh\nu + \eta^2/\cosh\nu],\ \eta/\cosh\nu,\ \tanh\nu,\ 1/\cosh\nu - \frac{1}{2}[\cosh\nu + \eta^2/\cosh\nu])$,
$-\infty < \nu < \infty, \quad -\infty < \eta < \infty.$

If we write the basis functions as $\Psi = \cosh\nu\ \Phi\, e^{i\tau\eta}$ then Φ satisfies

$$\frac{d^2\Phi}{d\nu^2} + \tanh\nu\,\frac{d\Phi}{d\nu} - \left[\tau^2\cosh^2\nu + \frac{s^2}{\cosh^2\nu} + \lambda\right]\Phi = 0, \tag{3.29}$$

where λ is the eigenvalue of L_2. The spectrum of λ is discrete and the corresponding eigenvalues are denoted by $\lambda_{\nu+2r}^{is}$, $r = 0, \pm 1, \cdots$. (This is the notation adopted by Meixner and Schäfke.[20]) We note that (3.29) is a form of the spheroidal equation. The basis functions are then

$$f_{\nu+2r,\tau}^{(23)} = \cosh\nu\ S_{\nu+2r}^{is\,(1)}(i\sinh\nu, \tau)\, e^{i\tau\eta},$$

where $-\infty < \tau < \infty$, $r = 0, \pm 1, \pm 2,$ and $\nu \neq \frac{1}{2} \pmod 1$. The eigenvalues of L_1 and L_2 are $-\tau^2$ and $\lambda_{\nu+2r}^{is}$.

24. $Y = (1/\sinh\nu + \frac{1}{2}[\sinh\nu + \eta^2/\sinh\nu],\ \eta/\sinh\nu,\ \coth\nu,\ \frac{1}{2}[\sinh\nu - \eta^2/\sinh\nu])$,
$-\infty < \nu < \infty, \quad -\infty < \eta < \infty.$

The basis functions have the form

$$f_{\nu+2r,\tau}^{(24)} = \sinh\nu\ S_{\nu+2r}^{is\,(1)}(\cosh\nu, \tau)\, e^{i\tau\eta}, \tag{3.30}$$

where $-\infty < \tau < \infty$, $r = 0, \pm 1, \pm 2, \cdots$, and $\nu \neq \frac{1}{2} \pmod 1$, and the eigenvalues of L_1 and L_2 are $-\tau^2$ and $\lambda_{\nu+2r}^{is}$, respectively.

25. $Y = (\frac{1}{2}\cosh\nu,\ \tanh\nu\cos\eta,\ \tanh\nu\sin\eta,\ 1/\cosh\nu - \frac{1}{2}\cosh\nu)$,
$-\infty < \nu < \infty, \quad 0 \leqslant \eta < 2\pi.$

The basis functions are

$$f_{km}^{(25)} = (2\pi^3)^{1/2}\,\Gamma\left(\frac{m + 1 + i\kappa + is}{2}\right)\Gamma\left(\frac{m + 1 + i\kappa - is}{2}\right)$$
$$\times [\Gamma(m + 1)\,\Gamma(1 + i\kappa)]^{-1}\,(\tanh\nu)^m\,(\cosh\nu)^{-ik}$$
$$\times {}_2F_1\left(\frac{m + 1 + i\kappa + is}{2}, \quad \frac{m + 1 + i\kappa - is}{2}, \quad m + 1;\ \tanh^2\nu\right)\begin{bmatrix}\cos m\eta \\ \sin m\eta\end{bmatrix}, \tag{3.31}$$

where $m = 0, 1, \cdots$ and $0 < \kappa < \infty$. The eigenvalues of the operators L_1 and L_2 are $-m^2$ and $-\kappa^2$, respectively,

$$\langle f_{\kappa m}^{(25)}, f_{\kappa' m'}^{(25)}\rangle = \delta(\kappa - \kappa')\,\delta_{mm'}.$$

26. $Y = (1/\sinh\nu + \frac{1}{2}\sinh\nu,\ \coth\nu\cos\eta,\ \coth\nu\sin\eta,\ \frac{1}{2}\sinh\nu)$,
$-\infty < \nu < \infty, \quad 0 \leqslant \eta < 2\pi.$

The basis functions are

$$f_{\kappa m}^{(26)} = \sqrt{2\pi}\,\pi^2\,(\tanh\nu)^{1+i(\kappa+s)}\,(\cosh\nu)^{-i\kappa}$$
$$\times \frac{\Gamma(\frac{1}{2}[m + 1 + i(s + \kappa)])}{\Gamma(\frac{1}{2}[m + 1 - i(s + \kappa)])}$$
$$\times \frac{\Gamma(\frac{1}{2}[m + 1 + i(\kappa - s)])\,\Gamma(-i\kappa)}{\Gamma(\frac{1}{2}[m + 1 + i(s - \kappa)])\,\Gamma(1 + i\kappa)}$$
$$\times {}_2F_1\left[\frac{-m + 1 + i(\kappa + s)}{2},\right.$$

$$\frac{m-1+i(\kappa+s)}{2};\quad 1+i\kappa;$$

$$\frac{1}{\cosh^2\nu}\Big)\begin{bmatrix}\cos m\eta\\ \sin m\eta\end{bmatrix},$$

where $m = 0, 1, 2 \cdots$ and $0 < \kappa < \infty$. The eigenvalues of the operators L_1 and L_2 are $-m^2$ and $-\kappa^2$ respectively where $m = 0, 1, \cdots$ and $0 < \kappa < \infty$. The normalization is the same as for system 25.

27. $Y = \left(\frac{\nu^4+4}{8\nu} + \frac{\eta^2}{2\nu}, \frac{\eta}{\nu}, \frac{1}{2}\nu, \frac{-\nu^4+4}{8\nu} - \frac{\eta^2}{2\nu}\right)$,

$-\infty < \nu < \infty, \qquad -\infty < \eta < \infty.$

28. From this point on the spectral problems that have to be solved involve more than one eigenvalue simultaneously. We therefore give only the coordinates on the cone in these cases,

$$t^2 = \frac{\nu\eta}{ab}, \qquad x^2 = \frac{(\nu-1)(\eta-1)}{(a-1)(b-1)},$$

$$y^2 = \frac{(\nu-b)(\eta-b)}{(a-b)(b-1)b}, \qquad z^2 = \frac{(\nu-a)(\eta-a)}{(a-b)(a-1)a},$$

$$1 < \eta < b < \nu < a.$$

29. $t^2 = -\frac{(\nu-1)(\eta-1)}{(a-1)(b-1)}, \qquad x^2 = -\frac{\nu\eta}{ab},$

$$y^2 = -\frac{(\nu-b)(\eta-b)}{(a-b)(b-1)b},$$

$$z^2 = \frac{(\nu-a)(\eta-a)}{(a-b)(a-1)a}, \qquad \eta < 0 < 1 < b < \nu < a.$$

30. $(x+it)^2 = \frac{2(\nu-a)(\eta-a)}{(a-b)(b-1)b},$

$$y^2 = \frac{(\nu-1)(\eta-1)}{(a-1)(b-1)}, \qquad z^2 = -\frac{\nu\eta}{ab},$$

$$\eta < 0 < \nu < 1.$$

31. $(t+x)^2 = \frac{\nu\eta}{a}, \qquad (t^2-x^2) = \frac{\nu+\eta}{a} - \frac{(a+1)\nu\eta}{a^2},$

$$y^2 = -\frac{(\nu-1)(\eta-1)}{(a-1)},$$

$$z^2 = \frac{(\nu-a)(\eta-a)}{a^2(a-1)}, \qquad 0 < \eta < 1 < \nu < a.$$

32. $(t+x)^2 = -\frac{\nu\eta}{a}, \quad (t^2-x^2) = \frac{\nu+\eta}{a} - \frac{(a+1)\nu\eta}{a^2},$

$$y^2 = -\frac{(\nu-1)(\eta-1)}{(a-1)},$$

$$z^2 = \frac{(\nu-a)(\eta-a)}{a^2(a-1)}, \qquad -\eta < 0 < 1 < \nu < a.$$

33. $(t+x)^2 = -\frac{\nu\eta}{a},$

$$(t^2-x^2) = -\frac{(\nu+\eta)}{a} + \frac{(a-1)\nu\eta}{a^2},$$

$$y^2 = \frac{(\nu-a)(\eta-a)}{a^2(a+1)}, \quad z^2 = -\frac{(\nu+1)(\eta+1)}{(a+1)},$$

$$\eta < -1 < 0 < \nu < a.$$

34. $(t-x)^2 = -\nu\eta, \quad 2y(x-t) = \nu+\eta-\nu\eta,$

$z^2 = (\nu-1)(\eta-1), \qquad \eta < 0 < \nu < 1.$

4. SEPARABLE BASIS FUNCTIONS ON $L^2(H_+)$

In this section we present the basis functions for $L^2(H_+)$ corresponding to the coordinate systems presented in Sec. 2. This is done using (3.5) and the spectral resolution of the operators L_1 and L_2 computed in the previous section. The basis functions are listed with the minimum of duplication necessary. Some of these basis functions which correspond to subgroup reductions were given by Vilenkin and Smorodinski[16] (see also Kalnins[17]). In each case the integral is relatively easy to evaluate because we know in advance the form of the separated solutions in the appropriate coordinate system such that variables separate in the integral. The 34 integral identities (3.5) corresponding to the separable coordinates for $\Delta\psi = \sigma(\sigma+2)\psi$ are nontrivial. Indeed many of the following results appear to be new.

1. $$F^{(1)}_{\tau m} = \frac{\pi\Gamma([|m|+1+i(s+\tau)]/2)}{\Gamma(|m|+1)} \times \frac{\Gamma([|m|+1+i(s-\tau)]/2)}{\Gamma(1+is)} \times e^{i\tau\nu}e^{im\eta}(\tanh\rho)^{|m|}(\cosh\rho)^{-1-is} \times {}_2F_1\left(\frac{|m|+1+i(s+\tau)}{2}, \frac{|m|+1+i(s-\tau)}{2}, \quad |m|+1; \tanh^2\rho\right).$$

2. $F^{(2)}_{\tau\kappa} = \frac{\sqrt{s\sinh\pi s}}{2\pi} e^{-\rho} K_{is}((\tau^2+\kappa^2)^{1/2} e^{-\rho}) e^{i\tau\nu}e^{i\kappa\eta}.$

3. $$F^{(3)qq'}_{em} = \frac{\Gamma(is)}{\Gamma(is-l)\sqrt{\sinh\rho}} \times P^{-l-1/2}_{-1/2+is}(\cosh\rho)\, E^{qq'}_{lm}(\nu)\, E^{qq'}_{lm}(\eta).$$

4. $F^{(4)qq'}_{\varepsilon jm} = C_\varepsilon H^s_q(\rho)\, E^{qq'}_{jm}(\nu)\, E^{qq'}_{jm}(\eta),$

where

$$H^s_q(\rho) = \frac{\Gamma(\frac{1}{2}+i(q+s))\,\Gamma(\frac{1}{2}+i(s-q))}{\Gamma(1+ip)\cosh\rho} \times P^{-is}_{-1/2+iq}(\varepsilon\tanh\rho).$$

5. $F^{(5),\varepsilon\varepsilon'}_{\pm jm} = C_{\pm\varepsilon\varepsilon'}H^s_q(\rho)\, F^m_j(\nu, k)\, F^m_j(\eta, k).$

6. $$F^{(10)}_{lm} = \left(\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}\right)^{1/2} \frac{\Gamma(is)}{\Gamma(is-l)\sqrt{\sinh\rho}} \times P^{-l-1/2}_{1/2+is}(\cosh\rho)\, P^m_l(\cos\nu)\, e^{im\eta}.$$

7. $F^{(11)}_{\varepsilon im} = C_\varepsilon H^s_q(\rho)\frac{\Gamma(\frac{1}{2}-m+iq)}{2\pi\Gamma(\frac{1}{2}+iq)} P^m_{-1/2+iq}(\cosh\nu)\, e^{im\eta}.$

8. $$F^{(12)}_{\varepsilon\varepsilon' jm} = C_{\varepsilon\varepsilon'}H^s_q(\rho)\frac{\Gamma(\frac{1}{2}+i(\tau+q))\,\Gamma(\frac{1}{2}+i(\tau-q))}{\Gamma(\frac{1}{2}+iq)} \times P^{-iq}_{-1/2+i\tau}(\varepsilon'\tanh\nu)\, e^{i\tau\eta}.$$

9. $F^{(13)}_{\varepsilon j\tau} = C_\varepsilon H^s_q(\rho)\frac{e^{-\nu/2}}{\Gamma(\frac{1}{2}+iq)} K_{iq}(\tau e^{-\nu})\, e^{i\tau\eta}.$

10. $F^{(14)}_{\chi m} = \sqrt{s\chi\sinh\pi s}\, e^{-\rho}K_{is}(\chi e^{-\rho})\, J_m(\chi\nu)\, e^{im\eta}.$

11. $F^{(15)}_{\chi\eta} = \sqrt{s\sinh\pi s}\, e^{-\rho}K_{is}(e^{-\rho})$

$$\times \begin{bmatrix} Ce_n(\nu, \chi^2/4)\, ce_n(\eta, \chi^2/4) \\ Se_n(\nu, \chi^2/4)\, se_n(\eta, \chi^2/4) \end{bmatrix}.$$

12. $F^{(16)}_{s\chi\lambda} = C_s \dfrac{\sqrt{s \sinh(\pi s/2)}}{\cosh \pi\lambda}\, e^{-\rho} K_{is}(\chi e^{-\rho})$
$\times [D_{-i\lambda-1/2}(\varepsilon\sigma\eta)\, D_{i\lambda-1/2}(\sigma\nu)$
$+ D_{-i\lambda-1/2}(-\varepsilon\sigma\eta)\, D_{i\lambda-1/2}(-\sigma\nu)]$.

13. $F^{(17)ss'}_{mn} = L^{ss'}_{smn}(\rho, k)\, L^{ss'}_{smn}(\nu, k) \begin{cases} \cos m\eta \\ \sin m\eta \end{cases}$.

14. $F^{(18)sP}_{mn} = [\mathrm{dn}(\nu, k)\, \mathrm{dn}(\rho, k)]^m\, K^{Ps}_{snm}(\mathrm{dn}(\nu, k))$
$\times K^{Ps}_{snm}(\mathrm{dn}(\rho, k)) \begin{cases} \cos m\eta \\ \sin m\eta \end{cases}$.

15. $F^{(21)}_{mn} = W^n_{sm}(\rho, k)\, W^n_{sm}(\nu, k)\, e^{im\eta}$.

16. $F^{(23)}_{\nu+2r,\tau} = T_{\nu,\tau} \cosh\nu \cos\rho\, S^{is\,(1)}_{\nu+2r}(i \sinh\nu, \tau)$
$\times Ps^{is}_{\nu}(\sin\rho, \tau^2)\, e^{i\tau\eta}$,

where $T_{\nu,\tau}$ is a normalization constant.

17. $F^{(24)}_{\nu+2r,\tau} = t_{\nu,\tau} \sinh\nu \sin\rho\, S^{is\,(1)}_{\nu+2r}(\cosh\nu, \tau)$
$\times Ps^{is}_{\nu}(\cos\rho, \tau^2)\, e^{i\tau\eta}$,

where $t_{\nu,\tau}$ is a normalization constant.

18. $F^{(25)}_{\kappa m} = M_{\kappa m}(\tan\rho \tanh\nu)^m (\cosh\nu)^{-i\kappa} (\cos\rho)^{1+m+is}$.
$\times {}_2F_1\left(\dfrac{m+1+i(\kappa+s)}{2}, \dfrac{m+1+i(\kappa-s)}{2}, m+1; \tanh^2\nu\right) {}_2F_1\left(\dfrac{m+1+i(\kappa+s)}{2}, \dfrac{m+1+i(s-\kappa)}{2}, m+1; -\sin^2\rho\right) \begin{cases} \cos m\eta \\ \sin m\eta \end{cases}$,

$M_{\kappa m} = \pi 2^{(3m+i(s-\kappa)+2)/2}$
$\times \dfrac{\Gamma(a)\Gamma(b)\Gamma(a+m-\frac{1}{2})\Gamma(1+is+m)}{\Gamma(1+i\kappa)\Gamma(\frac{1}{2}a+m+\frac{1}{4})\Gamma(\frac{1}{2}a+m+\frac{3}{4})\Gamma(1+is)}$
$\times {}_3F_2(a, b; a+m-\frac{1}{2}, \frac{1}{2}a+m+\frac{1}{4}, \frac{1}{2}a+m+\frac{3}{4}, \frac{1}{4})$

where

$a = [m+1+i(\kappa+s)]/2$,
$b = [m+1+i(\kappa-s)]/2$.

19. $F^{(26)}_{\kappa m} = M'_{\kappa m}(\tanh\nu)^{1+i(\kappa+s)} (\sinh\nu)^{-i\kappa} (\cot\rho)^m$
$\times (\sin\rho)^{1+m+is}\, {}_2F_1\left(\dfrac{-m+1+i(\kappa+s)}{2}, \dfrac{m+1+i(\kappa+s)}{2}; 1+is; \tanh^2\nu\right)$
$\times {}_2F_1\left(\dfrac{m+1+i(\kappa+s)}{2}, \dfrac{m+1+i(s-\kappa)}{2}; m+1; -\cos^2\rho\right) \begin{cases} \cos m\eta \\ \sin m\eta \end{cases}$,

where

$$M'_{\kappa m} = 2\pi^2 N \frac{\Gamma(1+is+i\kappa/2)\,\Gamma(a_1)\,\Gamma(a_2)\,\Gamma(a_3)}{\Gamma(b_1)\,\Gamma(b_2)} \times {}_3F_2\left(\begin{matrix} a_1, a_2, a_3 \\ b_1, b_2 \end{matrix}\middle| -1\right),$$

where

$$a_1 = \frac{-m+1+i(\kappa+s)}{2},\quad a_2 = \frac{m-1+i(s+\kappa)}{2},$$

$$a_3 = \frac{m+is}{2} + 1 + i(\kappa+s),$$

$$b_1 = 1 + i\kappa,$$

$$b_2 = \frac{m+i(s+\kappa)}{2} + 2 + 2is + i\kappa,$$

and

$$N = \frac{\Gamma([m+1+i(s+\kappa)]/2)\,\Gamma([m+1+i(\kappa-s)]/2)}{\Gamma([m+1-i(s+\kappa)]/2)\,\Gamma([m+1+i(s-\kappa)]/2)} \times \frac{\Gamma(-i\kappa)\,\Gamma(1+is+m)}{\Gamma(1+i\kappa)\,\Gamma(1+is)}.$$

Let $\{f_{\lambda\mu}\}$ be an ON basis for $L^2_s(C)$ consisting of (generalized) eigenfunctions corresponding to a commuting pair L_1, L_2,

$$L_1 f_{\lambda\mu} = \lambda f_{\lambda\mu}, \qquad L_2 f_{\lambda\mu} = \mu f_{\lambda\mu},$$

and let $\{F_{\lambda\mu}\}$ be the associated separable functions on H_+,

$$F_{\lambda\mu}(X) = \langle f_{\lambda\mu}, h(X, \cdot)\rangle. \tag{4.1}$$

It follows that

$$h(X, Y) = [X, Y]^{is-1} = \sum_{\lambda,\mu} f_{\lambda,\mu}(Y)\, \bar{F}_{\lambda,\mu}(X) \tag{4.2}$$

with convergence in $L^2_s(H_+)$ for each $X \in H_+$. A direct computation yields

$$\langle h(X_1, \cdot), h(X_2, \cdot)\rangle = 4\pi {}_2F_1(1-is, 1+is; \tfrac{3}{2}; \tfrac{1}{2} - \tfrac{1}{2}[X_1, X_2]) \tag{4.3}$$

and, from (4.2),

$$\langle h(X_1, \cdot), h(X_2, \cdot)\rangle = \sum_{\lambda,\mu} F_{\lambda,\mu}(X_2)\, \bar{F}_{\lambda,\mu}(X_1). \tag{4.4}$$

Thus, (4.3) is a bilinear generating function for products of separated solutions $F_{\lambda,\mu}$.

If $\{f'_{\alpha,\beta}\}$ is ON basis for $L^2_s(C)$ consisting of eigenfunctions of another commuting pair L'_1, L'_2 and $F'_{\alpha,\beta} = \langle f'_{\alpha,\beta}, h\rangle$ we have the pointwise convergent expansion

$$F'_{\alpha,\beta}(X) = \sum_{\lambda,\mu} C^{\alpha,\beta}_{\lambda,\mu} F_{\lambda,\mu}(X), \tag{4.5}$$

where the sum or integral is taken over the spectrum of L_1, L_2.[1] Furthermore the expansion coefficients can be computed in $L^2_s(C)$. Indeed

$$C^{\alpha,\beta}_{\lambda,\mu} = \langle f'_{\alpha,\beta}, f_{\lambda,\mu}\rangle, \tag{4.6}$$

so all overlaps can be expressed as integrals over a contour Γ on C.

A number of these coefficients can be found in the literature. In particular, systems 3 and 10 correspond to the subgroup reduction SO(3, 1) $\supset$ SO(3) and the overlaps relating these systems can be found in Ref. 12. Systems 4–9 and 11–13 correspond to the subgroup reduction SO(3, 1) $\supset$ SO(2, 1) and appropriate overlaps are computed in Ref. 13. Systems 2 and 14–16 correspond to the subgroup reduction SO(3, 1) $\supset$ E(2) and overlaps are contained in Ref. 18. The overlaps relating systems 1 and 3 can

be expressed in terms of Clebsch–Gordan coefficients for SO(2, 1).

5. EXPANSIONS ON $L^2(H_s)$

In this section we give the expansions on $L^2(H_s)$ for coordinate systems on the single sheeted hyperboloid H_s which cover one half of H_s. This is the imaginary Lobachevski space of Gelfand *et al.*[7] Only some of the coordinates given in Sec. 2 correspond in a natural way to such coordinates on H_s. The spectrum of Δ is both continuous and discrete for $L^2(H_s)$ and there are therefore two sets of basis functions for each of the coordinate systems we discuss. We now list the basis functions for the coordinate systems on $L^2(H_s)$ together with the coordinates. The orthonormalization is always Kronecker delta for discrete spectrum and Dirac delta for continuous spectrum. The discrete spectrum basis functions are obtained from (1.9) and (1.10) exactly as in the example worked out in Ref. 19.

1. $t = \frac{1}{2}[e^{\rho} - (1 + \nu^2 + \eta^2)\, e^{-\rho}], \quad x = e^{-\rho}\nu,$
$y = e^{-\rho}\eta, \quad z = \frac{1}{2}[e^{\rho} + (1 - \nu^2 - \eta^2)\, e^{-\rho}],$
$-\infty < \rho, \nu, \eta < \infty.$

The basis functions are

$$Fc^{(2)}_{\tau\kappa} = \frac{e^{-\rho}}{2\pi}\left(\frac{s}{2\sinh\pi s}\right)^{1/2}[J_{is}(\chi e^{-\rho}) + J_{-is}(\chi e^{-\rho})]\, e^{i\tau\nu}e^{i\kappa\eta},$$

where $\chi^2 = \tau^2 + \kappa^2$,

$$Fd^{(2)}_{\tau\kappa} = \frac{\sqrt{n}}{\pi}\, e^{-\rho}\, J_{2n}(\chi e^{-\rho})\, e^{i\tau\nu}e^{i\kappa\eta}.$$

The superscript refers to the system in Sec. 2 to which the coordinates correspond via analytic continuation.

2. $t = \sinh\rho, \quad x = (1/k')\cosh\rho\ \mathrm{dn}(\nu, k)\ \mathrm{dn}(\eta, k),$
$y = (ik/k')\cosh\rho\ \mathrm{cn}(\nu, k)\,\mathrm{cn}(\eta, k),$
$z = k\cosh\rho\ \mathrm{sn}(\nu, k)\,\mathrm{sn}(\eta, k),$
$-\infty < \rho < \infty, \quad \nu \in [-2K, 2K],$
$\eta \in [-K, -K + 2iK'].$

The basis functions are

$$Fc^{(3)ss'}_{ln} = \frac{4\pi s\Gamma(-is)}{\cosh\rho}\left[P^{is}_l(\tanh\rho) - \frac{\Gamma(l + 1 + is)}{\Gamma(l + 1 - is)}\, P^{-is}_l(\tanh\rho)\right] f^{(3)ss'}_{ln},$$

$$Fd^{(3)ss'}_{ln} = 2\left(n\frac{(2n - l)!}{(2n + l)!}\right)^{1/2}\frac{1}{\cosh\rho}\, P^{2n}_l(\tanh\rho) f^{(3)ss'}_{ln},$$

where for the discrete spectrum part $l = 0, 2, \ldots, 2n$.

3. $t = \sinh\rho, \quad x = \cosh\rho\sin\nu\cos\eta,$
$y = \cosh\rho\sin\nu\sin\eta, \quad z = \cosh\rho\cos\nu,$
$-\infty < \rho < \infty, \quad 0 \leqslant \nu \leqslant \pi, \quad 0 \leqslant \eta < 2\pi.$

The basis functions are

$$Fc^{(10)}_{lm} = \frac{4\pi s\Gamma(is)}{\cosh\rho}\left[P^{is}_l(\tanh\rho) - \frac{\Gamma(l + 1 + is)}{\Gamma(l + 1 - is)}\, P^{-is}_l(\tanh\rho)\right] f^{(10)}_{lm},$$

$$Fd^{(10)}_{lm} = 2\left(n\frac{(2n - l)!}{(2n + l)!}\right)^{1/2}\frac{1}{\cosh\rho}\, P^{2n}_l(\tanh\rho) f^{(10)}_{lm}.$$

4. $t = \frac{1}{2}[(e^{\rho} - (1 + \nu^2)\, e^{-\rho}], \quad x = e^{-\rho}\nu\cos\eta,$
$y = e^{-\rho}\nu\sin\eta, \quad z = \frac{1}{2}[e^{\rho} + (1 - \nu^2)\, e^{-\rho}],$
$-\infty < \rho < \infty, \quad 0 < \nu < \infty, \quad 0 \leqslant \eta < 2\pi.$

The basis functions are

$$Fc^{(14)}_{\chi m} = \left(\frac{s}{2\sinh\pi s}\right)^{1/2} e^{-\rho}[J_{is}(\chi e^{-\rho}) + J_{-is}(\chi e^{-\rho})] f^{(14)}_{\chi m},$$

$$Fd^{(14)}_{\chi m} = 2\sqrt{n}\ e^{-\rho}\, J_{2n}(\chi e^{-\rho}) f^{(14)}_{\chi m}.$$

5. $t = \frac{1}{2}[e^{\rho} - (1 + \cosh^2\nu - \sin^2\eta)\, e^{-\rho}],$
$x = e^{-\rho}\cosh\nu\cos\eta, \quad y = e^{-\rho}\sinh\nu\sin\eta,$
$z = \frac{1}{2}[e^{\rho} + (1 - \cosh^2\nu + \sin^2\eta)\, e^{-\rho}],$
$-\infty < \rho < \infty, \quad -\infty < \nu < \infty, \quad 0 \leqslant \eta < 2\pi.$

The basis functions are

$$Fc^{(15)}_{\chi n} = \left(\frac{s}{2\sinh\pi s}\right)^{1/2} e^{-\rho}[J_{is}(\chi e^{-\rho}) + J_{-is}(\chi e^{-\rho})] f^{(15)}_{\chi n},$$

$$Fd^{(15)}_{\chi n} = 2\sqrt{n}\ e^{-\rho}\, J_{2n}(\chi e^{-\rho}) f^{(15)}_{\chi n}.$$

6. $t = \frac{1}{2}[e^{\rho} - (1 + \frac{1}{4}(\eta^2 + \nu^2)^2)\, e^{-\rho}],$
$x = \frac{1}{2}\, e^{-\rho}(\eta^2 - \nu^2), \quad y = e^{-\rho}\eta\nu,$
$z = \frac{1}{2}[e^{\rho} + (1 - \frac{1}{4}(\eta^2 + \nu^2)^2)\, e^{-\rho}],$
$-\infty < \rho < \infty, \quad -\infty < \nu < \infty, \quad -\infty < \eta < \infty,$

$$Fc^{(16)\varepsilon}_{\chi\lambda} = \left(\frac{s}{2\sinh\pi s}\right)^{1/2} e^{-\rho}[J_{is}(\chi e^{-\rho}) + J_{-is}(\chi e^{-\rho})] f^{(16)\varepsilon}_{\chi\lambda},$$

$$Fd^{(16)\varepsilon}_{\chi\lambda} = 2\sqrt{n}\ e^{-\rho}\, J_{2n}(\chi e^{-\rho}) f^{(16)\varepsilon}_{\chi\lambda}.$$

7. The coordinates are as for coordinate system 17 with
$\rho \in [0, 2iK'], \quad \nu \in [iK', iK' + 2K], \quad 0 \leqslant \eta < 2\pi.$

The basis functions are

$$Fu^{(17)}_{mn} = L^{ss'}_{lmn}(\rho, k)\, L^{ss'}_{lmn}(\nu, k)\begin{bmatrix}\cos m\eta \\ \sin m\eta\end{bmatrix},$$

where $t = s$ for the continuous spectrum basis functions and $t = 2in$ for the discrete spectrum functions. These solutions are obtained by solving the recurrence relations for system (17) with s replaced by $2in$.

8. The coordinates are as in system 18 with
$\rho \in [0, 2iK'], \quad \nu \in [iK', iK' + 2K], \quad 0 \leqslant \eta < 2\pi.$

The basis functions are

$$Fu^{(18)}_{mn} = [\mathrm{dn}(\nu, k)\,\mathrm{dn}(\rho, k)]^m\, K^{Ps}_{lnm}(\mathrm{dn}(\nu, k)) \times K^{Ps}_{lnm}(\mathrm{dn}(\rho, k))\begin{bmatrix}\cos m\eta \\ \sin m\eta\end{bmatrix},$$

with t as in system 7.

9. The coordinates are as in system 19 with
$\rho \in [0, 2iK'], \quad \nu \in [iK', iK' + 2K],$
$0 \leqslant \eta < 2\pi.$

10. The coordinates are as in system 21 with
$\rho \in [0, 2iK'], \quad \nu \in [0, 2iK'], \quad 0 \leqslant \eta < 2\pi.$

The basis functions are

$$Fu^{(21)}_{nm} = W^n_{lm}(\rho, k)\, W^n_{lm}(\nu, k)\, e^{im\eta},$$

where t has the same significance as in system 7 of this section.

11. The coordinates are as in system 22 with

$$0 < \rho < \infty, \qquad 0 < \nu < \infty.$$

12. $$t = \frac{1}{2}\left[\frac{\cosh\nu}{\sinh\rho} - \frac{\sinh\rho}{\cosh\nu}\right] + \frac{\eta^2}{2\cosh\nu\,\sinh\rho},$$

$$x = \frac{\eta}{\cosh\nu\,\sinh\rho}, \qquad y = \tanh\nu\,\coth\rho,$$

$$z = \frac{1}{\cosh\nu\,\sinh\rho} + \frac{1}{2}\left[\frac{\sinh\rho}{\cosh\nu} - \frac{\cosh\nu}{\sinh\rho}\right] - \frac{\eta^2}{2\cosh\nu\,\sinh\rho},$$

$$-\infty < \rho < \infty, \quad -\infty < \nu < \infty, \quad -\infty < \eta < \infty.$$

The basis functions are

$$Fs^{(23)}_{r\tau} = K^t_{r\tau}\sinh\rho\,\cosh\nu\, S^{it(1)}_{\nu+2r}(\cosh\rho, \tau) \times S^{it(1)}_{\nu+2r}(\sinh\nu, \tau)\, e^{i\tau\eta},$$

with t as in system 7 and $K^t_{r\tau}$ a normalization constant.

13. $$t = \frac{1}{\sin\nu\,\sin\rho} + \frac{1}{2}\left[\frac{\sin\nu}{\sin\rho} + \frac{\sin\rho}{\sin\nu}\right] + \frac{\eta^2}{2\sin\rho\,\sin\nu},$$

$$x = \frac{\eta}{\sin\nu\,\sin\rho}, \qquad y = \cot\rho\,\cot\nu,$$

$$z = \frac{1}{2}\frac{\sin\nu}{\sin\rho} + \frac{\sin\rho}{\sin\nu} - \frac{\eta^2}{2\sin\nu\,\sin\rho},$$

$$Fs^{(24)}_{r\tau} = \kappa^t_{r\tau}\sin\nu\,\sin\rho\, Ps^{it}_{\nu}(\cos\nu, \tau^2) \times Ps^{it}_{\nu}(\cos\rho, \tau^2)\, e^{i\tau\eta},$$

with t as in system 7 and $\kappa^t_{r\tau}$ a normalization constant.

14. $$t = \frac{1}{2}\left[\frac{\cosh\nu}{\sinh\rho} - \frac{\sinh\rho}{\cosh\nu}\right], \quad x = \coth\rho\,\tanh\nu\,\cos\eta,$$

$$y = \coth\rho\,\tanh\nu\,\sin\eta,$$

$$z = \frac{1}{\cosh\nu\,\sinh\rho} + \frac{1}{2}\left[\frac{\sinh\rho}{\cosh\nu} - \frac{\cosh\nu}{\sinh\rho}\right],$$

$$-\infty < \rho < \infty, \qquad -\infty < \nu < \infty, \qquad 0 \leqslant \eta < 2\pi,$$

$$Fc^{(25)}_{\kappa m} = N_{\kappa m}(\tanh\nu)^m\,(\tanh\rho)^{1+i(\kappa+s)}\,(\cosh\nu\,\sinh\rho)^{-i\kappa} \times {}_2F_1\left(\frac{m+1+i(\kappa+s)}{2}, \frac{m+1+i(\kappa-s)}{2}; m+1; \tanh^2\nu\right) {}_2F_1\left(\frac{-m+1+i(\kappa+s)}{2}, \frac{m+1+i(\kappa+s)}{2}, 1+is; \tanh^2\rho\right)\begin{Bmatrix}\cos m\eta\\ \sin m\eta\end{Bmatrix},$$

where

$$N_{\kappa m} = 4\pi^2\, 2^{[1-m+i(s+\kappa]/2}\,\frac{\Gamma(1+is+m)}{\Gamma(1+is)\,\Gamma(1+i\kappa)} \times \sum_{r=0}^{\infty}\frac{(a_1)_r\,(a_2)_r\,(a_3)_r\,(a_4)_r}{(b_1)_r\,(b_2)_r\,(b_3)_r}\,\frac{1}{r!4^r},$$

$$a_1 = \tfrac{1}{2}[m+1+i(\kappa+s)],$$

$$a_2 = \tfrac{1}{2}[m+1+i(\kappa-s)], \qquad a_3 = -is,$$

$$a_4 = \tfrac{1}{2}[m-1+i(s-\kappa)], \qquad b_1 = m+1,$$

$$b_2 = \tfrac{1}{2}[m-1-i(s+\kappa)],$$

and

$$b_3 = \tfrac{1}{2}[m+1-i(s+\kappa)],$$

$$Fd^{(25)}_{lm} = (\tanh\nu\,\coth\rho)^{-1/2} d^{l+}_{|m-2n|,m+2n}(\cosh\nu) \times d^{l+}_{|m-2n|,m+2n}(i\sinh\rho)\begin{Bmatrix}\cos m\eta\\ \sin m\eta\end{Bmatrix}.$$

Here $l,m = 0, 1, 2, \cdots$ and $d^{l+}_{pq}(\cosh)$ is the matrix element of a hyperbolic rotation in the compact basis of the positive discrete series of SL(2, $\mathbf{R}$) as given by Bargmann.[11] Explicitly these functions are

$$d^{l+}_{pq}(\cosh z) = \bar{N}_{pq}(\sinh z)^{-(p+q)}(\cosh z)^{p-q} \times {}_2F_1(l+1-q, -l-q; 1+q-p; -\sinh^2 z), q > p$$

$$= \bar{N}_{pq}(\sinh z)^{-(p+q)}(\cosh z)^{q-p} \times {}_2F_1(l+1-p, -l-p; 1+p-q; -\sinh^2 z), \quad p > q,$$

where

$$\bar{N}_{pq} = (-1)^{p-q}\,\bar{N}_{qp} = \frac{1}{(p-q)!} \times \left(\frac{\Gamma(-l-q)\,\Gamma(l+1-q)}{\Gamma(-l-p)\,\Gamma(l+1-p)}\right)^{+1/2}$$

if $p \geqslant q$. (Note: We have at the time of writing not computed the normalization constant for discrete spectrum basis functions.)

15. $$t = \frac{1}{\sin\nu\,\cos\rho} - \frac{1}{2}\left[\frac{\sin\nu}{\sin\rho} + \frac{\sin\rho}{\sin\nu}\right],$$

$$x = \cot\rho\,\cot\nu\,\cos\eta, \qquad y = \cot\rho\,\cot\nu\,\sin\eta,$$

$$z = \frac{1}{2}\left[\frac{\sin\nu}{\sin\rho} + \frac{\sin\rho}{\sin\nu}\right],$$

$$Fc^{(26)}_{\kappa m} = N'_{\kappa m}(\cot\rho\,\cot\nu)^m(\sin\rho\,\sin\nu)^{1+m+is} \times {}_2F_1\left(\frac{m+1+i(\kappa+s)}{2}, \frac{m+1+i(s-\kappa)}{2}; m+1; -\cos^2\rho\right) {}_2F_1\left(\frac{m+1+i(\kappa+s)}{2}, \frac{m+1+i(s-\kappa)}{2}; m+1; -\cos^2\nu\right)\begin{Bmatrix}\cos m\eta\\ \sin m\eta\end{Bmatrix},$$

where

$$N'_{\kappa m} = 2\pi^2\, N\,\frac{\Gamma(a_1)\,\Gamma(a_2)\,\Gamma(a_3)\,\Gamma(-m+i\kappa)}{\Gamma(b_1)\,\Gamma(b_2)} \times {}_3F_2\left(\begin{matrix}a_1 & a_2 & a_3\\ b_1 & b_2\end{matrix}\,1\right)$$

with a_1, a_2, and b_1 as for system 12 on $L^2(H_s)$,

$$a_3 = i\kappa + \tfrac{1}{2}(is+m-1),$$

$$b_2 = 2i\kappa + \tfrac{1}{2}(is-3m-1),$$

and N is as given for system 26 on $L^2(H_+)$.

$$Fd^{(26)}_{lm} = N''_{lm}(\cos\nu\,\cos\rho)^m(\sin\nu\,\sin\rho)^{2n+1}$$

$$\times P_l^{(m,2n)}(\sin 2\nu)\, P_l^{(m,2n)}(\sin 2\rho)\begin{Bmatrix}\cos m\eta \\ \sin m\eta\end{Bmatrix},$$

where $P_l^{(\alpha,\beta)}(z)$ is a Jacobi polynomial and

$$N_{lm}'' = \left(\frac{n\Gamma(m+l+1)\,\Gamma(2n+l+1)\,2^{m+2n+1}}{2\pi l!\,\Gamma(m+2n+l+2)}\right)^{1/2} \Big/ P_l^{(m,2n)}(0).$$

16. $t = \dfrac{(\rho^2-\nu^2)^2+4}{8\rho\nu} + \dfrac{\eta^2}{2\rho\nu}, \qquad x = \dfrac{\eta}{\rho\nu},$

$$y = \frac{1}{2}\left[\frac{\nu}{\rho} + \frac{\rho}{\nu}\right], \qquad z = \frac{-(\rho^2-\nu^2)^2+4}{8\rho\nu} - \frac{\eta^2}{2\rho\nu},$$

$$-\infty < \nu, \rho, \eta < \infty.$$

We have not given any mention of the coordinate systems which require the solution of multiparameter eigenvalue problems.

[1]W. Miller Jr., "Symmetry, separation of variables and special functions" in *Theory and Application of Special Functions*, edited by R. Askey (Academic, New York, 1975).
[2]E. G. Kalnins and W. Miller Jr., J. Math. Phys. **17**, 356 (1976).
[3]E. G. Kalnins and W. Miller Jr., J. Math. Phys. **17**, 369 (1976).
[4]P. M. Olevski, Mat. Sb. **27**, 379–426 (1950).
[5]M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964).
[6]I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Rotation and Lorentz Groups and Their Applications* (Pergamon, New York, 1963).
[7]I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.
[8]J. S. Zmuidzinas, J. Math. Phys. **7**, 764–80 (1966).
[9]N. Limic, J. Niederle, and R. Raczka, J. Math Phys. **7**, 764 (1966).
[10]Ya. A. Smorodinski and I. I. Tugov, Zh, Eksp. Teor. Fiz, **50**, 653–59 (1966) [Sov. Phys. JETP **23**, 434–36 (1966)].
[11]V. Bargmann, Ann. of Math. **48**, 568–640 (1947).
[12]J. Patera and P. Winternitz, J. Math. Phys. **14**, 1130–39 (1973).
[13]E. G. Kalnins and W. Miller Jr., J. Math. Phys. **15**, 1263–74 (1974).
[14]These functions are defined in A. Erdélyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1955), Vol. 3.
[15]E. G. Kalnins, W. Miller Jr., and P. Winternitz, SIAM J. App. Math. **30**, 630–64 (1976).
[16]N. Ya. Vilenkin and Ya. A. Smorodinski, Zh. Eksp. Teor. Fiz. **46**, 1793 (1964) [Sov. Phys. JETP **46**, 1209 (1964)].
[17]E. G. Kalnins, "Subgroup reductions of the Lorentz group," Ph. D. thesis, University of Western Ontario (1972).
[18]W. Miller Jr., *Symmetry and separation of variables for linear differential equations* (Addison-Wesley, Reading, Mass., to appear).
[19]E. G. Kalnins, J. Math. Phys. **14**, 1316–19 (1973).
[20]J. Meixner and F. W. Schafke, *Mathieushe Funktionen und Spharoidfunktionen* (Springer, Berlin, 1965).
[21]N. Ya. Vilenkin, Mat. U.S.S.R. Sb. **3**, 109–21 (1967).

Structural stability of the phase transition in Dicke-like models*

R. Gilmore

Institut de Physique Théorique, Université Catholique de Louvain, B—1348 Louvain-la-Neuve, Belgium and Physics Department, University of South Florida, Tampa, Florida 33620 †

(Received 28 July 1976)

The free energy for a general class of Dicke models is computed and expressed simply as the minimum value of a potential function $\Phi = (E - TS)/N$. The function E/N is the image of the Hamiltonian under the quantum–classical correspondence effected by the atomic and field coherent state representations, and the function S is the logarithm of an SU(2) multiplicity factor. The structural stability of the second order phase transition under changes in the functional form of the Hamiltonian is determined by searching for stability changeovers along the thermal critical branch of Φ. The necessary condition for the presence of a second order phase transition is completely determined by the canonical kernel of the Hamiltonian. The sufficient condition is that a first order phase transition not occur at higher temperature. The critical temperature for a second order phase transition is given by a gap equation of Hepp–Lieb type.

1. INTRODUCTION

For sufficiently large values of the coupling constant λ, the Dicke model Hamiltonian[1] can support a second order phase transition.[2] Several recent studies[3–5] have shown that the phase transition disappears when the functional form of the Hamiltonian is changed. It is therefore of interest to inquire about the structural stability of the second order phase transition under changes in the functional form of the Hamiltonian describing the interaction between a simple field mode and N identical two-level atoms. Extension to finite mode systems is straightforward.

We describe the general class of nonlinear extensions[4] of the Dicke model under consideration in Sec. 2. In Sec. 3 the free energy F/N is computed for any model Hamiltonian in the class studied. The result is surprisingly simple. The free energy is expressed as the minimum of a potential function $\Phi = (E - TS)/N$. The function E/N is the P or Q representative of the Hamiltonian in the direct product atomic coherent state-field coherent state representation. The function S is the logarithm of an SU(2) multiplicity factor. The minimum is taken over the space of atomic and field coherent state parameters. Evaluated at the coherent state parameters which minimize Φ, the functions Φ, E/N, and S/N are equal to the free energy, energy, and entropy per particle.

In Sec. 4 we discuss first and second order transitons. Each model Hamiltonian must be investigated separately to determine whether or not first order phase transitions exist, since these involve global stability changeovers. Since second order phase transitions involve local properties, it is possible to give both necessary and sufficient conditions for their occurrence. The necessary condition does not involve the entire Hamiltonian, but only a particular piece of it. The critical temperature for a second order phase transition is determined from the "canonical kernel" by a gap equation of Hepp—Lieb type. The sufficient condition for a second order phase transition is that it is not preempted by a first order transition at higher temperature. The transition is extremely insensitive to the functional form of the Hamiltonian, provided the Hamiltonian possesses a particular kind of symmetry. In general, when the symmetry is violated, the second order transition disappears.

2. MODEL HAMILTONIANS

To study the effect which the structural form of the Hamiltonian has on the presence or absence of a second order phase transition in models of the Dicke type, we study the general class of Hamiltonians of the form

$$H/N = h_Q. \tag{2.1}$$

Throughout, $h = h(\mathbf{u}, \mathbf{v})$ is a function of the two three-component quantities $\mathbf{u} = (u_3, u_+, u_-)$ and $\mathbf{v} = (v_3, v_+, v_-)$. The operator h_Q is obtained from the function h through the operator substitutions indicated in Table I, and subsequent symmetrization, if necessary. The function h_C is obtained from h by the c-number substitutions shown in Table I.

We impose the following assumptions on h:

1. h is a finite multinomial in all arguments with finite coefficients;

2. h_Q is Hermitian;

3. h_C has a finite lower bound as a function of $\mu \in C$ for $\mathbf{v}$ in a sphere of radius $1/2$;

4. $h(u_3, u_+, u_-; v_3, v_+, v_-) = h(+u_3, -u_+, -u_-; +v_3, -v_+, -v_-)$;

5. h_C assumes its minimum value for $\mu = 0$ when $\mathbf{v} = 0$;

6. $\partial h/\partial v_3 > 0$ for $\mathbf{u} = 0$, $\mathbf{v} = 0$;

7. $h(0; -r, 0, 0)$ is a monotonic decreasing function

TABLE I. The operator substitutions and the c-number substitutions which transform $h(\mathbf{u}, \mathbf{v})$ to the operator h_Q and the c-number function h_C.

	$h \to h_Q \equiv H(a, a^\dagger; J)/N$	$h \to h_C \equiv E(\mu, r\theta\phi)/N$	Parameter properties
u_3	$n/N = a^\dagger a/N$	$\mu^*\mu = \alpha^*\alpha/N$	$\alpha \in \mathbb{C}, \mu \in \mathbb{C}$
u_+	$a^\dagger/\sqrt{N}$	$\mu^* = \alpha^*/\sqrt{N}$	$\alpha = \mu\sqrt{N}$
u_-	$a/\sqrt{N}$	$\mu = \alpha/\sqrt{N}$	
v_3	$J_3/N = (\sum_{j=1}^{N} \frac{1}{2}\sigma_j^3)/N$	$r\cos\theta$	$r \in [0, \frac{1}{2}]$
v_+	$J_+/N = (\sum_{j=1}^{N} \sigma_j^+)/N$	$\nu^* = r\sin\theta\, e^{+i\phi}$	$\theta \in [0, \pi]$
			$\phi \in [0, 2\pi]$
v_-	$J_-/N = (\sum_{j=1}^{N} \sigma_j^-)/N$	$\nu = r\sin\theta\, e^{-i\phi}$	$(\theta, \phi) \in S^2$

for $r \in [0, r_m]$, where r_m is the value of $r \in [0, \frac{1}{2}]$, where this function assumes its minimum value.

Examples: Multinomials h which yield Hamiltonians $H = Nh_Q$ previously studied include:

a. $h_D = u_3 + \epsilon v_3 + \lambda(u_+ v_- + u_- v_+)$, (Refs. 1, 2)

b. $h_{CR} = h_D + \lambda'(u_+ v_+ + u_- v_-)$, (Refs. 6, 7)

c. $h_{CR+A^2} = h_{CR} + \kappa(u_+ + u_-)^2$, (Refs. 3, 8)

d. $h = h_D + Q(\frac{1}{4} - v_3^2) + K(\frac{1}{2} + v_3)^2$, (Ref. 9)

e. $h = u_3 + f[\epsilon v_3 + \lambda(u_+ v_- + u_- v_+)]$, (Refs. 4, 10)

f. $h = u_3 + \epsilon v_3 + g[\lambda(u_+ v_- + u_- v_+)]$. (Ref. 10)

Notation: Through we set $\nabla_3^u h \equiv \partial h / \partial u_3$, $\nabla_+^v \equiv \partial h / \partial v_+$, etc. We also set $k_B = 1$.

3. FREE ENERGY

Lower and upper bounds on the free energy F/N determined from

$$\exp[-\beta N(F/N)] = \mathrm{Tr}_F \mathrm{Tr}_A \exp(-\beta N h_Q) \tag{3.1}$$

can be computed using the method introduced by Hepp and Lieb.[11,12] The Hilbert space $A = (\mathbb{C}^2)^{\otimes N}$ for the atomic subsystem is first decomposed into its $(2J+1)$-dimensional SU(2) irreducible invariant subspaces. Within each J-invariant subspace upper and lower bounds on the restricted partition function are obtained by replacing the irreducible spherical tensor operators $\mathcal{J}_M^L(\mathbf{J})$ in h_Q by their P and Q representatives in the atomic coherent state representation,[11,13] and replacing the trace in the $(2J+1)$-dimensional space by an integral over the atomic coherent state[14,15] parameters $(\theta, \phi) \in S^2$:

$$\mathrm{Tr}_J(\cdot) \to \frac{2J+1}{4\pi} \int d\Omega(\cdot) \tag{3.2a}$$

Here $d\Omega = \sin\theta \, d\theta \, d\phi$ is the invariant measure on the Bloch sphere.[14] For multinomial h (assumption 1, Sec. 2) the atomic coherent state P and Q representatives of h_Q are equal to $\mathcal{O}(1/N)$, and are obtained simply by making the c-number substitutions for $\mathbf{v}$ shown in Table I.

The summation over the different J-invariant subspaces, including multiplicity $Y(N,J) = N!\,(2J+1)[(N/2+J+1)!\,(N/2-J)!\,]^{-1}$ is estimated by the integral[12] over $r = J/N$:

$$\sum_{J=0 \text{ or } 1/2}^{N/2} Y(N,J)(\cdot) \to N \int_0^{1/2} dr \exp[Ns(r)](\cdot). \tag{3.2b}$$

The asymptotic logarithmic behavior for $Y(N,J)$ is the same as that of the binomial coefficient $\binom{N}{N/2+J}$ and is

$$s(r) = S(r)/N = -[(\tfrac{1}{2}+r)\ln(\tfrac{1}{2}+r) + (\tfrac{1}{2}-r)\ln(\tfrac{1}{2}-r)]. \tag{3.3}$$

The remaining trace over Fock space can be estimated using the method of Wang and Hioe.[6] In this method, the operator h_Q is replaced by its P or Q representative in the field coherent state[16] representation, and the Fock space trace is replaced by an integral over the field coherent state parameter α:

$$\mathrm{Tr}_F(\cdot) = \int \frac{d^2\alpha}{\pi}(\cdot) = \frac{N}{\pi}\int d^2\mu(\cdot). \tag{3.2c}$$

Here $d^2\alpha = d(\mathrm{Re}\alpha)d(\mathrm{Im}\alpha)$ is the invariant measure on the harmonic oscillator phase plane. For multinomial h, the field coherent state P and Q representatives of h_Q are equal to $\mathcal{O}(1/N)$, and are obtained simply by making the c-number substitutions for $\mathbf{u}$ shown in Table I.

For multinomial h, the Q and P representatives of h_Q in the direct product atomic-coherent-state–field-coherent-state representations are equal to $\mathcal{O}(1/N)$ and are simply given by h_C. From (3.1), (3.2), and (3.3), the free energy F/N is given by a five-dimensional integral with integrand $N^3 \exp(-\beta N\Phi)(\mu, r\theta\phi; \beta)$. The asymptotic form of such an integral is easily computed by Laplace's method[17]

$$F/N = \min_{\mu, r\theta\phi} \Phi(\mu, r\theta\phi; \beta) + \mathcal{O}(\ln N/N), \tag{3.4}$$

$$\Phi(\mu, r\theta\phi; \beta) = h_C - \beta^{-1} s(r) \tag{3.5a}$$

$$= \{E(\mu, r\theta\phi) - TS(r)\}/N. \tag{3.5b}$$

Since h is a multinomial and $(r\theta\phi)$ belong to a compact domain, h_C is bounded below (asuumption 3, sec. 2). Therefore, the minimum in (3.4) exists, is well defined, and is greater than $-\infty$ for any finite temperature T, $0 < T < \infty$.

When computed at the values of the parameters μ, $r\theta\phi$ which yield the minimum value in (3.4) for fixed β,

$$\begin{aligned}
&\lim_{N\to\infty}[F/N - \Phi(\mu, r\theta\phi; \beta)] = 0,\\
&\lim_{N\to\infty}[\langle h_Q\rangle - E(\mu, r\theta\phi; \beta)/N] = 0,\\
&\lim_{N\to\infty}[\langle S/N\rangle - s(r)] = 0,\\
&\lim_{N\to\infty}(\langle a/\sqrt{N}\rangle - \mu) = 0,\\
&\lim_{N\to\infty}(\langle \sigma_j^z\rangle - \nu) = 0.
\end{aligned} \tag{3.6}$$

Thus, on the minimal branch Φ, h_C, and $s(r)$ are the free energy, the energy, and the entropy per particle [cf. (3.5b)], μ is the intensive order parameter[18] for the field, and ν the intensive order parameter for the atomic system.

Remark 1: The function h_C is the classical representative of the operator h_Q under the quantum-classical correspondence effected by atomic and field coherent state representations. From (3.5) and (3.6) it is clear that the construction of the free energy F/N described above amounts simply to making the quantum–classical mapping $h_Q \to h_C$, and adding as entropy term the logarithm of the SU(2) multiplicity factor $Y(N,J)$.

Remark 2: To simplify the computation of the minimum in (3.4), either the field parameters μ, μ^* or the atomic parameters $(r\theta\phi)$ may be eliminated. When the field parameters are eliminated the resulting expression reduces to that obtained by Hepp and Lieb[12] using atomic coherent states[14] alone. When the atomic parameters are eliminated the resulting expression reduces to that obtained by Wang and Hioe[6] using field coherent states[16] alone.

Example: For the Dicke Hamiltonian (example a, Sec. 2)

$$\Phi(\mu, r\theta\phi; \beta) = \mu^*\mu + \epsilon r\cos\theta + \lambda r\sin\theta(\mu^*\exp(-i\phi) + \mu\exp(i\phi)) - \beta^{-1}s(r). \quad (3.7a)$$

Eliminating μ, μ^* results in [cf. Ref. 12, Eq. (2.12)]

$$\Phi_A(r\theta\phi, \beta) = \epsilon r\cos\theta - |\lambda r\sin\theta|^2 - \beta^{-1}s(r). \quad (3.7b)$$

Eliminating $(r\theta\phi)$ results in [cf. Ref. 6, Eq. (27)]

$$\Phi_F(\mu; \beta) = \mu^*\mu - \beta^{-1}\ln 2\cosh\beta\{(\epsilon/2)^2 + |\lambda\mu|^2\}^{1/2}. \quad (3.7c)$$

4. PHASE TRANSITIONS

The potential $\Phi(\mu, r\theta\phi; \beta)$ is a function of the five real state parameters μ, $(r\theta\phi)$ and the single control parameter β. At any temperature β^{-1} we define the critical points of Φ as the set of points in the five-dimensional state parameter space where the differential of Φ is zero. At fixed temperature the critical points are either isolated or form isolated submanifolds on which a gauge symmetry group acts transitively. Each isolated solution set $\mu(\beta)$, $\nu(\beta) = r(\beta)\sin\theta(\beta)\exp[-i\phi(\beta)]$ is called a critical branch.[8]

The critical points at which the Hessian[19] of Φ is nonsingular are called normal, nondegenerate, or Morse critical points.[19] Non-Morse critical points are isolated along any branch consisting of an isolated critical point at fixed β. This is also true, with minor modifications, for critical branches consisting of an isolated submanifold.

According to Thom's theorem,[20] for a potential depending on an arbitrary number of state parameters but only one control parameter, the generic non-Morse critical point is the A_2 or "fold catastrophe."[21] At such a critical point two critical branches coalesce and disappear as a function of decreasing (or increasing) control parameter. From another point of view, one critical branch "folds over" at the A_2 critical point to become, in its continuation, the other critical branch.[22,23]

However, if the potential has a symmetry which removes odd terms in its Taylor series expansion about some critical branch, the A_2 catastrophe is suppressed and the generic catastrophe along that particular branch is the $A_{3\pm}$ or "cusp catastrophe." At such a non-Morse critical point, a new critical branch bifurcates from the original branch, which suffers a changeover in its stability type as the control parameter passes through its critical value β_c. If the original branch is stable for $T > T_c$, it is unstable for $T < T_c$, and the bifurcating branch is stable or unstable depending on whether the catastrophe is A_{3+} or A_{3-} [cf. (6.8) and (6.9)]. The A_{3+} catastrophe is isomorphic to the Ginzberg–Landau potential.[24]

A phase transition occurs at β_c if the global minimum of Φ jumps from one critical branch $(\mu, r\theta\phi; \beta)_i$ to another $(\mu, r\theta\phi; \beta)_j$ at β_c. The transition is first order if

$$\lim_{\beta\to\beta_c}(\mu, r\theta\phi; \beta)_i \neq \lim_{\beta\to\beta_c}(\mu, r\theta\phi; \beta)_j; \quad (4.1)$$

otherwise it is second order. Second order phase transitions occur when an A_{3+} catastrophe occurs[8] on a critical branch which is globally stable for $T > T_c$.

From the invariance of h_C under the symmetry $\mu \to -\mu$, $\nu \to -\nu$ (assumption 4, Sec. 2) we see that $\mu = 0$, $\nu = 0$ ($\theta = 0$ and $\theta = \pi$) are critical points for all β. Further, the symmetry forbids the A_2 catastrophe along both these branches. Therefore, second order phase transitions are possible from either of these branches, but from no other critical branch, since no other symmetries are assumed for h.

In the high temperature limit the entropy term in (3.5) becomes more important than the energy term. As a result, the minimum of Φ is given by $r \to 0$ as $\beta \to 0$. To determine the values of the parameters $\mu, r\theta\phi$ which minimize Φ in this limit, we expand Φ in increasing powers of r about $r = 0$. The leading term, not containing r, is $h(\mathbf{u}, \mathbf{v})$, $\mathbf{v} = \mathbf{0}$. This is minimized at $\mu = 0$ (assumption 5, Sec. 2). The term linear in r can contain terms linear in v_3, but the $v_\pm$ terms must be multiplied by odd powers of $u_\pm$. As a result the term linear in r is uniquely $\nabla_3^v h(0, 0) r\cos\theta$. This is minimized at $\theta = \pi$ since $\nabla_3^v h(0, 0) \neq 0$ (assumption 6, sec. 2). The critical branch which minimizes Φ at high temperature is $\mu = 0$, $\theta = \pi$, ϕ arbitrary. Along this critical branch the value of $r(\beta)$ is obtained by minimizing Φ according to

$$0 = -\nabla_3^v h + \frac{1}{\beta}\ln\left(\frac{1+2r}{1-2r}\right). \quad (4.2)$$

The derivative is evaluated at $\mathbf{u} = \mathbf{0}$, $\mathbf{v} = (-r, 0, 0)$.

The model Hamiltonian (2.1) exhibits a phase transition if the global minimum of Φ leaves the thermal branch. In general, nonlocal jumps from one branch to another must be investigated for each model Hamiltonian separately. However, transition from the thermal branch to a bifurcating branch is a local phenomenon. It is sufficient to look for a changeover in stability along the thermal branch.

Stability of the thermal branch is investigated by expanding Φ in powers of μ, μ^* and δ, δ^*, where $\delta = (\theta - \pi)\exp(-i\phi)$. In terms of the matrices $M^\dagger = (\mu^*, \mu)$, $D^\dagger = (\delta^*, \delta)$ and their adjoints, the second order term in this expansion is

$$\tfrac{1}{2}M^\dagger I_2\nabla_3^u h M + \tfrac{1}{4}D^\dagger I_2 r\nabla_3^v h D + \tfrac{1}{2}M^\dagger A M - M^\dagger r B D + \tfrac{1}{2}D^\dagger r^2 C D. \quad (4.3)$$

Here A, B, C are 2×2 matrices composed of second derivatives of h with respect to $\mathbf{u}$, $\mathbf{v}$, e.g.,

$$B = \begin{bmatrix} \nabla_+^u\nabla^v h & \nabla_+^u\nabla_+^v h \\ \nabla_-^u\nabla_-^v h & \nabla_-^u\nabla_+^v h \end{bmatrix} = \begin{bmatrix} \nabla_+^v \\ \nabla_-^v \end{bmatrix}\begin{bmatrix} \nabla_+^v & ^\dagger \\ \nabla_-^v \end{bmatrix} h. \quad (4.4)$$

All derivatives are evaluated at β on the thermal branch: $\mu(\beta) = 0$, $\theta(\beta) = \pi$, $\phi(\beta) =$ undetermined, $r(\beta)$ determined by (4.2).

Eliminating either μ, μ^* or δ, δ^* from (4.3) leads to the following expressions for the quadratic term (4.3):

$$\tfrac{1}{2}rD^\dagger[\tfrac{1}{2}\nabla_3^v h D h I_2 + rC - rB^\dagger(\nabla_3^u h I_2 + A)^{-1}B]D \quad (4.5a)$$

or

$$\tfrac{1}{2}M^\dagger[\nabla_3^u h I_2 + A - rB^\dagger(\tfrac{1}{2}\nabla_3^v h I_2 + rC)^{-1}B]M. \quad (4.5b)$$

Stability on the thermal branch is determined by investigating the eigenvalues $\gamma_1(r)$, $\gamma_2(r)$ of either of the matrices in [ ]. Both eigenvalues are positive for $r \to 0^+$ and each has a finite number of zeroes. Let r_c be the smallest positive zero of either of the eigenvalues $\gamma_i(r)$. Then there can be a second order phase transition if $r_c \in (0, r_m)$. The parameter r actually assumes value r_c (assumption 7, Sec. 2). Without this assumption, r can "jump past" r_c along the thermal branch, and by so doing, bypass the non-Morse critical point, the bifurcation, and the second order phase transition altogether. In fact, assumption 7 can be relaxed simply to the requirement that r pass through value r_c at some critical temperature β_c.

The gap equation for the critical temperature β_c is simply obtained by setting $r = r_c$ in (4.2). The Hepp–Lieb gap equation immediately results:

$$r_c = \tfrac{1}{2}\tanh\tfrac{1}{2}\beta_c \nabla_3^v h, \quad \mathbf{u}=0, \quad \mathbf{v}=(-r_c, 0, 0). \tag{4.6}$$

Summary: From the preceeding analysis it is possible to formulate necessary and sufficient conditions for the existence of a second order phase transition in models (2.1) of the Dicke type obeying assumptions 1–7, Sec. 2.

1. The first positive zero for either determinant in (4.5) occurs for $r_c \in (0, r_m)$, where r_m is defined in assumption 7, Sec. 2;

2. A first order phase transition does not occur at a temperature $T > T_c$, where T_c is defined by (4.6).

Example: For the Hamiltonian of example c (Sec. 2):

$$\nabla_3^u = 1, \quad \nabla_3^v = \epsilon, \quad A = 2\kappa(I_2 + \sigma_x),$$
$$B = \lambda I_2 + \lambda'\sigma_x, \quad C = 0,$$
$$\gamma_1(r) = \tfrac{1}{2}\epsilon - r(\lambda+\lambda')^2/(1+4\kappa),$$
$$\gamma_2(r) = \tfrac{1}{2}\epsilon - r(\lambda-\lambda')^2.$$

This Hamiltonian has been studied extensively.[3,8] For $\lambda' = \lambda$, the only zero occurs for $2r_1 = \epsilon(1+4\kappa)/(2\lambda)^2$. The Thomas–Reich–Kuhn sum rule requires $(2\lambda)^2/\epsilon(1+4\kappa) < 1$.[3] Since r_1 has no zero in the interval $[0, \frac{1}{2}]$, there is no second order phase transition. For $\lambda' = 0$, γ_2 has a zero at $2r_2 = \epsilon/\lambda^2$. Thus, for $\lambda' = 0$ a second order phase transition is possible at a critical temperature determined from $1 = (\lambda^2/\epsilon)\tanh\frac{1}{2}\beta_c\epsilon$, contrary to the claims made in Ref. 3.

5. CANONICAL KERNEL

If there is a second order phase transition, then the only terms of h which are important for determining the critical temperature are those appearing in (4.5). Since all derivatives in (4.5) are evaluated at $\mathbf{u}=0$, $\mathbf{v}=(-r, 0, 0)$, it is sufficient, to determine the critical temperature, to reduce an arbitrary multinomial h to its canonical kernel

$$h^{CK} = u_3 f_1(v_3) + f_2(v_3) + \tfrac{1}{2}U^\dagger A U + U^\dagger B V + \tfrac{1}{2}V^\dagger C V. \tag{5.1}$$

Here $U = \mathrm{col}(u_-, u_+)$, $U^\dagger = (u_+, u_-)$, A, B, C are 2×2 matrix functions of v_3, and f_1, f_2 are functions of the single variable v_3. Then the value of r for which there is a changeover of stability on the thermal branch is determined from

$$\det[\tfrac{1}{2}f_2' I_2 + rC - rB^\dagger(f_1 I_2 + A)^{-1}B] = 0. \tag{5.2}$$

The functions and matrices in (5.2) are to be evaluated at $v_3 = -r$. The critical temperature is determined from (4.6).

Example: For example d (Sec. 2), $f_1 = 1$, $f_2 = \epsilon v_3 + Q(\frac{1}{4} - v_3^2) + K(\frac{1}{2} + v_3)^2$, $A = C = 0$, $B = \lambda I_2$, and (5.2) reduces to

$$\det[\tfrac{1}{2}(\epsilon + K + 2r[Q-K])I_2 - r\lambda^2 I_2] = 0,$$
$$r_m = \min\left[\frac{1}{2}, \frac{\epsilon+K}{2(K-Q)}\right], \quad \epsilon + K > 0, \quad K - Q \geq 0,$$
$$r_c = \frac{\frac{1}{2}(\epsilon+K)}{\lambda^2 + (K-Q)}.$$

There is a phase transition if $r_c < 1/2$.[9]

Example: For example e (Sec. 2), the canonical kernel is

$$h^{CK} = u_3 + f(\epsilon v_3) + \lambda f'(\epsilon v_3) U^\dagger I_2 V.$$

In the special case $f(x) = x + \alpha x^2$ $(\alpha > 0)$

$$r_m = \min\left(\frac{1}{2\alpha\epsilon}, \frac{1}{2}\right).$$

The smallest positive root of (5.2) is

$$r_c = \frac{1}{4\alpha\epsilon}\left[1 - \left(1 - \frac{4\alpha\epsilon^2}{\lambda^2}\right)^{1/2}\right].$$

There is a second order phase transition if[10]

$$4(\alpha\epsilon)(\epsilon/\lambda^2) < 1 \quad \text{and} \quad \alpha\epsilon \geq 1/2,$$
$$(\alpha\epsilon) + (\epsilon/\lambda^2) < 1 \quad \text{and} \quad \alpha\epsilon \leq 1/2.$$

Example: For example f (Sec. 2), the canonical kernel is

$$h^{CK} = u_3 + \epsilon v_3 + \lambda g'(0) U^\dagger I_2 V.$$

At the critical temperature, this model is equivalent to a Dicke model (example a, Sec. 2) with coupling constant $\lambda g'(0)$.

Remark: The canonical kernel h_Q^{CK} of an arbitrary h_Q is the inverse operator image, under the quantum–classical correspondence, of the 2-jet[25] of h_c around $\mu = 0$, $\nu = 0$.

6. DISCUSSION

We now discuss the three recent studies[3–5] of the structural stability of the phase transition under a change in the functional form of the Dicke Hamiltonian.

1. Rzażewski, Wódkiewicz, and Żakowicz[3] studied the Hamiltonian h_{CR+A}[2] (example c, Sec. 2) with $\lambda = \lambda'$. On the basis of physical arguments involving the Thomas–Reiche–Kuhn sum rule, they showed that, for this particular model a phase transition did not occur. They also claimed their results were valid for $\lambda' = 0$. This claim is incorrect,[8,26] as was shown in the Example of Sec. 4. Since resonant and nonresonant atom–field interaction terms are expected to couple to the thermal bath variables with different time constants or even through different relaxation mechanisms, it is expected that $\lambda' \ll \lambda$ for the description of any two-level system in thermal equilibrium with a single field mode.

2. Gambardella[4] studied a class of nonlinear extensions of the Dicke model obtained from the function $h(\mathbf{u},\mathbf{v})=u_3+[f(\mathbf{u},N\mathbf{v})/N]$. The lack of a phase transition, except in the case of the Dicke model itself, can be traced[10] directly to the inhomogeneous structure of the nonlinear extensions which were considered.

3. Provost *et al.*[5] studied the model obtained from

$$h=u_3+\epsilon v_3+\lambda(u_+v_-+u_-v_+)+c_+u_-+c_-u_+. \tag{6.1}$$

In this model the terms c_+u_- and c_-u_+ represent interactions between the field mode and external classical currents. With these extra terms (6.1) no longer obeys the symmetry assumption 4 (Sec. 2). Therefore, the generic non-Morse critical point is the A_2 catastrophe, and no second order phase transition is possible. Provost *et al.*[5] showed that there is no phase transition if $c_\pm\neq 0$.

A second order phase transition is possible when the symmetry breaking terms are added to a model Hamiltonian without the gauge invariance of the Dicke model, e.g., Example b, Sec. 2. However, the bifurcation is nongeneric[20,22] and can occur only under very special circumstances. Writing $u_\pm=(u_x\pm iu_y)/\sqrt{2}$ and similarly for $v_\pm$, $c_\pm$, the function h can be decomposed as follows

$$\begin{aligned} h&=u_3+\epsilon v_3+p_x+p_y,\\ p_x&=(\lambda+\lambda')u_xv_x+c_xu_x,\\ p_y&=(\lambda-\lambda')u_yv_y+c_yu_y. \end{aligned} \tag{6.2}$$

If $c_x=0$, h retains a reflection invariance under

$$(u_3,u_x,u_y;v_3,v_x,v_y)\to(+u_3,-u_x,+u_y;+v_3,-v_x,+v_y). \tag{6.3}$$

If, in the absence of currents, it is possible for the branch $u_x\neq 0$, $v_x\neq 0$ to bifurcate from the thermal branch at a higher temperature than the branch $u_y\neq 0$, $v_y\neq 0$, the addition of the particular (nongeneric) current with $c_x=0$ will not destroy this bifurcation.[8,26]

The effect of including linear terms of the form $c_+u_-+c_-u_+$ on the bifurcation diagram is summarized in Fig. 5.7 of Ref. 27. Thompson and Hunt[28] discuss other consequences of the structural instability of the bifurcation in the cusp catastrophe in the presence of linear symmetry breaking terms.

In addition to the three extended Dicke models which study the structural stability of the phase transition, three other model Hamiltonians have been studied which do not satisfy all assumptions 1–7 (Sec. 2) but which nevertheless possess second order phase transitions. Each of these three models involves multiple boson processes.

4. Hamiltonians describing double photon emission and absorption processes are obtained from the functions

$$h_2=u_3+\epsilon v_3+\lambda(u_+^2v_-+u_-^2v_+), \tag{6.4a}$$

$$h_{1+1}=\omega_1u_3+\omega_2t_3+\epsilon v_3+\lambda(u_+t_+v_-+u_-t_-v_+). \tag{6.4b}$$

The Hamiltonian obtained from h_2 describes absorption and emission of photons into the same mode; h_{1+1} describes processes involving two different modes. For both functions in (6.4), the canonical kernel is obtained by setting $\lambda=0$. The absence of interaction terms indicates the absence also of phase transitions. Nevertheless, both models can exhibit[18] second order phase transitions for λ sufficiently large. This comes about because the Fock space trace leading to the partition function is restricted. For (6.4a), the Fock space trace involves only states with an even or an odd number of photons. In the case (6.4b), the Fock space trace includes only those states in which the difference in the photon number in modes t, u has a constant integer value. The methods for computing the free energy presented in Sec. 3 are applicable only when the Fock space trace is unrestricted.

5. Thompson[29] has proposed the following model to describe phonon aided photon processes:

$$h=u_3+\Omega t_3+\epsilon v_3+\lambda(u_+v_-+u_-v_+)(1+\kappa t_++\kappa^*t_-). \tag{6.5}$$

Here $\mathbf{u}$, $\mathbf{v}$ refer to the photon mode and the atomic system, as usual, and $\mathbf{t}$ refers to the phonon mode. The function h is not invariant under the expected symmetry group:

$$\begin{aligned} &u_3,v_3,t_3\to+u_3,+v_3,+t_3,\\ &u_\pm,v_\pm,t_\pm\to-u_\pm,-v_\pm,-t_\pm. \end{aligned} \tag{6.6}$$

Instead, h and Φ are only invariant if $t_\pm\to+t_\pm$. Thus the order parameter $t_\pm$ is an even function of the order parameters $u_\pm$ and $v_\pm$. The canonical kernel of (6.5) is obtained by setting $\kappa=0$. As a result, there can be a second order phase transition at the usual Dicke temperature.

The potential Φ can easily be computed in terms of μ, μ^*, τ, τ^* and $(r\theta\phi)$. Eliminating the photon and phonon parameters leads to a simple expression for the reduced potential

$$\Phi_A(r\theta\phi;\beta)=\epsilon r\cos\theta-\frac{(\lambda r\sin\theta)^2}{1-\Omega^{-1}(2\kappa\lambda r\sin\theta)^2}-\beta^{-1}s(r). \tag{6.7}$$

When this is expanded about the thermal branch $\theta=\pi$, the first three terms in the Taylor expansion are ($\delta=\theta-\pi$)

$$\Phi_A(r\theta\phi;\beta)=-\epsilon r-\beta^{-1}s(r)+\left(\frac{\epsilon r}{2}-\lambda^2r^2\right)\delta^2+\left(-\frac{\epsilon r}{4!}+\frac{\lambda^2r^2}{3}-\frac{(2\kappa\lambda^2r^2)^2}{\Omega}\right)\delta^4. \tag{6.8}$$

At the bifurcation point $r=\epsilon/2\lambda^2<1$ and the coefficient of δ^4 is

$$\left(\frac{\epsilon}{4\lambda}\right)^2\left(1-\frac{(2\kappa\epsilon/\lambda)^2}{\Omega}\right). \tag{6.9}$$

Therefore, the non-Morse critical point is of type A_{3+} if $(2\kappa\epsilon/\lambda)^2<\Omega$ and the phase transition is of second order. If $(2\kappa\epsilon/\lambda)^2>\Omega$, the catastrophe is of type A_{3-} and there will be a first order phase transition for $r_1<\epsilon/2\lambda^2$ and $T_1>T_c$. In fact, it is possible for a first order phase transition to exist even if there is no bifurcation provided the minimum of $\Phi_A(r\theta\phi;\beta)$ occurs for $\theta\neq\pi$ at low temperatures.

7. CONCLUSIONS

We have presented a simple method for writing down the free energy F/N for a large class of Dicke-like models:

1. replace h_Q by its classical image h_C;

2. include the entropy term $-Ts(r)$;

3. minimize the resulting potential Φ over the atomic and field coherent state parameter space.

The critical properties associated with h_Q are then easily determined from the potential Φ, constructed as described above. First order phase transitions occur when there is a nonlocal transfer of stability from one critical branch of Φ to another. The phase transition is second order if the transfer is local. A necessary condition for the occurrence of a second order phase transition is given in (4.5). The sufficient condition is that a first order phase transition does not occur at a higher temperature. The critical temperature for any Hamiltonian h_Q exhibiting a second order phase transition is determined entirely from its canonical kernel h_Q^{CK}. The gap equation is given in (4.6).

The second order phase transition is remarkably structurally stable against changes in the functional form of the model Hamiltonian, provided a certain symmetry (assumption 4, Sec. 2) is conserved. When the symmetry is absent, the phase transition is not structurally stable and generically disappears.

The importance of the symmetry is not that it is broken at the phase transition (the potential Φ always retains the symmetry while the free energy F/N does not), but that it suppresses the A_2 catastrophe and allows the Ginzburg–Landau A_{3+} catastrophe to occur.

*Work supported by the U.S. Army Research Office, Durham, North Carolina, under Grant No. DAHC04-72-A-0001.
†Present address.

[1]R.H. Dicke, Phys. Rev. **93**, 99 (1954).
[2]K. Hepp and E.H. Lieb, Ann. Phys. (N.Y.) **76**, 360 (1973).
[3]K. Rzążewski, W. Wódkiewicz, and W. Żakowicz, Phys. Rev. Lett. **35**, 432 (1975).
[4]P.J. Gambardella, Physica A **80**, 550 (1975).
[5]J.P. Provost, F. Rocca, G. Vallee, and M. Sirugue (to be published).
[6]Y.K. Wang and F.T. Hioe, Phys. Rev. A **7**, 831 (1973).
[7]H.J. Carmichael, C.W. Gardiner, and D.F. Walls, Phys. Lett. A **46**, 47 (1973).
[8]R. Gilmore and C.M. Bowden, J. Math. Phys. **17**, 1617 (1976).
[9]Y.A. Kudenko, A.P. Slivinsky, and G.M. Zaslavsky, Phys. Lett. A **50**, 411 (1975).
[10]R. Gilmore, to be published.
[11]E.H. Lieb, Commun. Math. Phys. **31**, 327 (1973).
[12]K. Hepp and E.H. Lieb, Phys. Rev. **A 8**, 2517 (1973).
[13]R. Gilmore, J. Phys. A 9, L65 (1976).
[14]F.T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. **A 6**, 2211 (1972).
[15]J.M. Radcliffe, J. Phys. **A 4**, 313 (1971).
[16]R.J. Glauber, Phys. Rev. **130**, 2529 (1963); **131**, 2766 (1963).
[17]D.V. Widder, *An Introduction to Transform Theory* (Academic, New York, 1971).
[18]R. Gilmore and C.M. Bowden, Phys. Rev. **A 13**, 1898 (1976).
[19]J. Milnor, *Morse Theory*, Ann. Math. Stud. (Princeton U.P., Princeton, N.J., 1963).
[20]R. Thom, *Structural Stability and Morphogenesis* (Benjamin, New York, 1975).
[21]V.I. Arnol'd, Funct. Anal. Appl. **6**, 254 (1972).
[22]T. Poston and I. Stewart, "Thom's Classification Theorem (An Intuitive Approach)," University of Warwick preprint.
[23]E.C. Zeeman, Scientific American **234**(4), 65 (1976).
[24]L. Landau and E. Lifchitz, *Physique statisque* (Moscow, Editions Mir, 1967).
[25]M. Golubitsky and V. Guillemin, *Stable Mappings and Their Singularities* (Springer, New York, 1973).
[26]R. Gilmore, Phys. Lett. A **55**, 459 (1976).
[27]V.I. Arnol'd, Russian Math. Surveys **27**(5), 54 (1972).
[28]J.M.T. Thompson and G.W. Hunt, *A General Theory of Elasticstability* (Wiley, New York, 1973).
[29]B.V. Thompson, J. Phys. A 8, 126 (1975). The A_2 and A_{3-} catastrophes are clearly shown in Figs. 2b and 2e. The global stability changeover is shown in Fig. 2d.

Metastability and the analytic continuation of eigenvalues*

C. M. Newman†

Physics and Mathematics Departments, Technion, Haifa, Israel

L. S. Schulman

Physics Department, Technion, Haifa, Israel
and Physics Department, Indiana University, Bloomington, Indiana 47401
(Received 12 July 1976; revised manuscript received 3 September 1976)

A metastable analytic continuation of the Ising model free energy is conjectured to follow from certain analyticity properties of the eigenvalues of the transfer matrix. The resulting description of metastability is applicable to any system whose phase transition is associated with eigenvalue degeneracy. Motivation for the conjectures concerning the Ising model is provided by the study of eigenvalue continuation in several simpler systems.

1. INTRODUCTION

For finite systems, the partition function gives no evidence of metastability. Even if it should happen that the free energy, in the thermodynamic limit, can be analytically continued to a metastable region, such multivaluedness cannot appear in the finite system partition function which is entire. We accordingly replace the usual thermodynamic functions derived from the partition function with other functions which have the same thermodynamic limit but which already for finite systems have nontrivial multivalued branch and sheet structure. The analysis in this paper of the analytic structure of these functions is intended both to facilitate an eventual rigorous proof of the existence of such an analytic continuation in nontrivial models and also to provide some new motivation for the physical interpretation of the continuation as metastability.

Phase transitions, long range order and eigenvalue degeneracy have long been associated with one another.[1,2] The approach to metastability proposed here applies to those models in which the free energy can be obtained, in the thermodynamic limit, from the minimum eigenvalue of some linear operator and in which a phase transition corresponds to the asymptotic degeneracy of that eigenvalue; examples include Ising models and quantum field theories with broken symmetry. The metastable phase is found in the finite system, according to our view, by allowing certain thermodynamic variables to be complex and then analytically continuing a particular eigenvalue from a region where it is the minimum eigenvalue to a region where that is no longer the case. With the appropriate analytic continuation, the eigenvalue thus obtained determines the metastable free energy.

In Sec. 2 we explain our proposal in more detail by giving a series of conjectures for the Ising model. These conjectures are motivated by properties of simpler models which we present in Sec. 3. The relevance of these simpler models to the Ising model is discussed in Sec. 4, and finally a general discussion of metastability is given in Sec. 5.

2. ISING MODEL

We consider a two dimensional nearest neighbor Ising model with energy $-J\sum'\sigma_i\sigma_j - h\sum\sigma_i$ (where $\sum'$ denotes the sum over nearest neighbors and $J>0$ is fixed), at temperature T. For an $m\times n$ lattice with periodic boundary conditions, the self-adjoint transfer matrix $L_n(h)$ defined in the standard way,[3] is related to the partition function $Z_{m,n}(h)$ by $Z_{m,n}(h)=\mathrm{Tr}\{[L_n(h)]^m\}$ and the limiting free energy is

$$\frac{f^0(h)}{kT} = -\lim_{m,n\to\infty}\frac{1}{mn}\log Z_{m,n}(h) = -\lim_{n\to\infty}\frac{1}{n}\log\lambda_n^0(h),$$

where $\lambda_n^0(h) > \lambda_n^1(h) \geq \lambda_n^2(h) \geq \cdots \geq \lambda_n^{2^n}(h) \geq 0$ denote the eigenvalues of $L_n(h)$ listed in decreasing magnitude (for real h); we define $E_n^j(h) = -(1/n)\log\lambda_n^j(h)$.

It follows from the fact[4] that the zeros of $Z_{m,n}(h)$ all lie on the imaginary h axis that $E_n^0(h)$ and $f^0(h)$ are analytic in the right (or left) half-plane of complex h. The eigenvector associated with $E_n^0(h)$ is positively magnetized for $h>0$; our first conjecture concerns the analytic continuation of E_n^0 into the left half-plane in such a way that the positive magnetization is preserved when T is below the critical temperature T_c. Since we wish the continuation of E_n^0 to cease being the minimum eigenvalue of $-\log(L_n(h))/n$, it is necessary that the path of continuation avoid the neighborhood of certain real values of h, such as $h=0$.

Conjecture 1: For $T<T_c$, there are positive constants $\alpha, \eta > 0$ and $\gamma_n \to 0$ such that E_n^0 can be analytically continued to a function, $E_n^*(h)$, analytic and single valued on

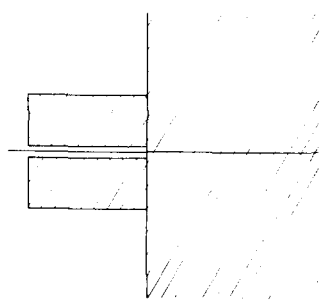


FIG. 1. Region $D_{(n)}$ (shaded) in which $E_n^0(h)$ can be analytically continued according to Conjecture 1.

$D_n = \{h: \mathrm{Re}h > 0\} \cup \{h: \mathrm{Re}h > -\alpha, \gamma_n < |\mathrm{Im}h| < \eta\}$ (see Fig. 1).

$E_n(h)$ converges uniformly on any fixed D_m as $\eta \to \infty$ to an analytic function $f^+(h)$; $f^+(h) = f^0(h)$ for $\mathrm{Re}h > 0$.

The restriction to $|\mathrm{Im}h| < \eta$ in D_n is due to the fact that $L_n(h + 2\pi i/\beta) = L_n(h)$ (where $\beta = 1/kT$) so that any singularities occurring near the negative real axis also appear near the lines, $\mathrm{Im}h = \pm 2\pi kT$. The restriction to $\mathrm{Re}h > -\alpha$ is based on the loss of metastability discussed below in Property 4. The restriction to $|\mathrm{Im}h| > \gamma_n$ is partially removed in the next conjecture which analyzes the singularity structure of $E_n^+(h)$ near the negative real axis.

Conjecture 2: There exist positive ρ_n's and $0 = x_1(n) > x_2(n) > \cdots > x_{M_n} > -\alpha$ such that $E_n^+(h)$ is analytic and single valued on

$\bar{D}_n = \{h: \mathrm{Re}h > 0\} \cup \{h: \mathrm{Re}h > -\alpha, |\mathrm{Im}h| < \eta, |h - x_k(n)| > \rho_n \ \forall k\}$; in addition, $\rho_n \to 0$ sufficiently rapidly that $\rho_n / \min_k\{|x_{k+1}(n) - x_k(n)|\} \to 0$. $E_n^+(h)$ can be further continued into $\{h: |h - x_k(n)| \le \rho_n\}$ for each k except for a pair of (square root) branch points at $x_k(n) \pm i y_k(n)$ ($|y_k(n)| \le \rho_n$) connected by a branch cut (see Fig. 2).

The intuitive picture behind Conjecture 2 is that, except for certain exceptional (real) values of h, the eigenvectors of $L_h(h)$, corresponding to E_n^k's within a finite (n-independent) distance from E_n^0, can be naturally divided into two groups according to the sign of their magnetization (order parameter). As h varies, one set of eigenvalues moves past the other much as two stacks of chips (of different thicknesses) can be moved vertically relative to each other. The exceptional values of h are those for which there would be exact degeneracy between one state from each group if there were no coupling between the groups. In the presence of coupling, no matter how small, the previously degenerate levels mix strongly resulting in two approximately degenerate actual eigenstates, each with a small value of the order parameter. Such an approximate (for real h) level crossing in an essentially two-level system gives rise to a pair of square root branch points at the location of actual level crossings in complex h. The size of the region in h around an exceptional value in which level crossing effects are important is determined by the matrix element of the relevant symmetry breaking operator between the two states or equivalently by the amount of "wrong" magnetization found in an eigenstate predominantly of the "right" magnetization. In order for the foregoing picture to be valid, it is necessary that the spacings in real h between (approximate) level crossings be much larger than the size of the level crossing regions themselves.

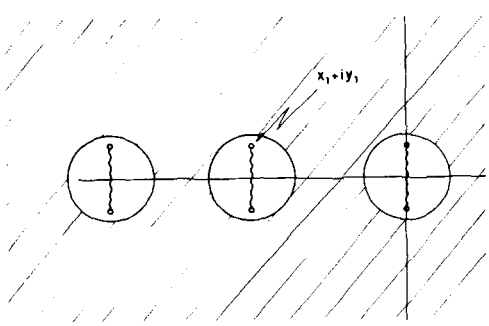


FIG. 2. $\bar{D}_n$ according to Conjecture 2 with branch cuts drawn to indicate where one can leave the metastable sheet. In particular the sheet reached through the cut at $h = 0$ takes one to the stable ground state for $h < 0$. The circles are $\{h: |h - x_k(n)| \le \rho_n\}$.

Conjecture 2 describes our prescription for following the bottom chip of one stack as it passes by successively higher chips of the second stack. The points $x_k(n)$ correspond to approximate level crossings and our method of analytic continuation is to make a short excursion into the complex plane each time the positively magnetized metastable eigenfunction is about to have one of these crossings with a state of negative magnetization. The conjecture that $\rho_n / \min_k\{|x_{k+1}(n) - x_k(n)|\} \to 0$ is the assumption that our picture of independent stacks of chips is correct for most real values of h (with $|h| < \alpha$).

Analytic continuation *through* one of the branch cuts of Fig. 2 would correspond to an adiabatic (infinitely slow) variation of h while our choice of analytic continuation *around* the branch cut is analogous to the "fast" approximation of quantum mechanics; it preserves the eigenvector rather than the eigenvalue label and thus $E_n^+(h)$ for real negative h away from the level crossing regions will equal $E_n^k(h)$ with k dependent on h. It consequently is natural to associate $E_n^+(h)$ with a "supercooled" or "superheated" metastable state which is formed and exists for times short in comparison to the relaxation time of the system so that the system tends to retain its former state (sign of magnetization) rather than to adiabatically move to the absolute minimum of free energy.

Before stating the next conjecture, we note that Conjecture 2 could be weakened to the extent of allowing more branch points (of perhaps higher than second order) in each level crossing region, $\{h: |h - x_k(n)| \le \rho_n\}$, without seriously affecting our picture of metastability. Of course, in that case, the two-level systems described above would be replaced with multilevel systems.

The relation of the next conjecture to our previous ones and its possible invalidity will be discussed in more detail in Sec. 5 below.

Conjecture 3: For any $-\alpha < x \le 0$, $\lim_{h \to x} f^+(h)$ exists, independent of the (complex) path along which $h \to x$; $f^+(h)$ is analytic and single valued on $\{h: \mathrm{Re}h > -\alpha, |\mathrm{Im}h| < \eta\}$ including $h = 0$.

We propose some further properties of $f^+(h)$ which, while not essential to our basic picture of metastability and less strongly implied by our later models, nevertheless may shed more light on the structure of metastability.

Property 4: The first singularity of $f^+(h)$ on the negative real axis corresponds to the asymptotic vanishing of n times the gap between $E_n^+(h)$ and the energy of the first excited metastable state [obtained by an analytic continuation of $E_n^1(h)$ from large positive h]. This singularity thus corresponds to infinite correlation length in the metastable state, a characteristic of second or-

der phase transitions. The location of this singularity as a function of T defines the spinodal line, $h_s(T)$, where metastability is lost; $h_s(T) \to 0$ as $T \to T_c$.

Property 5: The functions $f^+(h)$, defined for $h > h_s$ and the corresponding $f^-(h)$, defined for $h < |h_s|$ [by analytic continuation of $f^0(h)$ from $h < 0$] are branches of the same analytic function. The two branches can be connected by a (symmetric) analytic continuation path which winds around the spinodal singularities at $h = h_s$ (on the f^+ sheet) and $h = |h_s|$ (on the f^- sheet); on the "connecting" branch there are no real singularities for $h_s < h < |h_s|$.

Property 5 is suggested both by mean field theory models (see Sec. 3 below) and by the fact that, barring special factorization properties, all eigenvalues of the transfer matrix should be branches of a single analytic function of h. In the mean field models, there is a unique "connecting" branch which corresponds to a local maximum of the free energy; it seems that in the Ising case this uniqueness is lost.

3. MODELS

We present a series of models bearing varying degrees of resemblance to the Ising model and susceptible to varying degrees of explicit solution and rigorous analysis.

Each model consists of a sequence of (finite system) Hamiltonians, $H_n(h) = H_n(0) + hS$, with an infinite system (stable) free energy, $f^0(h) = \lim_{n\to\infty} E_n^0(h)$, where $E_n^0(h) < E_n^1(h) \leq E_n^2(h) \leq \cdots$ denote the eigenvalues of $H_n(h)$ (for real h). $H_n(h)$ is the analog of $-(1/n)\log[L_n(h)]$ for the Ising model with h the external magnetic field and S a symmetry breaking term. $Z_{m,n} = \mathrm{Tr}\{\exp[-mnH_n(h)]\}$ is the analog of the Ising partition function and $f^0 = \lim_{m,n\to\infty}(-1/mn)\log Z_{m,n}$. The first two models are further similar to the Ising model in that the zeros of $Z_{m,n}(h)$ are pure imaginary[5,6] so that $f^0(h)$ and $E_n^0(h)$ are analytic for $\mathrm{Re}h > 0$ (or $\mathrm{Re}h < 0$); they also share with the Ising model the feature that ground states for different temperatures (or different values of h) become asymptotically orthogonal as $n \to \infty$.

In each case an explicit formula for $f^0(h)$ implies that f^0 can be analytically continued from $h > 0$ past $h = 0$; the continuation $f^+(h)$ determines the metastable free energy for $h < 0$. Our main interest concerns the analyticity properties of the eigenvalues $E_n^k(h)$ and the validity of the analogs of the conjectures of Sec. 2 for these models.

A. The Lipkin model

Lipkin *et al.*[7-10] studied the "phase transition" associated with the onset of deformation in nucleii. Starting with a many-body problem they arrived at a Hamiltonian equivalent to the following:

$$H_n(h) = -(1/2n^2)J_z^2 - (1/n)TJ_x - (1/n)hJ_z, \tag{3.1}$$

where J_x, J_y, J_z are standard quantum mechanical spin operators acting on the $(2n+1)$-dimensional space of eigenstates of $J^2 = J_x^2 + J_y^2 + J_z^2$ with eigenvalue $n(n+1)$ and T is a parameter playing the role of temperature. For $T < 1$ there is a phase transition at $h = 0$, as $n \to \infty$, with a breaking of the symmetry described by the operator $\exp(i\pi J_y)$.

Using results of Lieb,[11] the limiting free energy can be written in terms of an integral over the surface of a sphere,

$$f^0(h) = \lim_{n\to\infty}(1/n^2)\log Z_{n,n}$$
$$= -\lim_{n\to\infty}(1/n^2)\log\int_0^{2\pi} d\phi \int_{-1}^{1} dz$$
$$\times \exp[n^2(z^2/2 + T(1-z^2)^{1/2}\cos\phi + hz)].$$

It follows that

$$f^0(h) = \inf_{-1\leq z\leq 1} V_h(z) \tag{3.2}$$

with

$$V_h(z) = -\tfrac{1}{2}z^2 - T(1-z^2)^{1/2} - hz. \tag{3.3}$$

For $T < 1$ and sufficiently small h, V_h has two local minima; the global minimum (for $h \neq 0$) is the one with $z/h > 0$. The analytic continuation of f^0 to f^+ clearly exists with f^+ being the value of V_h at the local minimum with $z > 0$. The first singularity encountered as h moves along the negative real axis occurs when this local minimum disappears, namely at

$$h_s = -(1 - T^{2/3})^{3/2} \quad (= -z_s^3),$$

marking the end of metastability (the spinodal line). Conjecture 3 and Property 4 of Sec. 2 are thus immediately verified for this model; Property 5 is also valid with the "connecting" branch of the free energy equal to the value of V_h at its local maximum.

We now turn to the eigenvalues and eigenfunctions of $H_n(h)$ so as to investigate the analog of Conjecture 2.

To study $E_n^0(h)$ near $h = 0$, we define vectors $\Psi(m,\theta)$ $(m = -n, -n+1, \ldots, n)$ which are normalized eigenvectors of $J_\theta = \exp(-i\theta J_y)J_z\exp(i\theta J_y)$, $J_\theta\Psi(m,\theta) = m\Psi(m,\theta)$, and seek θ so that $\Psi(m,\theta)$ is approximately an eigenvector of H_n. Writing $A(m,m')$ for $(\Psi(m',0), A\Psi(m,0))$, we have

$$(\Psi(m',\theta), H\Psi(m,\theta)) = -\delta_{mm'}\{(1/2n^2)m^2\cos^2\theta + T(m/n)\sin\theta$$
$$+ h(m/n)\cos\theta + (1/2n^2)\sin^2\theta[n(n+1) - m^2]\}$$
$$+ [J_x(m,m')/n]\{\sin\theta\cos\theta[(m+m')/2n]$$
$$- T\cos\theta + h\sin\theta\}$$
$$- (1/8n^2)\sin^2\theta(J_+^2 + J_-^2)(m,m'). \tag{3.4}$$

If $m = n - O(1)$ as $n \to \infty$, off diagonal elements in (3.4) vanish to order $1/n$ if we demand that

$$\sin\theta\cos\theta - T\cos\theta + h\sin\theta = 0. \tag{3.5}$$

This equation is identical to that for the location of the minimum of (3.3) (with $z = \cos\theta$), while the diagonal element of (3.4) yields, as $n \to \infty$, the same $f^0(h)$ as given by (3.2). We thus regard $\psi(n,\theta_1)$ and $\Psi(n,\theta_2)$ (for $h \neq 0$) as (approximately) the stable ground state and "metastable ground state" of H_n where θ_1 and θ_2 are the two solutions (for $T < 1$ and small $|h|$) of (3.5) corresponding to the global and local minima of V_h.

We now consider an idealized two level system ob-

tained by restricting $H_n(h)$ to the space spanned by $\{\Psi(n,\theta_+),\ \Psi(n,\pi-\theta_+)\}$ for small $|h|$; here $\theta_+=\sin^{-1}T < \pi/2$ corresponds to the minimum of V_0 with $z>0$. The 2×2 matrix approximation to $H_n(h)$ is

$$\left[-\frac{(1+T^2)}{2}+O\left(\frac{1}{n}\right)\right]\begin{pmatrix}1&0\\0&1\end{pmatrix}-\begin{pmatrix}0&\alpha\\\alpha&0\end{pmatrix}-h\begin{pmatrix}\cos\theta_+&0\\0&-\cos\theta_+\end{pmatrix},$$

where $\alpha=(\Psi(n,\theta_+),\ [J_z^2/2n^2+TJ_x/n+hJ_z/n]\Psi(n,\pi-\theta_+))$ can be calculated using rotation matrices[12] and is exponentially small in n:

$$\alpha\approx d^n_{nn}(\pi-2\theta_+)=(\sin\theta_+)^{2n}=\exp[-(-2\log T)n]$$

(to zeroth order in h).

The relevant h dependence of $E_n^0(h)$ and $E_n^1(h)$ near $h=0$ is approximated by the eigenvalues of

$$W_2=-\begin{pmatrix}0&\alpha\\\alpha&0\end{pmatrix}-h\begin{pmatrix}(1-T^2)&0\\0&-(1-T^2)^{1/2}\end{pmatrix},$$

where we now consider α to be a fixed, small, h-independent constant. The eigenvalues of W_2 are $\epsilon_0=-[\alpha^2+h^2(1-T^2)]^{1/2}$ and $\epsilon_1=-\epsilon_0$. In place of a level crossing for real h, there is at $h=0$ an eigenvalue gap $(\epsilon_1-\epsilon_0)$ of size 2α and a level crossing at $h=\pm i\alpha/(1-T^2)^{1/2}$. For real $h\gg\alpha$, the eigenvectors are approximately

$$u_0\approx\begin{pmatrix}1\\\alpha/2h(1-T^2)^{1/2}\end{pmatrix},\quad u_1\approx\begin{pmatrix}-\alpha/2h(1-T^2)^{1/2}\\1\end{pmatrix};$$

for real $h\ll-\alpha$, the formulas are interchanged. Thus as h varies along the real axis, u_0 changes from one polarity [corresponding to $\Psi(n,\theta_+)$] to the other [corresponding to $\Psi(n,\pi-\theta_+)$] with an associated change in the "magnetization," $-\partial\epsilon_0/\partial h$.

A central theme in this paper is the prescription for continuing an eigenvalue into the complex plane so as to minimize the change in the corresponding eigenvector. In the 2×2 matrix case, the analytic continuation goes around the branch points at $h=\pm i\alpha/(1-T^2)^{1/2}$, i.e., we fix the sheet of the continuation by connecting these two points by a branch cut (see Fig. 3). The eigenvalue $\epsilon_+(h)$ defined on this sheet starts on the positive real axis as ϵ_0 and ends on the negative real axis as ϵ_1, but the corresponding eigenvector u_+ retains (for $|h|\gg\alpha$) the form

$$u_+\approx\begin{pmatrix}1\\\alpha/2h(1-T^2)^{1/2}\end{pmatrix}$$

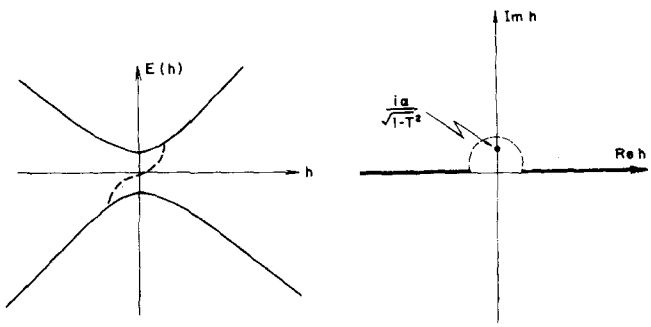


FIG. 3. Two level system eigenvalue crossing. (a) $E(h)$ for real h. (b) Passage around the branch point in the complex h plane.

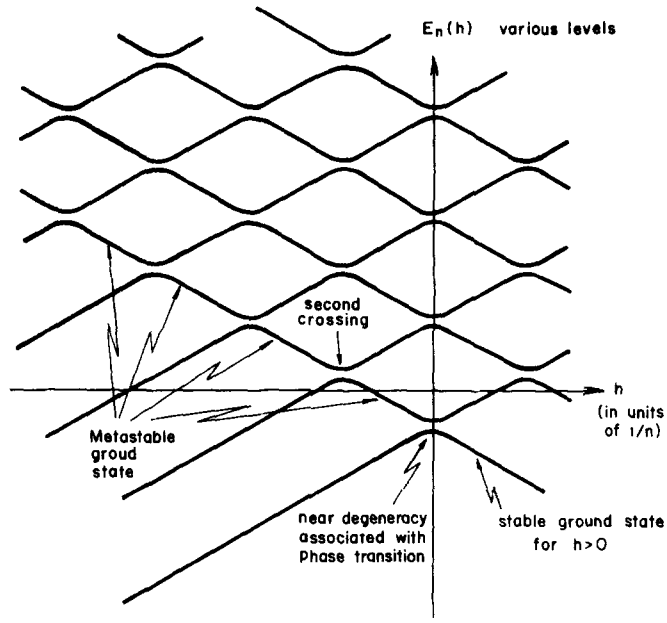


FIG. 4. Near level crossings for real h. The "metastable level" is the analytic continuation of $\Psi(n,\Theta_+)$ around the crossings.

As $n\to\infty$, the size of the necessary excursion into the complex h plane, being proportional to α, tends exponentially to zero. In order to justify our two-level system approximation, we must show that there is a range of $|h|\gg\alpha$ in which levels other than $\Psi(n,\theta_+)$ and $\Psi(n,\pi-\theta_+)$ can be ignored. This follows from the fact that the energy difference between (e.g.) $\Psi(n,\theta_+)$ and $\Psi(n-1,\theta_+)$ is $(1-T^2)/n+O(1/n^2)$, which is exponentially larger than α as $n\to\infty$.

The level crossing at $h=0$ is just the first of many. As we move to the left on the negative h axis the metastable level [approximately $\Psi(n,\theta_+)$] decreases in energy while the first excited state of opposite polarity [approximately $\Psi(n-1,\pi-\theta_+)$] increases until at some negative value of real h on the order of $1/n$ there is a second approximate level crossing analogous to the one described above between $\Psi(n,\theta_+)$ and $\Psi(n,\pi-\theta_+)$. Our picture of successive near level crossings for h real and negative is shown in Fig. 4. In each case, states of opposite polarity (sign of the expectation value of J_z) do not significantly mix unless their energies are close on the order of the matrix element of $S=J_z$ connecting the states. This matrix element is analogous to the α of the first crossing and decreases exponentially with n so that each level crossing is well approximated by a two-level system as above.

As $n\to\infty$ the number of crossings needed to continue to finite h also tends to infinity. Our earlier approximation for the "stable" excited states is not valid since $1-m/n$ does not approach zero. Instead we note that as $n\to\infty$ the eigenvalue equation for H has the form of a difference approximation to the differential equation

$$-\frac{1}{n^2}\frac{\partial}{\partial z}\left((1-z^2)\frac{\partial\psi}{\partial z}\right)+V_h(z)\psi(z)+\tilde{V}_n(z)\psi(z)=E_n\psi(z),\quad(3.6)$$

where $z=m/n$, m is the eivenvalue of J_z, $-1\le z\le1$,

$V_h(z)$ is given by (3.2), and $\tilde{V}_n(z) = O(1/n)$. This large n approximation is analogous to the WKB approximation in quantum mechanics with $1/n$ playing the role of Planck's constant. In the next model, we consider the WKB limit for an anharmonic oscillator and the qualitative conclusions obtained there apply here as well. Namely, the spacing between approximate level crossings (in Fig. 4) remains of order $1/n$, even for finite h, while the size of level crossing regions is of order e^{-n}; thus the two-level system approximations remain valid and the combination of Figs. 3 and 4 leads to Fig. 2—i.e., the validity of Conjecture 2. The absence of additional singularities away from the negative real axis (Conjecture 1) is based on a simple argument concerning polarization which we give below for the anharmonic oscillator.

B. The anharmonic oscillator

In this model, we take

$$H_n(h) = -\frac{1}{2n}\frac{d^2}{dx^2} + \frac{x^4}{4} - \frac{bx^2}{2} - hx,$$

acting on $L^2(\mathbb{R}, dx)$. This Hamiltonian arises as an approximation to Kac's model[2] of a one-dimensional ferromagnet with long range interaction. It may also be thought of as a simple model for symmetry breaking in φ^4 quantum field theories. It is clear that

$$f^0(h) = \inf_{-\infty<x<\infty} V_h(x), \quad V_h(x) = \tfrac{1}{4}x^4 - \tfrac{1}{2}bx^2 - hx \tag{3.7}$$

so that a phase transition occurs at $h=0$ (for $b>0$) corresponding to a breaking of the $x \to -x$ symmetry. For $b>0$, $f^0(h)$ can be analytically continued from $h>0$ past the origin to give $f^+(h)$, equal to the value of V_h at its right-hand local minimum. The spinodal line is $h_s = -2(b/3)^{2/3}$. The spinodal points $h = \pm\, h_s$ are square root branch points and the branches of $f^+(h)$ and the corresponding $f^-(h)$ are connected by an "unstable" sheet corresponding to the value of V_h at its local maximum. At $b=0$, these two square root branch points coalesce to form a cube root branch point at $h=0$ (the critical point), which then separates again into two square root branch points for $b<0$ (above the critical "temperature"), except that these are now located on the imaginary h axis.

Again for this model, we are primarily interested in the validity of the analogues of Conjectures 1 and 2. We base our arguments on the WKB approximation since this is equivalent to the $n \to \infty$ limit. The picture of two moving stacks of chips, described in Sec. 2, corresponds here to two groups of WKB eigenvalues—one group for each of the two "allowed" regions of V_h (for $b>0$ and small h) calculated as though the other region were absent. In this approximation there could be exact degeneracy between eigenfunctions in different regions. The degeneracy is broken because of the exponentially small (in n) tunneling probability between the regions (this has been rigorously shown for the first level crossing[13]). The spacing between level crossing regions as h varies along the negative real axis is seen by WKB calculation to be $O(1/\sqrt{n})$.

Thus Conjectures 1 and 2 and the corresponding pictures of Figs. 2 and 4 are valid provided that no singularities occur in the right half-plane away from the negative real axis. To verify this, we note that in the "independent wells" approximation discussed above, if we take $h = h_1 + ih_2$ with $|h_2|$ much greater than the tunneling probability, the energy eigenvalues of each well develop an imaginary part approximately proportional to $h_2\langle x\rangle$, where $\langle x\rangle$ denotes the expected value of x in the corresponding eigenfunction. Since $\langle x\rangle$ has different signs in the two wells, the (real h) degeneracies between levels from different wells is effectively broken and no degeneracies between levels of the same well can develop since the real parts of their energies are nondegenerate (away from the spinodal points).

C. The square well model

On the space $L^2([-\pi, \pi], dx)$, let

$$H_n(h) = -\frac{1}{n}\frac{d^2}{dx^2} + \lambda\delta(x) - h\,\mathrm{sgn}x \tag{3.8}$$

with Dirichlet boundary conditions at $x = \pm\pi$ and $\lambda > 0$. This Hamiltonian serves as a crude approximation to the anharmonic oscillator Hamiltonian discussed above, as well as to the Lipkin model via Eq. (3.6); the $\lambda\delta(x)$ term serves to separate the square well into two "allowed" regions and $S = \mathrm{sgn}x$ is the symmetry breaking term. Despite its simplicity, this model possess, as we shall show, most of the main features of the more complex models of phase transitions, including, as $n \to \infty$, two (asymptotically) independent sets of eigenvectors with all level crossings occurring in a small neighborhood of the real h axis. It thus provides a useful model of the relation between metastability and eigenvalue continuation.

The global analytic structure of the free energy is trivial for this model in that $f^0(h) = -|h|$, so that $f^+(h) = -h$ and $f^-(h) = h$. There is no spinodal point h_s so that f^+ and f^- are entire functions and are not branches of a single function; thus Property 5 of Sec. 2 is invalid and this model is most analogous to an Ising model at zero temperature whose free energy is also $-|h|$.

For this model, we can rigorously prove Conjecture 1 and parts of Conjecture 2 for $E_n^0(h)$, the minimum eigenvalue of the Hamiltonian of (3.8).

Theorem: For D any open bounded neighborhood of the origin in complex h space, there exists $K>0$ and functions $E_n^+(h)$ with the following properties:

(i) $E_n^+(h)$ is defined and analytic on $D \backslash B_n(K)$, where $B_n(K) = \{h : |h - (1-m^2)/2n| \le Km^2/n^2$ for some $m = 1, 2, 3, \cdots\}$,

(ii) $E_n^+(h) = E_n^0(h)$ for $h > 1/2n$,

(iii) $E_n^+(h) \to -h$ uniformly on compact subsets of $D \backslash (-\infty, 0]$.

Proof: Let $w = nE_n^0(h) - 1 + nh$ and $z = 2nh$; then, for $nh > \frac{1}{2}$, w is the unique solution of the eigenvalue equation

$$\sqrt{w+1}\,\cot\pi\sqrt{w+1} + \sqrt{z-w-1}\,\coth\pi\sqrt{z-w-1} + n\lambda = 0 \tag{3.9}$$

such that $-1 < w < 0$. This equation may be rewritten as a fixed point problem,

$$w = G(w,z) = g\left(\frac{-\sqrt{1+w}}{n\lambda} F(z-w)\right), \tag{3.10}$$

where $g(z) = [1 + (\arctan z)/\pi]^2 - 1$ and $F(z) = [1 + f(z)/n\lambda]^{-1}$ with $f(z) = \sqrt{z-1}\coth\pi\sqrt{z-1}$. Now $f(z)$ is analytic except for simple poles at $\{1 - m^2 : m = 1, 2, 3, \cdots\}$, and it can be shown by simple estimates that, for any bounded $\tilde{D}$, there is a $K' > 0$ such that $|f(z)| \leq n\lambda/2$ for $z/2n \in \tilde{D}\backslash B_n(K')$ and n sufficiently large. It then follows that for large n there exist positive δ and K so that if $|w| < \delta/n$ and $z/2n \in D\backslash B_n(K)$, then $|f(z-w)| \leq n\lambda/2$, $G(w,z)$ is jointly analytic, and $|G(w,z)| < \delta/n$. We now define $w_0(z) = 0$, $w_{k+1}(z) = G(w_k(z), z)$, and use some more simple estimates to show that $w_k(z)$ converges uniformly on $\{z : z/2n \in D\backslash B_n(K)\}$ to some $w(z)$ which solves (3.10), satisfies $-1 \leq w < 0$ for real z, is analytic on $\{z : z/2n \in D\backslash B_n(K)\}$, and is $O(1/n)$ there, thus implying all the conclusions of the theorem.

To investigate the validity of the remainder of Conjecture 2, we approximate $E_n^*(h)$ in the neighborhood of $\{(1-m^2)/2n\}$ by expanding $\cot\pi\sqrt{w+1}$ and $\coth\pi\sqrt{z-w-1}$ about their respective poles; letting $u = (1-m^2) - z$, we have that (3.9) is approximated by

$$\frac{1}{\pi} + \frac{2}{\pi w} + \frac{1}{\pi} + \frac{2m^2}{\pi(u+w)} + n\lambda = 0$$

or

$$w^2 + (u + \nu_n[m^2+1])w + \nu_n u = 0 \tag{3.11}$$

with

$$\nu_n = (1 + n\lambda\pi/2)^{-1}.$$

For each n the solution w of (3.11) has a pair of square root branch points, at $u = -\nu_n(m^2-1) \pm i2m\nu_n$; this indicates that $E_n^*(h)$ has square root branch points located at (approximately)

$$h = (1-\nu_n)\frac{1-m^2}{2n} \pm i\frac{m\nu_n}{n}, \quad m = 1, 2, \cdots,$$

which verifies Conjecture 2 of Sec. 1 with $x_k(n) \approx (1-\nu_n) \times (1-k^2)/2n$ and $y_k(n) \approx k\nu_n/n$. Since in this case $M_n = O(\sqrt{n})$ (i.e., there are $O(\sqrt{n})$ branch points in a finite region of the h plane) we have $\rho_n \equiv \sup_k |y_k(n)| = O(\sqrt{n}\,\nu_n/n) = O(1/n^{3/2})$ while $\min_k\{|x_{k+1}(n) - x_k(n)|\} \approx |x_2(n) - x_1(n)| \geq \text{const}/n$. The reason ρ_n is not exponentially smaller than $\min_k\{|x_{k+1}(n) - x_k(n)|\}$ as it is in the previous two models is that the $\lambda\delta(x)$ barrier does not prevent tunneling very effectively; to mimic this facet of the other models, one would have to take λ tending to infinity exponentially fast as $n \to \infty$.

4. RELEVANCE TO THE ISING MODEL

Do the features of the foregoing models apply to the Ising model? The essential elements present in the models of Sec. 3 are:

1. Existence of an order parameter, which is effectirely the expectation value of the symmetry breaking term S in the Hamiltonian.

2. Polarization, i.e., a division of the eigenvectors into two sets of opposite sign of the order parameter. Hence, as h varies, the picture of two stacks of chips in relative vertical motion is applicable. There may also be a set of unpolarized states, but these are of higher energy.

3. Smallness of the level crossing region (as the lowest level on one pile crosses levels in the other pile) compared to the spacing (in h) between level crossing regions. This justifies the picture that for "most" values of h levels in one stack move independently of the other stack.

4. Depletion of levels and narrowing of the spacing in the metastable stack as h approaches h_s.

The polarization is important because, as discussed for the anharmonic oscillator, it implies that for large enough $|\mathrm{Im}h|$ there should be no level crossings between levels in opposite "stacks" since the sign of $\mathrm{Im}E/\mathrm{Im}h$ is approximately proportional to the value of the order parameter. The absence of such level crossing indicates analyticity away from the negative real axis.

The actual size of the level crossing region in the complex h plane is determined by the size of the matrix elements of S between states of opposite polarity. This in turn depends on the amount of "wrong polarity" contained in an otherwise polarized state, which is closely releated to the tunneling probability for the models of Sec. 3.

The Ising model has an order parameter, the magnetization. It is reasonable to believe that the eigenvectors of the transfer matrix do arrange themselves in two stacks; this is known to happen in the thermodynamic limit when $h = 0$. A serious question that arises is whether element 3 above (or equivalently Conjecture 2 of Sec. 1) should be valid for the Ising model.

Our first guess concerning the spacing between level crossing regions is that they are on the average of order 2^{-n}; this corresponds to the assumption that a finite fraction of the eigenvalues of $-(1/n)\log L_n(h)$ lie within a finite distance of $E_n^0(h)$ and that they are approximately evenly spaced. To estimate the size of the level crossing regions, we note that (as for the Lipkin model) it is of the same order as the energy gap at the approximate (real h) degeneracy. For the approximate crossing at $h = 0$, it is known[14,15] that

$$E_n^1(0) - E_n^0(0) = O(e^{-cn}).$$

We suppose that crossings at $h \neq 0$ have a similar exponentially decreasing gap except that c is a function of h. The results of the previous section, e.g., the square well model, suggests that $c(h)$ decreases as h moves towards h_s. If our guess concerning the 2^{-n} spacing between level crossing regions is correct, then the validity of Conjecture 2 demands $c \equiv c(0) > \log 2$ while Property 4 of Sec. 2 suggests that $c(h) \to \log 2$ as $h \to h_s$ or $T \to T_c$.

A rough estimate for c, along the lines of a similar calculation by Kac,[15] suggests that it exceeds log2, at least for small T and probably for all $T < T_c$. The calculation, in the spirit of the 2×2 matrix discussion of Sec. 3, involves the transfer matrix $L_n(0)$, restricted to the two-dimensional subspace spanned by the vectors v_+, v_- corresponding to all spins $+1$ or all spins -1;

this leads to the estimate

$$e^{-cn} = O((v_+, L_n(0)v_-)/(v_+, L_n(0)v_+)) = e^{-2n\beta J}.$$

Since $2\beta_c J \approx 0.88$ and $\log 2 \approx 0.69$, we see that for $T < T_c$, $2\beta J > \log 2$ which supports the contention that $c > \log 2$. A different calculation, which can be found in the early work of Lassettre and Howe,[16] also suggests that $c > \log 2$ for small enough T.

5. DISCUSSION

Metastability has been a focus of controversy since the earliest studies of phase transitions.[16–18] First, the very existence of a metastable state can be questioned on the grounds that, as the system increases in size, the decay lifetime tends to zero so that it is problematic whether any metastable state can survive the thermodynamic limit. Second, there are certain models[19–21] which suggest that the approach to a first order phase transition involves an essential singularity in the thermodynamic functions. Finally, there is the problem of how the metastable state terminates[22–25]: Is there a well-defined spinodal line? Does the end of metastability correspond to a second order phase transition? In the paragraphs below we discuss these questions according to the point of view of this paper.

The possibility that $f^0(h)$ exhibits an essential singularity as $h \to 0^+$ has been proposed by several authors[19–21] based on various droplet or cluster models. They suggest that the singularity is associated with short range interaction models (such as the Ising model) and thus that the analytic continuation of mean field type models (such as those of Sec. 3) provides little evidence for the absence of a singularity in more realistic models.

The existence of an essential singularity at $h = 0$ is not totally inconsistent with our picture of metastability in that Conjecture 1 (and some of Conjecture 2) could be valid with Conjecture 3 invalid. This might occur, for example, if the size of level crossing regions, ρ_n, and the characteristic spacing between them, Δ_n, were of the same order of magnitude as $n \to \infty$ with the result that a branch cut develops for $f^+(h)$ along the negative real axis. However, we believe that Conjecture 3 is valid for the Ising model, with the difference between short and long range interaction models being manifested in a more subtle manner: namely, that in long range models $\rho_n = O(\exp(-1/\Delta_n)^p)$ for some $p > 0$ (as in the first two models of Sec. 3) while in short range models $\rho_n \sim (\Delta_n)^q$ for some $q > 1$ (as in the third model of Sec. 3).

The question as to whether and how critical droplets make an appearance in our picture is an intriguing one. It is tempting to suppose that they somehow correspond to the eigenstate of $L_n(h)$ lying at the common "peak" of the two groups of polarized states of our "stacks of chips" picture; in the Lipkin model, for example, this would be the state $\Psi(n, \theta_3)$, where θ_3 is the location of the local maximum of the V_h of (3.5). If this were so, then the value of the free energy along the connecting sheet(s) of Property 5 of Sec. 1 should be a parameter of the critical droplets.

However, it is just on the subject of critical droplets that mean field theories and local interaction theories differ most crucially. In mean field theory the critical droplet is a global excitation—in the work of Penrose and Lebowitz[26] it grows to infinity albeit more slowly then the volume—while for Ising and local theories it is expected to appear at a finite energy above the metastable ground state. It may be that critical droplets simply do not appear among the eigenvectors of the transfer matrix but only are manifested in the statistical mechanical states of the entire lattice, which are very different objects. Should our method prove successful, it may be precisely because we deal with the less physical eigenstates rather than with states of the entire lattice.

The end of metastability is a well-defined phenomenon according to Property 4 of Sec. 2. Thus, as the eigenvalue gap between the metastable ground state and *its* first excited state tends to zero, a second order phase transition occurs. This suggests a new numerical method to search for the spinodal line which is conceptually distinct from the Monte Carlo methods of Binder and Müller-Krumbhaar.[24]

As a final subject for this section, we consider the fact that the Ising model has no dynamics and the idea that it is for this reason that $f^+(h)$ has no singularity at $h = 0$ and remains real (with no branch cut) along the negative real axis. Study of the Ising model without dynamics is analogous to finding the energy of say the $2s$ level of hydrogen from the Coulomb force while neglecting the quantized electromagnetic field. Both systems are idealized since in the real world the same forces that provide binding (or ordering) also cause the decay. Nevertheless, solving for the $2s$ state is a good way to start work on hydrogen, and we expect that the idealized Ising metastable states have relevance to, say, metastable states for models with dynamics,[27] where perhaps the free energy develops a cut along the negative axis with the imaginary part of the free energy along that cut related to the metastable lifetime.[21]

ACKNOWLEDGMENTS

The authors are grateful for many useful conversations with M. Cassandro, C. Kuper, M. Revzen, and R. Shtokhamer. One of the authors (C.N.) would like to thank the Physics Department of the Technion for its hospitality during his visit there.

*Research supported in part by NSF Grant MPS 74-04870 A 01.
†NATO postdoctoral fellow on leave (1975-76) from the Department of Mathematics, Indiana University, Bloomington, Indiana, U.S.A.

[1] J. Ashkin and W.E. Lamb, Jr., Phys. Rev. 64, 159 (1943).
[2] M. Kac, "Mathematical Mechanisms of Phase Transitions," in *Brandeis Summer Institute* 1966: *Statistical Physics, Phase Transitions and Superfluidity*, edited by M. Chretien, E.P. Gross, and S. Deser (Gordon and Breach, New York, 1967).
[3] C.J. Thompson, *Mathematical Statistical Mechanics* (Macmillan, New York, 1972).
[4] T.D. Lee and C.N. Yang, Phys. Rev. 87, 410 (1952).
[5] M. Suzuki and M. Fisher, J. Math. Phys. 12, 235 (1971).
[6] B. Simon and R.B. Giffiths, Commun. Math. Phys. 33, 145 (1973).

[7]H.J. Lipkin, N. Meshkov, and A.J. Glick, Nucl. Phys. **62**, 188 (1965).
[8]N. Meshkov, A.J. Glick, and H.J. Lipkin, Nucl. Phys. **62**, 199 (1965).
[9]A.J. Glick, H.J. Lipkin, and N. Meshkov, Nucl. Phys. **62**, 211 (1965).
[10]D. Agassi, A.J. Lipkin, and N. Meshkov, Nucl. Phys. **86** 321 (1966).
[11]E.H. Lieb, Commun. Math. Phys. **31**, 327 (1973).
[12]A.R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1957), p. 59.
[13]Colin J. Thompson and Mark Kac, Studies Appl. Math. **48**, 257 (1969).
[14]C. Domb, Adv. Phys. **9**, nos. 34 & 35 (1960).
[15]M. Kac, in *Fundamental Problems in Statistical Mechanics, II*, edited by E.G.D. Cohen (North-Holland, Amsterdam, 1968).
[16]E.N. Lassettre and J.P. Howe, J. Chem. Phys. **9**, 747, 801 (1941).
[17]G.S. Rushbrooke, Proc. Roy. Soc. (London) A **166**, 296 (1938).
[18]J.G. Kirkwood, J. Phys. Chem. **43**, 97 (1939).
[19]A.F. Andreev, Zh. Eksp. Teor. Fiz. **45**, 2064 (1963) [Sov. Phys. JETP **18**, 1415 (1964)].
[20]M.E. Fisher, Physics **3**, 255 (1967).
[21]J.S. Langer, Ann. Phys. **41**, 108 (1967).
[22]B. Chu, F.J. Schoenes, and M.E. Fisher, Phys. Rev. **185**, 219 (1969).
[23]D.S. Gaunt and G.A. Baker, Jr., Phys. Rev. B **1**, 1184 (1970).
[24]K. Binder and H. Müller-Krumbhaar, Phys. Rev. B **9**, 2398 (1974).
[25]C. Domb, J. Phys. A **9**, 283 (1976).
[26]O. Penrose and J.L. Lebowitz, J. Stat. Phys. **3**, 211 (1971).
[27]D. Capocaccia, M. Cassandro, and E. Olivieri, Commun. Math. Phys. **39**, 185 (1974).

Factored irreducible symmetry operators

N. O. Folland

Physics Department, Kansas State University, Manhattan, Kansas 66502
(Received 22 June 1976; revised manuscript received 19 August 1976)

Irreducible symmetry operators (ISO's) are defined and their properties are displayed. A method for the systematic construction of ISO's in the form of products of interchangeable factors is described. The key to the method is a set of conditions for directly testing for the irreducibility of induced symmetry operators formulated in terms of the operators themselves. Detailed examples are presented for the tetrahedral group and double group. The tetrahedral double group provides an example in which the induction process in its simplest form fails, but which does yield to a modification of the induction techniques.

I. INTRODUCTION

There is an extensive and growing literature including many monographs[1–5] describing the application of group theoretical techniques to physical problems. Typically, a physical system is described by a Hamiltonian, H, and the collection of operators, G, which leave the Hamiltonian invariant form a group in the mathematical sense. The problem that group theory solves is that of finding a maximal set of mutually commuting operators. For finite groups the solution of the group theoretical problem may be formulated in terms of irreducible symmetry operators (ISO's) or equivalently, in terms of the irreducible unitary representations (IUR's) of the group.

The term "irreducible symmetry operator" or "ISO" defined in Sec. II is not universally used in the literature. Typically, assigning a specific name to these operators is avoided. Special cases of ISO's are projection operators. The page references in the monographs listed here (Refs. 1—5) refer to discussions focusing on what are here called ISO's. To avoid possible confusion the set of objects which form a group shall be referred to individually as group elements or operators and the term "symmetry operator" in this context will be avoided. It is peculiar that the monograph discussions of ISO's are so brief and incomplete in view of the fact that these operators are the basic working tools used in applying group theoretical methods to physical problems.

The present work was motivated by the empirical observation that for many finite groups of physical interest (including all nonrelativistic crystalline space group groups) the ISO's may be expressed as the product of interchangeable factors. The idea may best be conveyed to the uninitiated by a simple example. Consider a group G consisting of the identity element E and a two-fold operator C_i ($C_i^2 = E$). The ISO's for the group are $P_+ = \frac{1}{2}(E + C_i)$ and $P_- = \frac{1}{2}(E - C_i)$. It follows that $E = P_+ + P_-$, $P_+^2 = P_+$, $P_-^2 = P_-$, and $P_+P_- = P_-P_+ = 0$. Thus, any set of functions on which the operations of the group are defined can be decomposed as $\Phi = \Phi_+ + \Phi_-$, where $\Phi_+ = P_+\Phi$ and $\Phi_- = P_-\Phi$. And the eigenfunctions of H can be chosen to be simultaneous eigenfunctions of H, P_+, and P_-. The group theoretical problem is solved. For groups of small order there is little need for more formal procedures than those just described.

One purpose of the present work is to show that factored ISO's may be constructed systematically by induction techniques. The key to the induction-based construction procedure is a theorem given in Sec. III which gives necessary and sufficient conditions that induced operators are ISO's. In the sense that induction procedures are well known and tests for irreducibility of induced representations are treated at length in textbooks, the present results are not new and in certain respects are less refined. For example, the highly-refined techniques involving the "little group" can often produce the same results as the present methods (e.g., see Ref. 4, Sec. II. 6). The main advantages are practicality and simplicity. This becomes particularly apparent when time-reversal symmetry is considered. The present approach allows a high level of flexibility in choosing the form of degenerate irreducible representations.

A second purpose of the present work concerns situations in which standard induction techniques fail. In general for finite groups there may not exist subgroups from whose irreducible representations, the irreducible representations of the group may be induced. An example of such a group is considered, the tetrahedral double group. It is found in Sec. IV that a modified induction technique allows the construction of highly factored ISO's for the tetrahedral and cubic double groups.

The only extensive study on the process of expressing ISO's in factored form is that of Melvin.[6] Melvin assumes that an irreducible representation of the group is given and gives recipes for expressing the operators as a product of a "kernel" operator and a "quotient set" operator. Melvin does not explore the possibility of choosing the irreducible representations to facilitate the factorization process. The emphasis in the present work is to develop techniques for directly constructing the ISO's in factored form without knowledge of the irreducible representation.

II. SYMMETRY OPERATORS

In this section certain definitions and results of finite group theory will be briefly reviewed. These provide the basis for the definition of symmetry operators and irreducible symmetry operators (ISO's) and an exposition of their properties.

An element g of a group G is represented by a nonsingular, square matrix $D^A(g)$ of dimension n_A in that for every pair of group elements g, g' such that $g \cdot g' = g''$, then $D^A(g)D^A(g') = D^A(g'')$. The matrix elements

$D^A(g)_{ij}$ are complex numbers. The set of matrices $A=\{D^A(g)\}$ is called a representation of the group. The number of group elements (matrix representatives) in the group (representation) is symbolized G^0.

The "intertwining" of two representations A and B is characterized by the $(n_A\times n_B)$-dimensional, square super matrix $M(A,B)$ whose elements are defined

$$M(A,B)_{(i,m)(n,j)}=(n_A/G^0)\sum_g D^A(g^{-1})_{ij}D^B(g)_{mn}, \tag{1}$$

where the sum $\sum$ includes all elements of the group G. The rows (columns) of the super matrix are indicated by the double indices (i,m) and (n,j) with $i,j=1,\ldots,n_A$ and $n,m=1,\ldots,n_B$. The symbol g^{-1} denotes the group element inverse to g, $g\cdot g^{-1}=g^{-1}\cdot g=e$, where e is the identity element.

Representations A and B are said to be equivalent $(A\sim B)$ if $n_A=n_B$ and there exists a square matrix S such that $D^A(g)=SD^B(g)S^{-1}$ for each element g of G. The symbol S^{-1} denotes the matrix inverse to S, $SS^{-1}=S^{-1}S=E$, the identity matrix. Any representation A' of a finite group is equivalent to a representation A by unitary matrices. A matrix is unitary if $D^A(g)^{-1}=D^A(g)^\dagger$, where the symbol † denotes matrix transposition and complex conjugation. In element form, $[D^A(g)^{-1}]_{ij}=[D^A(g)_{ij}]^*$, where the symbol * denotes complex conjugation. Representations are said to be identical $(A\equiv B)$ if for each g included in group G, $D^A(g)=D^B(g)$. A representation A is irreducible if $M(A,A)$ is the unit supermatrix or

$$M(A,A)_{(i,m)(n,j)}=\delta_{in}\delta_{jm}. \tag{2}$$

If A and B are irreducible unitary representations (IUR's) and if $A\sim B$, then $A\equiv B$, then the "intertwining" matrix is evaluated to be

$$M(A,B)_{(i,m)(n,j)}=\delta_{AB}\delta_{in}\delta_{jm}. \tag{3}$$

In particular this means that if A and B have no IUR's in common, $M(A,B)=0$. Of course Eq. (3) is just a compact form of the orthogonality relations.

Symmetry operators are defined with respect to a group G and unitary representation A,

$$P(A)_{ij}=(n_A/G^0)\sum_g D^A(g^{-1})_{ij}\,g. \tag{4}$$

Symmetry operators have properties including:

$$[P(A)_{nm}]^\dagger=P(A)_{mn}, \tag{5}$$

$$gP(A)_{mn}=\sum_{k=1}^{n_A}P(A)_{mk}D^A(g)_{kn}, \tag{6}$$

$$P(A)_{ij}P(B)_{mn}=\sum_{k=1}^{n_A}M(A,B)_{(i,k)(n,j)}P(B)_{mk}. \tag{7}$$

The superscript dagger (†) in Eq. (5) indicates the adjoint operator. If an inner product is denoted by (ϕ_1,ϕ_2), then the adjoint of a general operator H is defined by $(\phi_1,H\phi_2)=(H^\dagger\phi_1,\phi_2)$.

If the symmetry operators are formed with respect to IUR's they are called irreducible symmetry operators (ISO's). For ISO's the properties above are altered only in the case of Eq. (7) which is simplified by Eq. (3) to

$$P(A)_{ij}P(B)_{mn}=\delta_{AB}\delta_{in}P(A)_{mj}. \tag{8}$$

The properties described by Eqs. (5)—(8) follow directly from the definition Eq. (4) and Eqs. (1)—(3). The following property of ISO's,

$$g=\sum_A\sum_{m=1}^{n_A}\sum_{n=1}^{n_A}P(A)_{mn}D^A(g)_{nm}, \tag{9}$$

is obtained most directly from consideration of the regular representation.

The regular representation R may be obtained by considering the G^0-dimensional carrier space $\phi^R=(g_1,g_2,\cdots)$, where each element of the group is a component of the carrier space. Then, $g\phi^R=\phi^RD^R(g)$ defines the regular representation matrix $D^R(g)$. The complete reduction of R by a similarity transform S to $R'=\{S^{-1}D^R(g)S\}$ transforms the space ϕ^R to $\phi^{R'}=\phi^RS$. One may choose $\phi^{R'}=(P(A)_{11},P(A)_{12},\ldots,P(B)_{ij},\cdots)$ including n_A^2 symmetry operators for each distinct irreducible representation A. In component form the relation $\phi^{R'}=\phi^RS$ is given by Eq. (4) and the inverse relation $\phi^R=\phi^{R'}S^{-1}$ is given by Eq. (9). Thus, the relation between the regular representation and the ISO's provides a broader understanding of the symmetry operators. The proof of Eq. (9) is left to the reader. For more discussion of the regular representation see any of the group theory monographs (Refs. 1—5).

The "resolution of the identity" or "spectral theorem" is a special case of Eq. (9).

$$e=\sum_A\sum_{m=1}^{n_A}P(A)_{mm} \tag{10}$$

in which $g=e$, the identity operator of the group. The operators $P(A)_{mm}$ which occur in Eq. (10) are projection operators because $[P(A)_{mm}]^\dagger=P(A)_{mm}$ and $P(A)_{mm}P(B)_{nn}=\delta_{AB}\delta_{mn}P(A)_{mm}$. The projection operators in Eq. (10) constitute the maximal set of commuting operators that can be formed with respect to the group. In this sense they constitute a solution of the group theoretical problem. Any set of basis functions on which the operations of the group are defined may be decomposed into irreducible subspaces according to Eq. (10).

In applications it is advantageous to use one projection operator for IUR A, say $P(A)_{11}$, and to generate "partner functions" using the property $P(A)_{1j}P(A)_{11}=P(A)_{1j}$. Thus, if $P(A)_{11}\phi_I$ has nonzero projection, then so will the "partners" $P(A)_{1j}\phi_I$. In general the relation

$$(P(A)_{1i}\phi_I,HP(B)_{1j}\phi_J)=\delta_{AB}\delta_{ij}(\phi_I,HP(A)_{11}\phi_J) \tag{11}$$

follows from Eqs. (5) and (8) and the assumption that H is invariant to group operations. Matrix elements between different IUR's and between different partners of the same IUR are zero. Matrix elements for each "partner" set $\{P(A)_{1j}\phi_I\}$ are identical to matrix elements for the projected set $\{P(A)_{11}\phi_I\}$.

One further remark concerning the definition and properties of symmetry operators is implicit in the proceeding paragraph. The "partner" operators $P(A)_{1j}$ with $j=1,\ldots,n_A$ contain all possible information about IUR A. The representation may be found from Eq. (6). The remaining ISO's may be found from the IUR and the definition, Eq. (4), or from the operators themselves because

$$P(A)_{ij}[P(A)_{ii}]^{\dagger} = P(A)_{ij}. \tag{12}$$

III. IRREDUCIBILITY AND INDUCED SYMMETRY OPERATORS

In this section the induction of symmetry operators for a group G is defined in terms of the ISO's of a proper subgroup H of G. The properties of the induced representations are described and a set of necessary and sufficient conditions for the irreducibility of the induced representations is formulated. These results provide the basis for the systematic construction of factored ISO's.

Suppose H is a proper subgroup of G. The group G may be expanded in left cosets of H;

$$G = SH, \tag{13}$$

where $S = \{e, s_2, \ldots, s_n\}$ is a set of $n = G^0/H^0$ coset generators and $s_1 = e$, the identity element. Let $a = \{D^a(h)\}$ be an IUR of H with ISO's defined by Eq. (4) to be

$$P(a)_{ij} = (n_a/H^0)\sum_h D^a(h^{-1})_{ij}\, h. \tag{14}$$

Induced symmetry operators for induced representation A are defined by the "partner" operators

$$P(A)_{(1i)(pj)} = s_p P(a)_{ij}, \tag{15}$$

where $p = 1, \ldots, n$ and $j = 1, \ldots, n_a$. The induced representation A is of dimension $n_A = n \times n_a$. The remaining symmetry operators for the representation A are found with the aid of Eq. (12). The adjoints of the partner operators of Eq. (15) are

$$P(A)_{(pj)(1i)} = P(a)_{ji} s_p^{-1}. \tag{16}$$

A general induced symmetry operator is given by

$$P(A)_{(qi)(pj)} = P(A)_{(1i)(pj)} P(A)_{(qi)(1i)} = s_p P(a)_{ij} s_q^{-1}. \tag{17}$$

It is easy to show that the induced representation A is unitary. Consider an arbitrary element g of G. In the double index notation of the induced representation Eq. (6) is

$$gP(A)_{(qi)(pj)} = \sum_{r=1}^{n}\sum_{k=1}^{n_a} P(A)_{(qi)(rk)} D^A(g)_{(rk)(pj)}. \tag{18}$$

It follows immediately that the induced matrix representative is

$$D^A(g)_{(rk)(pj)} = \sum_h D^a(h)_{kj}\, \Delta(g \circ s_p, s_r \circ h), \tag{19}$$

where the function $\Delta(g, g')$ has the value 1 if $g = g'$ and is 0 otherwise. Hence,

$$D^A(g^{-1})_{(pj)(rk)} = \sum_h D^a(h^{-1})_{jk}\, \Delta(g^{-1} \circ s_r, s_p \circ h^{-1})$$

$$= [D^A(g)_{(rk)(pj)}]^* \tag{20}$$

and the induced representation A is unitary. The right member of Eq. (20) follows from the unitarity assumed for IUR a and because $\Delta(g^{-1}s_r, s_p h^{-1}) = \Delta(gs_p, s_r h)$.

Thus, the induced symmetry operators defined by Eq. (15) or Eq. (17) carry a unitary representation A with matrix representatives given by Eq. (19). The induced unitary representation A need not be irreducible. The conditions for irreducibility may be expressed in terms of the symmetry operators. The induced symmetry operators will be ISO's if Eq. (8) holds or in the double index notation of the induced representation

$$P(A)_{(p'j')(pj)} P(A)_{(q'i')(qi)} = \delta_{(p'j')(qi)} P(A)_{(q'i')(pj)}. \tag{21}$$

The irreducibility conditions, Eq. (21), may be expressed in terms of the induced symmetry operators, Eq. (17), as

$$s_p P(a)_{j'j} s_p^{-1} s_q P(a)_{i'i} s_q^{-1} = \delta_{(p'j')(qi)}\, s_p P(a)_{i'j} s_q^{-1}. \tag{22}$$

Equation (22) is automatically satisfied for the case $p' = q$ because a is an IUR. The general irreducibility conditions may be simplified further. When $p' \neq q$, the right member of Eq. (22) is zero. The left member of Eq. (22) is a linear combination of group elements with coefficients which are complex numbers. Such an object will be zero only if each of the complex coefficients of group elements is zero and it cannot be made zero by multiplication from the left or right by group elements. Hence, the irreducibility conditions can be rewritten as

$$P(a)_{j'j} s_r h P(a)_{i'i} s_r^{-1} = 0, \quad r = 2, \ldots, n, \tag{23}$$

where Eq. (22) with $p' \neq q$ has been multiplied from the left by s_p^{-1} and from the right by $s_{q'} s_r^{-1}$ with $s_p^{-1} s_q = s_r h$. Using Eq. (6), Eq. (23) becomes

$$\sum_{m=1}^{n_a} [P(a)_{j'j} s_r P(a)_{i'm} s_r^{-1}] D^a(h)_{mi} = 0, \quad r = 2, \ldots, n. \tag{24}$$

The irreducibility conditions may be stated in the following theorem.

Irreducibility Theorem: Necessary and sufficient conditions that representation A of group G induced from IUR a of subgroup H be irreducible are that

$$P(a)_{ii} s_p P(a)_{jj} s_p^{-1} = 0, \tag{25}$$

where $i, j = 1, \ldots, n_a$ and $p = 2, \ldots, n = G^0/H^0$.

The proof is nearly complete with Eq. (24). Equation (25) are necessary conditions because they are special cases of Eq. (22), the general irreducibility conditions. Multiply Eq. (25) from the left by $P(a)_{ii'}$ and from the right by $s_p P(a)_{j'j} s_p^{-1}$ and use Eq. (8) to obtain

$$P(a)_{ii'} s_p P(a)_{j'j} s_p^{-1} = 0. \tag{26}$$

Now Eq. (26) is identical to the coefficients of $D^a(h)_{mi}$ (in square brackets) in Eq. (24) which is an alternative form of the general irreducibility conditions. Therefore, the conditions of Eq. (25) are sufficient and the proof is complete.

The conditions for irreducibility require that the projection operators for IUR a, $P(a)_{jj}$, annihilate with all conjugate projection operators $s_p P(a)_{jj} s_p^{-1}$. It might well appear that these conditions offer little improvement over the conventional character tests for irreducibility (see, e.g., Ref. 4, pp. 140–162). Certainly, it would be awkward and tedious to apply the conditions of Eq. (26) to a sizeable group when the projection operators are represented as in the definition, Eq. (4). The utility of the irreducibility theorem resides in that it provides the basis for the systematic construction of ISO's in the form of products of interchangeable factors.

TABLE I. Operations of the tetrahedral group and double group. The first column gives the symbol for the group operator. The effect of the group operator on a 3-space vector denoted by the three-tuple (x, y, z) is listed in the second column under the heading ${}^{s}R_{nam}$. Spin-space transformations are listed in the third column (labeled ${}^{\sigma}R_{nam}$) in the form ${}^{\sigma}R_{nam} = {}^{\sigma}E\cos(n\pi/m) + {}^{\sigma}R_{a2}\sin(n\pi/m)$ as discussed in the text. The last column lists the unit vector appropriate to the transformation (if any). ${}^{\sigma}E$ is the identity spin transformation and ${}^{\sigma}E_2$ a rotation in spin space by 360° about an arbitrary axis.

Symbol	${}^{s}R_{nam}$	${}^{\sigma}R_{nam}$	$\hat{R}_a$
E_{x2}	(x, y, z)	σ_E	–
E_2	(x, y, z)	${}^{\sigma}E_2 = -{}^{\sigma}E$	–
C_{x2}	$(x, -y, -z)$	${}^{\sigma}C_{x2}$	$\hat{C}_x$
C_{y2}	$(-x, y, -z)$	${}^{\sigma}C_{y2}$	$\hat{C}_y$
C_{z2}	$(-x, -y, z)$	${}^{\sigma}C_{z2}$	$\hat{C}_z$
T_{2w3}	(y, z, x)	$(-{}^{\sigma}E + \sqrt{3}\,{}^{\sigma}T_{w2})/2$	$\hat{T}_w$
T_{4w3}	(z, x, y)	$(-{}^{\sigma}E - \sqrt{3}\,T_{w2})/2$	$\hat{T}_w$
T_{2x3}	$(-y, z, -x)$	$(-{}^{\sigma}E + \sqrt{3}\,{}^{\sigma}T_{x2})/2$	$\hat{T}_x$
T_{4x3}	$(-z, -x, y)$	$(-{}^{\sigma}E - \sqrt{3}\,{}^{\sigma}T_{x2})/2$	$\hat{T}_x$
T_{2y3}	$(-y, -z, x)$	$(-{}^{\sigma}E + \sqrt{3}\,{}^{\sigma}T_{y2})/2$	$\hat{T}_y$
T_{4y3}	$(z, -x, -y)$	$(-{}^{\sigma}E - \sqrt{3}\,{}^{\sigma}T_{y2})/2$	$\hat{T}_y$
T_{2z3}	$(y, -z, -x)$	$(-{}^{\sigma}E + \sqrt{3}\,{}^{\sigma}T_{z2})/2$	$\hat{T}_z$
T_{4z3}	$(-z, x, -y)$	$(-{}^{\sigma}E - \sqrt{3}\,{}^{\sigma}T_{z2})/2$	$\hat{T}_z$

Often each factor itself is a projection operator and the application of the irreducibility theorem can be done by inspection. Trivial examples of factored symmetry operators are well known. For example, in the case where the structure of a group allows it to be expressed as a direct product $G = H \times K$ of two normal subgroups H and K, the ISO's for G are simply products of the ISO's for H and K separately. The factorization of an ISO must always be associated with structure inherent in the group. Conversely, the presentation of ISO's in factored form reveals structure inherent to the group.

In the next section the process of constructing ISO's as products of interchangeable factors will be discussed more generally and in the form of explicit applications to the tetrahedral group. The case of the tetrahedral double group is of interest in that it provides an example of the smallest subgroup of the cubic double groups for which the systematic construction process fails in principle. An alternative procedure is devised to treat this case.

IV. THE CONSTRUCTION OF FACTORED IRREDUCIBLE SYMMETRY OPERATORS

Suppose that a group G has a normal subgroup H with coset expansion $G = SH$. The ISO for the identity representation of H is $P(H_1) = (1/H^0)\sum_h h$. It is convenient to suppress the indices (11) on ISO's corresponding to one-dimensional IUR's. Since H is a normal subgroup of G, the conjugate operators $s_p P(H_1) s_p^{-1} = P(H_1)$ and an induced representation with respect to IUR H_1 of H must be reducible by the irreducibility theorem. However, even if the coset generators s_p cannot be chosen to form a group the set of operators $G_1 = \{s_p P(H_1)\}$ form a group isomorphic to the factor group G/H. Clearly, the ISO's for the group G_1 are expressible as a linear combination of coset generators with complex coefficients multiplied by $P(H_1)$. The connection with Melvin's work[6] is evident. $P(H_1)$ is Melvin's "kernel" operator and the linear combination of coset generators is Melvin's "quotient set." It may be possible that G_1 and H can be decomposed into subgroups thereby allowing further factorization. The failure of a given IUR of the subgroup to allow induced IUR's can often be a useful insight into the basic structure of the group.

More general construction procedures for factored ISO's using subgroup techniques can be formulated in the operator language. Suppose that a proper subgroup H of $G = SH$ has a one-dimensional IUR with ISO $P(H_2)$. The conjugate operators $s_p P(H_2) s_p^{-1}$ formed with respect to the coset generators S can be classified according to whether the quantity $Q(H_2)_p = P(H_2)\, s_p P(H_2)\, s_p^{-1}$ is zero or not. If $Q(H_2)_p = 0$ for all p except $s_1 = e$, then by the irreducibility theorem the induced representation is an IUR. When $Q(H_2)_p$ is nonzero for one or more $s_p \neq e$, several possibilities arise. First, if $s_p P(H_2) s_p^{-1} = P(H_2)$ for one or more $s_p \neq e$, then the set $\{s_p P(H_2) : s_p P(H_2) s_p^{-1} = P(H_2)\}$ which includes $P(H_2)$ is in general a multiplier group (see Ref. 5 for a discussion of multiplier groups). If s_p and s_q are included in the set and $s_p s_q = s_r h$, then s_r must also be in the set of coset generators which commute with ISO $P(H_2)$. It follows that $(s_p P(H_2))(s_q P(H_2)) = s_r h P(H_2) = (s_r P(H_2)) D^{H_2}(h)$, where the factor $D^{H_2}(h)$ is the one-dimensional matrix representative for element h in IUR H_2 of H. The previous statement confirms that the set of operators is a multiplier group. In this case the test for irreducibility of the induced representation indicated a larger subgroup to be considered for induction and provided information about the structure of this larger subgroup which may be used in finding its ISO's. The second possibility is not favorable. In the case $Q(H_2)_p$ is nonzero for one or more $s_p = e$ and no simple relation can be found between operator $P(H_2)$ and the conjugate operators $s_p P(H_2) s_p^{-1}$, the method fails. Although there are many groups of physical interest for which induction techniques can be used successfully, there are also groups for which the method fails.

The tetrahedral group and double group are now considered as illustrations of the induction methods. The groups operations are expressed in terms of symmetry axis unit vectors. There are three twofold axes: $\hat{C}_x = (k, 0, 0)$, $\hat{C}_y = (0, 1, 0)$, $\hat{C}_z = (0, 0, 1)$, and four threefold axes: $\hat{T}_w = (1, 1, 1)/\sqrt{3}$, $\hat{T}_x = (1, -1, -1)/\sqrt{3}$, $\hat{T}_y = (-1, 1, -1)/\sqrt{3}$, $\hat{T}_z = (-1, -1, 1)/\sqrt{3}$. Active rotations in the right-hand sense about an axis defined by a unit vector $\hat{R}_a$ by an angle $\theta = n2\pi/m$ are denoted by the symbol R_{nam}. Indices $n, m = 1$ are suppressed. A presuperscript $s(\sigma)$ distinguishes between operators on space (spin) coordinates in situations where an ambiguity might arise; otherwise the presuperscript is suppressed.

The operations of the tetrahedral group T are listed in Table I. Spin-space transformations are expressed in the form

$$ {}^{\sigma}R_{nam} = {}^{\sigma}E\cos(n\pi/m) + {}^{\sigma}R_{a2}\sin(n\pi/m), \tag{27} $$

where

$$ {}^{\sigma}R_{a2} = -i\hat{R}_a \cdot \boldsymbol{\sigma} \tag{28} $$

TABLE II. Multiplication and commutation relations between selected operators of the group DT, the tetrahedral double group. The double group relations of part A may be specialized to tetrahedral group relations by setting $E_2=1$. Hybrid operators are defined in the text, Eqs. (48)—(50).

A. Relations between selected tetrahedral double group operators:

$$C_{y2}C_{z2}=E_2C_{z2}C_{y2}=C_{x2},\quad C_{z2}C_{x2}=E_2C_{x2}C_{z2}=C_{y2},$$
$$C_{x2}C_{y2}=E_2C_{y2}C_{x2}=C_{z2},\quad C_{x2}C_{x2}=C_{y2}C_{y2}=C_{z2}C_{z2}=E_2,$$
$$C_{x2}T_{2w3}=T_{2w3}C_{y2},\quad C_{y2}T_{2w3}=T_{2w3}C_{z2},\quad C_{z2}T_{2w3}=T_{2w3}C_{x2},$$
$$C_{x2}T_{4w3}=T_{4w3}C_{z2},\quad C_{y2}T_{4w3}=T_{4w3}C_{x2},\quad C_{z2}T_{4w3}=T_{4w3}C_{y2}.$$

B. Relations between hybrid operators and selected double space group operators:

$$C_wC_w=C_dC_d=C_{wd}C_{wd}=E_2,\quad C_{wd}C_d=E_2C_dC_{wd}=C_w,$$
$$C_dC_w=E_2C_wC_d=C_{wd},\quad C_wC_{wd}=E_2C_{wd}C_w=C_d,$$
$$C_wT_{2w3}=T_{2w3}C_w,\quad C_wT_{4w3}=T_{4w3}C_w,$$
$$C_dT_{2w3}=T_{2w3}C_d(-E-\sqrt{3}C_w)/2,\quad T_{2w3}C_d=C_d(-E+\sqrt{3}C_w)T_{2w3}/2,$$
$$C_dT_{4w3}=T_{4w3}C_d(-E+\sqrt{3}C_w)/2,\quad T_{4w3}C_d=C_d(-E-\sqrt{3}C_w)T_{4w3}/2.$$

is given in terms of the matrices $\sigma|_j=\sigma_j$, $j=1,2,3$. The matrices σ_j are homomorphic to the Pauli spin matrices and have multiplication properties,

$$\sigma_i\sigma_j={}^{\sigma}E\delta_{ij}+i\sigma_{ijk}\sigma_k. \tag{29}$$

In listing the spin-space transformations an explicit choice of spin representation is not made. The multiplication properties of spin-space operators do not depend on the representation. The spin-space transformations are expressed in terms of rotations by 180° about the symmetry axes R_a because their multiplication properties are simple,

$$R_{a2}R_{b2}=-\hat{R}_a\cdot\hat{R}_b\,{}^{\sigma}E-(\hat{R}_a\times\hat{R}_b)\cdot(-i\sigma), \tag{30}$$

where $\hat{R}_a\times\hat{R}_b$ is the vector cross product and $\hat{R}_a\cdot\hat{R}_b$ is the vector dot product. Also, spin representations may be chosen to be compatible with the double group ISO's and simplify their application.

The choice of spin representation may be made on the basis of geometrical considerations. For the present purpose it is convenient to consider spin space to be two-dimensional. It is trivial to extend the method to deal with four-dimensional Dirac spinors. The usual representation for the Pauli spin matrices is

$$\pi_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \pi_2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \pi_3=\begin{pmatrix}1&0\\0&-1\end{pmatrix}, \tag{31}$$

where indices 1, 2, 3 correspond to the coordinate axes x, y, z. The Pauli spin matrices π_i satisfy Eq. (29). Alternatives to the conventional Pauli representation may be constructed with respect to any orthogonal set of real unit vectors $\hat{a},\hat{b},\hat{c}=\hat{a}\times\hat{b}$. Let

$$\sigma_n=\hat{n}\cdot\boldsymbol{\sigma}=\hat{n}_x\sigma_x+\hat{n}_y\sigma_y+\hat{n}_z\sigma_z, \tag{32}$$

where $\hat{n}$ is $\hat{a}$, $\hat{b}$, or $\hat{c}$. The multiplication properties of the operators σ_a, σ_b, σ_c are isomorphic to the operators σ_x, σ_y, σ_z which are also defined by Eq. (32) with $\hat{n}=\hat{x}$, $\hat{y}$, or $\hat{z}$, the coordinate axes unit vectors. Consequently, the matrices σ_a, σ_b, σ_c may be represented by the Pauli spin matrices

$$\sigma_a=\pi_1,\quad \sigma_b=\pi_2,\quad \sigma_c=\pi_3. \tag{33}$$

Then, Eq. (32) may be inverted to obtain the representation associated with the coordinate axes,

$$\sigma_n=\hat{n}\cdot\hat{a}\sigma_a+\hat{n}\cdot\hat{b}\sigma_b+\hat{n}\cdot\hat{c}\sigma_c, \tag{34}$$

where $n=x,y,z$ and $\hat{n}=\hat{x}$, $\hat{y}$, $\hat{z}$. The effect has been to choose the Pauli spin matrices to correspond to particular directions in space other than the coordinate axes.

Double group operators simultaneously transform coordinate space and spin space. They may be represented as $R_{nam}={}^{s}R_{nam}{}^{s}R_{nam}$. The tetrahedral group T has one normal subgroup $V=(E,C_{x2},C_{y2},C_{z2})$. The coset expansion of T with respect to V is $T=C_3V$ where the coset generators may be chosen to be the group $C_3=(E,T_{2w3},T_{4w3})$. In the case of T, the operator $T_{4w3}=(T_{2w3})^2=T_{w3}$. The tetrahedral double group DT has two normal subgroups, the trivial subgroup consisting of the identity and rotations by 360° (E,E_2) and the double subroup $DV=(E,C_{x2},C_{y2},C_{z2})$ (E,E_2). The coset expansion of DT with respect to DV is $DT=C_3DV$. In the case of DT the operator $T_{4w3}=(T_{2w3})^2=E_2T_{w3}$. Again C_3 is a group. Note that in Table I are listed all the operators for T plus E_2. The remaining operators for DT are obtained from the tabulated operators by multiplication by E_2.

Having defined the groups and characterized some aspects of their structure it remains to exhibit the commutation properties of selected operators before the induction-factorization methods may be used to obtain ISO's. Only those operators exhibited in the coset decomposition are needed. These are listed in Table II, Part A. The properties listed apply to the group T if E_2 is set to unity.

In seeking the ISO's for the group T the subgroups V and C_3 will be considered first. Group V is the direct product of normal subgroups of V, (E,C_{x2}), and (E,C_{y2}). Hence, the ISO's are immediately found to be

$$P(V_0)=\tfrac{1}{4}(E+C_{x2})(E+C_{y2}), \tag{35}$$
$$P(V_1)=\tfrac{1}{4}(E+C_{x2})(E-C_{y2}), \tag{36}$$
$$P(V_2)=\tfrac{1}{4}(E-C_{x2})(E+C_{y2}), \tag{37}$$
$$P(V_3)=\tfrac{1}{4}(E-C_{x2})(E-C_{y2}). \tag{38}$$

The group C_3 is a cyclic group of order three. The ISO's are

$$P(W_0)=(E+T_{2w3}+T_{4w3})/3, \tag{39}$$
$$P(W_1)=(E+\omega T_{2w3}+\omega^2T_{4w3})/3=P(W_2)^*, \tag{40}$$
$$P(W_2)=(E+\omega^2T_{2w3}+\omega T_{4w3})/3=P(W_1)^*, \tag{41}$$

where $\omega=\exp(i2\pi/3)$ is a cube root of unity. Next the conjugates ISO's for V are checked for possible induction. It is found using Table II, Part A that $T_{2w3}P(V_0)=P(V_0)\,T_{2w3}$, but $T_{4w3}P(V_1)\,T_{2w3}=P(V_2)$ and $T_{2w3}P(V_1)\,T_{4w3}=P(V_3)$. Inducing with respect to $P(V_0)$ produces reducible symmetry operators while inducing with respect to $P(V_1)$ gives an IUR by the irreducibility theorem. Now the operator set $C_3P(V_0)$ is isomorphic to C_3. Hence, the ISO's for T include three one-dimensional ISO's,

$$P(T_0) = P(W_0)P(V_0), \tag{42}$$

$$P(T_1) = P(W_1)P(V_0), \tag{43}$$

$$P(T_2) = P(W_2)P(V_0), \tag{44}$$

and one three-dimensional IUR with ISO's,

$$\begin{aligned} &P(T_3)_{11} = P(V_1), \quad P(T_3)_{12} = T_{4w3}P(T_3)_{11}, \\ &P(T_3)_{13} = T_{2w3}P(T_3)_{11}, \quad P(T_3)_{21} = T_{2w3}P(T_3)_{22}, \\ &P(T_3)_{22} = P(V_2), \quad P(T_3)_{23} = T_{4w3}P(T_3)_{22}, \\ &P(T_3)_{31} = T_{4w3}P(T_3)_{33}, \quad P(T_3)_{32} = T_{2w3}P(T_3)_{33}, \\ &P(T_3)_{33} = P(V_3). \end{aligned} \tag{45}$$

In treating the double group DT, the normal subgroups (E, E_2) and DV will be considered first. The ISO's for (E, E_2) are

$$Q_0 = \tfrac{1}{2}(E + E_2), \; Q_1 = \tfrac{1}{2}(E - E_2). \tag{46}$$

Both Q_0 and Q_1 commute with all operators of DT, and neither operator can be considered for induction. The coset expansion of DT with respect to (E, E_2) is $DT = C_T(E, E_2)$ where the set of coset generators is conveniently chosen to be the double group operators of T as listed in Table I. The operator sets $DT_1 = C_T Q_0$ and $DT_1 = C_T Q_1$ are orthogonal in the sense that any operator of DT_0 multiplied by an operator of DT_1 is zero because $Q_0 Q_1 = 0$. Since Q_0 is even with respect to E_2 the elements of DT_0 are isomorphic to T and the ISO's for DT_0 have the same form as the ISO's for T. In physical applications $E_2 = -E$ and $Q_0 = 0$, so these ISO's for DT are physically trivial. Of course, this situation is well known and always occurs with double groups. The operator set DT_1 is a multiplier group which submits to the coset decomposition $DT_1 = C_3 DV_1$, where $DV_1 = (E, C_{x2}, C_{y2}, C_{z2})Q_1$. The analogy to the treatment of T stops upon observing that although $DV_1 = ((E, C_{x2})Q_1)((E, C_{y2})Q_1)$ is the product of two multiplier groups, the ISO's cannot be expressed as the product of the ISO's for the multiplier groups. In fact the induction-based construction breaks down for this case. It is easy to find ISO's for DV_1,

$$\begin{aligned} &P(V_4)_{11} = \tfrac{1}{2}(E + iC_{z2})Q_1, \quad P(V_4)_{12} = C_{y2}P(V_4)_{11}, \\ &P(V_4)_{21} = -C_{y2}P(V_4)_{22}, \quad P(V_4)_{22} = \tfrac{1}{2}(E - iC_{z2})Q_1. \end{aligned} \tag{47}$$

The IUR of DV obtained from Eq. (47) is identical to that obtained from the Pauli spin matrices, Eqs. (28) and (31). The conjugate projection operators formed with respect to C_3 are not simple and induction fails.

However, it is possible to choose an IUR of DV_1 equivalent to that obtained from Eq. (47) which allows factored ISO's for DT. What is needed is an IUR with projection operators which commute with the operators of C_3. From the commutation properties in Table II, Part A it is seen that an operator proportional to $(C_{x2} + C_{y2} + C_{z2})$ will commute with C_3. Since double group operators are isomorphic to the corresponding spin-space operators, the construction of alternative spin representations described in connection with Eqs. (32)–(34) suggests the definition of the following "hybrid" double group operators:

$$C_{wd} = (-2C_{x2} + C_{y2} + C_{z2})/\sqrt{6}, \tag{48}$$

$$C_d = (-C_{y2} + C_{z2})/\sqrt{2}, \tag{49}$$

$$C_w = (C_{x2} + C_{y2} + C_{z2})/\sqrt{3}. \tag{50}$$

The multiplication properties of the "hybrid" operators are isomorphic to the operators, C_{x2}, C_{y2}, C_{z2}, as is shown in Table II, Part B. Of course, the above choice of "hybrid" operators is not unique. In general the coefficients of operators C_{x2}, C_{y2}, C_{z2}, must be orthogonal, real unit vectors in 3-space. The alternative ISO's for DV may be obtained by replacing C_{z2} with C_w and C_{y2} with C_d in Eqs. (47). In this IUR the hybrid operator matrix representatives are identical to those obtained from the Pauli spin matrices. The matrix representatives for C_{x2}, C_{y2}, and C_{z2} are most easily obtained by inverting Eqs. (48)–(50).

The form chosen for the hybrid operators allows the operator set DT_1 to be rearranged into a subset of commuting operators, $DW = (E, C_w)Q_1 C_3$, which is a multiplier group.

The ISO's for DW are

$$P(n, m) = \tfrac{1}{2}(E + imC_w)Q_1 P(W_n), \tag{51}$$

where $n = 0, 1, 2$ and $m = +1$ or -1. The ISO's have the property

$$P(n, m)P(n', m') = P(n, m)\delta_{nn'}\delta_{mm'}. \tag{52}$$

The operators of the rearranged set $DT_1 = (E, C_d)DW$ form a group algebra, but do not form a group or multiplier group. Nevertheless, the ISO's may be constructed as in the case of induction with respect to a subgroup,

$$\begin{aligned} &P(T_4)_{11} = P(0, +1), \quad P(T_4)_{12} = C_d P(T_4)_{11}, \\ &P(T_4)_{21} = -C_d P(T_4)_{22}, \quad P(T_4)_{22} = P(1, -1), \end{aligned} \tag{53}$$

$$\begin{aligned} &P(T_5)_{11} = P(1, +1), \quad P(T_5)_{12} = C_d P(T_5)_{11}, \\ &P(T_5)_{21} = -C_d P(T_5)_{22}, \quad P(T_5)_{22} = P(2, -1), \end{aligned} \tag{54}$$

$$\begin{aligned} &P(T_6)_{11} = P(2, +1), \quad P(T_6)_{12} = C_d P(T_6)_{11}, \\ &P(T_6)_{21} = -C_d P(T_6)_{22}, \quad P(T_6)_{22} = P(0, -1). \end{aligned} \tag{55}$$

In performing the inductionlike process, conjugate operators were calculated using the relations in Table II, Part B. The process may be extended easily to obtain ISO's for the double cubic group, DO, and the double full tetrahedral group including reflection planes, DT_d.

ACKNOWLEDGMENT

In conclusion, I wish to acknowledge my intellectual debt to Professor J.M. Keller who first introduced me to the notion of factored irreducible symmetry operators and gave me an appreciation of their utility in simplifying the application of group theoretical methods to physical problems.

[1] E.P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959), p. 118.
[2] M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley, Reading, Mass., 1962), pp. 111–13.
[3] M. Tinkham, *Group Theory and Quantum Mechanics* (McGraw-Hill, New York, 1964), pp. 40–43.
[4] L. Jansen and M. Boon, *Theory of Finite Groups* (Wiley, New York, 1967), pp. 223–33.
[5] M. Lax, *Symmetry Principles in Solid State and Molecular Physics* (Wiley, New York, 1974), pp. 99–101.
[6] M.A. Melvin, Rev. Mod. Phys. 28, 18 (1956).

The Wegner approximation of the plane rotator model as a massless, free, lattice, Euclidean field

J. Bricmont and F. Debacker-Mathot

Institut de Physique Théorique, Université de Louvain, 1348 Louvain-la-Neuve, Belgium
(Received 18 June 1976)

The approximate plane rotator model proposed by Wegner can be reinterpreted as a massless, free, lattice, Euclidean field model. Using this approach, we compute the two-point correlation function and the magnetization for the spin model as functions of the covariance of the fields.

INTRODUCTION

Stanley, Kaplan, and others[1] have shown on the basis of high-temperature expansion that the susceptibility of the two-dimensional plane rotator model is infinite below some critical temperature. As shown by Mermin,[2] this phenomenon cannot be due to spontaneous magnetization. As an explanation, Wegner[3] (see also Ref. 4) suggested an approximation of the plane rotator model in which he can compute explicitly the two-point function. His result agrees with what is known in one and three dimensions[7] and with infinite susceptibility at low temperatures in two dimensions.

Here we simply reinterpret this approximation: It is equivalent to a critical Gaussian spin model or to a massless free Euclidean (Bose) field on a lattice. Although the latter is not well defined, some quantities as differences of covariances remain well-defined when one passes from a massive field to a massless field (Sec. 2). We compute then the two-point function for the rotator model in the Wegner approximation in terms of differences of covariances of Euclidean fields. That allows us to easily recover Wegner's results (Sec. 3). When an external field is added, the model corresponds to a *massive* lattice field; in this case we compute the magnetization in terms of the covariance of the Euclidean field (Sec. 4).

I. THE PLANE ROTATOR MODEL AND THE WEGNER APPROXIMATION

To each point i of a d-dimensional lattice $\mathbb{Z}^d$, we attach a classical two-component spin of unit length, i.e., a vector: $\mathbf{s}_i=(\cos\phi_i,\sin\phi_i)$ $(i\in\mathbb{Z}^d)$. We consider a system of N such spins interacting with ferromagnetic, nearest neighbor interactions,

$$H=-J\sum_{\langle i,j\rangle}(\mathbf{s}_i\cdot\mathbf{s}_j)=-J\sum_{\langle i,j\rangle}\cos(\phi_i-\phi_j), \tag{1}$$

where $J>0$ and the sum is extended to all nearest neighbor lattice sites. (We use periodic boundary conditions.)

For any continuous function $f(\phi_1\cdots\phi_N)$, the average given by this model is

$$\langle f\rangle=\frac{\int_{-\pi}^{\pi}f(\phi_1\cdots\phi_N)\exp((\beta J)\sum_{\langle i,j\rangle}\cos(\phi_i-\phi_j))\prod^N d\phi_i}{\int_{-\pi}^{\pi}\exp((\beta J)\sum_{\langle i,j\rangle}\cos(\phi_i-\phi_j))\prod^N d\phi_i} \tag{2}$$

with

$\beta=(k_BT)^{-1}$, $k_B=$ Boltzmann constant, $T=$ temperature.

Since the measure in (2) is periodic, one may let $\phi_1\cdots\phi_N$ run from $-\infty$ to $+\infty$ and consider f as a periodic function in each variable separately. So in (2) we may change the integration limits to $-\infty$ and $+\infty$ both in the numerator and in the denominator simultaneously.

The approximation proposed by Wegner consists in assuming that for very low temperature, the difference $\phi_i-\phi_j$ between *neighboring* spins is very small. So in (1) we can expand the cosine up to second order: $\cos(\phi_i-\phi_j)\sim 1-\frac{1}{2}(\phi_i-\phi_j)^2$.

In that approximation, the average (2) becomes

$$\langle f\rangle=\frac{\int_{-\infty}^{\infty}f(\phi_1\cdots\phi_N)\exp((-\beta J/2)\sum_{\langle i,j\rangle}(\phi_i-\phi_j)^2)\prod^N d\phi_i}{\int_{-\infty}^{\infty}\exp((-\beta J/2)\sum_{\langle i,j\rangle}(\phi_i-\phi_j)^2)\prod^N d\phi_i}. \tag{3}$$

The denominator of (3) is the "partition function" Z_N^W of the Wegner model. It can also be written as

$$Z_N^W(\beta)=\int_{-\infty}^{\infty}\exp\left((-2d\beta J)\sum_{i=1}^N\phi_i^2+\beta J\sum_{\langle i,j\rangle}\phi_i\phi_j\right)\prod d\phi_i \tag{4}$$

which shows that the ϕ_i's are now Gaussian variables.

So the classical plane rotator model is approximated by a lattice Gaussian spin model. In fact, this Gaussian model is singular as we shall see in the following section.

II. MASSLESS EUCLIDEAN FIELDS

Let us begin with a massive free Euclidean field Φ of mass m and consider its lattice approximation Φ_i (the field Φ at the lattice point i[5]). (We take a unit lattice spacing, so the lattice is $\mathbb{Z}^d$). The measure associated to this lattice field is

$$\exp\left(-(2d+m^2)\sum_{i=1}^N\Phi_i^2+\sum_{\langle i,j\rangle}\Phi_i\Phi_j\right)\prod^N d\Phi_i. \tag{5}$$

If we express this in usual notation for Gaussian spin models[6] by letting $\Phi_i=q_i(4d+2m^2)^{-1/2}$ we get

$$\exp\left(-\tfrac{1}{2}\sum_{i=1}^N q_i^2+K\sum_{\langle i,j\rangle}q_iq_j\right)\prod^N dq_i \tag{6}$$

with $K=(4d+2m^2)^{-1}$.

For the particular value $K=(4d)^{-1}$ which corresponds to a massless Euclidean field, the respective measures can be rewritten as

$$\exp\left(-\frac{1}{8d}\sum_{\langle i,j\rangle}(q_i-q_j)^2\right)\prod^N dq_i$$

for the spins and

$$\exp\left(-\tfrac{1}{2}\sum_{\langle i,j\rangle}(\Phi_i-\Phi_j)^2\right)\prod^N d\Phi_i \tag{7}$$

for the massless Euclidean fields.

Comparing (7) and the denominator of (3) and putting $\phi_i=\Phi_i(\beta J)^{-1/2}$ we see that the Wegner approximation of the plane rotator model at *any* temperature β is equivalent to a model of free massless lattice Euclidean field [or to a Gaussian spin model at $K=(4d)^{-1}$].

Now, Berlin and Kac[6] have shown that Gaussian spin models cannot be defined for all values of K. Indeed the quadratic form appearing in (6) must be negative definite for the finite volume partition function to exist; this is the case for $K<(4d)^{-1}$. For such values of K, these authors compute the free energy in the infinite volume limit.[6] Moreover, the resulting expression (given by an integral) has only integrable singularities at $K=(4d)^{-1}$, thus the free energy still exists for $K=(4d)^{-1}$. However, for $K>(4d)^{-1}$ it becomes complex so that the model is no longer defined.

Thus $K=(4d)^{-1}$ is a critical value in the sense that the infinite volume limit free energy exists but the partition function (in a finite volume) does not. This is manifested in the expression (7): The quadratic form in the exponent of the measure is still negative but no longer negative definite: It vanishes when all the Φ_i's are equal.

However, from this N-spin singular model, we shall construct an $(N-1)$-spin nonsingular model: Consider the noncritical Gaussian model. The quadratic form appearing in (6) can be diagonalized by an orthogonal transformation[6] $t_i=\sum_{j=1}^N V_{ij}q_j$ where

$$V_{ij}=\frac{1}{\sqrt{N}}\left(\cos\frac{2\pi}{N}(i-1)(j-1)+\sin\frac{2\pi}{N}(i-1)(j-1)\right)$$

and we obtain as a new measure

$$\exp\left(-\sum_{i=1}^N(\tfrac{1}{2}-K\lambda_i)t_i^2\right)\prod_{i=1}^N dt_i. \tag{8}$$

The λ_i's depend on the dimension. They are computed explicitly in Ref. 6 but the only thing we need here is that for *all* dimensions, $\lambda_1=2d$ and $\lambda_i<2d$ for $i=2,\ldots,N$. Then when K tends to the critical value $(4d)^{-1}$, $K\lambda_1$ tends to $\frac{1}{2}$ and the coefficient of the variable t_1 in (8) vanishes. So the diagonalized form of the critical Gaussian measure is

$$dt_1\cdot\exp\left[-\sum_{i=2}^N\left(\frac{1}{2}-\frac{\lambda_i}{4d}\right)t_i^2\right]dt_2\cdots dt_N. \tag{9}$$

This explains in what sense the model is singular. The partition function is infinite because of the t_1 integral but the behavior in the $N-1$ other variables is a usual noncritical Gaussian model. When we compute average values of functions depending on $t_2,\ldots,t_N$ only, the t_1 integrals cancel in the numerator and the denominator. In particular, the differences q_i-q_j do not depend of t_1 (because of the explicit form of the V_{ij}'s). Thus q_i-q_j (resp. $\Phi_i-\Phi_j$) is a well-defined Gaussian random variable in the critical Gaussian model (resp. in the massless field model). That is not true for q_i or Φ_i themselves.

III. THE SPIN CORRELATION FUNCTION

The two-point function of the rotator model depends on differences $\phi_i-\phi_j$; therefore, it can be computed in a very simple way, using the equivalence between the Wegner approximation and the massless Euclidean free field model derived in Sec. 2.

$$\begin{aligned}\langle \mathbf{s}_i\cdot\mathbf{s}_j\rangle &= \lim_{N\to\infty}\langle\cos(\phi_i-\phi_j)\rangle_N=\lim_{N\to\infty}\left\langle\cos\left(\frac{\Phi_i-\Phi_j}{\sqrt{\beta J}}\right)\right\rangle_{N,m=0}\\ &=\lim_{N\to\infty}\left\langle\exp\left[i\left(\frac{\Phi_i-\Phi_j}{\sqrt{\beta J}}\right)\right]\right\rangle_{N,0}\\ &=\exp\left(-\frac{1}{2\beta J}\lim_{N\to\infty}\langle(\Phi_i-\Phi_j)^2\rangle_{N,0}\right)\\ &=\exp\left(-\frac{1}{\beta J}\lim_{N\to\infty}\lim_{m\to0}(\langle\Phi_i^2\rangle_{N,m}-\langle\Phi_i\Phi_j\rangle_{N,m})\right)\\ &=\exp\left(-\frac{1}{\beta J}\lim_{N\to\infty}\lim_{m\to0}(C_{ii}-C_{ij})_{N,m}\right),\end{aligned} \tag{10}$$

where $C_{ij}=\langle\Phi_i\Phi_j\rangle_{N,m}$ is the covariance of the massive free Euclidean field on a lattice, with periodic boundary conditions. An explicit computation (see the Appendix) gives the following result:

$$(C_{ii}-C_{ij})_{N,m}=\frac{1}{N}\sum_{k=k(1)}^{k(N)}\frac{1-\exp[i\mathbf{k}(\mathbf{i}-\mathbf{j})]}{2d+m^2-2\sum_{a=1}^d\cos k_a}$$

Taking now the limit $m\to0$ and then the infinite volume limit $(N\to\infty)$ we get

$$\langle\mathbf{s}_i\cdot\mathbf{s}_j\rangle=\exp\left(-\frac{1}{\beta J\pi^d}\int_0^\pi\frac{\sin^2(\mathbf{k}\cdot\mathbf{r}/2)}{4\sum_{a=1}^d\sin^2(k_a/2)}dk_1\cdots dk_d\right) \tag{11}$$

$(\mathbf{r}=\mathbf{i}-\mathbf{j})$.

This coincides with the result of Wegner. The integral (11) may be evaluated for $d=1,2,3$.[3]

$$\left.\begin{aligned}&\text{For } d=1,\quad \langle\mathbf{s}_i\cdot\mathbf{s}_j\rangle\sim\exp(-|\mathbf{r}|/2\beta J)\\ &\text{for } d=2,\quad \langle\mathbf{s}_i\cdot\mathbf{s}_j\rangle\sim|\mathbf{r}|^{-1/2\pi\beta J}\end{aligned}\right\}\text{for large } |\mathbf{r}|.$$

This quantity tends to zero very slowly as $|\mathbf{r}|\to\infty$ and therefore the susceptibility $\chi\sim(1/2T)\sum_{\mathbf{r}}|\mathbf{r}|^{-k_B/2\pi J}$ diverges for T sufficiently small.

$$\text{For } d=3,\quad \lim_{|\mathbf{i}-\mathbf{j}|\to\infty}\langle\mathbf{s}_i\cdot\mathbf{s}_j\rangle=\exp(-Cst/\beta J)\neq0.$$

Thus we have long range order in three dimensions (it is also true for higher dimensions).

IV. THE MODEL WITH AN EXTERNAL MAGNETIC FIELD

Consider now the plane rotator model with an external magnetic field. The measure is given by

$$\exp\left(\beta J\sum_{\langle ij\rangle}\cos(\phi_i-\phi_j)+\beta h\sum_{i=1}^N\cos\phi_i\right)\prod^N d\phi_i. \tag{12}$$

Let us see what happens if we expand not only $\cos(\phi_i-\phi_j)$ but also $\cos\phi_i$ in series up to second order, assuming that in an external field, the spins tend to take

the direction of the field ($\phi_i \ll 1$). This assumption is questionable in the limit $h \to 0$ but it allows us to find results coherent with those of Sec. 3 in particular for the value of the spontaneous magnetization in three dimensions.

With this approximation we have as a new measure

$$\exp\left(-\beta(2dJ+h)\sum_{i=1}^{N}\phi_i^2+\beta J\sum_{(ij)}\phi_i\phi_j\right)\prod^{N}d\phi_i,$$

or in terms of Euclidean fields

$$\exp\left(-(2d+h/J)\sum_{i=1}^{N}\Phi_i^2+\sum_{(ij)}\Phi_i\Phi_j\right)\prod^{N}d\Phi_i. \tag{13}$$

This measure describes a free lattice Euclidean field with mass $m=\sqrt{h/J}>0$; the addition of a magnetic field in the spin model corresponds to the addition of a mass in the Euclidean field model. We can now compute quantities depending on the Φ_i's themselves and not only on differences of Φ_i's as before, for instance, the magnetization

$$\lim_{N\to\infty}\langle\cos\phi_i\rangle_{N,h}$$

$$=\lim_{N\to\infty}\left\langle\cos\frac{\Phi_i}{\sqrt{\beta J}}\right\rangle_{N,m}=\lim_{N\to\infty}\left\langle\exp\left[i\left(\frac{\Phi_i}{\sqrt{\beta J}}\right)\right]\right\rangle_{N,m}$$

$$=\lim_{N\to\infty}\exp\left(-\frac{1}{2\beta J}\langle\Phi_i^2\rangle_{N,m}\right)=\exp\left(-\frac{1}{2\beta J}C_{ii}\right), \tag{14}$$

where C_{ii}, the covariance of the free lattice Euclidean field with mass m, is given by

$$C_{ii}=\frac{1}{\pi^d}\int_0^{\pi}\frac{1}{m^2+4\sum_{a=1}^{d}\sin^2(k_a/2)}\,dk_1\cdots dk_d$$

(see Appendix).

Now let the magnetic field go to zero. This corresponds to $m\to 0$ in the field model; in that limit C_{ii} diverges in one and two dimensions and converges for $d\geq 3$. (This is only true if we take the limit $m\to 0$ after the infinite volume limit. As can be seen in the Appendix the finite volume $m=0$ covariance is infinite in any dimension because of the $k=0$ term in the sum.)

Thus

$$\lim_{h\to 0}\langle\cos\phi_i\rangle_h=0 \quad \text{for } d=1,2,$$
$$\neq 0 \quad \text{for } d=3,4,\cdots.$$

So in three and more dimensions, there is a discontinuity in the magnetization at $h=0$, since by symmetry $\langle\cos\phi_i\rangle_{h=0}=0$ in all dimensions.

The value of the spontaneous magnetization in three dimensions equals the square root of the limit of $\langle\cos(\phi_i-\phi_j)\rangle$ when $|i-j|\to\infty$. This follows from Eqs. (10) and (14) and $\lim_{|i-j|\to\infty}C_{ij}=0$ by the Riemann–Lebesgue lemma. (See in the Appendix, C_{ij} with $m=0$ and $d=3$, $1/(6-2\sum_{a=1}^{3}\cos k_a)$ is integrable for $k_a\in[-\pi,\pi]$.)

APPENDIX: COVARIANCE OF THE FREE LATTICE EUCLIDEAN FIELD WITH PERIODIC BOUNDARY CONDITIONS

Consider the massive free Euclidean field on the lattice $\mathbb{Z}^d$, i.e.,[5] the Gaussian random process Φ_i indexed by $\mathbb{Z}^d$ with covariance $\langle\Phi_i\Phi_j\rangle=(-\Delta_1+m^2)^{-1}$, where Δ_1 is the finite difference operator which approximates the Laplace operator Δ in d dimensions.

Δ_1 is defined on the space $l^2(\mathbb{Z}^d)$ of square-summable sequences on $\mathbb{Z}^d$ by

$$(\Delta_1 f)_i=-2df_i+\sum_{j\,|i-j|=1}f_j.$$

The operator $A=-\Delta_1+m^2$ is a convolution operator with kernel

$$a(i-j)=\begin{cases}2d+m^2 & \text{if } i=j,\\ -1 & \text{if } |i-j|=1,\\ 0 & \text{otherwise.}\end{cases}$$

In order to describe the field with periodic boundary conditions we consider the subspace $l^2_{PN}(\mathbb{Z}^d)$ of $l^2(\mathbb{Z}^d)$ consisting in sequences periodic in each direction, the period being a cube of volume N. The periodic field is defined as before but we replace Δ_1 by Δ_1^P, the restriction of Δ_1 to $l^2_{PN}(\mathbb{Z}^d)$.

The Fourier transformation: $\hat{f}_k\equiv\sum_{j=1}^{N}f_j\exp(-ik\cdot j)$ is a unitary map from $l^2_{PN}(\mathbb{Z}^d)$ onto the space $\widehat{l^2_{PN}}$ of periodic square-summable sequences on the reciprocal lattice. In the first Brillouin zone, i.e., $[-\pi,\pi]^d$, there are exactly N allowed values of k denoted by $k^{(1)}\cdots k^{(N)}$ which are symmetrical around 0.

In $\widehat{l^2_{PN}}$ the Fourier transform of A is defined by

$$(\hat{A}\hat{f})_k\equiv\widehat{(Af)}_k.$$

$\hat{A}$ is the operator of multiplication by

$$\hat{a}_k=\sum_{j}^{N}a(j)\exp(-ik\cdot j)=2d+m^2-2\cos k_1$$
$$-2\cos k_2\cdots-2\cos k_d.$$

Now the covariance $C=A^{-1}$ of the field is a convolution operator in $l^2_{PN}(\mathbb{Z}^d)$ so that $\hat{C}=\hat{A}^{-1}$ is the operator of multiplication by $(\hat{a}_k)^{-1}$. Then by the inverse Fourier transform,

$$\langle\Phi_i\Phi_j\rangle=C_{ij}=\frac{1}{N}\sum_{k=k^{(1)}}^{k^{(N)}}\frac{\exp[ik\cdot(i-j)]}{2d+m^2-2\sum_{a=1}^{d}\cos k_a}.$$

When $N\to\infty$, the number of allowed values of k increases, so that in the limit k covers $[-\pi,\pi]^d$. Thus in the thermodynamical limit,

$$C_{ij} = \frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} \frac{\exp[ik\cdot(i-j)]}{2d + m^2 - 2\sum_{a=1}^{d} \cos k_a} d^d k$$

$$= \frac{1}{\pi^d} \int_{0}^{\pi} \frac{\cos(k\cdot(i-j))}{2d + m^2 - 2\sum_{a=1}^{d} \cos k_a} d^d k.$$

ACKNOWLEDGMENTS

We would like to thank J. R. Fontaine and Professor L. Landau for very interesting discussions and Dr. J.-P. Antoine for reading the manuscript.

[1] H. E. Stanley, *Introduction to Phase Transition and Critical Phenomena* (Clarendon, Oxford, 1971), and references therein.
[2] N. Mermin, J. Math. Phys. 8, 1061–64 (1967).
[3] F. Wegner, Z. Phys. **206**, 465–70 (1967).
[4] V. L. Berezinski, Zh. Eksp. Teor. Fiz. **59**, 907–20 (1970) [Sov. Phys. JETP **32**, 493–500 (1971)].
[5] B. Simon, *The $P(\phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U.P., Princeton, N.J., 1974).
[6] T. H. Berlin and M. Kac, Phys. Rev. **86**, 821–35 (1952).
[7] J. Fröhlich, B. Simon, and T. Spencer, Commun. Math. Phys. **50**, 79–85 (1976).

The positive energy conjecture and the cosmic censor hypothesis

Pong Soo Jang* and Robert M. Wald

Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637
(Received 28 June 1976;revised manuscript received 23 August 1976)

We show that a positive energy argument of Geroch can be modified to rule out a possible class of counterexamples to the cosmic censor hypothesis proposed by Penrose.

1. INTRODUCTION

In general relativity, global quantities such as total energy or total angular momentum of an isolated system do not arise as naturally as in special relativity. The essential reason why this is so is that in curved spacetime one does not have the action of the Poincaré group as a group of isometries. Nonetheless, if the spacetime is asymptotically flat, such quantities can be defined in general relativity as surface integrals in the asymptotic region. However, these global quantities cannot, in general, be converted into volume integrals over the interior region. Without such a direct relationship between the global and local quantities, it is difficult to determine any restrictions on the properties of the global quantities.

This situation leads one to speculate on the rather exotic possibility of extracting an unlimited amount of energy from an isolated system.[1] First, such a possibility arises if a nonsingular, asymptotically flat spacetime with physically reasonable matter (positive local energy density) is allowed to have negative total energy. Second, such a possibility arises if one can create a naked singularity starting with physically reasonable, nonsingular initial conditions. (Roughly speaking, a naked singularity is a singularity which can be "seen" from infinity. One cannot predict what will come out of a singularity, and thus one would not expect there to be a limit on energy extraction if a naked singularity were present.) However, there is evidence for the following conjecture: Spacetimes with either of the above described properties cannot occur in general relativity. In the first case this is known as the positive energy conjecture; in the second case, it is known as the cosmic censor hypothesis.

The common feature of the possibility of unlimited energy extraction suggests that these two conjectures might be related. In particular, it is even conceivable that the truth of the conjectures—if, indeed, they are true!—might just be two different aspects of one structural feature of general relativity. In this paper, we will find some evidence that this might be the case. We shall show that a slight modification of Geroch's[1] argument for the positive energy conjecture rules out a class of counter-examples to the cosmic censor hypothesis proposed by Penrose.[2]

In Sec. 2, we shall state the positive energy conjecture. Geroch's argument indicating that the total energy must always be positive as reviewed in Sec. 3. In Sec. 4 we describe a class of possible counterexamples to the cosmic censor hypothesis. We show in Sec. 5 that in the time symmetric case, these counterexamples are ruled out by a modification of Geroch's positive energy argument.

2. THE POSITIVE ENERGY CONJECTURE

An initial data set for a spacetime consists of a three-dimensional manifold S on which there are given a positive-definite metric q_{ab}, a symmetric tensor field p^{ab} (the extrinsic curvature), a local mass density μ, and a local current density J^a. These fields on S must obey the constraint equations

$$R - p^{ab}p_{ab} + p^2 = 2\mu, \tag{1}$$

$$D_a(p^{ab} - pq^{ab}) = J^b, \tag{2}$$

where R is the Ricci scalar of the metric q_{ab}, p is the trace of p^{ab}, and D_a is the covariant derivative operator with respect to the metric q_{ab}. Furthermore, μ and J^a must satisfy the local energy condition

$$\mu \geq (J^a J_a)^{1/2}. \tag{3}$$

An initial data set is said to be asymptotically flat if in the asymptotic region the metric q_{ab} approaches the Euclidean metric not slower than $1/r$ and $p^{ab}p_{ab}$ approaches zero not slower than $1/r^4$, where r is any typical radial distance. (A more precise definition of asymptotic flatness at spatial infinity is given by Geroch.[3]) From now on, an initial data set is always assumed to be asymptotically flat.

With any (asymptotically flat) initial data set there is associated a number E whose physical interpretation is the total mass-energy of the system, including the contributions both from matter and gravitational field. This number E is called the Arnowitt–Deser–Misner (ADM) energy. E is obtained by performing a flux integral over a topological 2-sphere in the asymptotically distant region. Although there are various forms of this flux integral, they all give the same value for an asymptotically flat initial data set. The most convenient form of this integral for our purpose is the following:

$$E = \frac{A^{1/2}}{64\pi^{3/2}} \int (2\tilde{R} - \tilde{p}^2)\, dA, \tag{4}$$

where A is the area of the integration surface which asymptotically approaches a metric sphere in the asymptotic region, $\tilde{R}$ is the intrinsic scalar curvature of the integration surface, and $\tilde{p}$ is the trace of the extrinsic curvature of the integration surface as a submanifold of S.

We now state the positive energy conjecture:

Positive energy conjecture: For a nonsingular asymptotically flat initial data set (i. e., all fields on S are smooth and S is complete in the Riemannian metric q_{ab}), the ADM energy E is nonnegative and vanishes if and only if the initial data set is that of Minkowski space.

3. GEROCH'S POSITIVE ENERGY ARGUMENT

Geroch[1] has given an argument to establish the validity of the positive energy conjecture for an initial data set with S having topology $\mathbb{R}^3$ and with $p=0$, i. e., vanishing trace of extrinsic curvature. In this case, Eqs. (1) and (3) imply that $R \geq 0$, and this is the only fact derivable from Eqs. (1)–(3) which we shall need.

We introduce a function t on S such that the two-dimensional surfaces $t=\text{const}$ in S are nested topological 2-spheres with the innermost surface reducing to a point. For each value of t, set

$$f(t)=\int (2\tilde{R}-\tilde{p}^2)\,dA, \tag{5}$$

where the integration extends over the surface $t=\text{const}$ and $\tilde{R}$ and $\tilde{p}$ denote the scalar curvature and the trace of the extrinsic curvature of the surface $t=\text{const}$, respectively. Note that by the Gauss–Bonnet theorem

$$\int \tilde{R}\,dA=8\pi. \tag{6}$$

We define a scalar field ϕ on S by $\phi r^a D_a t=1$, where r^a denotes the unit outward vector field normal to the 2-surfaces. Then, using the Gauss–Codazzi equation and the Gauss–Bonnet equation, the rate of change of $f(t)$ with respect to t is[1]

$$\frac{d}{dt}f(t)=\int [2\tilde{p}\tilde{D}^a\tilde{D}_a\phi+\phi\tilde{p}\tilde{p}^{ab}\tilde{p}_{ab}-\phi\tilde{p}\tilde{R}+\phi\tilde{p}R]\,dA, \tag{7}$$

where $\tilde{D}_a$ is the covariant derivative operator on the 2-surfaces with respect to the induced metric. Next, we choose[1] the 2-surfaces such that

$$\phi\tilde{p}=1. \tag{8}$$

Then, Eq. (7) becomes

$$\frac{d}{dt}f(t)=-\tfrac{1}{2}f(t)+\int [R+(\tilde{p}^{ab}\tilde{p}_{ab}-\tfrac{1}{2}\tilde{p}^2)+2\phi^{-2}\tilde{D}_a\phi\tilde{D}^a\phi]\,dA$$
$$\geq -\tfrac{1}{2}f(t). \tag{9}$$

But since $f(t)\to 0$ as the surfaces reduce to a point, Eq. (9) implies $f(t)\geq 0$ for all t. By Eq. (4) we have $E\geq 0$. Equality holds (i. e., $E=0$) only if $p^{ab}=0$ and q_{ab} is flat, i. e., initial data for flat space.

The above argument fails to be a full proof of positive energy conjecture (even in the case $S=\mathbb{R}^3$) for two important reasons: (1) The simplifying assumption $p=0$ has been made. However, there is some evidence[4] for believing that all nonsingular, asymptotically flat spacetimes must contain at least one asymptotically flat slice with $p=0$. If this is true, then a proof for the $p=0$ case would prove the positive energy conjecture for all nonsingular spacetimes. Even so, one would still want to establish the positive energy conjecture in the stronger form given above in Sec. 2 (i. e., one would like to prove that an initially regular spacetime must have $E\geq 0$ even if it develops singularities later). (2) At present, no proof has been given that the family of surfaces required in the above argument actually exists. Some intuitive arguments suggest that starting from a point the evolution equation $\phi=\tilde{p}^{-1}$ will always generate a smooth family of surfaces which asymptotically become metric spheres at infinity [so that Eq. (4) applies]. However, this has not been proven. Indeed, when minimal surfaces ($\tilde{p}=0$) are present on S, difficulties with the surface evolution can occur, though the modification of the argument given below in Sec. 5 appears to take care of these difficulties.

4. PENROSE'S PROPOSED COUNTEREXAMPLE TO THE COSMIC CENSOR HYPOTHESIS

Several years ago, Penrose[2] proposed a test of the cosmic censor hypothesis and other commonly held beliefs concerning gravitational collapse. His argument (generalized somewhat) runs as follows: Consider a spacetime in which gravitational collapse takes place. If cosmic censorship is right, the collapse should produce a black hole which should settle down to a stationary final state. From the theorems of Israel,[5] Hawking,[6] and Robinson,[7] the only stationary vacuum black holes are the Kerr solutions. The formula for the area of a Kerr black hole is

$$A_f=8\pi[M^2+(M^4-J^2)^{1/2}]\leq 16\pi M^2, \tag{10}$$

where M is the mass of the black hole and J is its angular momentum.

Consider, now, the initial data for this spacetime on some earlier spacelike slice S (e. g., at a dynamic phase of collapse). By the Hawking area theorem[8] (which is based on cosmic censorship) the initial surface area A_i of the horizon cannot be larger than the final area, so that

$$A_i\leq A_f. \tag{11}$$

Furthermore, since the energy carried off to infinity by radiation cannot be negative, we have

$$M\leq E, \tag{12}$$

where E is the ADM energy of the initial data set. Thus, we obtain

$$A_i\leq 16\pi E^2. \tag{13}$$

This inequality is not very useful since we must know the entire evolution before we can determine the location of the event horizon on S. However, we can obtain a more useful inequality as follows: The apparent horizon $\mathcal{H}$ is defined as the outer boundary of the region of S which contains trapped or marginally trapped surfaces. $\mathcal{H}$ itself must be a marginally trapped surface, and thus it satisfies

$$\tilde{p}+p^{ab}(q_{ab}-r_a r_b)=0, \tag{14}$$

where $\tilde{p}$ is the trace of the extrinsic curvature of $\mathcal{H}$ as a submanifold of S and r^a is the unit outward normal to $\mathcal{H}$ on S. One can show[9] that $\mathcal{H}$ must be a topological 2-sphere (or a disjoint union of spheres) and must necessarily lie inside of (or coincide with) the event horizon. Let $\mathcal{A}$ denote the greatest lower bound of the area of surfaces which enclose $\mathcal{H}$. Then clearly $\mathcal{A}\leq A_i$ and, thus,

$$\mathcal{A}\leq 16\pi E^2. \tag{15}$$

This equation relates quantities which are determined directly by the initial data on S.

Penrose's idea was to find an initial data set which violates Eq. (15). If such a data set exists, something must be seriously wrong with the assumptions which went into the above derivation of the inequality. The cosmic censor hypothesis is by far the weakest link in the above derivation, so if a data set is found which violates Eq. (15), it almost certainly will provide us with a counterexample to the cosmic censor hypothesis. Of course, the validity of Eq. (15) for all initial data sets would not imply that the cosmic censor hypothesis must be true.

Penrose[2] originally proposed the above test of the cosmic censor hypothesis in the context of the collapse of a shell composed of null fluid, with the spacetime flat inside the shell and with the shell coinciding with the apparent horizon at some instant of time. He and Gibbons ruled out the existence of violations of the above inequality in a wide class of cases, though they did not succeed in ruling out all possible counterexamples of this type.

The general context outlined above is difficult to analyze because there is no simple expression for $\mathcal{A}$. This situation improves considerably in the case of time symmetric initial data ($p^{ab}=0$). By Eq. (14), the apparent horizon is a minimal surface, $\tilde{p}=0$. Hence, in this case $\mathcal{A}$ is simply the area of the apparent horizon.

The case of time symmetric, vacuum, initial data with q_{ab} conformally flat was analyzed in detail by Gibbons,[10] who showed that it is very unlikely that one could obtain a violation of Eq. (15) with such data sets. In the next section we show that a modification of the argument of Sec. 3 rules out—subject to the existence of the family of surfaces used in the proof—the possibility of obtaining a violation of Eq. (15) with any time symmetric initial data set whose apparent horizon consists of a single component.

5. NONEXISTENCE OF A CLASS OF PENROSE COUNTEREXAMPLES

Consider a time symmetric initial data set whose apparent horizon H has only one component. As previously remarked, H must have the topology of a sphere. Consider a nested family of 2-spheres analogous to those used in Sec. 3, but now with the property that the surface defined by $t=0$ is H. The family of surfaces for $t>0$ is again defined by the equation

$$\phi\tilde{p}=1. \tag{16}$$

As in Sec. 3, we assume such a family of spheres exists and that as $t\to\infty$ the surfaces asymptotically become metric spheres in the asymptotic region.

Setting

$$f(t)=\int(2\tilde{R}-\tilde{p}^2)\,dA, \tag{17}$$

we find again

$$\frac{df}{dt}\geq-\tfrac{1}{2}f, \tag{18}$$

i.e.,

$$\frac{d}{dt}\left[\exp(t/2)f\right]\geq 0. \tag{19}$$

On the other hand, the rate of change of the area $A(t)$ of the $t=\text{const}$ surfaces is given by

$$\frac{d}{dt}A(t)=\int\tilde{p}\phi\,dA=\int dA=A(t). \tag{20}$$

Thus, we find

$$A(t)=C\exp(t). \tag{21}$$

Since, the surface $t=0$ is just H, we have

$$C=\mathcal{A} \tag{22}$$

where $\mathcal{A}$ is the area of H. Equation (19) now implies

$$\lim_{t\to\infty}\frac{A^{1/2}(t)}{\mathcal{A}^{1/2}}f(t)\geq f(0). \tag{23}$$

Since H is extremal ($\tilde{p}=0$), by the Gauss–Bonnet theorem we have

$$f(0)=16\pi. \tag{24}$$

On the other hand, by Eq. (4), the left-hand side is just

$$\lim_{t\to\infty}\frac{A^{1/2}(t)}{\mathcal{A}^{1/2}}f(t)=\frac{64\pi^{3/2}}{\mathcal{A}^{1/2}}E, \tag{25}$$

where E is the ADM energy. Thus, we obtain

$$\frac{64\pi^{3/2}E}{\mathcal{A}^{1/2}}\geq 16\pi, \tag{26}$$

i.e.,

$$\mathcal{A}\leq 16\pi E^2 \tag{27}$$

which is just Eq. (15)!

Thus, we have found that a modification of Geroch's positive energy argument shows Penrose's inequality, Eq. (15), cannot be violated in this case. An exactly similar modification of Jang's[11] positive energy proof for the spherically symmetric case establishes Penrose's inequality for that case also. It will be interesting to see if further relationships exist between the positive energy conjecture and the cosmic censor hypothesis.[12]

ACKNOWLEDGMENTS

We wish to thank Robert Geroch for many interesting discussions. One of us (P.S.J) wishes to thank Joshua Goldberg for reading an early version of the manuscript. The research reported in this paper has in part been supported by the National Science Foundation under Grant MPS 74-17456 with the University of Chicago.

*Present address: Syracuse University, Department of Physics, 201 Physics Bldg., Syracuse, New York 13210.

[1]R. Geroch, Ann. N.Y. Acad. Sci. **224**, 108 (1973).
[2]R. Penrose, Ann. N.Y. Acad. Sci. **224**, 125 (1973).
[3]R. Geroch, J. Math. Phys. **13**, 956 (1972).
[4]Y. Choquet-Bruhat, C.R. Acad. Sci. Paris **280**, 169 (1975); M. Cantor, A. Fischer, J. Marsden, N.O. O'Murchadha, and J.W. York (to be published).

[5]W. Israel, Phys. Rev. **164**, 1776 (1967); Comm. Math. Phys. **8**, 245 (1968).
[6]S. Hawking, Comm. Math. Phys. **25**, 152 (1972).
[7]D. Robinson, Phys. Rev. Lett. **34**, 905 (1975); B. Carter, Phys. Rev. Lett. **26**, 331 (1971).
[8]S. Hawking, Phys. Rev. Lett. **26**, 1344 (1971).
[9]S. Hawking, in *Black Holes*, edited by Dewitt and DeWitt (Gordon and Breach, New York, 1973).
[10]G. Gibbons, Comm. Math. Phys. **27**, 87 (1972).
[11]P.S. Jang, J. Math. Phys. **17**, 141 (1976).
[12]Gary Gibbons has pointed out to us that the argument of Sec. 5 also applies to an axisymmetric initial data set which is invariant under simultaneous reflection of the axial Killing field and the normal to the hypersurface. In this case the apparent horizon $\mathcal{H}$ is again a minimal surface, $\tilde{p} = 0$, and so again one can show $\mathcal{A} \leq 16\pi E^2$. Actually, since angular momentum J cannot be radiated away in an axisymmetric spacetime, one would like to prove the stronger relation $\mathcal{A} \leq 8\pi[E^2 + (E^4 - J^2)^{1/2}]$ if the spacetime is vacuum outside the apparent horizon; however, we have not succeeded in showing this.

Canonical parameters of the 6*j* coefficient*

Loren A. Lockwood

Departamento de Física, Universidad de Oriente, Cumaná, Venezuela
(Received 18 August 1976)

A set of six canonical parameters is derived for the 6*j* coefficient in a way analogous to that used previously for the 3*j* coefficient. The procedure involves implicitly a twelve-to-one homomorphism which reduces the known 144-element symmetry group to a 12-element group. The symmetries are completely and elegantly described by only five of the six parameters. A further property of the canonical parameters is their complete independence.

In a previous note,[1] a set of canonical parameters was demonstrated for Wigner's 3*j* coefficient. The parameters were shown to possess unique properties, in comparison with other known parameters. In the following, an entirely analogous result will be shown for the 6*j* coefficient. Needless details of the derivation will be omitted, inasmuch as it runs quite parallel to that of Ref. 1.

Wigner's 6*j* coefficient may be considered to be defined by the equation[2]

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ j_4 & j_5 & j_6 \end{Bmatrix} \equiv (6\mathrm{J}) = \Delta(j_1,j_2,j_3)\Delta(j_1,j_5,j_6) \times \Delta(j_2,j_4,j_6)\,\Delta(j_3,j_4,j_5) \times \sum_t (-1)^t (t+1)! \big[(j_1+j_2+j_4+j_5-t)! \times (j_1+j_3+j_4+j_6-t)!\,(j_2+j_3+j_5+j_6-t)! \times (t-j_1-j_2-j_3)!\,(t-j_1-j_5-j_6)! \times (t-j_2-j_4-j_6)!\,(t-j_3-j_4-j_5)!\big]^{-1}, \quad (1)$$

where the triangular function Δ is defined as

$$\Delta(x,y,z) = \left(\frac{(x+y-z)!\,(z+x-y)!\,(y+z-x)!}{(x+y+z+1)!}\right)^{1/2} \quad (2)$$

if the argument of each of the factorials is a nonnegative integer, and $\Delta(x,y,z)=0$ otherwise. It should be noted that the j's, while in a sense independent, are nonetheless mutually constrained by the four triangular functions occuring in Eq. (1).

To obtain the canonical parameters, a set of seven intermediate parameters is defined first, as follows: $p_1=j_2+j_3+j_5+j_6$, $p_2=j_1+j_3+j_4+j_6$, $p_3=j_1+j_2+j_4+j_5$, $q_1=j_1+j_5+j_6$, $q_2=j_2+j_4+j_6$, $q_3=j_3+j_4+j_5$, and $q_4=j_1+j_2+j_3$. In terms of the p's and q's, Eq. (1) may be cast into the following particularly symmetrical form:

$$(6\mathrm{J}) = \left(\prod_{i=1}^{3}\prod_{k=1}^{4}\frac{(p_i-q_k)!}{(q_k+1)!}\right)^{1/2} \times \sum_t (-1)^t (t+1)! \left(\prod_{i=1}^{3}\prod_{k=1}^{4}(p_i-t)!\,(t-q_k)!\right)^{-1}. \quad (3)$$

Each of the p's and q's must be a nonnegative integer, and none of the p's may be smaller than any of the q's. The index t runs from the largest q up to the smallest p.

The obvious invariance of Eq. (3) under all permutations of the three p's and of the four q's shows the 6*j* coefficient to possess a 144-element symmetry group, as was first noted by Regge,[3] who also pointed out that the symmetry group is isomorphic to the direct product of the permutation groups of three and four objects. A 24-element subgroup is simply represented in terms of physically significant interchanges among the j's. The remainder involve replacing certain j's by algebraic expressions involving the original ones.

Before the canonical parameters can be defined, the p's and q's must be placed in ascending order. Therefore, let $p_x \le p_y \le p_z$ and $q_w \le q_x \le q_y \le q_z$.[4] Then the canonical parameters are defined as follows: $n=p_x-q_z$, $a=q_z-q_y$, $b=q_z-q_x$, $c=q_z-q_w$, $d=p_y-p_x$, and $e=p_z-p_x$. These six defining equations may be used to eliminate all but one of the p's and q's in Eq. (3); suppose that only q_z remains. The summation index t is then replaced by q_z+s. The limits of the sum over s are thus from zero to n. The parameter q_z may then be eliminated by taking advantage of the fact that

$$p_x+p_y+p_z = q_w+q_x+q_y+q_z. \quad (4)$$

Hence $q_z=3n+a+b+c+d+e$. Therefore, finally, the 6*j* coefficient may be expressed as follows:

$$(6\mathrm{J}) = PRT, \quad (5)$$

where

$$P = (-1)^{n+a+b+c+d+e}, \quad (6)$$

$$R = [n!\,(n+a)!\,(n+b)!\,(n+c)!\,(n+d)!\,(n+e)! \times (n+a+d)!\,(n+b+d)!\,(n+c+d)! \times (n+a+e)!\,(n+b+e)!\,(n+c+e)!]^{1/2} \times [(3n+a+b+c+d+e+1)!\,(3n+b+c+d+e+1)! \times (3n+a+c+d+e+1)!\,(3n+a+b+d+e+1)!]^{-1/2}, \quad (7)$$

and

$$T = \sum_{s=0}^{n}(-1)^s (3n+a+b+c+d+e+s+1)!\,[s!\,(s+a)! \times (s+b)!\,(s+c)!\,(n-s)!\,(n+d-s)!\,(n+e-s)!]^{-1}. \quad (8)$$

Equations (5)–(8) show that the $6j$ coefficient, when expressed in terms of the canonical parameters n, a, b, c, d, and e, is invariant under any permutation of a, b, and c, or of d and e. Therefore, a 12-element symmetry group is evident, which may be expressed as follows:

$$
\begin{aligned}
&(6J)(n; a, b, c; d, e) \\
&\quad = (6J)(n; a, b, c; e, d) = (6J)(n; b, a, c; d, e) \\
&\quad = (6J)(n; b, a, c; e, d) = (6J)(n; a, c, b; d, e) \\
&\quad = (6J)(n; a, c, b; e, d) = (6J)(n; c, b, a; d, e) \\
&\quad = (6J)(n; c, b, a; e, d) = (6J)(n; b, c, a; d, e) \\
&\quad = (6J)(n; b, c, a; e, d) = (6J)(n; c, a, b; d, e) \\
&\quad = (6J)(n; c, a, b; e, d). \qquad (9)
\end{aligned}
$$

By means of arguments similar to those used in Ref. 1, it may be shown that this group is entirely equivalent to the 144-element symmetry group shown to exist by Regge. Thus the ordering of the p's and q's, and the subsequent definition of the canonical parameters, gives rise to a twelve-to-one homomorphism, in a way similar to that shown previously for the $3j$ coefficient. Moreover, as before, the parameter n does not enter into the symmetry operations.

Another important property of the canonical parameters is that they are completely independent. Each must be a nonnegative integer, but there exist no relationships among them which must be satisfied, as was the case with the j's, as well as with the p's and q's.

It can be seen that the $6j$ coefficient will vanish only if the factor T does so. Inasmuch as the terms are of alternating sign, the sum may be equal to zero; however, there are no obvious simple relationships among the parameters which will necessarily give rise to this result.

The physical significance of the canonical parameters remains unclear at present. Their relative mathematical elegance, however, in comparison with other known parameters, is obvious.

*Supported in part by Consejo de Investigación, Universidad de Oriente.

[1]L. Lockwood, J. Math. Phys. **17**, 1671 (1976).
[2]See, e.g., A. de-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963), p. 139.
[3]T. Regge, Nuovo Cimento **11**, 116 (1959).
[4]At this point it is necessary only that p_x be the smallest of the p's and that q_z be the largest of the q's.

Nonstationary multiple scattering

A. P. Wang

Department of Mathematics, Arizona State University, Tempe, Arizona 85281
(Received 12 July 1976)

This paper is devoted to the abstract formulation of the general nonstationary multiple scattering problem in radiative transfer. We consider the linear operators on curved surfaces. Therefore, the problem is attacked from the unified and general point of view. Special geometric considered are spherical shells and slabs. Special cases lead to stationary, instantaneous, and time invariant cases. The extension from stationary to nonstationary involves the distribution theory and the concept of nonpredictive operators. Many well-known physical problems in astrophysics are solved in the unified way.

I. INTRODUCTION

Nonstationary radiative transfer problems are relatively unexplored because of the difficulty of their mathematical treatment. Nevertheless, it is generally recognized that nonstationary radiative transfer in terrestrial plantary and stellar atmospheres should be taken into account for the treatment of particular transient phenomena. A comprehensive list of references concerning previous results on nonstationary transfer problems up to 1970 can be found in Refs. 1–3.

In 1926 work on nonstationary radiative transfer was carried out by Milne and Chandrasekhar.[4] In these studies the nonstationary radiation field is considered to be governed principally by various decay rates. At the same time the problem of radiation through an absorbing medium in motion was treated by Rosseland.[5] The general nonstationary multiple-scattering problem was relatively unexplored until around 1960.

A paper by Preisendorfer[6] in 1958 indicates that the method of the principle of invariance initiated by Ambarzumian and Chandrasekhar can be extended to the nonstationary case in a slab or plane-parallel media. At the same time, Bellman, Kalaba, and Ueno have systematically applied the invariant imbedding technique to a wide variety of transport problems of practical interest in the field of radiative transfer, neutron diffusion, wave propagation, and so on. Some of their results on nonstationary transport problems in a slab are presented in Refs. 1, 2, 7, and 8. A more recent work by Redheffer and Wang[3] presented a formal extension from stationary to nonstationary scattering from a unified point of view. It introduced the nonstationary star product, the nonpredictive operators, and some mathematical properties of such systems. However, it is still limited to the slab or plane-parallel media.

There are many physical problems in which curved surfaces must be taken into consideration. This has led to the study of stationary radiative transfer in spherical media, by Bellman, Ueno, *et al.*, (1968)[9] and Wang (1970).[10] A comprehensive list of references can be found in those two papers. From the astrophysics point of view these two papers have succeeded in solving the Chandrasekhar problem in a spherical shell, Schuster's problem in the theory of line formation, and the Milne problem of the diffusion of light from a central star.

It is obvious that the time is due to consider the nonstationary multiscattering problem by taking curvature into consideration. This paper is devoted merely to the formulation of those problems. As in Refs. 3 and 10, we attack the problem from a unified viewpoint, i.e., the linear nonstationary operators. General results are presented for curved surfaces. Special geometrics considered are spherical shells and slabs. In case the linear operators are stationary, the former leads to the results of Refs. 9 and 10. The latter leads to the results of Ref. 3. The general nonstationary operator consists of three special cases, namely stationary, instantaneous nonstationary, and time invariant. The mathematical properties and their solutions will be discussed in our next paper.

II. FORMULATION

Let us consider a one-parameter (in space) family of general radiation field Ω. In this field $I(x, \pm u, t) \in \Omega$ is the intensity[4] in a specific frequency at position x, in direction $\pm u$, and in time t. For a linear system of radiative transfer, input intensity I_i and output intensity I_0 are related by $I_0 = L \circ I_i$, $I_0, I_i \in \Omega$, where $\circ$ signifies the bounded operation under the linear operator L. In nonstationary radiative transfer, if the input is given by $I(x,u,t)$, the diffused outputs are $I^*(y,v,t)$ and $I^*(x,-v,t)$, and specular output is $II(y,v,t)$, then in our operator notation we write:

$$I^*(y,v,t) = \mathcal{T}(x,y;v,u;t,\tau) \cdot I(x,y,\tau),$$

$$I^*(x,-v,t) = \mathcal{R}(x,y;-v,u;t,\tau) \cdot I(x,u,\tau),$$

and

$$II(y,v,t) = \mathcal{P}(x,y;v,u;t,\tau) \cdot I(x,u,\tau). \tag{II.1}$$

It was introduced in Ref. 3 that such nonstationary operators may be formally represented by integral forms. For example,

$$I^*(y,v,t) = \int_{-0}^{1} \int_{-\infty}^{\infty} P(x,y;v,u;t,\tau) I(x,y,\tau)\, d\tau \frac{du}{v}, \tag{II.2}$$

where $P(x,y;v,u;t,\tau)$ is the kernel associated with the operator $\mathcal{T}$. Likewise, associated with operators $\mathcal{R}$ and $\mathcal{P}$, we have kernels

$$P(x,y;-v,u,t,\tau) \text{ and } A(x,y;v,u;t)\delta(v^* - u)\delta(t-\tau), \tag{II.3}$$

where δ is the Dirac delta, $v^* = v^*(x,y,u)$, $v^* \to v$ as $y \to x$, and

$$(\mathcal{P}\cdot I)(y,v,t)=\int_0^1 A(x,y;v,u;t)I(x,y,\tau)\delta(v^*-u)\frac{dv}{v}. \tag{II.4}$$

In a more careful discussion of such operators, mathematical distribution theory should be used. For example, Eq. (II. 2) may be interpreted as a regular (double) distribution with respect to variables to τ and u. Furthermore, it is necessary that $\mathcal{T}$ be a δ-generating function, i. e., take the identity operator as a limit in space. More precisely, $P(x,y;v,u;t,\tau)\to\delta(v^*-u)\delta(t-\tau)$ as $y\to x$. And $\mathcal{R}$ and $\mathcal{P}$ are zero operators as $y\to x$. For more details, see Refs. 3 and 10. With this in mind, there are certain properties in the distribution theory used in the analysis which will be pointed out in the derivation. We assume that all kernels are differentiable in the sense of distribution theory.

In Ref. 3 the "nonpredictive operators" were introduced. A pair of linear operators $(\mathcal{F},\mathcal{G})$ are nonpredictive if $\mathcal{F}_t\cdot\mathcal{G}+\mathcal{F}\cdot\mathcal{G}_\tau=0$, where the subscript denotes the partial derivative with respect to time t and τ, and the product is defined by

$$(\mathcal{F}\cdot\mathcal{G})(x,y;v,u;t,\tau)$$
$$=\int_{-1}^1\int_{-\infty}^{\infty}F(x,y;v,w;t,\xi)G(x,y;w,u;\xi,\tau)\,d\xi\,dw.$$

It has been pointed out that an operator, with the property that output depends only on the past input, has the nonpredictive property. Usage of such properties also will be pointed out in the derivation.

As in the stationary case the derivatives of P, Q, and A with respect to the space variable x, at $x=y$, are *phase functions* $\tilde{p}(x;\pm v,\pm u;t,\tau)$ and *extension coefficients* $\alpha(x;t)\delta(v^*-u)$. The parameters t and τ signify nonstationary. The optional depth used in the stationary case cannot be easily extended; therefore, it defeats its original purpose. In view of the integral form (II. 2), we can introduce a function which measures the total diffused output versus its input up to time t,

$$C(x,y;u,t)=\int_{-\infty}^t\int_{-1}^1 P(y,x;v,u;t,\tau)I(x,y,\tau)\,dv\,d\tau$$
$$\times\Big[\int_{-\infty}^t\int_0^1 I(x,u,\tau)\,du\,d\tau\Big]^{-1}. \tag{II.5}$$

The nonstationary *albedo of single scattering* is defined as

$$\sigma(x,u,t)=\lim_{\Delta\to0}(1/\Delta)C(x,x+\Delta,u,t). \tag{II.6}$$

As a consequence a new phase function p is defined,

$$\sigma(x,\pm u,t)p(y;\pm v,\pm u;t,\tau)=P(y;\pm v,\pm u;t,\tau) \tag{II.7}$$

for $\sigma(x,\pm u,t)\neq0$; otherwise $p=0$.

III. EQUATIONS OF TRANSFER AND EXAMPLES

Let us consider the transfer of radiation taking place in a thin medium where x and $y=x+\Delta x$ are two boundary points. For Δx small, the input intensity at x is $I(x,u,t)$; then the total intensity at y is $I(x+\Delta x,u+\Delta u,t+\Delta t)$, where $I=I^*+II$. The difference in total intensities is due to attenuation, absorbing and scattering.

As in Refs. 2 and 3, Δu and Δt are functions of Δx. Taking the limit as $\Delta x\to0$, the *equation of transfer* for the total radiation field is

$$u\frac{d}{dx}I(x,\pm u,t)=-\alpha(x,\pm u;t,\tau)\cdot I(x,\pm u,\tau)$$
$$=\int_{-1}^1\sigma(x,v,t)p(x;\pm u,v;t,\tau)\cdot I(x,v,\tau)\,dv. \tag{III.1}$$

The right-hand side of the equation in the abstract form involves the operation $\cdot$ as introduced in (II. 2).

To give some feeling about the abstract operator and the equation transfer, we shall present a few solid examples which include some well-known ones in the field of radiative transfer. The following examples are presented for two physical geometrics, namely a slab and a symmetric spherical shell, along with some special interpretations of the $\cdot$ operation.

For the spherical shell geometry, $I=I(r,u,t)$, where r denotes a point on the sphere with radial coordinate r, $y_0\le r\le x_0$. The intensity in the total radiation field at r directed toward the outside surface at time t is denoted $I(r,-u,t)$, and the intensity of the inward directed radiation is $I(r,u,t)$. And

$$v^*=v^*(x,y,u)=[1-(x/y)^2(1-v^2)]^{1/2},$$

as introduced in Ref. 2, along with

$$v_*=v_*(x,y,v)=[1-(y/x)(1-v^2)]^{1/2}$$

and

$$(v_*)^*=(v^*)_*.$$

Then we have the *equation of transfer* for a *spherical shell*

$$\frac{dI}{dr}=I_r\pm\frac{1-u^2}{ru}I_u+\mu_\pm I_t, \tag{III.2}$$

where $I=I(r,\pm u,t)$ and the subscripts of I denote partial differentiation. The μ_+ denotes the inverse of the outward propagation constant and the μ_- denotes that of the inward one. The detailed analysis is an easy extension of that given in the stationary case[10]; therefore, it is not presented here.

For a slab geometry case $I=I(z,\pm u,t)$, we merely replace the radius by the physical thickness $z, x_0\le z\le y_0$ and $u=u^*=u_*$. There, the second term in (III. 2) disappears. The result is, for $I=I(x,\pm u,t)$,

$$\frac{dI}{dz}=I_z+\mu_\pm I_t\ \text{(slab)}. \tag{III.3}$$

With the above geometry in mind, we are ready to formulate the equation of transfer under various interpretations of the $\cdot$ operation.

A. Case 1. Stationary case

The case under consideration is that $I=I(x,\pm u)$ which is independent of time t. Thus, there is no time delay between any input and output pair, and the kernels have the form

$$\alpha(x,\pm u,t,\tau)=\alpha(x,u)\delta(t-\tau)$$

and

$$P(x,\pm u,\pm v;t,\tau)=p(x,\pm u,\pm v)\delta(t-z),$$

where the δ acts like the identity under the integral representation. Therefore, we can write

$$\alpha(x,\pm u;t,\tau)\cdot I(x,\pm u,\tau)=\alpha(x,\pm u)I(x,\pm u),$$

$$p(x;\pm u,\pm v;t,\tau)\cdot I(x,\pm v,\tau)=p(x;\pm u,\pm v)I(x,\pm v),$$

and $\sigma(x,t)=\sigma(x)$. Then the equation of transfer for a *spherical shell* reduces to

$$\left(u\frac{\partial}{\partial r}\pm\frac{1-u^2}{r}\frac{\partial}{\partial u}\right)I(r,\pm u)$$
$$=-\alpha(r,\pm u)I(r,\pm u)+\sigma(r)\int_{-1}^{1}p(r,\pm u,v)I(r,v)\,dv, \tag{III. 4}$$

and the equation of transfer for a *slab* is

$$\frac{\partial}{\partial z}I(z,\pm u)=-\alpha(z,\pm u)I(z,\pm u)$$
$$+\sigma(r)\int_{-1}^{1}p(z,\pm u,v)I(z,v)\,dv. \tag{III. 5}$$

Where α, σ, and p are customarily called the attenuation (or extinction) coefficient, the abortion coefficient and the phase function in stationary case. If Eqs. (III. 4) and (III. 5) are subject to uniform boundary conditions and α, σ, and p are homogeneous with respect to r or z whichever the case may be, then (III. 4) and (III. 5) have the exact form and meaning as given by Ueno[9] and Chandrasekhar.[4]

B. Case 2. Instantaneous nonstationary case

The extinction coefficients, and absorption coefficients phase functions, in this case are functions of time t. They are products of $\delta(t-\tau)$ and nonstationary functions, i. e.,

$$\alpha(x,\pm u;t,\tau)=\alpha(x,\pm u,t)\delta(t-\tau),$$

$$p(x,\pm v,\pm u,t,\tau)=p(x,\pm v,\pm u,t)\delta(t-\tau).$$

As in the previous case, the operations $\cdot$ reduce to the multiplication of functions. Therefore, the integral representations with respect to time disappear as in the stationary case. The analysis is an easy extension of that for the stationary case. The equation of transfer has the form

$$\left(u\frac{\partial}{\partial r}+\frac{1-u^2}{r}\frac{\partial}{\partial u}+\mu_{\pm}\frac{\partial}{\partial t}\right)I(r,\pm u,t)$$
$$=\alpha(r,t)I(r,\pm u,t)+\sigma(r,t)\int_{-1}^{1}p(r,\pm u,v,t)I(r,v,t)\,dv \tag{III. 6}$$

for the *spherical shell* geometry and

$$\left(u\frac{\partial}{\partial z}+\frac{\partial}{\partial t}\right)I(z,\pm u,t)$$
$$=\alpha(z,t)I(z,\pm u,t)+\sigma(z,t)\int_{-1}^{1}p(z,\pm u,v,t)I(z,v,t)\,dv \tag{III. 7}$$

for the *slab geometry*. Equation (IV. 7) agrees with results obtained in Refs. 2 and 6. The result of (III. 6) is new, as far as we know.

C. Case 3. The time-invariant case

Operators in our general formulations depend on the absolute values of two times. In this special case, we assume that these operators are functions of the difference of two times, i. e.,

$$\alpha(x,\pm u,t,\tau)=\alpha(x,\pm u,t-\tau)$$

and

$$p(x,\pm u,v,t,\tau)=p(x,\pm u,v,t-\tau).$$

The equations of transfer can be obtained immediately by replacing the linear operation $\cdot$ by the convolution operation $*$. The slab geometry case was discussed in Ref. 10, and the spherical shell case is again a new result.

IV. TRANSMISSION AND REFLECTION

The differential equations for the transmission and reflection operators may be obtained when a thin layer is added to a given medium. This can be accomplished either by the method of particle counting or by the method of medium coefficients. The approach used here follows from the latter method.

The scattering matrix for the nonstationary case has the form

$$\begin{pmatrix}\mathcal{T} & \mathcal{W}\\ \mathcal{R} & \mathcal{V}\end{pmatrix}+\begin{pmatrix}\mathcal{P} & 0\\ 0 & \mathcal{Q}\end{pmatrix}, \tag{IV. 1}$$

where the first term is the *diffuse part* and second is the *specular part*. They are different from the stationary case in the sense that they involve the pair (t,τ) as parameters and they admit an integral representation (II. 2). The operators $\mathcal{T}$, $\mathcal{R}$, and $\mathcal{P}$ have been introduced in (II. 1) while operators $\mathcal{W}$, $\mathcal{V}$, and $\mathcal{Q}$ have kernels

$$P(x,y;-v,-u;t,\tau),\ P(x,y;v,-u;t,\tau)$$

and

$$A(x,y;-v,-u)(u-u),$$

respectively.

The method we shall use is to compare a given scattering matrix to that with a thin layer attached, then take the limit. It is noted that we do not assume that operators are commutative and the δ operator is involved in our integral representation. Therefore, we cannot assume that the limits of a left-hand and a right-hand operator under the composite operation [e. g., (IV. 2)] are the same. For example, let $\mathcal{H}$ be an arbitrary operator; we shall compute limits of $\delta(v^*-u)\cdot\mathcal{H}$ and $\mathcal{H}\cdot\delta(v^*-u)$, where $v^*=v^*(y,z,v)$, and $\mathcal{H}$ is independent of y. We compute

$$\frac{\partial}{\partial z}[\delta(v^*-u)\cdot\mathcal{H}]=\frac{\partial}{\partial z}(v^*-u)\cdot\mathcal{H}+\delta(v^*-u)\cdot\frac{\partial}{\partial z}\mathcal{H}.$$

As in Ref. 11, the symbolic function δ is differentiable and changing of variables is allowed; therefore,

$$\frac{\partial}{\partial z}\delta(v^*-u)\cdot\mathcal{H}\Big|_{z=y}=\delta'(v^*-u)\frac{\partial v^*}{\partial z}\cdot\mathcal{H}\Big|_{y=z}$$
$$=-\frac{\partial v^*}{\partial z}\Big|_{z=y}\delta(v-u)\cdot\frac{\partial}{\partial v}\mathcal{H}, \tag{IV. 2}$$

where we also used the property that $\delta(v^*-u)\to\delta(v-u)$ as $y\to x$.

For the spherical geometry case, see the previous section. After some trival computation, the above results (IV. 2) leads to

$$\frac{\partial}{\partial z}[\delta(v^*-u)\cdot H]\Big|_{z=y} = -\delta(v-u)\cdot\left(\frac{1-v^2}{yv}\frac{\partial}{\partial v}H\right) \qquad \text{(IV. 3)}$$

For the slab geometry case, the specular part of the intensity travels in a straight line. Its parameter u remains the same. Therefore, v^* is independent of y and (IV. 2) reduces to

$$-\delta(v-u)\cdot\frac{\partial}{\partial v}H. \qquad \text{(IV. 4)}$$

To compute the derivative of $H\cdot\delta(v^*-u)$, we observe that apart from $O(|v^*-v|^2)$,

$$H\cdot\delta(v^*-u)=H\cdot\left(\delta(v-u)+(v^*-v)\frac{\partial}{\partial v}\delta(v-u)\right),$$

where we used the limiting property of v^* as stated in Sec. II. Since $(v-u)$ is independent of z, the first term in the right-hand side of the above equation disappears. Therefore,

$$\begin{aligned}\frac{\partial}{\partial z}[H\cdot\delta(v^*-u)] &= \frac{\partial}{\partial z}[H\cdot(v^*-v)\frac{\partial}{\partial v}\delta(v-u)]\\ &= \frac{\partial}{\partial z}\left[-\frac{\partial}{\partial v}H\cdot(v^*-v)\delta(v-u)\right.\\ &\quad \left.-H\cdot\left(\frac{\partial}{\partial v}(v^*-v)\right)\delta(v-u)\right]\\ &= -\left(\frac{\partial}{\partial v}H\cdot\frac{\partial}{\partial z}(v^*-v)\right.\\ &\quad \left.+H\cdot\frac{\partial^2}{\partial z\,\partial v}(v^*-v)\right)\cdot\delta(v-u).\end{aligned}$$

It should be noted that the analysis also applies when v^* is replaced by v_*.

For the spherical case, we obtain

$$\frac{\partial}{\partial z}[H\cdot\delta(v^*-u)]\Big|_{z=y} = -\delta(v-u)\left[\frac{1+u^2}{yu^2}+\frac{1-u^2}{yu}\frac{\partial}{\partial u}H\right]; \qquad \text{(IV. 5)}$$

details are not presented here. When v^* is replaced by v_*, the right-hand side merely changes sign. For the slab geometry, the right-hand side of the above equation reduces to the same expression as (IV. 4), because $v^*=v_*=v$ and $\delta(v-u)$ commutes with every operator under consideration.

Let us construct the differential operator equation which governs the system. This is done by taking two scattering matrices, one extended from $(x, y+\Delta)$ and the other from (x,y), and then taking limit as $\Delta\to 0$. To avoid boring the reader with long and complicated, but straightforward details, we merely present our results.

The specular operators P and Q satisfies the differential equations

$$\left(\frac{\partial}{\partial y}+\mu\frac{\partial}{\partial t}\right)P = B\cdot P$$

and $\qquad \text{(IV. 6)}$

$$\left(\frac{\partial}{\partial y}-\nu\frac{\partial}{\partial \tau}\right)Q = Q\cdot\tilde{D},$$

where $\mu=\mu_+$ and $\nu=\mu_-$ and operators $B=B(y;v,u)$ and $D=D(y,-v;u)$ are

$$B\cdot = \frac{\partial}{\partial z}[\delta(v^*-u)\cdot]\Big|_{z=y},$$

$$\cdot\tilde{D} = \frac{\partial}{\partial z}[\cdot\,\delta(v^*-u)]\Big|_{z=y}.$$

As for the diffuse operators, we use

$$H\frac{\partial}{\partial t}\cdot = -\frac{\partial}{\partial\tau}H\cdot,$$

the nonpredictive properties as introduced in Ref. 10. These diffuse operators are governed by

$$\begin{aligned}&\left(\frac{\partial}{\partial y}+\mu\frac{\partial}{\partial t}\right)T = (b+P\cdot c)\cdot(T+P)-B\circ P,\\ &\left(\frac{\partial}{\partial y}+\mu\frac{\partial}{\partial t}-\nu\frac{\partial}{\partial\tau}\right)W = a+b\cdot W+W\cdot\tilde{d}+W\cdot c\cdot W, \qquad \text{(IV. 7)}\\ &\frac{\partial}{\partial y}R = (V+Q)\cdot c\cdot(t+P),\end{aligned}$$

and

$$\left(\frac{\partial}{\partial y}-\nu\frac{\partial}{\partial\tau}\right)V = (V+Q)\circ(d+c\circ W)-Q\circ\tilde{D},$$

where

$$\begin{pmatrix} b & \tilde{b}\\ d & \tilde{d}\end{pmatrix} = \begin{pmatrix} B+b^* & \tilde{B}+b^*\\ D+d^* & \tilde{D}+d^*\end{pmatrix}\begin{pmatrix} b^* & a\\ c & d^*\end{pmatrix}$$

$$= \frac{\sigma}{2v}\begin{pmatrix} p^{++} & p^{+-}\\ p^{-+} & p^{--}\end{pmatrix}$$

and

$$\tilde{B}=\tilde{B}(y,v,u), \quad D=D(y,-v,-u),$$

$$\cdot\tilde{B} = \frac{\partial}{\partial z}[\cdot\,\delta(v_*-u)]_{z=y},$$

$$D\cdot = \frac{\partial}{\partial z}[\delta(v_*-u)\cdot]_{z=y},$$

$$p^{\pm\pm}=p(y;\pm v,\pm u;t,\tau).$$

Special cases

Case 1. Stationary case

The associative operators P and intensities I are independent of time, and we may assume that the kernel of P has the form $P(y,x;v,u)\delta(t-\tau)$ and all operations can be expressed as a single integration with respect to w, see (II. 2). Since operators are independent of time, all terms involving $\partial/\partial t$ and $\partial/\partial\tau$ in (IV. 6) disappear. By use of (IV. 4), (IV. 5) and remarks following (IV. 5), we obtain the results for spherical symmetry presented in Ref. 10. This in turn leads to the well-known results for slab geometry in the stationary case, in which $D=\tilde{D}$ and $B=\tilde{B}$.

Case 2. Instantaneous nonstationary case

The kernel associated with the operator P has the form $P(x,y;v,u;t-\tau)\delta(t-\tau)$. Similar to case 1, all operations are reduced to a single integration. But since P depends only on the difference of t and τ, say h, the partial differential in (IV. 6), can be reduced by observing

$$\frac{\partial \rho}{\partial t} = \frac{\partial \rho}{\partial h} = -\frac{\partial \rho}{\partial \tau}.$$

By the results of (IV.3) and (IV.4), the equations of a spherical shell for ρ and Q have the form

$$\left(\frac{\partial}{\partial y} + \mu\frac{\partial}{\partial h}\right)P = -\left(\frac{\alpha(y)}{\partial v} + \frac{1-v^2}{yv}\frac{\partial}{\partial v}\right)P,$$

$$\left(\frac{\partial}{\partial y} - \nu\frac{\partial}{\partial h}\right)Q = -Q\left(\frac{\alpha(y)}{u} - \frac{1+u^2}{u^2 y} - \frac{1-u^2}{yu}\frac{\partial}{\partial u}\right),$$

where the kernels $P = P(x,y;v,u;h)$ and $Q = Q(x,y;-v,-u;h)$. As for diffusion, we merely write down those for T and W in the spherical shell geometry:

$$\frac{2}{\sigma(y,h)}\left(\frac{\partial}{\partial y} + \mu\frac{\partial}{\partial h} + \frac{\alpha(y)}{u} - \frac{1+u^2}{yu^2} + \frac{1-u^2}{yu}\frac{\partial}{\partial u}\right)T(x,y;v,w;h)$$

$$= \int_0^1 p(y;v,w;h)T(x,y;w,u;h)\frac{dw}{w}$$

$$+ \int_0^1\int_0^1 W(x,y;v,w;h)p(y;w,\bar{w};h)$$

$$\times T(x,y;\bar{w},v;h)\frac{dw}{w}\frac{d\bar{w}}{\bar{w}}$$

$$+ \int_0^1 p(y;v,w;h)P(x,y;w,v;h)\frac{dw}{w}$$

$$+ \int_0^1\int_0^1 P(x,y;v,w;h)p(y;w,\bar{w};h)$$

$$\times P(x,y;\bar{w},u;h)\frac{dw}{w}\,d\bar{w}$$

and

$$\frac{2}{\sigma(y,h)}\left(\frac{\partial}{\partial y} + (\mu+\nu)\frac{\partial}{\partial h} - \frac{1+u^2}{yu^2} + \frac{1-v^2}{yv}\frac{\partial}{\partial v} + \frac{1-u^2}{yu}\frac{\partial}{\partial u}\right)$$

$$\times W(x,y;v,w;h)$$

$$= p(y;v,-u;h) + \int_0^1 p(y;v,w;h)W(x,y;w,-u;h)\,dw$$

$$+ \int_0^1 W(x,y;v,-w;h)p(y;-w,-u;h)\frac{dw}{w}$$

$$+ \int_0^1\int_0^1 W(x,y;-v,-w;h)p(y;-w,\bar{w})$$

$$\times W(x,y;\bar{w},u;h)\,d\bar{w}\frac{dw}{w}.$$

As for the slab geometry, $v_* = v^* = v$ and in the above equations the left-hand sides contain only the first and second terms, the other parts of each equation remain the same. As for the specular equation, it only contains the first term on the right-hand side and the left-hand side remains the same.

Case 3. Time-invariant case

In this case

$$T(x,y;v,u;t,\tau) = T(x,y;v,u,t-\tau)$$

and

$$W(x,y;v,-u;t,\tau) = W(x,y;v,-u;t-\tau).$$

The diffuse equations for the spherical and slab geometrices remain the same. However, all operations in the general form for the diffuse operator now can be expressed as the convolution operation $*$ with respect to time. For example,

$$\frac{2}{\sigma(y,h)}\left(\frac{\partial}{\partial y} + \mu\frac{\partial}{\partial h} + \frac{\alpha(y)}{u} - \frac{1+u^2}{yu^2} + \frac{1-u^2}{yu}\frac{\partial}{\partial u}\right)$$

$$\times T(x,y;v,w;h)$$

$$= \int_0^1 p(y;v,w;h) * T(x,y;v,w;h)$$

$$+ \int_0^1\int_0^1 W(x,y;v,w;h) * p(y,w,\bar{w};h)$$

$$* T(x,y;\bar{w},v;h)\frac{dw}{w}\,d\bar{w}$$

$$+ \int_0^1 p(y;v,w;h) * P(x,y;w,v;h)\frac{dw}{w}$$

$$+ \int_0^1\int_0^1 P(x,y;v,w;h) * p(y;w,\bar{w};h)$$

$$* P(x,y;\bar{w},w;h)\frac{dw}{w}\,d\bar{w}$$

and

$$\frac{2}{\sigma(y,h)}\left(\frac{\partial}{\partial y} + (\mu+\nu)\frac{\partial}{\partial h} - \frac{1+u^2}{yu^2} + \frac{1-v^2}{yv} + \frac{1-u^2}{yu}\frac{\partial}{\partial u}\right)$$

$$\times W(x,y;v,w;h)$$

$$= p(y;v,-u;h) + \int_0^1 p(y;v,w;h) * W(x,y;w,-u;h)\,dw$$

$$+ \int_0^1 W(x,y;v,-w;h) * p(y;-w,-u;h)\frac{dw}{w}$$

$$+ \int_0^1\int_0^1 W(x,y;-v,-w;h) * W(x,y;\bar{w},u;h)\,d\bar{w}\frac{dw}{w},$$

and likewise and true for the other equations.

[1] S. Ueno, J. Math. Appl. **4**, 310—21 (1962).
[2] R. Bellman, R. Kalaba, and S. Ueno, Icarus 1 (3), 191—99 (1962).
[3] R.M. Redheffer and A.P. Wang, J. Math. Mech. **19**, 765—81 (1970).
[4] S. Chandrasekhar, *Radiative Transfer* (Oxford U.P., Oxford, 1950), pp. 354—64.
[5] S. Rosseland, Astrophys. J. **63**, 342—67 (1926).
[6] R.W. Preisendorfer, Proc. Nat. Acad. Sci. USA **44**, 328—32 (1958).
[7] R. Bellman, R. Kalaba, and S. Ueno, "Invariant imbedding and scattering of light in a one-dimensional medium with a moving boundary," The Rand Corp. RM 3960-ARPA (March 1964).
[8] R. Bellman and P.T.Y. Poon, J. Math. Phys. Sci. 5, 459—74 (1974).
[9] R. Bellman, S. Ueno, *et al.*, J. Math. Phys. **9**, 909—12 (1968).
[10] A.P. Wang, J. Math. Phys. **14**, 855—62 (1973).
[11] Bernard Friedman, *Lecture on Application-Oriented Mathematics* (Holden-Day, San Francisco, 1967).

Geometry of superspaces with Bose and Fermi coordinates and applications to graded Lie bundles and supergravity*

Freydoon Mansouri

Department of Physics, Yale University, New Haven, Connecticut 06520
(Received 9 August 1976)

The geometry of superspaces with Bose- and Fermi-type coordinates is presented from a coordinate independent point of view. Various geometrical quantities of conventional manifolds are generalized so as to be applicable to superspaces. It is shown that these generalizations can be basically arrived at algebraically by replacing, in the definitions of various geometrical quantities, the Lie derivative of the conventional manifolds with a generalized graded Lie bracket. Explicit expressions for connection coefficients, Riemann curvature tensor, etc., are derived. The general formalism is then applied to graded Lie bundles the relevance of which to supergravity theories is demonstrated.

I. INTRODUCTION

The understanding of supersymmetries in four-dimensional space–time[1–3] has been the subject of many recent papers.[4] Although the manner in which such symmetries are to be made use of in particle physics is far from clear, a number of their mathematical properties have been elucidated. In particular, it has become clear that their group theoretic and algebraic properties fall within the domain of graded Lie groups and graded Lie algebras.[4–7] Several authors[8–14] have also discussed some of the geometrical properties of manifolds involving both Fermi and Bose coordinates. For example, expressions in a coordinate induced basis for connection coefficients and contracted curvature tensor of a superspace are given by Nath and Arnowitt.[9]

The main purpose of this paper is to give a general coordinate independent discussion of the geometry of manifolds involving both Fermi-type and Bose-type coordinates.[15] We refer to such manifolds as superspaces and extend various notions of modern differential geometry so that they will be applicable to them. Among these notions are the metric, connection, curvature, wedge product, exterior derivatives, etc. An important step in generalizing these concepts turns out to be the generalization of the Lie derivative, or the commutator, to a more general Lie bracket implicit in graded Lie algebras.

The geometry of the superspaces is presented in Sec. II. Starting from general definitions, we derive expressions for various geometrical quantities of the superspace and often illustrate them with simple examples. Although it is not the purpose of this paper to give detailed field theoretic applications of such geometries, a simple example is given in Sec. III of how one constructs a fiber bundle with four-dimensional space–time as its base manifold but with a graded Lie group as its structural group. It is shown that it is possible to construct a locally supersymmetric Lagrangian from the gauge potentials (the connection coefficients) associated with the graded Lie bundle. It leads to a supergravity theory involving spin 2 and 3/2 fields, which was obtained by Freedman, van Nieuvenhuizen, and Ferrara[16] and then in its first order form by Deser and Zumino.[17] In fact, the Lagrangian we obtain is identical with that of Deser and Zumino. Locally supersymmetric theories arising more consistently from graded Lie bundle geometries will be discussed elsewhere.[18]

II. GEOMETRY OF SUPERSPACES WITH BOSE AND FERMI COORDINATES

Unless a change or generalization is necessary, we follow the notation and conventions of Refs. 12 and 19, except that boldface characters are used in place of characters with overarrows. To aid the reader who is familiar with conventional concepts of differential geometry, each concept is first written down for Bose-type manifolds, and then its generalization to Bose–Fermi-type superspaces are given. As far as we know, many of these generalizations are new even in mathematical literature.

(i) *Vector fields, bases, and 1-forms*: We begin by noting that in modern differential geometry the basis vectors are taken to be identical with directional derivatives. For example, let, in some neighborhood, X^μ, $\mu=1,\ldots,n$, be quantities whose values $X^\mu(P)$ are the coordinates of the point P. The operator ∂_μ defined by

$$\partial_\mu f=\frac{\partial}{\partial X^\mu}f(X^1,\ldots,X^n)$$

is the vector tangent to the lines $X^k=\text{const}$ $(k\neq\mu)$. The n operators

$$\{\mathbf{e}_\mu\}=\{\partial_\mu\},$$

where μ indicates which vector fields, and not which components, are the prototypes of basis vectors. Since they are partial derivatives of coordinates, such a basis is called coordinate induced. The general feature of a coordinate induced basis $\{\mathbf{e}_\mu\}$ is that

$$[\mathbf{e}_\mu,\mathbf{e}_\nu]=0\quad\text{(coordinate basis)}.$$

More generally, however, this commutator is not zero:

$$[\mathbf{e}_\mu,\mathbf{e}_\nu]=C^\lambda_{\mu\nu}\mathbf{e}_\lambda\quad\text{(general basis)},$$

where the $C^\lambda_{\mu\nu}$ are the commutation coefficients of the basis $\{\mathbf{e}_\mu\}$.

These ideas readily generalize to superspace. Let

$$Y^i=\{X^\mu,\theta^\alpha\},\quad \mu=1,\ldots,n,\quad \alpha=1,\ldots,m,\qquad (2.1)$$

determine, in some neighborhood, the coordinates of a point P in a superspace with n Bose-type and m Fermi-type coordinates. Then we define a basis

$$\{e_i\}=\{\partial_i\},\quad i=1,\ldots,m+n, \tag{2.2}$$

where

$$\partial_i f=\frac{\partial}{\partial Y^i} f(X^1,\ldots,X^n;\ \theta^1,\ldots,\theta^m). \tag{2.3}$$

Since the order of various terms now does matter, in this paper by ∂_i we always mean differentiation from the left. We also define the *generalized Lie brackets*

$$[e_i, e_j\}\equiv e_i e_j-(-)^{\alpha_i\alpha_j}e_j e_i, \tag{2.4}$$

$$\alpha_i=\begin{cases}0 & \text{if } e_i \text{ is a Bose-type operator}\\ 1 & \text{if } e_i \text{ is a Fermi-type operator.}\end{cases}$$

Then the basis vectors (2.2) are the simplest prototypes of those satisfying the generalized commutation relations

$$[e_i, e_j\}=C_{ij}^k e_k. \tag{2.5}$$

It is clear from (2.4) that except when both e_i and e_j are of Fermi-type (2.4) is just the commutator $[e_i, e_j]$; when e_i and e_j are both Fermi-type, (2.4) becomes an anticommutator. Therefore, the generalized commutation coefficients C_{ij}^k in (2.5) are, respectively, antisymmetric and symmetric in i and j.

The 1-forms $\{\omega^\mu\}$ are defined to be the duals to the elements of the basis $\{e_\nu\}$:

$$\langle\omega^\mu, e_\nu\rangle=\delta_\nu^\mu.$$

The simplest prototypes of 1-forms are the coordinate differentials $\{dX^\mu\}$. The 1-forms in superspace are defined similarly:

$$\langle\omega^i, e_j\rangle=\delta_j^i. \tag{2.6}$$

Given a set of basis vectors $\{e_i\}$, one can expand an arbitrary vector in the form

$$U=U^i e_i, \tag{2.7}$$

where the U^i are the expansion coefficients (contravariant components) of U. The Bose or Fermi nature of the expansion coefficients depends on whether U is Bose or Fermi type. Since again the order of various terms is important, we use the convention that expansion coefficients appear to the left of basis vectors.

Similarly one can expand an arbitrary 1-form in terms of basis 1-forms:

$$V=V_i\omega^i, \tag{2.8}$$

the Bose or Fermi nature of V_i depending on V.

(ii) *The metric*: In the mathematical literature, the metric is defined as a bilinear nonsingular operator which acts on pairs of vectors to produce c-numbers, the components of the metric tensor:

$$g=g_{\mu\nu}\omega^\mu\otimes\omega^\nu$$

where, *by definiton*,

$$g_{\mu\nu}=g_{\nu\mu}=e_\mu\cdot e_\nu=g(e_\mu, e_\nu),$$

$e_\mu=$ basis vectors,

$\omega^\nu=$ corresponding 1-forms.

It is well known[12,20] that this definition is equivalent to the one used commonly in physics. Taking our clue from these expressions, we take for the metric operator of the superspace

$$g=g_{ij}\omega^i\otimes\omega^j \tag{2.9}$$

where now

$$g_{ij}=(-)^{\alpha_i\alpha_j}g_{ji}=e_i\cdot e_j. \tag{2.10}$$

It is, of course, possible to further generalize (2.9) such that g_{ij} would not satisfy (2.10). But much of the elegance of the geometry described in this paper depends on (2.10). Moreover, there is no loss of generality since the antisymmetric properties (or symmetric for Fermi coordinates) of bases are absorbed into the generalized bracket (2.5). In the special case where ω^i are duals of coordinate bases, i.e., when

$$\omega^i=dY^i$$

then one gets the familiar line element

$$(\Delta S)^2=g_{ij}dY^i dY^j.$$

The contravariant components of g are defined such that

$$g^{ij}g_{jk}=\delta_k^i. \tag{2.11}$$

(iii) *The directional covariant derivative operator* ∇_U: The important defining properties of this operator are that (a) acting on a tensor field it produces a tensor of the same rank, (b) it is linear in U,

$$\nabla_{(f_1U_1+f_2U_2)}=f_1\nabla_{U_1}+f_2\nabla_{U_2}, \tag{2.12}$$

and (c) it acts as a derivative operator on functions and tensors

$$\nabla_U f=Uf, \tag{2.13}$$

$$\nabla_U(S\otimes T)=\nabla_U(S)\otimes T+S\otimes\nabla_U(T). \tag{2.14}$$

In a similar fashion we require that the directional covariant derivatives in superspace have the properties (2.12) and (2.13), but replace (2.14) by

$$\nabla_U(S\otimes T)=\nabla_U(S)\otimes T+(-)^{\alpha_S\alpha_\nabla}S\otimes\nabla_U(T). \tag{2.15}$$

(iv) *Generalized Lie derivatives and torsion*: As we shall see below, to explicitly calculate the connection coefficients, we need the concept of a "generalized Lie derivative." Conventionally, the Lie derivative $\mathcal{L}_U(V)$ is defined as the commutator

$$\mathcal{L}_U(V)=[U, V]. \tag{2.16}$$

As a most natural generalization of this we take

$$\mathcal{L}_U(V)\to\mathcal{H}_U(V)=[U, V\}, \tag{2.17}$$

where the generalized bracket is given by (2.4). For example, with respect to basis vectors $\{e_i\}$

$$\mathcal{H}_{e_i}(e_j)=[e_i, e_j\}=e_i e_j-(-)^{\alpha_i\alpha_j}e_j e_i. \tag{2.18}$$

For conventional manifolds a torsion tensor is defined as

$$T(U, V)=\nabla_U(V)-\nabla_V(U)-\mathcal{L}_U(V). \tag{2.19}$$

Then the vanishing of this tensor, or its coefficients in a coordinate basis is the mark of a torsion-free manifold:

$$\nabla_U(V)-\nabla_V(U)=\mathcal{L}_U(V)=[U, V]\quad\text{(torsion-free manifolds).} \tag{2.20}$$

Analogously, we introduce a tensor field $F(\mathbf{U},\mathbf{V})$ in superspace as follows:

$$F(\mathbf{U},\mathbf{V})=\nabla_{\mathbf{U}}(\mathbf{V})-(-)^{\alpha_U\alpha_V}\nabla_{\mathbf{V}}(\mathbf{U})-\mathcal{M}_{\mathbf{U}}(\mathbf{V}). \tag{2.21}$$

Clearly, the nonvanishing of this tensor indicates the presence of torsion. The converse is not true, however. That is, the vanishing of $F(\mathbf{U},\mathbf{V})$ allows for a particular type of torsion which arises as a result of the noncommutativity of the Fermi-type coordinates.

Just as torsion-free manifolds for which $T(\mathbf{U},\mathbf{V})$ vanishes are an important subclass of general Bose-type manifolds, consider the subclass of superspaces for which the tensor $F(\mathbf{U},\mathbf{V})$ vanishes. Then we have

$$\nabla_{\mathbf{U}}(\mathbf{V})-(-)^{\alpha_U\alpha_V}\nabla_{\mathbf{V}}(\mathbf{U})=\mathcal{M}_{\mathbf{U}}(\mathbf{V})\equiv[\mathbf{U},\mathbf{V}]. \tag{2.22}$$

In this case, as in the case of torsion-free manifolds, the connection coefficients can be determined uniquely. In more general cases when (2.22) is not satisfied to determine the geometry completely, one must also specify the torsion coefficients F^i_{jk} or supply an equivalent set of information.

(v) *Connection coefficients*: Since $\nabla_{\mathbf{e}_i}$ carries a tensor field it acts on into a tensor field of the same rank, then the quantity $\nabla_{\mathbf{e}_i}(\mathbf{e}_j)$ is expandable in terms of the basis set $\{\mathbf{e}_i\}$:

$$\nabla_{\mathbf{e}_i}(\mathbf{e}_j)=\Gamma^k_{ji}\mathbf{e}_k. \tag{2.23}$$

The quantities Γ^k_{ji} are called the connection coefficients. They are formally obtained from the expression

$$\Gamma^k_{ji}=\langle\omega^k,\nabla_{\mathbf{e}_i}(\mathbf{e}_j)\rangle. \tag{2.24}$$

These are immediately applicable to superspace without any modification.

To uniquely determine the connection coefficients when (2.22) is applicable, we need one more requirement, that of the vanishing of the directional covariant derivative of metric tensor:

$$\nabla_{\mathbf{U}}(\mathbf{g})=0. \tag{2.25}$$

For Bose-type manifolds, this is, of course, the usual requirement. Applying the operator $\nabla_k\equiv\nabla_{\mathbf{e}_k}$ to $\mathbf{g}$, we get

$$\nabla_k(g_{ij}\omega^i\otimes\omega^j)=[g_{ij,k}-(-)^{(\alpha_i+\alpha_j)\alpha_k}(\Gamma_{kij}+(-)^{\alpha_i\alpha_j}\Gamma_{kji})]\,\omega^i\otimes\omega^j. \tag{2.26}$$

Then (2.25) leads to

$$g_{ij,k}=(-)^{(\alpha_i+\alpha_j)\alpha_k}[\Gamma_{kij}+(-)^{\alpha_i\alpha_j}\Gamma_{kji}]. \tag{2.27}$$

Next we note from (2.22) and (2.23) that

$$\Gamma^l_{jk}-(-)^{\alpha_k\alpha_j}\Gamma^l_{kj}=-C^l_{jk}. \tag{2.28}$$

Solving these two equations for the connection coefficients, we get

$$\Gamma^l_{ki}=g^{jl}\Gamma_{kij}, \tag{2.29}$$

where

$$\Gamma_{kij}=\tfrac{1}{2}(-)^{(\alpha_i+\alpha_j)\alpha_k}[(g_{ij,k}+C_{ij,k})+(-)^{\alpha_i\alpha_j+\alpha_i\alpha_k+\alpha_j\alpha_k}\times(g_{kj,i}+C_{kji})-(-)^{\alpha_k\alpha_j+\alpha_i\alpha_k}(g_{ki,j}+C_{kij})], \tag{2.30}$$

$$C_{jki}\equiv g_{li}C^l_{jk}. \tag{2.31}$$

For manifolds with torsion one must use instead of (2.22), the general expression (2.21). Then instead of (2.28) one gets

$$\hat{\Gamma}^l_{jk}-(-)^{\alpha_k\alpha_j}\hat{\Gamma}^l_{kj}=-C^l_{jk}+F^l_{jk}, \tag{2.28a}$$

where F^l_{jk} are torsion coefficients. Then, instead of (2.30) one obtains

$$\hat{\Gamma}^l_{ki}=g^{jl}\hat{\Gamma}_{kij}=\tfrac{1}{2}g^{jl}[\Gamma_{kij}-F_{ijk}-(-)^{\alpha_i\alpha_j+\alpha_i\alpha_k+\alpha_j\alpha_k}F_{kji}-(-)^{\alpha_k\alpha_j}F_{ikj}], \tag{2.30a}$$

where the coefficients Γ_{kij} are given by (2.28).

It is to be emphasized that the order of the indices in raising and lowering various tensor coefficients associated with superspace are important and must be adhered to in our convention.

(vi) *Wedge products and exterior derivatives*: To discuss other geometrical quantities such as curvature 2-forms, we must have the analogs of wedge products in superspaces. The ordinary wedge product of two 1-forms is defined to be

$$\omega^j\wedge\omega^k=\tfrac{1}{2}(\omega^j\otimes\omega^k-\omega^k\otimes\omega^j), \tag{2.32}$$

i.e., the antisymmetric part of $\omega^j\otimes\omega^k$. A useful quantity in superspace which reduces to the above definition is the generalized wedge product

$$\omega^i\diamond\omega^j=\tfrac{1}{2}(\omega^i\otimes\omega^j-(-)^{\alpha_i\alpha_j}\omega^j\otimes\omega^i) \tag{2.33}$$

when one or both of ω^i are Bose type this reduces to (2.32). But when both ω^i and ω^j are of Fermi type, we get

$$\omega^\alpha\vee\omega^\beta=\tfrac{1}{2}(\omega^\alpha\otimes\omega^\beta+\omega^\beta\otimes\omega^\alpha), \tag{2.34}$$

i.e., the symmetric part of $\omega^\alpha\otimes\omega^\beta$.

Exterior derivative or the curl operator is the generalization of the familiar concept of the differential of a function, e.g., df. Given a basis $\{\mathbf{e}_i\}$ and the corresponding 1-forms $\{\omega^i\}$, the action of the operator d on a function is defined to be

$$df=\mathbf{e}_i f\,\omega^i. \tag{2.35}$$

In a coordinate basis this is just

$$df=\partial_i f\,d\mathbf{X}^i. \tag{2.36}$$

More generally, d is an operator which (a) acting on an n-form gives an $(n+1)$-form, (b) $dd\omega=0$ for any ω, and

$$\text{(c)}\quad d(\omega^i\wedge\omega^j)=d\omega^i\wedge\omega^j+(-)^n\omega^i\wedge d\omega^j \tag{2.37}$$

if ω^i is an n-form.

From (2.36) it is easy to see why $dd=0$ on a function f,

$$ddf=\partial_\mu\partial_\nu f\,d\mathbf{X}^\mu\wedge d\mathbf{X}^\nu,$$

$\partial_\mu\partial_\nu$ is symmetric while $d\mathbf{X}^\mu\wedge d\mathbf{X}^\nu$ is anitsymmetric.

We define the d operator in superspace with essentially the same properties as it acts on generalized wedge products:

(a) d acting on n-forms gives $(n+1)$ forms,

(b) $dd\omega=0$ for any ω,

(c) Given one forms ω^1 and ω^2,

$$d(\omega^1 \diamond \omega^2) = d\omega^1 \diamond \omega^2 + (-)^{\alpha_1 \alpha_d} \omega^1 \diamond d\omega^2. \tag{2.38}$$

In particular dd in a coordinate basis acting on functions in superspace gives

$$ddf = \partial_j \partial_i f d\mathbf{Y}^j \diamond d\mathbf{Y}^i$$
$$= \partial_\mu \partial_\nu f d\mathbf{X}^\mu \wedge d\mathbf{X}^\nu + \partial_\alpha \partial_\beta f d\theta^\alpha \vee d\theta^\beta = 0.$$

The action of d on higher forms can be worked out by induction.

(vii) *Curvature operator and the curvature tensor of the superspace*: The curvature operator R is a mixed tensor of rank (1,1). Acting on a pair of vector fields $\mathbf{U}, \mathbf{V}$, it gives another vector field:

$$R(\mathbf{U}, \mathbf{V}) = \nabla_{\mathbf{U}} \nabla_{\mathbf{V}} - \nabla_{\mathbf{V}} \nabla_{\mathbf{U}} - \nabla_{[\mathbf{U}, \mathbf{V}]}$$
$$= [\nabla_{\mathbf{U}}, \nabla_{\mathbf{V}}] - \nabla_{[\mathbf{U}, \mathbf{V}]}. \tag{2.39}$$

This is then used to define the Riemann curvature tensor with three covariant ranks and one contravariant rank. That is, given $R(\mathbf{U}, \mathbf{V})$, another vector $\mathbf{W}$, and a 1-form ω we have for the Riemann curvature tensor

$$\mathcal{R}(\omega; \mathbf{W}, \mathbf{U}, \mathbf{V}) = \omega R(\mathbf{U}, \mathbf{V})(\mathbf{W}). \tag{2.40}$$

From (2.39) it is clear that $R(\mathbf{U}, \mathbf{V})$ acting on $\mathbf{W}$ produces another vector field. So, given a basis $\{\mathbf{e}_\mu\}$, one has

$$R(\mathbf{e}_\mu, \mathbf{e}_\nu)(\mathbf{e}_\lambda) = R^\rho_{\lambda\mu\nu} \mathbf{e}_\rho. \tag{2.41}$$

The coefficients on the rhs are the components of the Riemann curvature tensor. The generalization of these concepts to superspace again involves the replacement of the commutators with the generalized brackets (2.4). Thus the curvature operator of the superspace has the form

$$R(\mathbf{U}, \mathbf{V}) = \nabla_{\mathbf{U}} \nabla_{\mathbf{V}} - (-)^{\alpha_U \alpha_V} \nabla_{\mathbf{V}} \nabla_{\mathbf{U}} - \nabla_{[\mathbf{U}, \mathbf{V}]}$$
$$= [\nabla_{\mathbf{U}}, \nabla_{\mathbf{V}}] - \nabla_{[\mathbf{U}, \mathbf{V}]}. \tag{2.42}$$

Using this, the curvature tensor of the superspace is defined as in (2.40), and the components of the curvature tensor in a basis $\{\mathbf{e}_i\}$ are given by

$$R(\mathbf{e}_i, \mathbf{e}_j)(\mathbf{e}_k) = R^l_{kij} \mathbf{e}_l. \tag{2.43}$$

From these one can then proceed to derive Cartan's equations for the superspace.

(viii) *Components of the curvature tensor of superspace*: As is clear from (2.41) for bosonic mumfolds the components of the Riemann curvature tensor are given in the basis $\{\mathbf{e}_\mu\}$:

$$R^\mu_{\nu\rho\lambda} = \langle \omega^\mu, R(\mathbf{e}_\rho, \mathbf{e}_\lambda) \mathbf{e}_\nu \rangle$$
$$= \langle \omega^\mu, ([\nabla_\rho, \nabla_\lambda] - \nabla_{[\mathbf{e}_\rho, \mathbf{e}_\lambda]}) \mathbf{e}_\nu \rangle$$
$$= \Gamma^\mu_{\nu\lambda,\rho} - \Gamma^\mu_{\nu,\rho\lambda} + \Gamma^\delta_{\nu\lambda} \Gamma^\mu_{\delta\rho} - \Gamma^\delta_{\nu\rho} \Gamma^\mu_{\delta\lambda} - C^\delta_{\rho\lambda} \Gamma^\mu_{\nu\delta},$$

where $\{\omega^\mu\}$ are the 1-forms dual to the bases $\{\mathbf{e}_\nu\}$. Accordingly we write for the components of the curvature tensor of the superspace

$$R^i_{jkl} = \langle \omega^i, ([\nabla_k, \nabla_l\} - \nabla_{[\mathbf{e}_k, \mathbf{e}_l]}) e_j \rangle$$
$$= \langle \omega^i, R(\mathbf{e}_k, \mathbf{e}_l) \mathbf{e}_j \rangle \tag{2.44}$$

$$= \Gamma^i_{jl,k} - (-)^{\alpha_k \alpha_l} \Gamma^i_{jk,l} + (-)^{\alpha_k \alpha_m + \alpha_k \alpha_j + \alpha_l \alpha_k} \Gamma^m_{jl} \Gamma^i_{mk}$$
$$(-)^{\alpha_j \alpha_l + \alpha_m \alpha_l} \Gamma^m_{jk} \Gamma^i_{ml} - C^m_{kl} \Gamma^i_{jm}. \tag{2.45}$$

From this we get from the Ricci tensor of superspace

$$R_{jl} \equiv R^i_{jil} = \Gamma^i_{jl,i} - (-)^{\alpha_i \alpha_l} \Gamma^i_{ji,l} + (-)^{\alpha_i \alpha_m + \alpha_i \alpha_j + \alpha_i \alpha_l} \Gamma^m_{jl} \Gamma^i_{mi}$$
$$- (-)^{\alpha_i \alpha_l + \alpha_m \alpha_l} \Gamma^m_{ji} \Gamma^i_{ml} - C^m_{il} \Gamma^i_{jm}. \tag{2.46}$$

Finally, the scalar curvature of the superspace is given by

$$R = R^j_j = g^{lj} R_{jl}. \tag{2.47}$$

III. A GRADED LIE BUNDLE AND SUPERGRAVITY THEORY

As a simple application of the superspace geometries developed in the previous section, we consider a simple graded Lie bundle, i.e., one which has the space–time as base manifold and a graded Lie group as structural group. The geometrical invariants of such a fiber bundle do not as a rule lead to locally supersymmetric scalars. The reason is that in such a scheme the covariance is defined only with respect to space–time coordinate transformations. But it is shown that by using a particular combination of available gauge potentials it is possible to construct a locally supersymmetric scalar which is identical with the Deser–Zumino Lagrangian.
In this case, as well as the Deser–Zumino case, it is not clear why one should pick such a combination of the gauge fields involved. A logically more consistent approach to locally supersymmetric theories will be given elsewhere.[18]

The main strategy for the computation of various geometrical quantities will be the same as that described in Refs. 12 and 19. But now we will use the appropriate formulas derived in the previous section for a superspace.

The basis: We begin by writing down the algebra of the graded Lie group which we want to use as a fiber. It is the 14-parameter group with generators

$$\{P_A, J_{AB}, S_\alpha\}, \quad A = 0, \ldots, 3, \quad \alpha = 1, \ldots, 4, \tag{3.1}$$

where P_A are the generators of translations, J_{AB} those of homogeneous Lorentz transformations, and S_α are the supersymmetry generators. Since this algebra has a Poincaré subalgebra, most of the necessary computations have already been carried out in Ref. 19. Therefore, whenever no confusion can arise we only give the additional parts which involve the supersymmetry generators S_α.

The fiber bundle of interest to us is an 18-dimensional manifold, and to describe it we need to specify a set of 18 basis vectors in the tangent space to a point of the bundle. We take the basis in the vertical sector of the tangent space to be isomorphic to the algebra (3.1). For the horizontal tangent space we can take either the coordinate basis

$$\{\mathbf{h}_\mu\} = \{\partial_\mu\} \tag{3.2}$$

or the gauge covariant basis

$$\mathbf{E}_\mu = \mathbf{h}_\mu + N^A_\mu \mathbf{h}_A + N^{AB}_\mu \mathbf{h}_{AB} + N^\alpha_\mu \mathbf{h}_\alpha. \tag{3.3}$$

To specify a basis completely, we must specify the generalized commutators of its elements. Thus in the gauge covariant basis

$$\{\mathbf{E}_\mu, \mathbf{E}_A, \mathbf{E}_{AB}, \mathbf{E}_\alpha\} \tag{3.4}$$

we have

$$[\mathbf{E}_\mu, \mathbf{E}_\nu] = -F^A_{\mu\nu}\mathbf{E}_A - F^{AB}_{\mu\nu}\mathbf{E}_{AB} - F^\alpha_{\mu\nu}\mathbf{E}_\alpha,$$
$$0 = [\mathbf{E}_\mu, \mathbf{E}_A] = [\mathbf{E}_\mu, \mathbf{E}_{AB}] = [\mathbf{E}_\mu, \mathbf{E}_\alpha] = [\mathbf{E}_A, \mathbf{E}_B] = [\mathbf{E}_A, \mathbf{E}_\alpha], \tag{3.5}$$

$$[\mathbf{E}_A, \mathbf{E}_{BB'}] = f^C_{A\,BB'}\mathbf{E}_C, \quad [\mathbf{E}_\alpha, \mathbf{E}_{AB}] = f^B_{\alpha\,AB}\mathbf{E}_B,$$
$$[\mathbf{E}_{AA'}, \mathbf{E}_{BB'}] = f^{CC'}_{AA'\,BB'}\mathbf{E}_{CC'}, \tag{3.6}$$
$$\{\mathbf{E}_\alpha, \mathbf{E}_\beta\} = f^A_{\alpha\beta}\mathbf{E}_A.$$

In the direct product basis we take the set

$$\{\mathbf{h}_\mu, \mathbf{h}_A, \mathbf{h}_{AB}, \mathbf{h}_\alpha\}, \tag{3.7}$$

where

$$[\mathbf{h}_\mu, \mathbf{h}_\nu] = [\mathbf{h}_\mu, \mathbf{h}_A] = [\mathbf{h}_\mu, \mathbf{h}_{AB}] = [\mathbf{h}_\mu, \mathbf{h}_\alpha] = 0$$

and the algebra of the bases in vertical tangent space is the same as that in (3.6). As demonstrated in Refs. 12 and 19, the connection coefficients N^A_μ, N^{AB}_μ, and N^α_μ in (3.3) are the "gauge potentials" of the generators specified by their upper index. Similarly, the quantities $F^A_{\mu\nu}$, $F^{AB}_{\mu\nu}$, $F^\alpha_{\mu\nu}$ are the corresponding field tensors. Since $\mathbf{E}_\mu$ is a Bose-type operator, it is clear from (3.3) and (3.5) that N^α_μ and $F^\alpha_{\mu\nu}$ are anticommuting objects.

The metric: The computations are most easily carried out in the gauge covariant basis. So we shall specify the components of the metric tensor in this basis in which, by definition, the metric tensor is block diagonal. So we have for the metric tensor components

$$\begin{bmatrix} g_{\mu\nu} & & & \\ & \eta_{\hat{A}\hat{B}} & & \\ & & \eta_{AB} & \\ & & & \eta_{\alpha\beta} \end{bmatrix}, \tag{3.8}$$

where

$$g_{\mu\nu} = \mathbf{E}_\mu \cdot \mathbf{E}_\nu = g_{\nu\mu} = \mathbf{E}_\nu \cdot \mathbf{E}_\mu,$$
$$\eta_{\hat{A}\hat{B}} = \mathbf{E}_{\hat{A}} \cdot \mathbf{E}_{\hat{B}} = \eta_{\hat{B}\hat{A}} = \mathbf{E}_{\hat{B}} \cdot \mathbf{E}_{\hat{A}}, \tag{3.9}$$
$$\eta_{AB} = \mathbf{E}_A \cdot \mathbf{E}_B = \eta_{BA} = \mathbf{E}_B \cdot \mathbf{E}_A,$$
$$\eta_{\alpha\beta} = \mathbf{E}_\alpha \cdot \mathbf{E}_\beta = -\eta_{\beta\alpha} = -\mathbf{E}_\beta \cdot \mathbf{E}_\alpha,$$

and index $\hat{A} = AA'$, $g_{\mu\nu}$ is the metric tensor of the base manifold, and η_{AB} and $\eta_{\hat{A}\hat{B}}$ are Euclidean metrics.

Motion along fiber: By using the Jacobi identities as well as other commutators, equations of variation of various quantities such $F^A_{\mu\nu}$, N^α_μ, etc., along the fiber can be derived. Many of these which are due to the Poincaré subgroup have been given in Ref. 20. We therefore record here those which are new or which are modified:

$$\partial_{\hat{B}} F^\alpha_{\mu\nu} = -f^\alpha_{\hat{B}\beta} F^\beta_{\mu\nu},$$
$$\partial_A F^\alpha_{\mu\nu} = \partial_\alpha F^{\hat{B}}_{\mu\nu} = 0,$$
$$\partial_\alpha F^B_{\mu\nu} = f^B_{\alpha\beta} F^\beta_{\mu\nu}, \tag{3.10}$$
$$\partial_\alpha F^\beta_{\mu\nu} = -f^\beta_{\alpha\hat{B}} F^{\hat{B}}_{\mu\nu},$$

$$\partial_{\hat{B}} N^\alpha_\mu = -f^\alpha_{\hat{B}\beta} N^\beta_\mu,$$
$$\partial_A N^\alpha_\mu = \partial_\alpha N^{\hat{B}}_\mu = 0,$$
$$\partial_\alpha N^B_\mu = f^B_{\alpha\beta} N^\beta_\mu,$$
$$\partial_\alpha N^\beta_\mu = -f^\beta_{\alpha\hat{A}} N^{\hat{A}}_\mu. \tag{3.11}$$

Relation between field tensors and potentials: These are derived[12,19] by considering the relation between the basis (3.4) and (3.7). One gets

$$F^{\hat{A}}_{\mu\nu} = N^{\hat{A}}_{\mu,\nu} - N^{\hat{A}}_{\nu,\mu} + f^{\hat{A}}_{\hat{B}\hat{C}} N^{\hat{B}}_\mu N^{\hat{C}}_\nu, \tag{3.12}$$

$$F^A_{\mu\nu} = N^A_{\mu,\nu} - N^A_{\nu,\mu} + f^A_{\hat{B}C}(N^{\hat{B}}_\mu N^C_\nu - N^C_\nu N^{\hat{B}}_\nu) - f^A_{\alpha\beta} N^\alpha_\mu N^\beta_\nu, \tag{3.13}$$

$$F^\alpha_{\mu\nu} = N^\alpha_{\mu,\nu} - N^\alpha_{\nu,\mu} + f^\alpha_{\beta\hat{A}}(N^{\hat{A}}_\mu N^\beta_\nu - N^{\hat{A}}_\nu N^\beta_\mu). \tag{3.14}$$

The invariance, or the lack thereof, of various geometrical invariants under local supersymmetry transformations can be surmised at this stage. The quantity $N^{\hat{A}}_\mu$ appearing in the three $F_{\mu\nu}$'s above is related via our parallel transport requirement to the connection coefficients in the base manifold. Covariance with respect to such a connection is the covariance with respect to general space–time coordinate transformations with or without and *a priori* does not involve local supersymmetry. The proper way of ensuring local supersymmetry is to ensure covariance with respect to a suitable connection in the base manifold.

Let us ignore for the moment the logical basis of writing down locally supersymmetric scalar densities. Given $F^\alpha_{\mu\nu}$ and N^β_ρ, we can construct the contracted object

$$\epsilon_A{}^{\rho\mu\nu} f^A_{\alpha\beta} N^\beta_\rho F^\alpha_{\mu\nu} / \det(N^A_\mu).$$

Also given $F^{AB}_{\mu\nu}$ and N^A_μ, we can construct the contracted object

$$N^\mu_A N^\nu_B F^{AB}_{\mu\nu}.$$

Clearly, the sum of these two expressions multiplied by $\det(N^\mu_A)$ is just the first order Lagrangian density.[17] What is lacking is a logical basis for making this choice.

The connection coefficients: We now specialize the problem to torsion-free manifolds. These coefficients can be directly calculated from (2.29) in the gauge covariant basis. The ones which are nonzero are listed below. With $a = A$, $\hat{A}$, or α, we have

$$\Gamma^a_{bc} = -\tfrac{1}{2} f^a_{bc},$$
$$\Gamma^\nu_{\mu a} = \Gamma^\nu_{a\mu} = -\tfrac{1}{2} g^{\rho\nu} \eta_{ab} F^b_{\mu\rho}, \tag{3.15}$$
$$\Gamma^a_{\mu\nu} = -\Gamma^a_{\nu\mu} = \tfrac{1}{2} F^\alpha_{\mu\nu},$$
$$\Gamma^\lambda_{\mu\nu} = \left\{ {\lambda \atop \mu\nu} \right\}_{\text{base}},$$

where $\left\{ {\lambda \atop \mu\nu} \right\}_{\text{base}}$ are the connection coefficients of the base manifold. In a coordinate basis they reduce to the Christoffel symbols.

Scalar curvature of the bundle: It is now straightforward to compute various geometrical covariants and invariants of the bundle. Here we just write down the expression for the scalar curvature of the bundle:

$$R = R_G + R_{\text{base}} - \tfrac{1}{4} g^{\mu\rho} g^{\nu\lambda} \eta_{AB} F^A_{\mu\lambda} F^B_{\nu\rho}$$
$$- \tfrac{1}{4} g^{\mu\rho} g^{\nu\lambda} \eta_{\hat{A}\hat{B}} F^{\hat{A}}_{\mu\lambda} F^{\hat{B}}_{\nu\rho} - \tfrac{1}{4} g^{\mu\rho} g^{\nu\lambda} \eta_{\alpha\beta} F^\alpha_{\mu\lambda} F^\beta_{\nu\rho}, \tag{3.16}$$

where

R_G = scalar curvature of the group manifold,

R_{base} = scalar curvature of the base manifold.

Now we require that the homogeneous part of the connection in the fiber bundle be related to the connection in base manifold by the principle of parallel transport,[19]

$$N_\mu^{AB} = \eta^{BC}\Gamma_{\mu C}^{A} \tag{3.17}$$

and that

$$g_{\mu\nu} = \eta_{AB}N_\mu^A N_\nu^B. \tag{3.18}$$

Then $\nabla_U(N_\mu^A)=0$ would give $F_{\mu\nu}^A=0$ so that the contribution of translation field tensors to R vanishes. By (3.17) the $F_{\mu\nu}^{\hat{A}}$ are then expressed in terms of $g_{\mu\nu}$ and the connection coefficients of the base manifold. The last term in (3.16) is the contribution of the "graded" part of the structural group. Since supersymmetry transformations are not incorporated in the base manifold, the graded part acts more or less like an internal symmetry group, a feature which is not desirable. The proper way of incorporating the graded part is discussed elsewhere.[18]

IV. FINAL REMARKS

We have given a self-contained discussion of the geometry of superspaces from a fairly general point of view and have derived explicit formulas for the computation of various geometrical covariants and invariants. The general formalism is then illustrated by studying the geometry of a simple graded Lie bundle and by writing down locally supersymmetric Lagrangian densities. It is hoped that this work will help resolve the mathematical problems involved in dealing with supersymmetries.

The primary aim of this paper has been to clarify questions which will pave the way for useful physical interpretations of supersymmetries in particle physics. But the generalized concepts, several of which are new, are also of intrinsic mathematical interest. It is quite remarkable that a simple extension of commutator to a more general Lie bracket plays such a crucial role in shaping all the geometrical characteristics of superspace. Aside from mere elegance and generality, the coordinate independent approach adopted in this work was strongly motivated by the desire to display how the passage from conventional to superspace geometries could be envisaged algebraically. Once this algebraic approach to arriving at new geometries is recognized, one can imagine the possibility of constructing other sophisticated geometries based on more complicated algebras. In particular, we conjecture that this point of view will prove useful in the study of geometries related to nonassociative algebras.

ACKNOWLEDGMENTS

I thank F. Gürsey and Y. Nambu for their encouragement and S.W. MacDowell for many helpful discussions. I would like to express my appreciation to the Aspen Center for Physics for their hospitality in the summer of 1976 where this work was written up. In particular, stimulating conversations with the helpful suggestions by D. Z. Freedman, M. Gell-Mann, and Y. Ne'eman are gratefully acknowledged.

*Research supported in part by ERDA.

[1]J. Wess and B. Zumino, Nucl. Phys. B 70, 39 (1974).
[2]D.V. Volkov and V.P. Akulov, Phys. Lett. B 46, 109 (1973); D.V. Volkov and V.A. Soroka, Zh. Eksp. Teor. Fiz. Pis. Red. 18, 529 (1973) [JETP Lett. 18, 312 (1973); Y.A. Golfand and E.P. Liktman, JETP Lett. 13, 323 (1971).
[3]A. Salam and J. Strathdee, Nucl. Phys. B 76, 477 (1974); Phys. Rev. D 11, 1521 (1975).
[4]For a recent review see S. Ferrara, Frascati Preprint LNF-75/46 (P), to appear in Nuovo Cimento.
[5]F.A. Berezin, *The Method of 2nd Quantization* (Academic, 1966); F.A. Berezin and G.I. Katz, Math. Sb. (USSR) 82, 343 (1970) [English Transl., Vol. II]; V.G. Kac, Funct. Anal. Appl. 9, 91 (1975) (Russian).
[6]L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. 47, S73 (1975), and Tel Aviv Report TAUP 448-74 (1974).
[7]A. Pais and V. Rittenberg, J. Math. Phys. 16, 206 (1975); P.G.O. Freund and I. Kaplansky, J. Math. Phys. 17, 228 (1976); B. Durand, University of Wisconsin preprint, 1976.
[8]L.N. Chang, K. Macrae, and F. Mansouri, Phys. Lett. B 57, 59 (1975).
[9]P. Nath and R. Arnowitt, Phys. Lett. B 56, 177 (1975).
[10]R. Arnowitt, P. Nath, and B. Zumino, Phys. Lett. B 56, 81 (1975).
[11]P.G.O. Freund and Y.M. Cho, Phys. Rev. D 12, 1711 (1975).
[12]L.N. Chang, K. Macrae, and F. Mansouri, Phys. Rev. D 13, 235 (1976).
[13]B. Zumino, in *The Proceedings of the Conference on Gauge Theories and Modern Field Theory, Northeastern University, Sept.*, 1975, edited by R. Arnowitt and P. Nath (MIT Press, Cambridge, Mass., 1976).
[14]R. Casalbuoni, G. Domokos, and S. Kovesi-Domokos, Nuovo Cimento A 31, 423 (1976); CERN Preprint TH-2150-CERN (1976).
[15]Similar programs are being carried out by G. Domoks and S. Kovesi-Domokos (private communication); by S.W. MacDowell (private communication); and by M. Gell-Mann and Y. Ne'eman (private communication).
[16]D.Z. Freedman and P. van Nieuwenhuizen, Phys. Rev. D 13, 3214 (1976).
[17]S. Deser and B. Zumino, CERN Preprint TH2164 (1976).
[18]S.W. MacDowell and F. Mansouri, in preparation.
[19]F. Mansouri and L.N. Chang, Phys. Rev. D 13, 3192 (1976).
[20]See, for example, C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

The Kähler structure of asymptotic twistor space

M. Ko and E. T. Newman*

Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

R. Penrose

Mathematics Institute, Oxford University, Oxford, England
(Received 2 September 1976)

Asymptotic twistor space $\mathcal{T}$ is a 4-complex-dimensional Kähler manifold (of signature $++--$) which can be constructed from an asymptotically flat space-time containing gravitational radiation. The properties of this Kähler structure are investigated, the Kähler metric being of a particular type, arising from a scalar Σ with special homogeneity properties. The components of the Kähler curvature $K^{\alpha\beta}_{\gamma\delta}$ are found explicitly in terms of the asymptotic Weyl curvature of the space-time. When gravitational radiation is present, $K^{\alpha\beta}_{\gamma\delta} \neq 0$, whereas for a stationary field $K^{\alpha\beta}_{\gamma\delta} = 0$. The "Ricci-flat" condition $K^{\alpha\beta}_{\alpha\gamma} = 0$ is found always to hold.

INTRODUCTION

The theory of twistors provides a novel approach to the treatment of special-relativistic physics which is particularly well suited to the description of massless fields (of any spin) and other conformally invariant aspects of physics.[1,2] The theory can also be used to treat situations in which conformal invariance does not hold and when the space–time is not flat but, as yet, little work has been published on this. In the present paper we examine the curved twistor space $\mathcal{T}$ which arises naturally from an asymptotically flat space–time $\mathcal{M}$ in which gravitational radiation may be present. It is the presence of this radiation which modifies the structure of the twistor space from the standard one arising in the case of flat Minkowski space M. This standard twistor space T is a complex four-dimensional vector space, on which is defined a Hermitian form Σ of signature $(++--)$. [This structure for T is invariant under the fifteen parameter conformal group $C^{\uparrow}_{+}(1,3)$ for M, in accordance with the local isomorphism $SU(2,2) \to C^{\uparrow}_{+}(1,3)$.] But when (outgoing) radiation is present the vector space structure of the twistor space is lost. However, $\mathcal{T}$ remains a complex manifold; furthermore it retains a real scalar field Σ defined analogously to the standard Σ for T, but which cannot now be regarded as a Hermitian form owing to the lack of linear structure for $\mathcal{T}$.[1] The second derivatives of Σ (once holomorphic and once antiholomorphic) with respect to the coordinates in $\mathcal{T}$ define the components of a Kähler metric on $\mathcal{T}$ [of signature $(++--)$]. The curvature $K^{\alpha\beta}_{\gamma\delta}$ of this Kähler metric had been computed earlier, using a combination of algebraic and geometric arguments, the result being given in barest outline in Ref. 1. It turns out that only five of the 100 independent curvature components survive, that $K^{\alpha\beta}_{\gamma\delta}$ is Ricci-flat (i.e., $K^{\alpha\beta}_{\alpha\delta} = 0$), and that these surviving components can all be expressed explicitly in terms of standard asymptotic curvature quantities for the space–time $\mathcal{M}$ which describe the radiation behavior of the gravitational field. These earlier arguments, somewhat involved conceptually and difficult to express explicitly, have not been described in print. Here we present a much more direct approach to the problem, which shows how these curvature components may be obtained by a straightforward computation.

The discussion given here, as regards twistor theory, will be essentially self-contained. Some familiarity with the standard notation for asymptotically flat space–times will be assumed.

1. KÄHLER MANIFOLDS

The definition of a Kähler manifold K can be given in various different equivalent forms.[3–5] For our present purposes, K is a complex n-dimensional manifold[6] on which is defined a nondegenerate Hermitian metric $\underline{G}$, obtainable (locally) as a second derivative (once holomorphic and once antiholomorphic) of a real scalar field Σ. Thus, if z^{α} $(\alpha = 1, \ldots, n)$ are local holomorphic coordinates for K, with complex conjugates $\bar{z}^{\alpha'}$ $(\alpha' = 1', \ldots, n')$, then the components of $\underline{G}$ in this coordinate system constitute the nonsingular Hermitian matrix[7]

$$\begin{bmatrix} G_{11'} & \cdots & G_{1n'} \\ \vdots & & \vdots \\ G_{n1'} & \cdots & G_{nn'} \end{bmatrix} = G_{\alpha\beta'} = \frac{\partial^2 \Sigma}{\partial z^{\alpha} \partial \bar{z}^{\beta'}} . \tag{1.1}$$

Any vector field $\underline{V}$ (real or complex) can be expressed in these coordinates as

$$\underline{V} = V^{\alpha} \frac{\partial}{\partial z^{\alpha}} + W^{\alpha'} \frac{\partial}{\partial \bar{z}^{\alpha'}} . \tag{1.2}$$

The vector field is *real* if

$$W^{\alpha'} = \bar{V}^{\alpha'} \quad (= \overline{V^{\alpha}}). \tag{1.3}$$

It is called *holomorphic* if $W^{\alpha'} = 0$ and V^{α} are holomorphic functions of z^{β} (i.e., $\partial V^{\alpha} / \partial \bar{z}^{\beta'} = 0$); it is called *antiholomorphic* if $V^{\alpha} = 0$ and $W^{\alpha'}$ are holomorphic in $\bar{z}^{\beta'}$ (so the complex conjugate of a holomorphic vector field is antiholomorphic). A scalar field ϕ is holomorphic if simply $\partial \phi / \partial \bar{z}^{\alpha'} = 0$ and antiholomorphic if $\partial \phi / \partial z^{\alpha} = 0$.

For many purposes it is convenient to use a single index for the entire set of coordinates $z^1, \ldots, z^n, \bar{z}^{1'}, \ldots, \bar{z}^{n'}$ and write

$$V^{\Gamma} = (V^{\alpha}, W^{\alpha'}) \quad (\Gamma = 1, \ldots, n, 1', \ldots, n'). \tag{1.4}$$

Then the components

$$G_{\Gamma\Lambda}=\begin{bmatrix}0 & G_{\gamma\lambda'}\\ G_{\lambda\gamma'} & 0\end{bmatrix} \tag{1.5}$$

define a (real) Riemannian metric on K. This metric is real because

$$G_{\Gamma\Lambda}V^{\Gamma}V^{\Lambda}=G_{\gamma\lambda'}V^{\gamma}W^{\lambda'}+G_{\lambda\gamma'}W^{\gamma'}V^{\lambda} \tag{1.6}$$

is real whenever $\underline{V}$ is real. A Riemann curvature tensor $\underline{R}$ is thereby defined, with components $R_{\Gamma\Lambda\Sigma\Theta}$, the only nonvanishing members turning out to be those of the form

$$R_{\gamma\lambda'\xi\theta'},\quad R_{\gamma\lambda'\theta'\xi},\quad R_{\gamma'\lambda\xi\theta'},\quad R_{\gamma'\lambda\xi'\theta}. \tag{1.7}$$

The Kähler metric components $G_{\alpha\beta'}$, together with their inverse matrix $G^{\alpha\beta'}$ for which

$$G_{\alpha\beta'}G^{\gamma\beta'}=\delta^{\gamma}_{\alpha},\quad G^{\alpha\beta'}G_{\alpha\gamma'}=\delta^{\beta'}_{\gamma'}, \tag{1.8}$$

can be used to eliminate all primed indices from a tensor expression by pushing them all into the opposite position, e.g.,

$$T_{\alpha\beta\gamma'}{}^{\rho\sigma'}\rightarrow T_{\alpha\beta}{}^{\gamma\rho}{}_{\sigma}=T_{\alpha\beta\mu'}{}^{\rho\nu'}G^{\gamma\mu'}G_{\sigma\nu'}. \tag{1.9}$$

The surviving Riemann tensor components (1.7) can then be expressed in terms of the quantity $K^{\alpha\beta}_{\lambda\mu}$

$$K^{\alpha\beta}_{\lambda\mu}=R_{\lambda}{}^{\alpha}{}_{\mu}{}^{\beta}=-R_{\lambda}{}^{\alpha\beta}{}_{\mu}=-R^{\alpha}{}_{\lambda\mu}{}^{\beta}=R^{\alpha}{}_{\lambda}{}^{\beta}{}_{\mu}. \tag{1.10}$$

It turns out (as we shall see shortly) that $K^{\alpha\beta}_{\lambda\mu}$ enjoys a symmetry and a Hermiticity property:

$$\overline{K^{\alpha\beta}_{\lambda\mu}}=\overline{K^{\alpha'\beta'}_{\lambda'\mu'}}=K^{\alpha\beta}_{\lambda\mu}=K^{(\alpha\beta)}_{(\lambda\mu)}. \tag{1.11}$$

Calculations can be made independent of coordinates by the introduction of an abstract index formulation. Indices $\alpha,\beta,\ldots,\omega,\alpha_0,\ldots,\alpha_1,\cdots$ are used here abstractly (so V^{α} *is* a vector, namely the vector with components V^{α}—the "primed" components being taken as zero). The corresponding *conjugate* abstract indices $\alpha',\beta',\cdots$ are *not* needed since we can write W_{α} for the vector with components $W^{\beta'}G_{\alpha\beta'}$. Thus, instead of giving rise to primed abstract indices, compex conjugation entails the interchange of all upper and lower indices, e.g.,

$$\overline{T^{\alpha\beta}}_{\gamma}=\overline{T}_{\alpha\beta}{}^{\gamma}, \tag{1.12}$$

which expresses the fact that $\overline{T^{\alpha\beta}}_{\gamma}=\overline{T}^{\alpha'\beta'}{}_{\gamma'}$, corresponds to $\overline{T}_{\alpha\beta}{}^{\gamma}=\overline{T}^{\alpha'\beta'}{}_{\gamma'}G_{\alpha\alpha'}G_{\beta\beta'}G^{\gamma\gamma'}$. The abstract version of the Kähler metric tensor is, with this notation, simply the Kronecker tensor δ^{β}_{α}, as follows from (1.8). The Kähler Hermitian scalar product between vectors U^{α}, V^{α} is thus

$$U^{\alpha}\overline{V}_{\alpha}=G_{\alpha\beta'}U^{\alpha}\overline{V}^{\beta'}. \tag{1.13}$$

The notation is thereby brought into line with that of standard twistor theory.[1,2]

The information concerning the curvature, etc., of the Kähler structure goes into the formal properties of covariant derivative. In components we have

$$V^{\Theta}{}_{;\Phi}=V^{\Theta}{}_{,\Phi}+\Gamma^{\Theta}_{\Psi\Phi}V^{\Psi}, \tag{1.14}$$

where the $\Gamma^{\Theta}_{\Psi\Phi}$ are the ordinary Christoffel symbols for the metric (1.5). From (1.1) and (1.5) one obtains

$$\Gamma^{\theta}_{\phi\phi}=G^{\theta\chi'}\frac{\partial^3\Sigma}{\partial z^{\phi}\partial z^{\phi}\partial\bar{z}^{\chi'}},\quad \Gamma^{\theta'}_{\phi'\phi'}=G^{\chi\theta'}\frac{\partial^3\Sigma}{\partial z^{\chi}\partial\bar{z}^{\phi'}\partial\bar{z}^{\phi'}},$$

$$\Gamma^{\theta}_{\phi\phi'}=\Gamma^{\theta}_{\phi'\phi}=\Gamma^{\theta'}_{\phi\phi}=\Gamma^{\theta'}_{\phi'\phi'}=\Gamma^{\theta}_{\phi'\phi}=\Gamma^{\theta'}_{\phi\phi'}=0 \tag{1.15}$$

so that

$$V^{\theta}{}_{;\phi}=\frac{\partial V^{\theta}}{\partial z^{\phi}}+\Gamma^{\theta}_{\phi\phi}V^{\phi},\quad V^{\theta;\phi}=\frac{\partial V^{\theta}}{\partial\bar{z}^{\phi'}}G^{\phi\phi'},$$

$$W_{\theta;\phi}=\frac{\partial W_{\theta}}{\partial z^{\phi}}-\Gamma^{\phi}_{\theta\phi}W_{\phi},\quad W_{\theta}{}^{;\phi}=\frac{\partial W_{\theta}}{\partial\bar{z}^{\phi'}}G^{\phi\phi'}. \tag{1.16}$$

The four expressions (1.16) are the component forms of the respective abstract expressions[9]

$$\Box_{\phi}V^{\theta},\quad \Box^{\phi}V^{\theta},\quad \Box_{\phi}W_{\theta},\quad \Box^{\phi}W_{\theta}. \tag{1.17}$$

The two derivative operators $\Box_{\phi}$, $\Box^{\phi}$ extend so as to apply to any tensor quantity $T^{\lambda\cdots\nu}_{\alpha\cdots\gamma}$ in the usual way, so that the normal additivity and Leibniz properties hold. The relation to the curvature $K^{\alpha\beta}_{\lambda\mu}$ is achieved via

$$(\Box_{\lambda}\Box^{\alpha}-\Box^{\alpha}\Box_{\lambda})V^{\beta}=K^{\alpha\beta}_{\lambda\mu}V^{\mu},$$

$$(\Box^{\lambda}\Box_{\alpha}-\Box_{\alpha}\Box^{\lambda})W_{\beta}=K^{\lambda\mu}_{\alpha\beta}W_{\mu}, \tag{1.18}$$

as follows from (1.10) and the relation

$$V^{\Theta}{}_{;\Phi;\Psi}-V^{\Theta}{}_{;\Psi;\Phi}=R^{\Theta}{}_{\Xi\Phi\Psi}V^{\Xi}. \tag{1.19}$$

Alternatively, the coordinate expressions

$$K_{\alpha\beta'\lambda\mu'}=-G_{\lambda\mu'},_{\alpha\beta'}+G^{\rho\sigma'}G_{\rho\mu'},_{\beta'}G_{\lambda\sigma'},_{\alpha}, \tag{1.20}$$

$$K_{\alpha\beta'\lambda}{}^{\mu}=-\Gamma^{\mu}_{\alpha\lambda,\beta'} \tag{1.21}$$

may be employed.

We have the vanishing torsion properties

$$\Box_{\alpha}\Box^{\beta}f=\Box^{\beta}\Box_{\alpha}f \tag{1.22}$$

and

$$\Box_{[\alpha}\Box_{\beta]}f=0,\quad \Box^{[\alpha}\Box^{\beta]}f=0 \tag{1.23}$$

for any scalar f. In fact, relations (1.23) generalize to the commutation properties

$$\Box_{\alpha}\Box_{\beta}=\Box_{\beta}\Box_{\alpha},\quad \Box^{\alpha}\Box^{\beta}=\Box^{\beta}\Box^{\alpha} \tag{1.24}$$

applied to *any* tensor quantity. Equation (1.24) follows because (1.7) are the only surviving Riemann tensor components. But a simple "abstract" proof of these facts can also be given. Choose f to be an arbitrary holomorphic scalar on K. Then $\Box^{\alpha}f=0$ (since the component version of $\Box^{\alpha}f$ is $G^{\alpha\beta'}\partial f/\partial\bar{z}^{\beta'}$), whereas $\Box_{\alpha}f$ is arbitrary at any one point. Since the torsion vanishes, we would have a curvature quantity $Q^{\alpha\beta\gamma}_{\lambda}$ for which

$$(\Box^{\alpha}\Box^{\beta}-\Box^{\beta}\Box^{\alpha})\Box_{\lambda}f=Q^{\alpha\beta\gamma}_{\lambda}\Box_{\gamma}f. \tag{1.25}$$

But

$$\Box^{\alpha}\Box^{\beta}\Box_{\lambda}f=\Box^{\alpha}\Box_{\lambda}\Box^{\beta}f=0,$$

and similarly if α and β are interchanged. Thus (1.25) vanishes, and so $Q^{\alpha\beta\gamma}_{\lambda}=0$ because of the arbitrariness at each point of $\Box_{\gamma}f$. The second of relations (1.24) follows—and the proof of the first is similar.

The symmetry properties of $K^{\alpha\beta}_{\lambda\mu}$ also follow directly from the formal properties of $\Box_\alpha$ and $\Box^\alpha$. The "Hermiticity"

$$\bar{K}^{\alpha\beta}_{\lambda\mu} = K^{\alpha\beta}_{\lambda\mu} \tag{1.26}$$

[cf. (1.11)] is a direct consequence of (1.18) and the fact that $\Box_\alpha$ and $\Box^\alpha$ are complex conjugates of one another. (In components, $\overline{\Box_\alpha} = G_{\beta\alpha'}\Box^{\beta'}$.) The symmetry of $K^{\lambda\mu}_{\alpha\beta}$ in α, β follows because

$$\begin{aligned} K^{\lambda\mu}_{\alpha\beta}\Box_\mu f &= (\Box^\lambda\Box_\alpha - \Box_\alpha\Box^\lambda)\Box_\beta f \\ &= \Box^\lambda\Box_\alpha\Box_\beta f - \Box_\alpha\Box_\beta\Box^\lambda f, \end{aligned}$$

each term of which is symmetric in α and β, by (1.24) (whether or not f is chosen holomorphic). The symmetry in λ, μ follows similarly, so that

$$K^{\lambda\mu}_{\alpha\beta} = K^{(\lambda\mu)}_{(\alpha\beta)}. \tag{1.27}$$

The "Bianchi identities" are correspondingly derived:

$$\begin{aligned} \Box_{[\alpha}(K^{\lambda}_{\beta]\gamma}{}^{\mu}W_\mu) &= \Box_{[\alpha}(\Box^\lambda\Box_{\beta]} - \Box_{\beta]}\Box^\lambda)W_\gamma \\ &= \Box^\lambda\Box_{[\alpha}\Box_{\beta]}W_\gamma - K^{\lambda\mu}_{[\alpha\beta]}\Box_\mu W_\gamma - K^{\lambda\mu}_{\gamma[\alpha}\Box_{\beta]}W_\mu - 0 \end{aligned}$$

gives, on expanding the left-hand side and using (1.24), (1.27),

$$\Box_{[\alpha}K^{\lambda}_{\beta]\gamma}{}^{\mu} = 0 \tag{1.28}$$

and, similarly

$$\Box^{[\lambda}K^{\mu}_{\alpha}{}^{\nu]}_{\beta} = 0. \tag{1.29}$$

Note that by (1.27) we can re-express these relations as

$$\Box_\alpha K^{\lambda\mu}_{\beta\gamma} = \Box_{(\alpha}K^{\lambda\mu}_{\beta\gamma)}, \quad \Box^\lambda K^{\mu\nu}_{\alpha\beta} = \Box^{(\lambda}K^{\mu\nu)}_{\alpha\beta}. \tag{1.30}$$

2. KÄHLER MANIFOLDS WITH HOMOGENEOUS SCALAR

All the above properties hold in a general Kähler manifold. The scalar Σ need not be globally defined and is not, itself, considered to be part of the Kähler structure of K. However, in the situation that concerns us here, Σ *has geometric meaning*. Furthermore, it has a certain homogeneity property which can be described as follows: There exist coordinates z^α, $\bar{z}^{\alpha'}$ (possibly defined only locally) for which

$$\Sigma = \Sigma(z^\alpha, \bar{z}^{\alpha'}) \tag{2.1}$$

is separately homogeneous of degree unity in z^α and in $\bar{z}^{\alpha'}$, i.e.,

$$\Sigma(\lambda z^\alpha, \bar{\lambda}\bar{z}^{\alpha'}) = \lambda\bar{\lambda}\Sigma(z^\alpha, \bar{z}^{\alpha'}). \tag{2.2}$$

Euler's theorem gives

$$z^\alpha\Sigma_{,\alpha} = \Sigma = \bar{z}^{\alpha'}\Sigma_{,\alpha'} \tag{2.3}$$

and, by taking one more derivative [cf. (1.1)],

$$z^\alpha G_{\alpha\beta'} = \Sigma_{,\beta'}, \quad \bar{z}^{\alpha'}G_{\beta\alpha'} = \Sigma_{,\beta}. \tag{2.4}$$

We introduce vectors Z^α and $\bar{Z}_\alpha$, whose components are z^α and $\bar{z}^{\beta'}G_{\alpha\beta'}$, respectively, whence by (2.4) we have

$$Z^\alpha = \Box^\alpha\Sigma, \quad \bar{Z}_\alpha = \Box_\alpha\Sigma. \tag{2.5}$$

Equation (2.3) gives

$$\Sigma = Z^\alpha\bar{Z}_\alpha. \tag{2.6}$$

A tensor $T^{\rho\cdots\tau}_{\alpha\cdots\gamma}$ will be called *homogeneous of degree* (p, q) if its components satisfy

$$T^{\rho\cdots\tau}_{\alpha\cdots\gamma}(\lambda z^\theta, \bar{\lambda}\bar{z}^{\theta'}) = \lambda^p\bar{\lambda}^q T^{\rho\cdots\tau}_{\alpha\cdots\gamma}(z^\theta, \bar{z}^{\theta'}). \tag{2.7}$$

By referring to (1.16) and using the facts

$$\begin{aligned} &z^\phi\Gamma^\theta_{\psi\phi} = (G^{\theta\chi'}\Sigma_{,\psi\chi'})_{,\phi}z^\phi = 0, \\ &G^{\phi\psi'}{}_{,\chi}z^\chi = 0 \end{aligned} \tag{2.8}$$

(which follow, again, from Euler's theorem) we have the "covariant" version of Euler's theorem

$$\begin{aligned} Z^\mu\Box_\mu T^{\rho\cdots\tau}_{\alpha\cdots\gamma} &= pT^{\rho\cdots\tau}_{\alpha\cdots\gamma}, \\ \bar{Z}_\mu\Box^\mu T^{\rho\cdots\tau}_{\alpha\cdots\gamma} &= qT^{\rho\cdots\tau}_{\alpha\cdots\gamma}, \end{aligned} \tag{2.9}$$

expressing the homogeneity (2.7). Since, by (2.5), Z^α and $\bar{Z}_\alpha$ are defined explicitly from Σ, $\Box_\alpha$, and $\Box^\alpha$, we see that (2.9) expresses the homogeneity (2.7) in a coordinate independent fashion. One readily obtains the homogeneity degrees

$$\begin{aligned} &\Sigma\colon (1,1), \quad Z^\alpha\colon (1,0), \quad \bar{Z}_\alpha\colon (0,1), \\ &\Box_\alpha\colon (-1,0), \quad \Box^\alpha\colon (0,-1), \quad K^{\alpha\beta}_{\rho\sigma}\colon (-1,-1), \end{aligned} \tag{2.10}$$

where the homogeneity of the operators $\Box_\alpha$, $\Box^\alpha$ is to be interpreted as the fact that when acting on a tensor of homogeneity degrees (p, q), $\Box_\alpha$ produces one of degrees $(p-1, q)$ and $\Box^\alpha$ produces one of degrees $(p, q-1)$.

Since $Z^\alpha{}_{;\beta} = z^\alpha{}_{,\beta} = \delta^\alpha_\beta$ [by (1.16) and (2.8)] and $Z^\alpha{}_{;\beta'} = z^\alpha{}_{,\beta'} = 0$ (together with their conjugates), we have

$$\Box_\beta Z^\alpha = \delta^\alpha_\beta = \Box^\alpha\bar{Z}_\beta \tag{2.11}$$

and

$$\Box^\alpha Z^\beta = 0, \quad \Box_\alpha\bar{Z}_\beta = 0. \tag{2.12}$$

Equations (2.11), (2.12) tell us that $(Z^\alpha, \bar{Z}_\alpha)$ behaves like a "position vector" for the space K. Indeed, writing $z^\Theta = (z^\theta, \bar{z}^{\theta'})$, we can re-express (2.11), (2.12) as

$$z^\Theta{}_{;\Phi} = \delta^\Theta_\Phi, \tag{2.13}$$

the semicolon now denoting the full operation (1.14).

This analogy with the concept of position vector is brought out more strongly if we adopt a somewhat different attitude to the space K which is useful in some contexts. According to this alternative view we allow the coordinates z^α and $\bar{z}^{\alpha'}$ to vary independently of one another—so instead of using the notation $\bar{z}^{\alpha'}$ for the second set of coordinates, we introduce new independent quantities[10] $w^{\alpha'}$, where $w^{\alpha'}$ ranges over some suitably small neighborhood of $\bar{z}^{\alpha'}$, for each z^α. The manifold now becomes $2n$-complex-dimensional (i.e., $4n$-real-dimensional) with local holomorphic coordinates $z^\alpha, w^{\alpha'}$. We denote this new space by CK; it is a complexification ("second complexification"!) of K considered as a *real* $2n$-dimensional manifold. The "real slice" K of CK is given when $w^{\alpha'} = \bar{z}^{\alpha'}$. We suppose that Σ is real-analytic on K, i.e., analytic in the real and imaginary parts of z^α so that its definition extends *holomorphically* (i.e., complex-analytically) to CK, which is assumed to be a sufficiently small complex extension of K that this is uniquely possible. Thus,

$$\Sigma = \Sigma(z^\alpha, w^{\alpha'}) \tag{2.14}$$

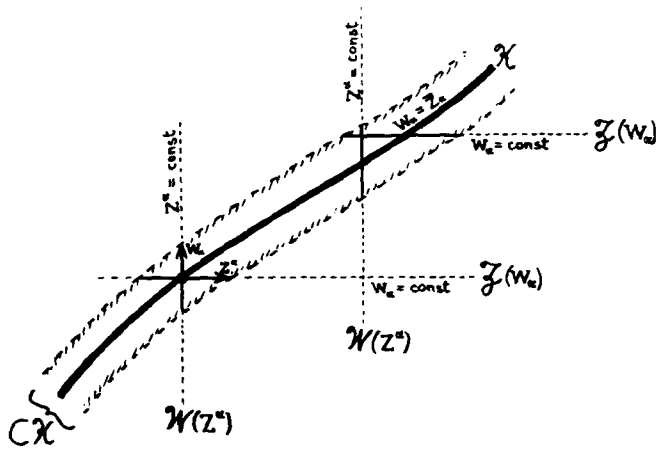


FIG. 1.

is a holomorphic function (for each coordinate chart) of $2n$ complex variables.

The Kähler metric and the covariant derivative operators $\Box_\alpha$, $\Box^\alpha$ also extend uniquely to $C\mathcal{K}$ with the same formal properties as before, but now we write W_α instead of $\bar{Z}_\alpha$:

$$Z^\alpha = \Box^\alpha \Sigma, \quad W_\alpha = \Box_\alpha \Sigma, \quad Z^\alpha W_\alpha = \Sigma, \tag{2.15}$$

$$\Box_\alpha Z^\beta = \delta^\beta_\alpha = \Box^\beta W_\alpha, \quad \Box^\alpha Z^\beta = 0, \quad \Box_\alpha W_\beta = 0. \tag{2.16}$$

We may think of

$$Z^\alpha \oplus W_\alpha \tag{2.17}$$

as denoting the "position vector" of a point of $C\mathcal{K}$. As we hold W_α fixed and vary Z^α, we obtain a space which has a well-defined structure as a vector-space $\mathcal{Z}(W_\alpha)$. The existence of this vector space structure follows because the operators $\Box_\alpha$ *commute* with each other (1.24) and satisfy the first relation (2.16). The vector Z^α is now legitimately a position vector (in the normal sense) for the space $\mathcal{Z}(W_\alpha)$ for each given W_α. Similarly, as we hold Z^α fixed and vary W_α, we obtain a vector space $\mathcal{W}(Z^\alpha)$ for which W_α is a normal position vector and $\Box^\alpha$ the corresponding gradient operator (see Fig. 1). The space $C\mathcal{K}$ itself does not, however, have a vector space structure, in general. Although (2.17) behaves formally like a position vector, with respect to the operator

$$\Box_\alpha \oplus \Box^\alpha, \tag{2.18}$$

[cf. (2.13)], the operators (2.18) do not commute with one another, the commutation giving rise to the curvature $K^{\alpha\beta}_{\rho\sigma}$. Thus, $K^{\alpha\beta}_{\rho\sigma}$ expresses the way in which the flat structure of each $\mathcal{Z}(W_\alpha)$ gets "twisted around" in relation to its neighbors, i.e., as W_α varies, and the way in which the flat structure of each $\mathcal{W}(Z^\alpha)$ gets "twisted around" as Z^α varies.

As a final remark in this section, we note

$$(\Box_\rho \Box^\alpha - \Box^\alpha \Box_\rho) Z^\beta = 0 - \Box^\alpha \delta^\beta_\rho = 0$$

so that

$$K^{\alpha\beta}_{\rho\sigma} Z^\sigma = 0$$

and similarly

$$K^{\alpha\beta}_{\rho\sigma} W_\beta = 0,$$

which takes the form

$$K^{\alpha\beta}_{\rho\sigma} \bar{Z}_\beta = 0$$

on K.

3. ASYMPTOTIC TWISTOR SPACE

We now specialize the preceding theory to the case of the space $\mathcal{T}$ of asymptotic twistors. This space has been introduced and briefly discussed in Refs. 1 and 2. Here we give an independent derivation of the properties of $\mathcal{T}$.

Let $\mathcal{M}$ be an asymptotically flat space–time, with future[11] null conformal infinity $\mathcal{I}^+$. Suppose that $\mathcal{M}$ and its metric g_{ab} are such that a conformally related metric $\hat{g}_{ab} = \Omega^2 g_{ab}$ exists for $\bar{\mathcal{M}}$ $(= \mathcal{M} \cup \mathcal{I}^+)$ which is analytic in a neighborhood of every point of $\mathcal{I}^+$, the conformal factor Ω being also analytic. Let $C\mathcal{M}$ and $C\mathcal{I}^+$ denote complexifications of $\bar{\mathcal{M}}$ and $\mathcal{I}^+$, respectively. (At the present state of understanding it is unfortunately necessary to assume analyticity for the space–time $\mathcal{M}$. It is to be hoped that future developments will allow this restriction to be removed. In any case, a nonanalytic space–time can be approximated arbitrarily closely by analytic ones. It is not necessary that the complexification $C\bar{\mathcal{M}}$ extend very "far" from the real section $\bar{\mathcal{M}}$; the radius of convergence of the Taylor series at each point P of $\bar{\mathcal{M}}$ need only be some positive quantity $\epsilon(P)$, and then $\epsilon(P)$ will also define how far into the complex $C\bar{\mathcal{M}}$ extends at P.)

Certain cross-sections of $C\mathcal{I}^+$ have been referred to[12,13] as "good cuts". These are the intersections of $C\mathcal{I}^+$ with null hypersurfaces $\mathcal{N}$ in $C\mathcal{M}$ whose asymptotic *shear*—taking the appropriate measure of shear—vanishes. There are, in fact, two measures of asymptotic shear for $\mathcal{N}$, denoted σ^0 and $\tilde{\sigma}^0$. For a real null hypersurface $\mathcal{N}$, we should have had $\tilde{\sigma}^0 = \bar{\sigma}^0$, but for a complex $\mathcal{N}$, σ^0 and $\tilde{\sigma}^0$ are independent quantities. We have the choice to make either $\sigma^0 = 0$ or $\tilde{\sigma}^0 = 0$ (but not both, if outgoing gravitational radiation is present). The choice $\sigma^0 = 0$ has been normally made in earlier work. Here we select $\tilde{\sigma}^0$, instead, since this fits in rather better with the notation for twistor space.

Let u, ζ, and $\tilde{\zeta}$ be complex *Bondi-type coordinates* (cf. Refs. 12, 13) for $C\mathcal{I}^+$, the points of $\mathcal{I}^+$ being given by $\tilde{\zeta} = \bar{\zeta}$, u real ($\zeta = \infty$ being allowed). A "good" cross section $\mathcal{X}$ of $C\mathcal{I}^+$ is defined by a "cut function" $X = X(\zeta, \tilde{\zeta})$, satisfying the equation

$$\tilde{\delta}^2 X = \tilde{\sigma}^0(X, \zeta, \tilde{\zeta}), \tag{3.1}$$

where $\mathcal{X}$ is defined by

$$u = X(\zeta, \tilde{\zeta}). \tag{3.2}$$

The quantity ζ, in Eq. (3.1), appears only as a parameter; for each fixed value ζ_0 of ζ, (3.1) defines a two-complex-parameter family of curves in $(u, \tilde{\zeta})$-space which are, in fact, null geodesics on $C\mathcal{I}^+$. As ζ_0 varies, we therefore get a three-complex-parameter family. Each such null geodesic γ defines an *asymptotic twistor* Z^α, but only *up to proportionality*.

For the *twistor* Z^α proper (as opposed to the projective twistor) we require also a "scaling" for γ. This scaling will be necessary in order to assign a meaning to $\Sigma = Z^\alpha \bar{Z}_\alpha$. We have one more complex parameter for

the scaling, so the space $\mathcal{T}$ of asymptotic twistors Z^α will be four-complex-dimensional. This scaling is actually achieved by assigning a spinor $\pi_{A'}$ at each point of γ, which is parallelly propagated along γ, and for which $\iota^A\pi^{A'}$ is a tangent vector to γ. The tangent vectors to the generators of $C\mathcal{I}^+$ (corresponding to the parameter u) are n^a where $n^a = \iota^A\tilde{\iota}^{A'}$. For any standard choice of ι^A and corresponding $\tilde{\iota}^{A'}$ (where we normally take ι^A and $\tilde{\iota}^{A'}$ parallelly propagated along generators of $C^+\mathcal{I}$, with $\tilde{\iota}^{A'} = \bar{\iota}^{A'}$ on $\mathcal{I}^+$), the complex number

$$\pi_{1'} = \pi_{A'}\tilde{\iota}^{A'} \tag{3.3}$$

can be used to measure the required scaling for Z^α.

To obtain the definition of Σ, we need to locate $\bar{\gamma}$, the complex conjugate geodesic to γ on $C\mathcal{I}^+$. Now any locus on $C\mathcal{I}^+$ will have a uniquely defined complex conjugate locus, namely that which is obtained from the original locus when the point transformation of complex conjugation is applied to $C\mathcal{I}^+$. If real coordinates had been chosen for $\mathcal{I}^+$, this would simply be the transformation sending each point of $C\mathcal{I}^+$ into the one with the complex conjugated coordinates. But with the present choice, the complex conjugate of the point labelled $(u, \zeta, \tilde{\zeta})$ is labelled $(\bar{u}, \bar{\tilde{\zeta}}, \bar{\zeta})$. Thus, $\bar{\gamma}$ lies on the (u, ζ)-plane defined by $\tilde{\zeta} = \zeta_0$. Furthermore, assuming that γ is such that u attains a well-defined value u_0 on it when $\tilde{\zeta} = \bar{\zeta}_0$, then γ and $\bar{\gamma}$ will each meet a (unique) common generator μ of $C\mathcal{I}^+$, namely that defined by $\zeta = \zeta_0$, $\tilde{\zeta} = \bar{\zeta}_0$. The two points of intersection of γ and $\bar{\gamma}$ with μ will then have respective u values u_0 and $\bar{u}_0$. [We note also that $\bar{\gamma}$ is a member of the two complex-parameter family of curves in the (u, ζ)-plane which are defined by solutions $\hat{X}(\tilde{\zeta}, \zeta)$, at $\tilde{\zeta} = \bar{\zeta}_0$, of $\eth^2\hat{X}(\tilde{\zeta}, \zeta) = \sigma^0(\hat{X}, \tilde{\zeta}, \zeta)$.] We now define[1]

$$\Sigma = i(u_0 - \bar{u}_0)\, n^{AA'}\bar{\pi}_A\pi_{A'}, \tag{3.4}$$

which, by (3.3) and $n^a = \iota^A\bar{\iota}^{A'}$, can be written

$$\Sigma = i(u_0 - \bar{u}_0)\, \bar{\pi}_1\pi_{1'}. \tag{3.5}$$

We shall require a holomorphic parametrization for the system of Z^α's, so that we can take derivatives of Σ with respect to these parameters. The different solutions of (3.1) for fixed $\zeta = \zeta_0$ can be parametrized by pairs of complex numbers μ^0, μ^1 [constant of integration for the second-order ordinary differential equation (3.1)] and we assume that as ζ_0 varies, the solutions (i.e., the curves γ) vary holomorphically as μ^0, μ^1 vary holomorphically. Thus, we have a single holomorphic function

$$X = X(\tilde{\zeta}; \mu^0, \mu^1, \zeta_0) \tag{3.6}$$

of four variables satisfying (3.1), the different solutions of (3.1) being functions of $\tilde{\zeta}$ parametrized by μ^0, μ^1, and ζ_0. Thus μ^0, μ^1, and ζ_0 are holomorphic coordinates for *projective* asymptotic twistor space $P\mathcal{T}$. To assign coordinates to the nonprojective space $\mathcal{T}$, we need to make sure that the scaling for Z^α is also appropriately parametrized. The quantity $\pi_{1'}$ in (3.5) cannot be used as it stands, since it is actually not a holomorphic coordinate, being dependent on the value of $\bar{\iota}^{A'}$ at the interaction of γ with μ (which depends on $\bar{\gamma}$). However, the spinor $\pi_{A'}$ itself depends holomorphically on Z^α. What is needed is to choose a *component* of $\pi_{A'}$ in a holomorphic way (i.e., in a way "independent" of $\bar{Z}_\alpha$).

To achieve this, we may conformally rescale the metric on $C\mathcal{I}^+$ so that it becomes flat. The original metric on $C\mathcal{I}^+$ is that induced by $\hat{g}_{ab}$ and may be taken to be

$$d\hat{s} = -\frac{4d\zeta\, d\tilde{\zeta}}{(1+\zeta\tilde{\zeta})^2} + 0\cdot du^2 \tag{3.7}$$

(cf. Refs. 8, 12). We rescale this to flatness by introducing $d\hat{\hat{s}}^2 = (1+\zeta\tilde{\zeta})^2\, d\hat{s}^2$, which is induced on $C\mathcal{I}^+$ by the 4-metric

$$\hat{\hat{g}}_{ab} = (1+\zeta\tilde{\zeta})^2\,\hat{g}_{ab}. \tag{3.8}$$

The replacement of the vector n^a by

$$k^a = (1+\zeta\tilde{\zeta})^{-1}n^a = (1+\zeta\tilde{\zeta})^{-1}\iota^A\tilde{\iota}^{A'} = \kappa^A\tilde{\kappa}^{A'} \tag{3.9}$$

naturally accompanies this rescaling (cf. Ref. 12), and k^a becomes covariantly constant on $C\mathcal{I}^+$ with respect to $\hat{\hat{g}}_{ab}$. We can arrange that κ^A and $\tilde{\kappa}^{A'}$ are also covariantly constant on $C\mathcal{I}^+$ with respect to $\hat{\hat{g}}_{ab}$ (cf. Ref. 14). Now $\pi_{A'}$ is constant along γ with respect to $\hat{\hat{g}}_{ab}$ as well as with respect to $\hat{g}_{ab}$ (see Ref. 1, 2), so the scalar product

$$r = \pi_{A'}\tilde{\kappa}^{A'} \tag{3.10}$$

is constant along γ and does not now depend on the intersection of γ with $\tilde{\zeta} =$ constant plane. In fact, r is a holomorphic coordinate defining the scaling of Z^α (r being now "independent" of $\bar{\gamma}$ or of $\bar{Z}_\alpha$); substituting (3.10) and (3.9) into (3.5), we get (dropping the subscripts on u_0, $\bar{u}_0$, ζ_0 and $\bar{\zeta}_0$)

$$\Sigma = \Sigma(Z^\alpha, \bar{Z}_\alpha) = ir\bar{r}(1+\zeta\tilde{\zeta})(u - \bar{u}) \tag{3.11}$$

for the Kähler scalar on $\mathcal{T}$, or

$$\Sigma = \Sigma(Z^\alpha, W_\alpha) = ir\tilde{r}(1+\zeta\tilde{\zeta})(u - \tilde{u}) \tag{3.12}$$

for the complexified version $C\mathcal{T}$ of this (with $W_\alpha = \tilde{Z}_\alpha$), where

$$u = X(\tilde{\zeta}; \mu^0, \mu^1, \zeta), \tag{3.13a}$$

$$\tilde{u} = \tilde{X}(\zeta; \tilde{\mu}^{0'}, \tilde{\mu}^{1'}, \tilde{\zeta}). \tag{3.13b}$$

In the case of Minkowski space M we can choose

$$X(\tilde{\zeta}; \mu^0, \mu^1, \zeta) = (1+\zeta\tilde{\zeta})^{-1}(\mu^0 + \mu^1\tilde{\zeta}), \tag{3.14a}$$

$$\tilde{X}(\zeta; \tilde{\mu}^{0'}, \tilde{\mu}^{1'}, \tilde{\zeta}) = (1+\tilde{\zeta}\zeta)^{-1}(\tilde{\mu}^{0'} + \tilde{\mu}^{1'}\zeta) \tag{3.14b}$$

so that

$$\Sigma(Z^\alpha, W_\alpha) = ir\tilde{r}(\mu^0 + \mu^1\tilde{\zeta} - \tilde{\mu}^{0'} - \tilde{\mu}^{1'}\zeta). \tag{3.15}$$

Then, introducing new coordinates

$$z^0 = ir\mu^0, \quad z^1 = -ir\mu^1, \quad z^2 = r, \quad z^3 = -r\zeta \tag{3.16a}$$

$$w^{0'} = -i\tilde{r}\tilde{\mu}^{0'}, \quad w^{1'} = i\tilde{r}\tilde{\mu}^{1'}, \quad w^{2'} = \tilde{r}, \quad w^{3'} = -\tilde{r}\tilde{\zeta}, \tag{3.16b}$$

we obtain the standard flat-space twistor form[1] of the scalar product

$$\Sigma(Z^\alpha, W_\alpha) = z^0w^{2'} + z^1w^{3'} + z^2w^{0'} + z^3w^{1'}. \tag{3.17}$$

The coordinates z^α, as given by (3.16a), may also be used in the general case for a curved twistor space $\mathcal{T}$, in place of r, μ^0, μ^1, ζ. Notice that μ^0, μ^1, and ζ are defined by the ratios of z^0, z^1, z^2, z^3 while r scales all four proportionally. A similar remark holds for $w^{0'}$, $w^{1'}$, $w^{2'}$, $w^{3'}$ in relation to the tilde variables.

Hence, Σ is indeed homogeneous of degrees (1,1) in the two sets of variables as is required for (2.2). The results of Sec. 2 may therefore be applied.

The signature of (3.17), given when $w^{\alpha'} = \bar{z}^{\alpha'}$ is $++--$. So the signature of $G_{\alpha\beta'}$ is $++--$ in the case of flat twistor space. The signature remains constant under continuous deformations for which the Kähler metric does not degenerate. The signature of the Kähler metric of $\mathcal{T}$ must therefore also be $++--$.

It is possible to obtain the Kähler curvature of $\mathcal{T}$ in a completely coordinate-free manner, but the details of this argument are hard to express. Consequently we summarize here only the direct coordinate calculation of the components of the curvature tensor. It is convenient to do the calculation in a new holomorphic coordinate system defined on $\mathcal{T}$ as follows:

$$x^0 = \mu^0, \quad x^1 = \mu^1, \quad x^2 = r, \quad x^3 = \zeta. \tag{3.18}$$

(Note that Σ is not homogeneous in these coordinates.) From (3.11) and (1.5), we obtain the components of the metric tensor

$$G_{\Gamma\Lambda} = \begin{array}{c|cccc|cccc} & x^0 & x^1 & x^2 & x^3 & \bar{x}^{0'} & \bar{x}^{1'} & \bar{x}^{2'} & \bar{x}^{3'} \\ \hline x^0 & & & & & 0 & & B & \\ x^1 & & & 0 & & & & & \\ x^2 & & & & & & \bar{B}^T & & A \\ x^3 & & & & & & & & \\ \hline \bar{x}^{0'} & & & & & & & & \\ \bar{x}^{1'} & & & 0 & & \bar{B} & & & \\ \bar{x}^{2'} & & & & & & & 0 & \\ \bar{x}^{3'} & & & B^T & & & A^T & & \end{array} \tag{3.19}$$

with

$$A = \begin{pmatrix} i(1+\zeta\bar{\zeta})(X-\bar{X}) & i\bar{r}[(1+\zeta\bar{\zeta})(X-\bar{X})]_{,\bar{\zeta}} \\ ir[(1+\zeta\bar{\zeta})(X-\bar{X})]_{,\zeta} & ir\bar{r}[(1+\zeta\bar{\zeta})(X-\bar{X})]_{,\zeta\bar{\zeta}} \end{pmatrix}, \tag{3.20}$$

$$B = \begin{pmatrix} ir(1+\zeta\bar{\zeta})X_{,\mu^0} & ir\bar{r}[(1+\zeta\bar{\zeta})X_{,\mu^0}]_{,\bar{\zeta}} \\ ir(1+\zeta\bar{\zeta})X_{,\mu^1} & ir\bar{r}[(1+\zeta\bar{\zeta})X_{,\mu^1}]_{,\bar{\zeta}} \end{pmatrix},$$

where ^{T} denotes matrix transposition, $^-$ denotes complex conjugation, and commas denote partial derivatives as before.

Using the facts that $\tilde{\delta}^2 X(\bar{\zeta}; \mu^0, \mu^1, \zeta) = \tilde{\sigma}^0(X, \zeta, \bar{\zeta})$ and $\delta^2 \bar{X}(\zeta; \bar{\mu}^{0'}, \bar{\mu}^{1'}, \bar{\zeta}) = \sigma^0(\bar{X}, \bar{\zeta}, \zeta)$, the components of the curvature tensor given by (1.20) can be written in the form

$$K_{A3'B3'} = \frac{-ir\ddot{\bar{\sigma}}^0}{1+\zeta\bar{\zeta}} \frac{\partial X}{\partial x^A} \frac{\partial X}{\partial x^B},$$

$$K_{A3'33'} = \frac{-ir\bar{r}}{(1+\zeta\bar{\zeta})^2} (\ddot{\tilde{\sigma}}^0 \delta X + \delta\dot{\tilde{\sigma}}^0) \frac{\partial X}{\partial x^A},$$

$$K_{33'33'} = \frac{-ir\bar{r}}{(1+\zeta\bar{\zeta})^3} \times \{\ddot{\tilde{\sigma}}^0(\delta X)^2 + 2\delta X \delta\dot{\tilde{\sigma}}^0 + \delta^2\tilde{\sigma}^0 + \sigma^0\ddot{\tilde{\sigma}}^0$$

$$- [\ddot{\sigma}^0(\bar{\delta}\bar{X})^2 + 2\bar{\delta}\bar{X}\bar{\delta}\dot{\sigma}^0 + \bar{\delta}^2\sigma^0 + \dot{\sigma}^0\tilde{\sigma}^0]$$

$$+ \dot{\sigma}^0\dot{\tilde{\sigma}}^0(X-\bar{X})\}, \tag{3.21}$$

where A and B take on values of 0 and 1. The $\tilde{\sigma}^0$ quantities are evaluated at the intersection of γ with μ and the σ^0 quantities at the intersection of $\bar{\gamma}$ with μ, i.e., $\tilde{\sigma}^0(u, \zeta, \bar{\zeta})$ and $\sigma^0(\bar{u}, \bar{\zeta}, \zeta)$. The dots denote $\partial/\partial u$ and the δ ($\bar{\delta}$) operator is applied to $\tilde{\sigma}^0$, $\dot{\tilde{\sigma}}^0$, and $\delta\tilde{\sigma}^0$ (σ^0, $\dot{\sigma}^0$, and $\bar{\delta}\sigma^0$) keeping u and $\bar{\zeta}$ ($\bar{u}$ and ζ) fixed. All other nonzero components of the curvature tensors can be obtained by using (1.10) and (1.11).[15] (3.21) shows that the components of the curvature tensor are combinations of quantities K, K_1, K_2, which have the striking form

$$K = \bar{\delta}^2\sigma^0 + \dot{\sigma}^0\tilde{\sigma}^0 - \delta^2\tilde{\sigma}^0 - \dot{\tilde{\sigma}}^0\sigma^0 - \dot{\sigma}^0\dot{\tilde{\sigma}}^0(X-\bar{X}),$$
$$K_1 = \delta\dot{\tilde{\sigma}}^0, \quad K_2 = -\ddot{\tilde{\sigma}}^0, \tag{3.22}$$

when $\Sigma = 0$, γ and $\bar{\gamma}$ intersect μ at the same point $(u = \bar{u})$, and K reduces to

$$K_0 = \bar{\delta}^2\sigma^0 + \dot{\sigma}^0\tilde{\sigma}^0 - \delta^2\tilde{\sigma}^0 - \dot{\tilde{\sigma}}^0\sigma^0. \tag{3.23}$$

The quantities K_2, K_1, K_0 are the respective R^{-1}, R^{-2}, and portion of the R^{-3} part of the gravitational (Weyl tensor) field of $\mathcal{M}$ in a standard notation (cf. Refs. 16, 17), i.e., ψ_4^0, ψ_3^0 and $\psi_2^0 - \bar{\psi}_2^0$, R being an affine parameter on outgoing null geodesics.

The components of the Kähler "Ricci tensor" can be obtained either from contraction of (3.21) or from the coordinate expression

$$K_{\mu'\nu} = K_{\mu'\lambda\nu}{}^{\lambda} = \frac{1}{2} \frac{\partial^2 \ln|G|}{\partial x^{\mu'} \partial x^{\nu}}, \tag{3.24}$$

where $|G|$ is the determinant of the matrix $G_{\Gamma\Lambda}$. The Ricci tensor turns out to be identically zero.

As a final remark, we note that if the Kähler scalar Σ is replaced by $\log\Sigma$, then we obtain a Kähler metric defined on the *three-dimensional projective* asymptotic twistor space. This metric becomes singular, however, on $\Sigma = 0$. It has signature $(+\,+\,-)$ on the region $\Sigma > 0$ and $(+\,-\,-)$ on $\Sigma < 0$. In this form, the Kähler metric bears strong resemblance to the Bergmann metric[18] for bounded domains in C^n. This may have relevance particularly when $\mathcal{T}$ is the twistor space defining a nonlinear graviton,[19] since then the full extent of $\mathcal{T}$ is the region $\Sigma > 0$, with boundary $\Sigma = 0$. It seems likely that the metric studied here should also have significance for the study of such "gravitons."

*This research was supported in part by a grant from the National Science Foundation.

[1]R. Penrose and M. A. H. MacCallum, Phys. Repts. C 6, No. 4 (1973).
[2]R. Penrose, "Twistor theory, its aims and achievements," in *Quantum Gravity*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Oxford U. P., Oxford, 1975).
[3]S. Kobayashi and K. Nomizu, *Foundation of Differential Geometry* (Interscience, New York, 1963).
[4]J. A. Schouten, *Ricci-Calculus* (Springer-Verlag, Berlin, 1954).
[5]K. Yano, *Differential Geometry on Complex and Almost Complex Spaces* (MacMillan, New York, 1965).
[6]Thus K is $2n$-dimensional as a real manifold with a preferred set of complex coordinate charts related to one an-

other by holomorphic (i.e., complex analytic) coordinate transformations.

[7]Boldface types are used to denote ordinary numerical tensor component indices in order to distinguish them from the corresponding abstract indices.[8] Thus, $\alpha, \beta, \cdots, a, b, \cdots, \Gamma, \Theta, \cdots$ are the abstract versions of corresponding component indices $\boldsymbol{\alpha}, \boldsymbol{\beta}, \cdots, \mathbf{a}, \mathbf{b}, \cdots, \boldsymbol{\Gamma}, \boldsymbol{\Theta}, \cdots$.

[8]R. Penrose, "Structure of Space-Time," in *Battelle Rencontres*, edited by C.M. DeWitt and J.A. Wheeler (Benjamin, New York, 1968).

[9]There is complete symmetry (via the operation of complex conjugation) between the operators $\Box_\alpha$ and $\Box^\alpha$ even though the coordinate expressions (1.16) look very different from one another. The asymmetry in the coordinate expressions arises merely from the fact that the unprimed indices have been notationally singled out.

[10]For many purposes it is convenient to adopt a standard procedure of replacing "bars" over quantities by "tildes," so while a quantity $\bar{\xi}$ would initially have been the complex conjugate of ξ, upon complexification we would introduce a new quantity $\tilde{\xi}$ which is independent of ξ (cf. Sec. 3). With this notation w would be written $\tilde{z}$.

[11]The construction to be described will, of course, equally well refer to $\mathscr{I}^-$ as to $\mathscr{I}^+$. But for retarded gravitational fields the structure of $\mathscr{I}^-$ is identical to that for Minkowski space.

[12]R. Penrose, "Relativistic symmetry groups," in *Group Theory in Non-Linear Problems*, edited by A.O. Barut (Reidel, Dordrecht, 1972).

[13]R.O. Hansen and E.T. Newman (to be published).

[14]R. Penrose, Proc. Roy. Soc. London A 284, 159 (1965).

[15]The curvature tensor of $C\overline{I}$ can be calculated in the same way starting from (3.12). The results are similar with all barred quantities replaced by tilded quantities.

[16]E.T. Newman and R. Penrose, J. Math. Phys. 3, 566 (1962).

[17]E.T. Newman and T. Unti, J. Math. Phys. 3, 891 (1962).

[18]S. Bergman, Math. Surveys, No. 5 (Amer. Math. Soc., Providence, R.I., 1968), 2nd ed.

[19]R. Penrose, General Relativity Gravitation 7, 31 (1976).

On the existence of localized solutions in nonlinear chiral theories*

Jarmo Hietarinta

Department of Physics, The Ohio State University, Columbus, Ohio 43210
(Received 17 June 1976)

We study the existence of localized solutions to theories based on Weinberg's nonlinear realization of chiral SU(2)⊗SU(2). The analysis is done by using specific variations of the action integral and then checking the ensuing global conditions. The following cases are studied: (i) π fields only and without time dependence, (ii) π fields with simple time dependence, (iii) π fields coupled to gauge fields, (iv) the above with certain chiral symmetry breaking potentials. We find that only in certain special cases could there be localized solutions. In most cases the intrinsic nonlinearity of the system does not seem to be enough to guarantee their existence.

I. INTRODUCTION

Recently there has been renewed interest in attempts to incorporate localized classical solutions into quantum field theoretical calculations.[1] This is hoped to give in some respects better results than the usual method of quantization, because the essential nonlinearities of the system are then taken into account nonperturbatively.

Generally speaking, these nonlinearities that support localized classical solutions can be divided into the following two types: (a) The nonlinearity is added by hand (e.g., by adding a spontaneously symmetry breaking potential) and (b) there is intrinsic nonlinearity in the system itself due to some symmetry principle (e.g., in Yang–Mills[2] theory). We would, of course, prefer a theory which has natural nonlinearities as in case (b). A nonlinear theory could also be thought of as following from linear theory with a constraint, if after solving for the constrained fields we can express the linear Lagrangian of all fields as a nonlinear Lagrangian of the unconstrained fields.

Among theories that have intrinsic nonlinearity we can include those based on nonlinear realization of chiral SU(2)⊗ SU(2).[3] There has already been some interest in classical localized solutions of these systems[4–6] and of the associated linear system[7] where the constraint is $\sum_\alpha \phi_\alpha^2 = 1$. The purpose of this paper is to study further the existence of localized solutions to the nonlinear system.

By a classical localized *solution* we mean a configuration which extremizes the relevant classical action integral. From the semiclassical point of view one expects that quantization is then taken care of by small fluctuations about that extremum configuration. To find this classical solution, the standard methods in calculus of variations[8] lead to the Euler–Lagrange equations. In this paper we in general do not analyze the Euler–Lagrange equations but instead work directly with the action integral and get from it some qualitative results.

The functionals that we are considering here are of the type

$$E[\phi] = \int_\Omega d^n x \mathcal{L}(\phi^a, \partial_i \phi^a). \tag{1.1}$$

To start with, we assume $\phi^a \in \mathcal{H}$, where $\mathcal{H}$ is the space of bounded continuous functions on Ω, with continuous mth derivatives. To fully set up the variational problem, we have to define in addition (i) the set of admissible functions M. There are various types of constraints that could be relevant. For example, usually we assume the values of ϕ^a and $\partial_i \phi^a$ to be fixed on the boundary of Ω. Often some function(al) will be assumed to have a given value; recently many problems have led to situations where the constraint has been given by specifying a topological charge[9] for ϕ.

To give meaning to *local* extremum, we also have to define (ii) nearness for functions. To do that, one usually starts with $\mathcal{H}$ and defines a norm $\|\cdot\|$ there and then restricts to M. Then one says that $\phi_c \in M$ gives, e.g., a local minimum if there is a positive real number δ such that

$$E[\phi_c] < E[\phi], \text{ whenever } \phi \in M \text{ and } \|\phi_c - \phi\| < \delta. \tag{1.2}$$

The following norms often appear in mathematical literature[8]:

$$\|\phi\|_0 = \max_{x\in\Omega} |\phi(x)|, \tag{1.3a}$$

$$\|\phi\|_1 = \max_{x\in\Omega} |\phi(x)| + \max_{x\in\Omega} \left\{ \sum_i [\partial_i \phi(x)]^2 \right\}^{1/2}. \tag{1.3b}$$

Obviously the set of functions near ϕ_c of (1.2) is smaller if we use (1.3b). It could be expected that the existence of solutions to (1.2) depends on the definition of the norm as sometimes nearby functions ψ, which otherwise could give $E[\psi] < E[\phi_c]$, could be eliminated by a stricter choice of the norm.

In the particle interpretation of the solution a global minimum is more favored. In the quantized version it is expected that the particle corresponding to a local minimum would decay to a lower energy configuration through tunneling, provided there are no conserved quantum numbers to forbid this. However, if this tunneling can be considered to take some time, then we still could consider the solution to correspond to a particle of finite lifetime. [The ease of tunneling might suggest a proper definition for the norm.]

Most of our analysis on the existence of classical solutions is done using certain specific variations. For these variations the nearby functions depend on a continuous parameter η: $\phi_\eta = \phi(x;\eta)$ with $\phi(x;1) = \phi_c(x)$. Consequently, E will be a function of η with $E(\eta)_{\eta=1} = E[\phi_c]$. If

$$\|\phi(x;\eta)-\phi_c(x)\|\to 0, \quad \text{as } \eta\to 1, \tag{1.4}$$

then a necessary condition for $\phi_c(x)$ to give an extremum of $E[\phi]$ is

$$\left.\frac{\partial E(\eta)}{\partial\eta}\right|_{\eta=1}=0. \tag{1.5}$$

The most useful variational conditions are based on (i) $\phi_\lambda=\phi_c(\lambda x)$ and (ii) $\phi_\xi=\xi\phi_c(x)$. They have been used extensively before.[10] Of course, one may use these variations only if they do not violate the constraints of the problem, e.g., if $\phi(\infty)=C\neq 0$, then (ii) is not allowed. Both variations do conserve topological charges that depend on the number of zeroes of ϕ. In addition to constraints, we also have to check that the condition (1.4) is satisfied for the particular norm we are using. For (1.3) this is done in Appendix A.

In this paper we consider theories based on Weinberg's nonlinear realization of chiral SU(2)⊗ SU(2).[3,11] [For concise definitions, see Appendix B. Explicit forms in three different but equivalent formulations are given in Appendix C.] In Secs. II and III we consider theories with π fields only, first the time independent case (Sec. II) and then with simple time dependence (Sec. III). In Sec. II we also compare the results for the nonlinear realization with those for the associated linear but constrained theory. In Sec. IV coupling to gauge fields is added. We find that there are no classical finite energy solutions except possibly such that $\lim_{x\to\infty}(\pi^a)^2=C\neq 0$. In Sec. V we add some simple chiral symmetry breaking potential (without spontaneous symmetry breaking), but again our results are in general negative. Chiral fields with fermions are not considered.

II. TIME INDEPENDENT SOLUTIONS

Consider the chiral Lagrangian

$$L=-\int \tfrac{1}{2}(\partial_\mu\pi_a)(\partial^\mu\pi_b)F_{ab}(\pi^2)\,d^4x. \tag{2.1}$$

We are looking for time independent solutions and therefore our object is to extremize

$$E[\pi]=\tfrac{1}{2}\int(\partial_i\pi_a)(\partial_i\pi_b)F_{ab}(\pi^2)d^nx. \tag{2.2}$$

Let us now apply the scale variation $\pi^a_\lambda=\pi^a(\lambda\mathbf{x})$ on (2.2). We first get

$$E[\pi_\lambda]=\lambda^{2-n}E[\pi]. \tag{2.3}$$

Since we are only looking for nontrivial solutions with finite energy, we have $0<E<\infty$. Therefore, we can use (1.5) to get the necessary condition $n=2$ for a solution to exist.

If we assume that $\pi^2\to 0$ as $r\to\infty$, we can also use the variation $\pi_\xi(\mathbf{x})=\xi\pi(\mathbf{x})$. Using this on the form (C8) gives

$$\left.\frac{\partial E}{\partial\xi}\right|_{\xi=1}=\int\{(\partial_i\pi_a)(\partial_i\pi_a)+\tfrac{1}{4}(\partial_i\pi^2)^2\times[2(f_\pi^2-\pi^2)^{-1}+\pi^2(f_\pi^2-\pi^2)^{-2}]\}\,d^nx. \tag{2.4}$$

Since $|\pi^a(x)|\leq f_\pi\ \forall\ \mathbf{x}\in\mathbf{R}^n$ we see that the first order variation does not vanish unless $\pi^a\equiv 0$, so under the assumption $\pi(\infty)=0$ the functional (2.2) has no extremizing localized solutions.

The two-dimensional case that passes the scaling test has also been studied by Duff and Isham,[5] who report of finding a solution. This is in conflict with our arguments of nonexistence in the previous paragraph. We are now going to analyze this further. Duff and Isham use the ansatz

$$\pi^i=f_\pi\hat{x}^i\alpha(\rho),\quad i=1,2,$$
$$\pi^3=0, \tag{2.5}$$

where $\rho=(x_1^2+x_2^2)^{1/2}$ and $\hat{x}^i=x^i/\rho$. Using the form (C8) gives the energy as

$$E[\pi^a]=f_\pi^2\pi\int_0^\infty\rho\,d\rho[\alpha'^2/(1-\alpha^2)+\alpha^2/\rho^2]. \tag{2.6}$$

A change of variables,

$$\alpha(\rho)=\sin\theta(\rho),\quad \theta(\rho)=\tfrac{1}{2}\psi(\ln\rho), \tag{2.7}$$

is then used to give for ψ the familiar sine–Gordon equation

$$\psi''(z)-\sin\psi(z)=0 \tag{2.8}$$

as the Euler–Lagrange equation. The solution to this is

$$\psi(z)=4\arctan e^z, \tag{2.9}$$

so that

$$\alpha(\rho)=\sin(2\arctan\rho)=2\rho/(1+\rho^2). \tag{2.10}$$

The problem with this method is that the change of variables (2.7) is only allowed if its Jacobian $\partial\alpha/\partial\theta$ is nonzero. Now, as ρ goes from 0 to ∞, the argument θ in (2.7) goes from 0 to π, and we find that for $\theta=\pi/2$, $\partial\alpha/\partial\theta=\cos\theta=0$.

The nonexistence of solutions to (2.6) can be also seen by using integral form for the Euler–Lagrange equation, which now is

$$2\rho\alpha'/(1-\alpha^2)=\int_0^\rho\rho'\,d\rho'\,2\alpha[\alpha'^2(1-\alpha^2)^{-2}+\rho'^{-2}]+C. \tag{2.11}$$

For some fixed C the solution (2.10) should satisfy this for all ρ. This is clearly impossible, and instead of (2.11), the following equation is satisfied formally:

$$2\rho\alpha'/(1-\alpha^2)=\int_0^\rho\rho'\,d\rho'\,2\alpha[\alpha'^2(1-\alpha^2)^{-2}+\rho'^{-2}]-C\theta(\rho-1), \tag{2.12}$$

where θ is the step function and C is an infinite constant

$$C=\int_0^\infty 2\alpha\rho'[\alpha'^2(1-\alpha^2)^{-2}+\rho'^{-2}]\,d\rho'. \tag{2.13}$$

Therefore, the Euler–Lagrange equation for α actually has a δ-function singularity.

The clearest way to see that happens is to use the W formalism (C7). Then we want to extremize

$$E[\pi^a]=f_\pi^2\pi\int_0^\infty d\rho\,\rho[\alpha_W'^2+\alpha_W^2\rho^{-2}](1+\alpha_W^2)^{-2}. \tag{2.14}$$

The first integral to the corresponding Euler–Lagrange equation is immediately found to be

$$(-\rho^2\alpha_W'^2+\alpha_W^2)(1+\alpha_W^2)^{-2}=C. \tag{2.15}$$

The constant C is determined by $\alpha_W^2\to 0$ and $\alpha_W'^2\to 0$ as $\rho\to\infty$, and so $C=0$. For this value of C the solutions without corners are

$$\alpha_{1W}=a\rho,\quad \alpha_{2W}=a\rho^{-1}, \tag{2.16}$$

neither of which is bounded. If the solution (2.10) is transformed into W representation using (C1) with (C14), we get

$$\alpha_W = \alpha \Phi_{BW}(\alpha^2)$$
$$= [(1+\rho^2)/2\rho][1 - |(1-\rho^2)/(1+\rho^2)|]$$
$$= \begin{cases} \rho & (\rho \leq 1), \\ 1/\rho & (\rho \geq 1). \end{cases} \tag{2.17}$$

This and all the other continuous solutions to (2.15) have corners. However, such solutions do not extremize (2.14), because they do not satisfy the Weierstrass–Erdmann corner conditions[8] at the corner point ρ_0:

$$\left.\frac{\delta \mathcal{E}}{\delta \tilde{\alpha}'}\right|_{\rho \to \rho_0^+} = \left.\frac{\delta \mathcal{E}}{\delta \tilde{\alpha}'}\right|_{\rho \to \rho_0^-}. \tag{2.18}$$

To avoid the restriction to $n=2$ that comes from the scale variation the following modified Lagrangian has been proposed[6]

$$L = -\int [-\tfrac{1}{2} \partial_\mu \pi^a \partial^\mu \pi^b F_{ab}(\pi)]^p \, d^n x. \tag{2.19}$$

The condition from scale variation is $2p=n$. However, the analysis with variation $\pi \to \xi\pi$ can still be carried through and leads to the same result of nonexistence.

When we derived (2.4), we had to assume that $\pi^2(\infty) = 0$. This is not very restrictive, since from (2.6) we see that it is also required in order to have finite energy for this ansatz. The same applies for (2.19) with $n=3$, $p=\frac{3}{2}$ and if the hedgehog ansatz is used.[6]

So far we have been considering the nonlinear problem, but since there is an associated linear problem, let us see how these results would apply to that case. In the linear theory we would have four fields π_α with a constraint

$$\sum_{\alpha=0}^{3} \pi_\alpha^2 = f_\pi^2. \tag{2.20}$$

We could formally solve for π_0, say, and then express the linear Lagrangian nonlinearly in terms of the π_i's only, as in (C8). The difference between the linear and nonlinear methods comes from the fact that the sign of π_0 cannot be uniquely determined from (2.20). In the nonlinear theory this sign does not appear in the Lagrangian and cannot be recreated by using the π_i's, which are now assumed to give a complete description of the system. In the linear theory π_0 will continue to carry dynamical information in its sign, although its magnitude is fixed.

If we look for an interpretation that guarantees the existence of a localized solution, then it is crucial to require π_0 to change sign at $\rho=1$, where $\sum_{i=1}^3 \pi_i^2 = f_\pi^2$. In that case the variation $\pi_i \to \xi\pi_i$ is forbidden, because for $\xi < 1$, $\sum_i \xi^2 \pi_i^2 < f_\pi^2$ always and then π_0 would not have a zero. This sign change was utilized by Honerkamp *et al.*[7] Their solution to the linear problem has π_i as given in (2.5) and (2.10), but they also give

$$\pi_0 = f_\pi (1-\rho^2)/(1+\rho^2), \tag{2.21}$$

which gives the required sign change.

Another possibility is to take the unbounded solution we gave in (2.16). This has been done in the Appendix of Ref. 6, but since we assumed boundedness for the fields we will not discuss it any further.

In conclusion then we have shown that the nonlinear realization of SU(2)⊗ SU(2) does not have bounded solutions although the linear, constrained, system does. The extra field, although fixed in magnitude, is needed for topological reasons.

III. SIMPLE TIME DEPENDENCE

For convenience, let us define

$$\phi = 2^{-1/2}(\pi_1 + i\pi_2), \quad \phi^* = 2^{-1/2}(\pi_1 - i\pi_2). \tag{3.1}$$

The simple time dependence that we are considering here can then be described by $(x^0 = t)$

$$\phi(\mathbf{x}, t) = \exp(i\omega t)\Psi(\mathbf{x}), \tag{3.2}$$

where Ψ can be complex. We still assume $\pi_3 = 0$. Substituting this into the Lagrangian (C8) gives the following action:

$$A = -\tfrac{1}{2} \int d^n x [(\partial_i \pi^a)^2 - \omega^2 \pi^2 + \tfrac{1}{4}(\partial_i \pi^2)^2 (f_\pi^2 - \pi^2)^{-1}] \tag{3.3}$$

and charge

$$Q = -\omega \int d^n x \, \pi^2. \tag{3.4}$$

Now A should be stationary with respect to all variations (keeping ω constant). What amounts to the same is to require the energy $(= -A - \omega Q)$ to be stationary with respect to variations that keep Q fixed. Lagrange multiplier technique leads then to (3.3).[12]

To study the existence of solutions to this problem, we use the variation

$$\pi^a(\mathbf{x}) \to \pi_\lambda = \lambda^{n/2} \pi^a(\lambda \mathbf{x}). \tag{3.5}$$

With this variation Q is conserved automatically. After a change in the integration variable we get

$$A_\lambda = -\tfrac{1}{2}\lambda^2 \int d^n x \{(\partial_i \pi^a)^2 + \tfrac{1}{4}(\partial_i \pi^2)^2 \lambda^n (f_\pi^2 - \lambda^n \pi^2)^{-1}\} + \tfrac{1}{2}\omega^2 \int d^n x \pi^2. \tag{3.6}$$

The condition of stationary action gives

$$0 = \left.\frac{\partial A}{\partial \lambda}\right|_{\lambda=1}$$
$$= -\tfrac{1}{2} \int d^n x \{2(\partial_i \pi^a)^2 + \tfrac{1}{4}(\partial_i \pi^2)^2 [2(f_\pi^2 - \pi^2) + n f_\pi^2](f_\pi^2 - \pi^2)^{-2}\}. \tag{3.7}$$

Now the integrand is positive so the condition (3.7) is never satisfied with a nontrivial solution.

We can also use ordinary scale variation $\pi^a \to \pi^a(\lambda x)$ on (3.3). This gives the condition

$$(2-n) \int d^n x [(\partial_i \pi^a)^2 + \tfrac{1}{4}(\partial_i \pi^2)^2 (f_\pi^2 - \pi^2)^{-1}] + n\omega^2 \int d^n x \pi^2 = 0. \tag{3.8}$$

Thus $n > 2$ is a necessary condition for the existence of any localized solutions.

For the three-dimensional case we could take again[6]

$$(\partial_t \pi^a)^2 = \omega^2 (\pi^a)^2, \tag{3.9}$$

with π^2 and $(\partial_t \pi^a)^2$ time independent. Using this and the Lagrangian (2.19) with (C8) gives the action

$$S = -\int d^3 x \{\tfrac{1}{2}[(\partial_i \pi^a)^2 - \omega^2 \pi^2 + \tfrac{1}{4}(\partial_i \pi^2)^2 (f_\pi^2 - \pi^2)^{-1}]\}^p. \tag{3.10}$$

The above variational analysis can again be carried out. If instead of (3.5) we choose

$$\pi^a(\mathbf{x}) \to \pi^a_\lambda = \lambda^{3/2p}\pi^a(\lambda\mathbf{x}), \tag{3.11}$$

then we get a result similar to (3.7), and the ordinary scale variation gives a necessary condition $p < \frac{3}{2}$. We would like to point out that from the ordinary scale variation (which should be allowed irrespective of topological consideration) we *always* get the result that if the theory is set up so that it has time independent solutions, then it does not have solutions of the described time dependence and vice versa.

IV. COUPLING TO GAUGE FIELDS

In order to couple the pion fields to gauge fields we have to define the gauge covariant derivative to replace ∂_i. (We consider only time independent cases and also take the time component of the gauge field to vanish.) When the gauge field is Abelian we can define[5]

$$\mathcal{D}\pi^a = \partial_j\pi^a + e\epsilon^{ab}\pi^b A_j, \quad a = 1, 2, \tag{4.1}$$

$$G_{ij} = \partial_i A_j - \partial_j A_i, \tag{4.2}$$

with $\pi^3 = 0$.[13] (Here $\epsilon^{12} = -\epsilon^{21} = 1$). The energy which is to be extremized is

$$E = \int d^n x\left[\tfrac{1}{4}G_{ij}G_{ij} + \tfrac{1}{2}\mathcal{D}_j\pi^a\mathcal{D}_j\pi^b F_{ab}(\pi)\right]. \tag{4.3}$$

For a non-Abelian gauge field we define

$$\mathcal{D}_j\pi^a = \partial_j\pi^a + e\epsilon^{abc}A^b_j\pi^c \tag{4.4}$$

$$G^a_{ij} = \partial_i A^a_j - \partial_j A^a_i + e\epsilon^{abc}A^b_i A^c_j \tag{4.5}$$

$$E = \int d^n x\left[\tfrac{1}{4}G^a_{ij}G^a_{ij} + \tfrac{1}{2}\mathcal{D}_j\pi^a\mathcal{D}_j\pi^b F_{ab}(\pi)\right]. \tag{4.6}$$

[This does not represent a gauging of the full chiral SU(2)⊗ SU(2), but just its linear part.]

For variational study we can in all of these cases use the scale variations

$$\pi^a(\mathbf{x}) \to \pi^a_\lambda = \pi^a(\lambda\mathbf{x}), \tag{4.7}$$

$$A^{(a)}_j(\mathbf{x}) \to A^{(a)}_{j\lambda} = \lambda A^{(a)}_j(\lambda\mathbf{x}). \tag{4.8}$$

The extra λ in (4.8) is chosen so that $\mathcal{D}$ and G change according to (dropping indices)

$$\mathcal{D}\pi(\mathbf{x}) \to \lambda\mathcal{D}\pi(\lambda\mathbf{x}), \quad G(\mathbf{x}) \to \lambda^2 G(\lambda\mathbf{x}) \tag{4.9}$$

as can be easily seen by using the definitions before. The choice (4.8) guarantees also that $\lim_{r\to\infty} rA(r)$ is invariant as is required by some conserved charge arguments.

Substituting (4.7) and (4.8) into (4.3) and (4.6) gives in each case

$$E(\lambda) = \lambda^{4-n}I_G + \lambda^{2-n}I_D, \tag{4.10}$$

where I_G and I_D are λ-independent. I_G contains the pure gauge term G^2, and I_D the covariant derivative part. I_G and I_D are obviously positive for a nontrivial solution, and they must both be finite to give finite energy. The variational condition is

$$0 = \left.\frac{\partial E}{\partial\lambda}\right|_{\lambda=1} = (4-n)I_G + (2-n)I_D, \tag{4.11}$$

from which we get the necessary condition $n = 3$ for a finite energy solution.[14]

As is evident from the derivation, the result (4.11) is not limited to chiral theories. For example, for pure gauge theories ($I_D \equiv 0$), (4.11) indicates that for $n = 3$[2] there are no finite energy solutions,[15] but for $n = 4$ such solutions are not excluded.[16] The nonchiral case $F_{ab}(\pi^2) = \delta_{ab}$ has been studied for $n = 3$ by Prasad and Sommerfield,[17] who obtain solutions in closed form.

We can also apply the variation $\pi^a(\mathbf{x}) \to \xi\pi^a(\mathbf{x})$ on (4.3). This leads to (2.4) (with ∂_i's replaced by $\mathcal{D}_i$) and a negative result to the existence of a finite energy solution. This variation does not exclude solutions with $\pi^2(\infty) \neq 0$. [The solution of Prasad and Sommerfield indeed has $\phi^2(\infty) \neq 0$.]

Our conclusion from this section is that the nonlinearity of the chiral system does not change the results of existence or nonexistence from those of nonchiral theories.

V. ADDITIONAL POTENTIAL

In the previous sections the Lagrangian density was chirally invariant. We now add an extra potential energy term $V(\pi^2)$ and check its implications. Since we are mainly interested in whether chiral theories could have localized solutions due to their intrinsic nonlinearity, we do not consider potentials that support spontaneous symmetry breaking,[18] because in that case even standard theories seem to have localized solutions.[1,12,17,19] We therefore assume

(i) $\lim_{r\to\infty} \pi^2(r, \Omega) = 0$,

(ii) $V(\pi^2 = 0) = 0$,

We also want the theory to describe pions with positive mass, so we want

(iii) $V'(\pi^2)\big|_{\pi^2=0} = \frac{1}{2}m_\pi^2 > 0$.

The energy to be extremized is now

$$E[\pi^2] = \int \left[\tfrac{1}{2}(\partial_i\pi^a)(\partial_i\pi^b)F_{ab}(\pi) + V(\pi^2)\right]d^n x \tag{5.1}$$

in the time independent case of Sec. II. In the W formulation the now allowed $\pi \to \xi\pi$ variation leads to the following condition:

$$\int \left[f_\pi^4(\partial_i\pi^a)^2(f_\pi^2 - \pi^2)/(f_\pi^2 + \pi^2)^3 + 2\pi^2 V'(\pi^2)\right]d^n x = 0, \tag{5.2}$$

so that we must have at least

$$\int V'(\pi^2)\pi^2\, d^n x < 0. \tag{5.3}$$

The same condition results for gauge coupling. [In (5.2) ∂_i's would be replaced by $\mathcal{D}_i$, but the first term would still be positive.]

With the simple time dependence of Sec. III we want to extremize [B formulation]

$$A = -\int d^n x\left\{\tfrac{1}{2}\left[(\partial_i\pi^a)^2 - \omega^2\pi^2 + \tfrac{1}{4}(\partial_i\pi^2)^2(f_\pi^2 - \pi^2)^{-1}\right] + V(\pi^2)\right\}. \tag{5.4}$$

For the variation we again take (3.5) and the condition then is

$$\int d^n x\left\{(\partial_i\pi^a)^2 + \tfrac{1}{4}(\partial_i\pi^2)^2\left[f_\pi^2 - \pi^2 + \tfrac{1}{2}nf_\pi^2\right](f_\pi^2 - \pi^2)^{-2}\right\} + n\int d^n x\left\{-V(\pi^2) + \pi^2 V'(\pi^2)\right\} = 0, \tag{5.5}$$

leading to the necessary condition

$$\int d^n x\{\pi^2 V'(\pi^2) - V(\pi^2)\} < 0. \tag{5.6}$$

Let us now take a look at some of the possible potentials. We assume that the symmetry breaking term transforms according to $(N/2, N/2)$ under $SU(2)\otimes SU(2)$. This leads to a differential equation for the potential.[3,20] The solution to this equation can be expressed as Chebyshev polynomials of the second kind[20] in the variable $[f(\pi^2)^2/f^2(\pi^2)+\pi^2)]^{1/2}$. If we add a constant $(0,0)$ term so that V will satisfy the condition (ii), we get, for example, $[V_{N/2}]$:

$$V_{1/2} \propto 1 - z, \tag{5.7a}$$

$$V_1 \propto 1 - z^2, \tag{5.7b}$$

$$z = \{f^2(\pi^2)/[f^2(\pi^2)+\pi^2]\}^{1/2}. \tag{5.8}$$

For use in (5.3) we write [W formulation]

$$V_{1/2} = \tfrac{1}{2} m_\pi^2 \pi^2 (1 + \pi^2/f_\pi^2)^{-1} \tag{5.9a}$$

$$V_1 = \tfrac{1}{2} m_\pi^2 \pi^2 (1 + \pi^2/f_\pi^2)^{-2}. \tag{5.9b}$$

Now

$$\pi^2 V'_{1/2}(\pi^2) = \tfrac{1}{2} m_\pi^2 \pi^2 (1 + \pi^2/f_\pi^2)^{-2} \geq 0,$$

$$\pi^2 V'_1(\pi^2) = \tfrac{1}{2} m_\pi^2 \pi^2 \frac{1 - \pi^2/f_\pi^2}{(1 + \pi^2/f_\pi^2)^3} \geq 0,$$

and so neither of these potentials satisfy the condition (5.3). For (5.6) we use B formulation and get

$$V_{1/2} = m_\pi^2 f_\pi^2 (1 - (1 - \pi^2/f_\pi^2)^{1/2}), \tag{5.10a}$$

$$V_1 = \tfrac{1}{2} m_\pi^2 \pi^2; (\pi^2) \tag{5.10b}$$

hence

$$\pi^2 V'_{1/2}(\pi^2) - V_{1/2}(\pi^2)$$
$$= m_\pi^2 f_\pi^2 (1 - \tfrac{1}{2}\pi^2/f_\pi^2 - (1 - \pi^2/f_\pi^2)^{1/2})(1 - \pi^2/f_\pi^2)^{-1/2}$$
$$\leq 0,$$

$$\pi^2 V'_1(\pi^2) - V_1(\pi^2) = 0.$$

Therefore, the only action that passes this test for $N=1,2$ is (5.4) with (5.10a). Potentials with higher N could have symmetry breaking solutions $\pi^2(\infty) \neq 0$.

For a system with such a high symmetry we could also try to use variations related to that symmetry. Namely define

$$\delta\pi^a = \epsilon[X^p, \pi^a]; \tag{5.11}$$

then the chirally invariant part of the Lagrangian density is automatically invariant under this variation. To use this for the symmetry breaking potential, we note that

$$\delta(\pi^2) = \epsilon 2\pi^a [X^p, \pi^a]$$
$$= \epsilon 2\pi^p [f(\pi^2) + \pi^2 g(\pi^2)]. \tag{5.12}$$

We could actually use a linear combination of X^p's with constant coefficient, so that we get

$$\delta E = 2\epsilon \int d^n x\, k_p \pi^p [f(\pi^2) + \pi^2 g(\pi^2)] V'(\pi^2). \tag{5.13}$$

[When applying this to (5.4), the explicit time dependence of π_i has already been taken into account, therefore, the term $-(\pi^2)$ should be included to the symmetry breaking potential.] In all of the standard *ansätze* π^a satisfies, however, the condition

$$\int d\Omega\, \pi^a = 0, \quad \forall\, a, \tag{5.14}$$

in which case the condition (5.13) is empty. If (5.14) is not satisfied, then (5.13) could be useful for further analysis.

VI. CONCLUSIONS

We have studied the existence of localized solutions for theories based on Weinberg's nonlinear realization of chiral $SU(2)\otimes SU(2)$. We have used the standard variations $\phi(x) \to \phi(\lambda x)$ and $\phi(x) \to \xi\phi(x)$ on the action integral. In general our result is that the intrinsic nonlinearity of the system is not enough to support confined solutions. Topological arguments would help in the case studied in Sec. II, if the nonlinear theory was considered arising from a linear theory with a constraint. Some of the systems passing the test could have localized solutions even if the chiral nonlinearity is removed. An additional chiral symmetry breaking potential seems to help in a special case in Sec. V. Although the results of this paper are rather negative, it would still be interesting to study whether some other nongauge-type symmetry schemes could by themselves support localized solutions.

ACKNOWLEDGMENTS

The author would like to acknowledge discussions with K. Tanaka and E. Takasugi.

APPENDIX A

Whichever norm we use, we will always assume that $\|\phi_c\| < \infty$. Then it is easy to see that $\phi_\xi = \xi\phi_c(x)$ satisfies the condition (1.4):

$$\|\phi_\xi - \phi_c\| = |\xi - 1| \cdot \|\phi_c\| \to 0, \quad \text{as } \xi \to 1.$$

To discuss $\phi_\lambda = \phi(\lambda \mathbf{x})$ we first prove the following simple lemma.

Lemma: Let $r = [\sum_i (x^i)^2]^{1/2}$ and assume $\Psi(\mathbf{x}) \equiv \Psi(r, \Omega)$ (where Ω stands for angular variables) satisfies the following three conditions:

(1) Ψ is continuous and differentiable with respect to r for all $r > 0$ and all Ω,

(2) for any R, $r\, d\Psi/dr$ is bounded for $0 < r < R < \infty$,

(3) $\lim_{r\to\infty} \Psi(r, \Omega) = C(\Omega)$, where $C(\Omega)$ is bounded for all Ω.

Then for each $\epsilon > 0$ there is a positive real number δ such that

$$|\psi(\lambda \mathbf{x}) - \psi(\mathbf{x})| < \epsilon, \quad \forall\, \mathbf{x} \in \mathbf{R}^n - \{0\}, \quad \text{if } |\lambda - 1| < \delta.$$

Proof: For simplicity assume $\lambda > 1$. From assumption (3) above we get by Cauchy's criteria that there is a real number R such that

$$|\psi(\lambda r, \Omega) - \psi(r, \Omega)| < \epsilon, \quad \forall\, r > R$$

On the other hand since Ψ is continuous and differentiable in $\mathbf{R}^n - \{0\}$ we can use the mean value theorem and get

$$|\psi(\lambda r, \Omega) - \psi(r, \Omega)| = |\lambda - 1|\, r\, |d\psi(\xi, \Omega)/d\xi|,$$

for some $\xi \in [r, \lambda r]$. We need this only for $0 < r \leq R$. By

(2), $r\,d\Psi/dr$ was bounded, so we can estimate

$$|\lambda-1|r\left|\frac{d\Psi}{d\xi}\right|<|\lambda-1|M.$$

Therefore, by choosing $\delta=\epsilon/M$ we complete the proof.

We can now use this lemma to check (1.4). If the conditions of the lemma are satisfied by ϕ [ϕ and $\partial_i\phi$], then we can use this variation with the norm (1.3a) [(1.3b)]. The conditions (1) and (3) must be satisfied by any candidate for a classical solution (the point $r=0$ is left out to allow monopole-type discontinuities). One might, however, argue that boundedness of $r\,d\phi/dr$ and $r(d/dr)(\partial^i\phi)$ could be a too strict condition. In such a case results from the variation $\phi\to\phi_\lambda$ do not apply.

APPENDIX B

We follow Weinberg's definitions[3] throughout. The algebra of generators of chiral SU(2)⊗SU(2) is

$$\begin{aligned}&[T_a,T_b]=i\epsilon_{abc}T_c,\\&[T_a,X_b]=i\epsilon_{abc}X_c,\\&[X_a,X_b]=i\epsilon_{abc}T_c.\end{aligned}\tag{B1}$$

The pion field π_a is assumed to transform nonlinearly under X_a,

$$[X_a,\pi_b]=-if_{ab}(\pi^2),\tag{B2}$$

and linearly under T_a,

$$[T_a,\pi_b]=i\epsilon_{abc}\pi_c.\tag{B3}$$

If f_{ab} is assumed to be even, then the Jacobi identities lead to the form

$$f_{ab}=\delta_{ab}f(\pi^2)+\pi_a\pi_b g(\pi^2),\tag{B4}$$

where

$$g(\pi^2)=[1+2f(\pi^2)f'(\pi^2)]/[f(\pi^2)-2\pi^2f'(\pi^2)],\tag{B5}$$

and f is an arbitrary (reasonably regular) function $\{f'\equiv[d/d(\pi^2)]f(\pi^2)\}$.

Other fields are assumed to transform according to

$$[X_a,\psi]=v_{ab}(\pi^2)t_b\psi,\tag{B6}$$

$$[T_b,\psi]=-t_b\psi,\tag{B7}$$

where t_b is an Hermitian matrix. The function v_{ab} is given by

$$v_{ab}(\pi^2)=\epsilon_{abc}\pi_c v(\pi^2)\tag{B8}$$

where

$$v(\pi^2)=\{f(\pi^2)+[f^2(\pi^2)+\pi^2]^{1/2}\}^{-1}.\tag{B9}$$

The covariant derivative

$$D_\mu\pi_a\equiv d_{ab}(\pi^2)\partial_\mu\pi_b\tag{B10}$$

is assumed to transform according to (B6) and (B7). This determines d_{ab} and

$$\begin{aligned}D_\mu\pi_a\propto{}&[f^2(\pi^2)+\pi^2]^{-1/2}\partial_\mu\pi_a\\&-[f^2(\pi^2)+\pi^2]^{-1}\cdot[f'(\pi^2)+\tfrac{1}{2}v(\pi^2)]\pi_a\partial_\mu\pi^2.\end{aligned}\tag{B11}$$

The kinetic part of the Lagrangian is then

$$\mathcal{L}=-\tfrac{1}{2}D_\mu\pi_aD^\mu\pi_a=-\tfrac{1}{2}\partial_\mu\pi_a\,\partial^\mu\pi_bF_{ab},\tag{B12}$$

$$\begin{aligned}F_{ab}&=f^2(0)(\delta_{ab}\bar{F}+\pi_a\pi_b\bar{G})\\&=f^2(0)\{\delta_{ab}[f^2(\pi^2)+\pi^2]^{-1}+\pi_a\pi_b[f^2(\pi^2)+\pi^2]^{-2}\\&\quad\times[4\pi^2(f')^2-4ff'-1]\}.\end{aligned}\tag{B13}$$

APPENDIX C

The nonlinear realization is covariant under the redefinition

$$\pi_a^*=\pi_a\Phi(\pi^2),\tag{C1}$$

provided we also change f and g by

$$f^*((\pi_a^*)^2)=f(\pi^2)\Phi(\pi^2),\tag{C2}$$

$$\begin{aligned}g^*((\pi_a^*)^2)=[&g(\pi^2)\Phi(\pi^2)+2f(\pi^2)\Phi'(\pi^2)\\&+2\pi^2g(\pi^2)\Phi'(\pi^2)]\Phi(\pi^2)^{-2}.\end{aligned}\tag{C3}$$

From (C2) we see that specific choices of $f(\pi^2)$ can be connected by suitable Φ, therefore giving equivalence between these realizations. The following definitions are common [they will be referred to as W, B, and I formulations, respectively]

$$f_W(\pi^2)=(1/2f_\pi)(f_\pi^2-\pi^2),\tag{C4}$$

$$f_B(\pi^2)=(f_\pi^2-\pi^2)^{1/2},\tag{C5}$$

$$f_I(\pi^2)=(\pi^2)^{1/2}\cot((\pi^2)^{1/2}\kappa),\tag{C6}$$

where $\kappa=3.14159.../(2f_\pi)$. For these choices of f we get the following Lagrangians using (B12) and (B13):

$$\mathcal{L}_W=-\tfrac{1}{2}(\partial_\mu\pi_a)(\partial^\mu\pi_a)[f_\pi^2/(f_\pi^2+\pi^2)]^2,\tag{C7}$$

$$\mathcal{L}_B=-\tfrac{1}{2}[(\partial_\mu\pi_a)(\partial^\mu\pi_a)+\tfrac{1}{4}(\partial_\mu\pi^2)^2(f_\pi^2-\pi^2)^{-1}],\tag{C8}$$

$$\begin{aligned}\mathcal{L}_I=-\tfrac{1}{2}&((\partial_\mu\pi_a)(\partial^\mu\pi_a)\sin^2[(\pi^2)^{1/2}\kappa]/\pi^2\kappa^2\\&+\tfrac{1}{4}(\partial_\mu\pi^2)^2\{1-\sin^2[(\pi^2)^{1/2}\kappa]/\pi^2\kappa^2\}/\pi^2).\end{aligned}\tag{C9}$$

Let us define the transforming Φ_{AB} by

$$f_B^*=f_A\Phi_{AB}.\tag{C10}$$

Then

$$f_B^*(\pi^2\Phi_{AB}^2(\pi^2))=f_A(\pi^2)\Phi_{AB}(\pi^2),\tag{C11}$$

from which we can solve for Φ_{AB}. In (C4)–(C6) the parameter f_π was chosen so that in each case $f(\pi^2=f_\pi^2)=0$. We want the transformation (C1) to keep the points $\pi^2=0$ and $\pi^2=f_\pi^2$ fixed and this condition determines the functions Φ as follows:

$$\Phi_{WB}=2f_\pi^2/(f_\pi^2+\pi^2),\tag{C12}$$

$$\Phi_{WI}=2\arctan((\pi^2)^{1/2}/f_\pi)/(\pi^2)^{1/2}\kappa,\tag{C13}$$

$$\Phi_{BW}=[f_\pi^2-f_\pi(f_\pi^2-\pi^2)^{1/2}]/\pi^2,\tag{C14}$$

$$\Phi_{BI}=\arcsin[(\pi^2)^{1/2}/f_\pi]/(\pi^2)^{1/2}\kappa,\tag{C15}$$

$$\Phi_{IW}=\tan[(\pi^2)^{1/2}\kappa/2]f_\pi/(\pi^2)^{1/2},\tag{C16}$$

$$\Phi_{IB}=\sin[(\pi^2)^{1/2}\kappa]f_\pi/(\pi^2)^{1/2}.\tag{C17}$$

The Jacobian matrix of the transformation (C1) is

$$\frac{\partial\pi_a^*}{\partial\pi_b}=\delta_{ab}\Phi+\pi_a\pi_b2\Phi,\tag{C18}$$

so

$$\det\frac{\partial\pi_a^*}{\partial\pi_b}=\Phi^2(\Phi+2\pi^2\Phi').\tag{C19}$$

Since $\Phi \neq 0$ in the open interval $(0, f_\pi^2)$, we only need to check that $\Phi(\pi^2) + 2\pi^2\Phi'(\pi^2) \neq 0$ for all $0 < \pi^2 < f_\pi^2$. From (C12)—(C17) this is easily shown.

*Research supported in part by the U.S. Energy Research and Development Administration under Contract No. AT(11-1)-1545.

[1]For a review, see R. Rajaraman, Phys. Reports **21C**, 227 (1975).
[2]T.T. Wu and C.N. Yang, in *Properties of Matter under Unusual Conditions*, edited by H. Mark and S. Fernbach (Interscience, New York, 1969).
[3]S. Weinberg, Phys. Rev. **166**, 1568 (1968).
[4]Y. Dothan, Univ. of Helsinki preprint (unpublished).
[5]M.J. Duff and C.J. Isham, Nucl. Phys. B **108**, 130 (1976).
[6]S. Deser, M.J. Duff, and C.J. Isham, Nucl. Phys. (to be published).
[7]J. Honerkamp, A. Patani, M. Schlindwein, and Q. Shafi, Lett. Nuovo Cimento **15**, 97 (1976).
[8]N.I. Akheizer, *The Calculus of Variation* (Blaisdell, New York, 1962); I.M. Gel'fand and S.V. Fomin, *The Calculus of Variations* (Prentice-Hall, Englewood Cliffs, N.J., 1963).
[9]T.H.R. Skyrme, Proc. Roy. Soc. (London) A **260**, 127 (1961); J. Arafune, P.G.O. Freund, and G.J. Goebel, J. Math. Phys. **16**, 433 (1975).
[10]G.H. Derrick, J. Math. Phys. **5**, 1252 (1964); G. Rosen, *ibid.* **7**, 2066, 2071 (1966).
[11]See also, P. Chang and F. Gürsey, Phys. Rev. **164**, 1752 (1967); **169**, 1397 (1968).
[12]For other equivalent methods, see R. Friedberg, T.D. Lee, and A. Sirlin, Phys. Rev. D **13**, 2739 (1976). The existence of time dependent solutions has also been studied by G. Rosen, J. Math. Phys. **9**, 996 (1968).
[13]Using (3.1), we can write (4.1) as
$$D_j\phi = \partial_j\phi - ieA_j\phi.$$
[14]In particular, the case $n = 2$ considered in Ref. 5 has no finite energy solutions.
[15]This has been shown by G. Rosen, J. Math. Phys. **13**, 595 (1972); Quart. Appl. Math. **27**, 133 (1969). When using (4.8) it is assumed that $A_j^{(a)}(\infty) = 0$, and therefore solutions with $A_j^{(a)}(\infty) \neq 0$ are not excluded.
[16]For a solution see A.A. Belavin, A.M. Polyakov, A.S. Schwartz, and Yu.S. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[17]M.K. Prasad and C.M. Sommerfield, Phys. Rev. Lett. **35**, 760 (1975).
[18]For a spontaneously symmetry breaking potential a solution has been obtained in closed form by Dothan, Ref. 4. This reference uses *I* formulation.
[19]G. 't Hooft, Nucl. Phys. B **79**, 276 (1974); A.M. Polyakov, Zh. Eksp. Theor. Fiz. Pis. Red. **20**, 430 (1974) [JETP Lett. **20**, 194 (1974)].
[20]W. Sollfrey, Phys. Rev. **173**, 1805 (1968); C.J. Liu and Y. Chow, Nuovo Cimento A **1**, 369 (1971).

Subgroups of the Euclidean group and symmetry breaking in nonrelativistic quantum mechanics*

J. Beckers,† J. Patera, M. Perroud, and P. Winternitz

Centre de recherches mathématiques, Université de Montréal, Montréal, Canada
(Received 14 July 1976)

A systematic study of explicit symmetry breaking in the nonrelativistic quantum mechanics of a scalar and a spinor particle is presented. The free Schrödinger (or Pauli) equation is invariant under the Euclidean group E(3); an external field will break this symmetry to a lower one. We first find all continuous subgroups of E(3) and then for each subgroup construct the most general (within certain restrictions) external field that breaks the symmetry from E(3) to the corresponding subgroup. For a scalar particle the interaction term is assumed to be of the form $V(\mathbf{r})+\mathbf{A}(\mathbf{r})\mathbf{P}$, where $\mathbf{P}$ is the momentum, i.e., it involves an arbitrary scalar and vector potential. For a spinor particle it is of the form $V(\mathbf{r})+\mathbf{A}(\mathbf{r})\mathbf{P}+\mathbf{B}(\mathbf{r})\boldsymbol{\sigma}+M_{ik}(\mathbf{r})\sigma_i P_k$ (σ_i are the Pauli matrices). A one-to-one correspondence between subgroups of E(3) and classes of "symmetry breaking potentials" is established. The remaining subgroup symmetry is then used to solve or at least simplify the obtained Schrödinger equation. The existence of a one-dimensional invariance group (for a particle in a field) leads to the partial separation of variables and determines the functional dependence of the wavefunction on one variable. A two-dimensional group implies the complete separation of variables and the functional dependence on two variables. A higher dimensional invariance group implies the separation of variables in one or more systems of coordinates and in some cases specifies the wavefunction completely.

1. INTRODUCTION

Several recent publications have been devoted to the problem of classifying all continuous subgroups of Lie groups that are of interest in physics.[1-5] In particular, all continuous subgroups of the Poincaré group, the similitude group (the Poincaré group extended by dilations), the de Sitter groups, and some further groups are now known.

One of the physical motivations for undertaking a study of the subgroup structure of a given Lie group is the interest in symmetry breaking in physics. Indeed, consider a physical system that has a symmetry, described by a certain Lie group G. This symmetry group, once found, can be used to determine some, and in certain cases all, of the properties of the system under consideration. The system itself may then be modified by taking into account further interactions, by being placed into an external field or into an environment imposing certain boundary conditions, etc. Typically such a modification will change the symmetry properties of the system, often lowering the symmetry or destroying it completely.

We are interested in a systematic study of the case when the additional "influence" does lower the symmetry group from the original G to a subgroup $G_i \subset G$. A classification of all subgroups G_i thus provides a classification of all symmetry breaking interactions (or other influences) for a given system. For each subgroup G_i we thus wish to find the most general interaction that breaks the symmetry from G to G_i. The subgroup G_i can then be used to study the new modified system. Thus, the generators and invariants of G_i will provide integrals of motion and quantum numbers for the system. The representation theory of the group G_i will provide wavefunctions of the system, or at least many properties of the wavefunctions.

We intend to pursue this line of thought in connection with various types of symmetry breaking for a variety of physical systems. In particular we plan to investigate systematically from this point of view many of the differential equations and Hamiltonians of classical and quantum physics, both in the relativistic and nonrelativistic cases.

The present article is devoted to a particularly simple and intuitively clear case, namely that of a stable nonrelativistic scalar (spin-0) or spinor (spin-$\frac{1}{2}$) particle in an external field. We restrict ourselves to a study of explicit symmetry breaking in the Schrödinger (or Pauli) equation, due to the introduction of scalar and vector potentials into the Hamiltonian.

The geometric invariance group of the Hamiltonian for a free particle is the Euclidean group E(3), generated by the rotations and translations of a three-dimensional real Euclidean space. The continuous subgroups of E(3) are known,[6] as is its representation theory.[7-9] We shall modify the Schrödinger equation by introducing a scalar and a vector potential, both depending on the particle coordinates and the Pauli equation by introducing potentials depending on the particle coordinates, momenta (linearly) and spins. For each subgroup G_i of E(3) we shall find the most general potential (within the considered Ansatz) reducing the symmetry from E(3) to G_i.

In Sec. 2 we shall discuss some properties of the group E(3), construct a lattice of its subgroups (see Fig. 1) and present some relevant results on its representation theory. In Sec. 3 we consider a scalar particle and for each subgroup of E(3) explicitly construct the symmetry breaking potentials. The results are summarized in Table I and discussed in the same section. The same problem for a spinor particle is solved in Sec. 4 in which we present and discuss a somewhat cumbersome list of "subgroup invariant" potentials. In Sec. 5 we demonstrate, first for a scalar, then for a spinor particle, how the remaining symmetry group can be used to investigate the obtained Schrödinger (or Pauli) equations. The existence of a one-dimensional

invariance group leads to the partial separation of variables in some system of coordinates and to the explicit form of the dependence of the wavefunctions on one variable. Higher-dimensional invariance groups lead to the complete separation of variables in one or more coordinate systems and provide the dependence of the wavefunctions on at least two of the variables, in some cases on all three of them. The conclusions and future outlook are presented in the final Sec. 6.

2. THE EUCLIDEAN GROUP E(3) AND ITS SUBGROUPS

The group E(3) is defined as the group of transformations of the three-dimensional real vector space, leaving the Euclidean distance between two points invariant. Thus, we can write an element of the group as $(\Lambda, \mathbf{a})$, where

$$x_i' = \sum_{k=1}^{3} \Lambda_{ik} x_k + a_i, \quad i = 1, 2, 3. \tag{1}$$

Here Λ_{ik} is an O(3) matrix and a_i are the components of a real vector. The Lie algebra of E(3) is generated by three infinitesimal rotations J_i and three infinitesimal translations P_i, satisfying the commutation relations

$$[J_i, J_k] = i\epsilon_{ikl} J_l, \quad [J_i, P_k] = i\epsilon_{ikl} P_l, \quad [P_i, P_k] = 0. \tag{2}$$

We shall also have the opportunity to use the parity operator Π and time reversal operator T, satisfying

$$\Pi J_i \Pi^{-1} = J_i, \quad \Pi P_i \Pi^{-1} = -P_i,$$
$$TJ_i T^{-1} = -J_i, \quad TP_i T^{-1} = -P_i. \tag{3}$$

The finite transformations $(\Lambda, \mathbf{a})$ obey the composition law

$$(\Lambda_1, \mathbf{a}_1)(\Lambda_2, \mathbf{a}_2) = (\Lambda_1 \Lambda_2, \mathbf{a}_1 + \Lambda_1 \mathbf{a}_2). \tag{4}$$

All subalgebras of the algebra (2) can be found by a direct application of a classification algorithm presented earlier.[2] Indeed, the algebra L of E(3) has an Abelian ideal, namely the translations $T = \{P_1, P_2, P_3\}$.

The factor algebra $F = L/T$ is isomorphic to the rotation algebra O(3). According to the algorithm we proceed as follows: 1. Find all conjugacy classes [under O(3)] of subalgebras F_a of F. These can be represented by $F_1 = \{J_1, J_2, J_3\}$, $F_2 = \{J_3\}$ and $F_3 = \{0\}$. 2. For each subalgebra F_a $(a = 1, 2, 3)$ find all invariant subspaces $T_{a,k}$ in T. Use the normalizer of F_a in E(3) (i.e., all Euclidean transformations leaving the subalgebra F_a invariant) to classify the invariant subspaces into conjugacy classes. Taking a representative of each subspace, we obtain all "splitting subalgebras" of L as the

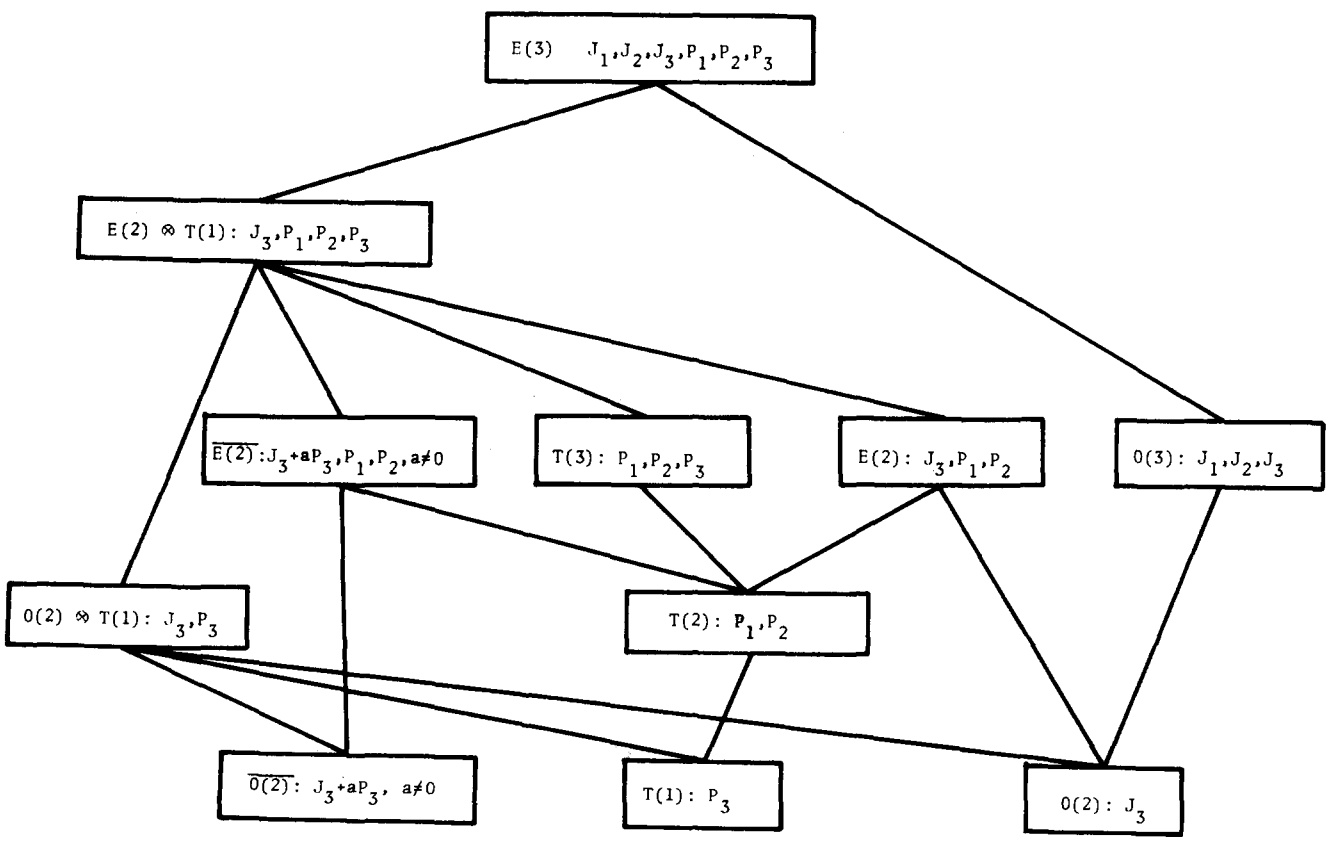


FIG. 1. Subgroups of E(3); E(n), T(n), and O(n) are groups of Euclidean transformations, translations, and rotations, respectively; $\overline{E(2)}$ and $\overline{O(2)}$ are the universal covering groups of E(2) and O(2).

algebraic sums $F_a \dot{+} T_{a,k}$. 3. Find all nonsplitting subalgebras of L, i.e., such subalgebras that contain elements of the type $B+X$ with $B \in F$, $X \in T$, that are not conjugate under E(3) to an element of F or of T. To do this, we consider each subalgebra F_a separately. We take all generators of F_a and add to each of them an arbitrary linear combination of all generators of T. We then use the translations $\exp T$ and the outer automorphism of F_a in E(3) to simplify the general linear combinations as far as possible. Finally we must assure that the generators thus obtained, together with the generators of a chosen invariant subalgebra $T_{a,k} \in T$, form an algebra.

This procedure leads directly to the lattice of subalgebras shown on Fig. 1. In each box we give the generators of the Lie subalgebra and also comment on the corresponding Lie group [$E(n), O(n)$, and $T(n)$ are the groups of Euclidean transformations, rotations, or translations of an n-dimensional real Euclidean space, respectively]. The parameter a, occuring twice in Fig. 1, satisfies $a \neq 0$, $-\infty < a < \infty$ if conjugation is considered under the proper Euclidean group and $a > 0$ if parity and/or time reversal are included.

The representation theory of the group E(3) is well known,[7-9] and we need not go into it here. Irreducible unitary representations can be labelled by a pair of real numbers, namely the eigenvalues of the two invariant operators

$$\mathbf{P}^2 \text{ and } (\mathbf{J}, \mathbf{P})/\mathbf{P}^2, \tag{5}$$

corresponding to the energy and helicity of a free particle.

We shall actually make use of a certain type of reducible representation of E(3). Consider the Hilbert space $L^2(R^3, C^{2j+1})$, of normalizable $2j+1$ component spinors, each component of which is a complex valued function of the space coordinates x_1, x_2, x_3. Consider now an element $g \equiv (\Lambda, \mathbf{a})$ of the universal covering group $\overline{E(3)}$ of E(3) and represent it by the operator $U(g)$ acting on $\psi \in L^2(R^3, C^{2j+1})$ as

$$[U(g)\psi](\mathbf{r}) = D^j(\Lambda)\psi(g^{-1}\mathbf{r}). \tag{6}$$

Here $D^j(\Lambda)$ is a $(2j+1)\times(2j+1)$ matrix realizing an irreducible unitary representation of SU(2) (its matrix elements are Wigner D functions). It can actually be proven that more generally any representation of the form

$$[U(g)\psi](\mathbf{r}) = S(g, \mathbf{r})\psi(g^{-1}\mathbf{r}),$$

where $S(g, \mathbf{r})$ is a unitary operator acting on the space C^{2j+1}, is unitarily equivalent to (6).

In this article we shall make use of representation (6) for $j=0$ and $j=\frac{1}{2}$. For $j=0$ we simply have the "quasiregular representation,"[10] and $\psi(\mathbf{r})$ is a function satisfying

$$\int |\psi(\mathbf{r})|^2 \, dx\, dy\, dz < \infty. \tag{7}$$

For $j=\frac{1}{2}$, $\psi(\mathbf{r})$ is a two-component spinor satisfying

$$\psi(\mathbf{r}) = \begin{pmatrix} \psi_1(\mathbf{r}) \\ \psi_2(\mathbf{r}) \end{pmatrix}, \int \{|\psi_1(\mathbf{r})|^2 + |\psi_2(\mathbf{r})|^2\} dx\, dy\, dz < \infty. \tag{8}$$

3. SCALAR PARTICLE IN AN EXTERNAL FIELD

Let us consider the Hamiltonian

$$H = -\tfrac{1}{2}\Delta + V(\mathbf{r}) + \mathbf{A}(\mathbf{r})\mathbf{P}, \tag{9}$$

where $V(\mathbf{r})$ and $\mathbf{A}(\mathbf{r})$ are a scalar and vector potential and $P_a = -i\partial/\partial x_a$ are the components of the linear momentum operator.

If we require that the Hamiltonian be a Hermitian (or at least symmetric) operator, satisfying $H = H^*$, we find

$$\mathbf{A}(\mathbf{r}) = \mathbf{A}^*(\mathbf{r}), \quad \operatorname{Im} V(\mathbf{r}) = -\tfrac{1}{2}\operatorname{div}\mathbf{A}(\mathbf{r}). \tag{10}$$

Thus, $\mathbf{A}$ is a real vector function and the imaginary part of V is determined by $\mathbf{A}$.

If parity conservation is assumed, we must have

$$\Pi H \Pi^{-1} = H. \tag{11}$$

This implies

$$V(-\mathbf{r}) = V(\mathbf{r}), \quad \mathbf{A}(-\mathbf{r}) = -\mathbf{A}(\mathbf{r}). \tag{12}$$

Time reversal can be represented by an antiunitary operator $T = K$ (where K is complex conjugation). The time reversal transformation implies

$$THT^{-1} = \tilde{H}, \tag{13a}$$

where $\tilde{H}$ is the time reversed Hamiltonian. The manner in which $\tilde{H}$ is related to H depends on the physics of the problem. Thus $\tilde{V} = V$ but $\tilde{\mathbf{A}} = -\mathbf{A}$ if we are studying a particle in an electromagnetic field (the electric field vector $\mathbf{E}$ goes into itself, the magnetic field $\mathbf{H}$ changes sign). Thus, if $\mathbf{A}$ is the electromagnetic vector potential, time reversal invariance implies

$$V(\mathbf{r}) = V^*(\mathbf{r}), \quad A_i(\mathbf{r}) = A_i^*(\mathbf{r}). \tag{13b}$$

Let us now consider invariance under subgroups of the Euclidean group E(3). Consider the representation (6) with $j=0$, and let g be a transformation belonging to a subgroup of E(3). The invariance condition is

$$U(g)HU(g)^{-1} = H, \tag{14}$$

which implies

$$V(\mathbf{r}) = V(g\mathbf{r}), \quad D^1_{ik}(\Lambda)A_k(\mathbf{r}) = A_i(g\mathbf{r}). \tag{15}$$

Algebraically the "global" invariance conditions (15) are equivalent to the requirement

$$[H, X_a] = 0, \tag{16}$$

i.e., that the Hamiltonian should commute with all generators X_a of the subgroup.

Let us now consider each of the subgroups of E(3), proceeding from lower dimensions to higher ones. In each case it is a simple matter to solve the appropriate equations (15) or (16) and thus to find the invariant scalar and vector potentials $V(\mathbf{r})$ and $\mathbf{A}(\mathbf{r})$. We simply list the results of these calculations in Table I.

Several comments on the results of Table I should be made.

1. We make use of several different types of coordinate systems in Euclidean 3-space, namely,

TABLE I. Subgroups of E(3) and invariant potentials in spin-0 Schrödinger equation.

No.	Generators of symmetry group	Coordinates used in potentials	$V(\mathbf{r})$	$\mathbf{A}(\mathbf{r})\mathbf{P}$
1	P_3	Cartesian	$V(x_1, x_2)$	$\mathbf{A}(x_1, x_2)\mathbf{P}$
2	J_3	Cylindrical	$V(\rho, z)$	$a(\rho, z)\mathbf{rP} + b(\rho, z)P_3 + c(\rho, z)J_3$
3	$J_3 + aP_3$ $a \neq 0$	Helical	$V(\rho, v)$	$f(\rho, v)(\cos uP_1 + \sin uP_2)$ $+ g(\rho, v)(-\sin uP_1 + \cos uP_2) + h(\rho, v)P_3$
4	P_1, P_2	Cartesian	$V(x_3)$	$\mathbf{A}(x_3)\mathbf{P}$
5	J_3, P_3	Cylindrical	$V(\rho)$	$a(\rho)\rho(\cos\phi P_1 + \sin\phi P_2) + b(\rho)P_3 + c(\rho)J_3$
6	J_1, J_2, J_3	Spherical	$V(r)$	$a(r)\mathbf{rP}$
7	J_3, P_1, P_2	Cylindrical	$V(z)$	$a(z)P_3$
8	$J_3 + aP_3, P_1, P_2$ $a \neq 0$	Helical	$V = \text{const}$	$f[\cos(u+v)P_1 + \sin(u+v)P_2]$ $+ g[\sin(u+v)P_1 - \cos(u+v)P_2] + hP_3$ $f, g, h = \text{const}$
9	P_1, P_2, P_3	Cartesian	$V = \text{const}$	$\mathbf{AP}$, $A_i = \text{const}$
10	J_3, P_1, P_2, P_3	Cartesian	$V = \text{const}$	aP_3, $a = \text{const}$
11	J_1, J_2, J_3 P_1, P_2, P_3	Cartesian	$V = \text{const}$	$\mathbf{A} = 0$

(a) Cartesian: x_1, x_2, x_3

(b) Cylindrical: $\rho = (x_1^2 + x_2^2)^{1/2}$, $\phi = \arctan(x_2/x_1)$, $z = x_3$,

(c) Spherical: $r = (x_1^2 + x_2^2 + x_3^2)^{1/2}$, $\theta = \arccos(x_3/r)$, $\phi = \arctan(x_2/x_1)$,

(d) Helical: $\rho = (x_1^2 + x_2^2)^{1/2}$, $u = (1/2a)(z + a\phi)$, $v = (1/2a)(z - a\phi)$,

where $a > 0$ is a parameter and z and ϕ are cylindrical coordinates.

2. There is a one-to-one correspondence between subgroups of E(3) and the general form of a Hamiltonian of type (9) left invariant by the subgroup.

3. A scalar potential $V(\mathbf{r})$ alone is not sufficient to distinguish between all subgroups. Indeed, the groups E(3), E(2)⊗T(1), $\overline{\text{E(2)}}$, and T(3) all imply $V = \text{const}$. Similarly T(2) and E(2) both imply $V = V(z)$. The reason for this is that the requirement of homogeneity (translational invariance) is so restrictive for a scalar potential, that little space remains for manifestations of anisotropy.

4. The Hamiltonian of a spinless particle in an electromagnetic field can be written as

$$H = \tfrac{1}{2}[\mathbf{P} - e\mathbf{A}_E(\mathbf{r})]^2 + \phi(\mathbf{r}), \tag{17}$$

i.e., is of the form (9) with

$$\mathbf{A}(\mathbf{r}) = -e\mathbf{A}_E(\mathbf{r}),$$
$$V(\mathbf{r}) = \phi(\mathbf{r}) + \tfrac{1}{2}e^2\mathbf{A}_E^2(\mathbf{r}) + (ie/2)\text{div}[\mathbf{A}_E(\mathbf{r})], \tag{18}$$

(we have put $\hbar = m = 1$). A gauge transformation

$$\mathbf{A}_E(\mathbf{r}) \to \mathbf{A}_E(\mathbf{r}) + \text{grad}[f(\mathbf{r})], \tag{19}$$

where $f(\mathbf{r})$ is an arbitrary real function, leads to a new wavefunction, differing from the original one by a phase only:

$$\psi(\mathbf{r}) \to \psi(\mathbf{r})e^{if(\mathbf{r})}. \tag{20}$$

If we are only interested in $|\psi(\mathbf{r})|^2$, then gauge invariance can be used to simplify the obtained Hamiltonians further. In particular all constant scalar and vector potentials can be transformed away, many of the different potentials obtained become equivalent, etc.

5. Notice that invariance under T(3), E(2), E(2)⊗T(1), or E(3) specifies H up to certain constants. The groups T(1), O(2), or $\overline{\text{O(2)}}$ imply that H can depend on two variables in an arbitrary manner. The groups T(2), O(2)⊗T(1), O(3), or E(2) allow for an arbitrary dependence on one variable. By the same token, if T(3), E(2), E(2)⊗T(1), or E(3) is the invariance group then group representation theory will completely determine the wavefunctions, energy levels, etc. In all other cases group theory will only provide some properties of these physical quantities.

6. We have not imposed parity or time reversal invariance. Applying relations (12) and (13) to the entries of Table I we see that we would obtain further restrictions on the functions and constants in columns 4 and 5. Thus, e.g., in the case of spherical symmetry the pseudoscalar ($\mathbf{rP}$) is excluded by parity conservation, unless the coefficient $a(\mathbf{r})$ is itself a pseudoscalar. We shall not go into these considerations here.

4. PARTICLE WITH SPIN ½ IN EXTERNAL FIELD

Let us now consider the Hamiltonian

$$H = -\tfrac{1}{2}\Delta + V(\mathbf{r}) + \mathbf{A}(\mathbf{r})\boldsymbol{\sigma} + \mathbf{B}(\mathbf{r})\mathbf{P} + M_{ik}(\mathbf{r})\sigma_i P_k, \tag{21}$$

where V, A_i, B_i and M_{ik} are arbitrary functions of the coordinates, $\mathbf{P}$ is again the linear momentum, and σ_i are the Pauli matrices.

The hermiticity condition $H=H^+$ implies

$$B_i(\mathbf{r})=B_i^*(\mathbf{r}),\quad M_{ik}(\mathbf{r})=M_{ik}^*(\mathbf{r}),$$

$$\mathrm{Im}[V(\mathbf{r})]=-\frac{1}{2}\frac{\partial B_k(\mathbf{r})}{\partial x_k},\quad \mathrm{Im}[A_i(\mathbf{r})]=-\frac{1}{2}\frac{\partial M_{ik}(\mathbf{r})}{\partial x_k}. \tag{22}$$

Parity conservation in this case implies

$$V(-\mathbf{r})=V(\mathbf{r}),\quad \mathbf{A}(-\mathbf{r})=\mathbf{A}(\mathbf{r}),$$
$$B(-\mathbf{r})=-B(\mathbf{r}),\quad M_{ik}(-\mathbf{r})=-M_{ik}(\mathbf{r}). \tag{23}$$

The antiunitary operator representing time reversal can in this case be written as

$$T=-i\sigma_2 K, \tag{24}$$

where K is the operator of complex conjugation. The time reversal transformation gives

$$THT^{-1}=-\tfrac{1}{2}\Delta+V^*(\mathbf{r})-\mathbf{A}^*(\mathbf{r})\boldsymbol{\sigma}-\mathbf{B}^*(\mathbf{r})\mathbf{P}$$
$$+M_{ik}^*(\mathbf{r})\sigma_i P_k=\tilde{H}. \tag{25}$$

The form of the Hamiltonian $\tilde{H}$ for the time reversed process depends on the properties of the external quantities V, $\mathbf{A}$, $\mathbf{B}$, and M_{ik}, and we leave this question open.

We now make use of the representation (6) with $j=\frac{1}{2}$ to study the implications of the invariance of H under subgroups of E(3). The invariance condition $U(g)HU(g)^{-1}=H$ in this case implies

$$V(\mathbf{r})=V(g\mathbf{r}),\quad D^1_{ik}(\Lambda)A_k(\mathbf{r})=A_i(g\mathbf{r}),$$
$$D^1_{ik}(\Lambda)B_k(\mathbf{r})=B_i(g\mathbf{r}),\quad D^1_{ik}(\Lambda)D^1_{lj}(\Lambda)M_{kj}(\mathbf{r})=M_{il}(g\mathbf{r}). \tag{26}$$

The global invariance conditions (26) are again equivalent to the requirement $[H,X_a]=0$, i.e., that H should commute with all generators of the subgroup. The generators in this case are realized as

$$P_i=-i\frac{\partial}{\partial x_i},\quad J_i=L_i+\tfrac{1}{2}\sigma_i,\quad L_i=\epsilon_{ikl}x_kP_l. \tag{27}$$

Let us again consider each subgroup separately.

A. Translations T(1): P_3

Invariance under translations generated by P_3 implies that V, $\mathbf{A}$, $\mathbf{B}$ and M_{ik} all depend on x_1 and x_2 only:

$$V(\mathbf{r})=V(x_1,x_2),\quad A_i(\mathbf{r})=a_i(x_1,x_2),$$
$$B_i(\mathbf{r})=b_i(x_1,x_2),\quad M_{ik}(\mathbf{r})=m_{ik}(x_1,x_2). \tag{28}$$

In addition, if $H=H^+$ we have:

$$b_i(x_1,x_2)=b_i^*(x_1,x_2),\quad m_{ik}(x_1,x_2)=m_{ik}^*(x_1,x_2),$$
$$\mathrm{Im}[v(x_1,x_2)]=-\frac{1}{2}\left(\frac{\partial b_1(x_1,x_2)}{\partial x_1}+\frac{\partial b_2(x_1,x_2)}{\partial x_2}\right), \tag{29}$$
$$\mathrm{Im}[a_i(x_1,x_2)]=-\frac{1}{2}\left(\frac{\partial m_{i1}(x_1,x_2)}{\partial x_1}+\frac{\partial m_{i2}(x_1,x_2)}{\partial x_2}\right).$$

B. Rotations O(2): J_3

Requiring that $[J_3,H]=0$ and using cylindrical coordinates, in which $J_3=-i\partial/\partial\phi+\frac{1}{2}\sigma_3$, we find

$$\frac{\partial A_a}{\partial\phi}-\epsilon_{3ba}A_b=0,\quad \frac{\partial B_a}{\partial\phi}-\epsilon_{3ba}B_b=0,$$
$$\frac{\partial V}{\partial\phi}=0,\quad \frac{\partial M_{ab}}{\partial\phi}-\epsilon_{3kb}M_{ak}-\epsilon_{3ia}M_{ib}=0. \tag{30}$$

These equations can be immediately solved to give

$$V=V(\rho,z),\quad A_1=a_1(\rho,z)\cos\phi-a_2(\rho,z)\sin\phi,$$
$$A_2=a_1(\rho,z)\sin\phi+a_2(\rho,z)\cos\phi, \tag{31}$$
$$A_3=a_3(\rho,z),$$

$$B_1=b_1(\rho,z)\cos\phi-b_2(\rho,z)\sin\phi,$$

$$B_2=b_1(\rho,z)\sin\phi+b_2(\rho,z)\cos\phi,$$

$$B_3=b_3(\rho,z),$$

$$M_{13}=m_1(\rho,z)\cos\phi-m_2(\rho,z)\sin\phi,$$

$$M_{23}=m_1(\rho,z)\sin\phi+m_2(\rho,z)\cos\phi,$$

$$M_{31}=n_1(\rho,z)\cos\phi-n_2(\rho,z)\sin\phi, \tag{32}$$

$$M_{32}=n_1(\rho,z)\sin\phi+n_2(\rho,z)\cos\phi,$$

$$M_{33}=m_3(\rho,z),$$

$$M_{11}=M(\rho,z)\cos2\phi+N(\rho,z)\sin2\phi+D_1(\rho,z),$$

$$M_{22}=-M(\rho,z)\cos2\phi-N(\rho,z)\sin2\phi+D_1(\rho,z),$$

$$M_{12}=-N(\rho,z)\cos2\phi+M(\rho,z)\sin2\phi+D_2(\rho,z),$$

$$M_{21}=-N(\rho,z)\cos2\phi+M(\rho,z)\sin2\phi-D_2(\rho,z).$$

The hermiticity conditions (22) in this case imply that b_i, m_i, n_i, M, N, and D_i are real functions and that

$$\mathrm{Im}\,V=-\frac{1}{2}\left(\frac{\partial b_1}{\partial\rho}+\frac{1}{\rho}b_1+\frac{\partial b_3}{\partial z}\right),$$

$$\mathrm{Im}a_1=-\frac{1}{2}\left(\frac{\partial M}{\partial\rho}+\frac{\partial D_1}{\partial\rho}+\frac{2}{\rho}M+\frac{\partial m_1}{\partial z}\right),$$
$$\mathrm{Im}a_2=-\frac{1}{2}\left(\frac{\partial N}{\partial\rho}+\frac{\partial D_2}{\partial\rho}+\frac{2}{\rho}N-\frac{\partial m_2}{\partial z}\right), \tag{33}$$

$$\mathrm{Im}a_3=-\frac{1}{2}\left(\frac{\partial n_1}{\partial\rho}+\frac{1}{\rho}n_1+\frac{\partial m_3}{\partial z}\right).$$

Putting (32) into (21), we obtain a Hamiltonian that is explicitly O(2) invariant. It can be written, e.g., in terms of O(3) scalars (or pseudoscalars) like $(\mathbf{r}\boldsymbol{\sigma})$, $(\mathbf{r}\mathbf{P})$, $(\boldsymbol{\sigma}\mathbf{L})$, and $(\boldsymbol{\sigma}\mathbf{P})$, the third components of vectors (or axial vectors), like σ_3, P_3, L_3, $(\mathbf{r}\times\boldsymbol{\sigma})_3$, $(\boldsymbol{\sigma}\times\mathbf{P})_3$, and the appropriate components of higher order tensors, like $[M(\rho,z)/\rho^2][2x_1x_2(-\sigma_1P_1+\sigma_2P_2)+(x_1^2-x_2^2)(\sigma_1P_2+\sigma_2P_2)]$, etc.

C. Universal covering group $\overline{\mathrm{O(2)}}$: J_3+aP_3

We use helical coordinates and reduce the problem to that of O(2). The resulting Hamiltonian is given by

formula (21), in which all the functions V, A_i, B_i, M_{ik} are given as in (32) with the replacements

$$\rho \to \rho, \quad z \to v = (1/2a)(z - a\phi), \quad \phi \to u = (1/2a)(z + a\phi). \tag{34}$$

The hermiticity conditions are now somewhat different, namely

$$\begin{aligned}
\mathrm{Im}V &= -\frac{1}{2}\left[\cos v\left(\frac{\partial b_1}{\partial \rho} - \frac{1}{2\rho}\frac{\partial b_2}{\partial v} + \frac{1}{2\rho}b_1\right)\right. \\
&\quad \left. - \sin v\left(\frac{\partial b_2}{\partial \rho} + \frac{1}{2\rho}\frac{\partial b_1}{\partial v} + \frac{1}{2\rho}b_2\right) + \frac{1}{2a}\frac{\partial b_3}{\partial v}\right], \\
\mathrm{Im}a_3 &= -\frac{1}{2}\left[\cos v\left(\frac{\partial n_1}{\partial \rho} - \frac{1}{2\rho}\frac{\partial n_2}{\partial v} + \frac{1}{2\rho}n_1\right)\right. \\
&\quad \left. - \sin v\left(\frac{\partial n_2}{\partial \rho} + \frac{1}{2\rho}\frac{\partial n_1}{\partial v} + \frac{1}{2\rho}n_2\right) + \frac{1}{2a}\frac{\partial m_3}{\partial v}\right], \\
\mathrm{Im}a_1 &= -\frac{1}{2}\left[\left(-\frac{\partial N}{\partial \rho} - \frac{N}{\rho} + \frac{1}{2\rho}\frac{\partial M}{\partial v} - \frac{\partial D_2}{\partial \rho} - \frac{1}{2\rho}\frac{\partial D}{\partial v}\right)\cos v\right. \\
&\quad + \left(\frac{\partial M}{\partial \rho} + \frac{M}{\rho} + \frac{1}{2\rho}\frac{\partial N}{\partial v} - \frac{\partial D_1}{\partial \rho} + \frac{1}{2\rho}\frac{\partial D_2}{\partial v}\right)\sin v \\
&\quad \left. + \frac{1}{2a}\left(m_1 + \frac{\partial m_2}{\partial v}\right)\right], \\
\mathrm{Im}a_2 &= \frac{1}{2}\left[\left(\frac{\partial M}{\partial \rho} + \frac{M}{\rho} - \frac{1}{2\rho}\frac{\partial N}{\partial v} + \frac{\partial D_1}{\partial \rho} - \frac{1}{2\rho}\frac{\partial D_2}{\partial v}\right)\cos v\right. \\
&\quad - \left(-\frac{\partial N}{\partial \rho} - \frac{1}{\rho}N + \frac{1}{2\rho}\frac{\partial M}{\partial v} + \frac{\partial D_2}{\partial \rho} + \frac{1}{2\rho}\frac{\partial D_1}{\partial v}\right)\sin v \\
&\quad \left. - \frac{1}{2a}\left(m_2 - \frac{\partial m_1}{\partial v}\right)\right].
\end{aligned} \tag{35}$$

All arbitrary functions on the right-hand side of (35) depend on ρ and v only.

D. Translations T(2): P_1, P_2

The Hamiltonian is (21) with V, A_i, B_i, and M_{ik} depending on z only. Hermiticity implies

$$\mathrm{Im}[V(z)] = -\frac{1}{2}\frac{dB_3(z)}{dz}, \quad \mathrm{Im}[A_i(z)] = -\frac{1}{2}\frac{dM_{i3}(z)}{dz}. \tag{36}$$

E. Cylindrical group O(2) ⊗ T(1): J_3, P_3

The Hamiltonian is given by (21) with the coefficients as in (32) with the added condition that $a_i, b_i, m_i, n_i, M, N, D_i$, and V depend on ρ only.

F. Translations T(3): P_1, P_2, P_3

The Hamiltonian is given by (21) with all coefficients constant and real.

G. The group of rotations O(3): J_1, J_2, J_3

Making use of some elementary tensor algebra or, equivalently, requiring that say J_1 commutes with the Hamiltonian given by (32), we obtain the general O(3) invariant Hamiltonian:

$$\begin{aligned}
H = &-\tfrac{1}{2}\Delta + V(r) + a(r)(\boldsymbol{\sigma}\mathbf{P}) + b(r)(\boldsymbol{\sigma}\mathbf{r}) \\
&+ c(r)(\mathbf{r}\mathbf{P}) + d(r)(\boldsymbol{\sigma}\mathbf{L}).
\end{aligned} \tag{37}$$

Note that the $a(r)$ and $c(r)$ terms violate parity conservation. Hermiticity implies that $a(r)$, $c(r)$, and $d(r)$ are real and that

$$\mathrm{Im}[V(r)] = -\tfrac{1}{2}[3c(r) + rc'(r)], \quad \mathrm{Im}[b(r)] = -(1/2r)a'(r). \tag{38}$$

H. Euclidean group E(2): J_3, P_1, P_2

The Hamiltonian in this case is obtained by considering the conditions (32) and imposing $[P_1, H] = [P_2, H] = 0$. The result is

$$\begin{aligned}
H = &-\tfrac{1}{2}\Delta + V(z) + a(z)(\boldsymbol{\sigma}\mathbf{P}) + b(z)\sigma_3 + c(z)P_3 \\
&+ d(z)\sigma_3 P_3 + e(z)(\boldsymbol{\sigma}\times\mathbf{P})_3.
\end{aligned} \tag{39}$$

Hermiticity implies that $b(z)$, $c(z)$, $d(z)$, and $e(z)$ are real and

$$\mathrm{Im}[V(z)] = -\tfrac{1}{2}b'(z), \quad \mathrm{Im}[a(z)] = -\tfrac{1}{2}[c'(z) + d'(z)]. \tag{40}$$

I. The group $\overline{E(2)}$: $J_3 + aP_3, P_1, P_2, a \neq 0$

Invariance with respect to $J_3 + aP_3$ gives relations (32) with the substitution (34). Adding the requirement of translational invariance, we find

$$\begin{aligned}
&V = \text{const}, \\
&A_1 = a_1\cos(z/a) - a_2\sin(z/a), \quad A_3 = a_3 \\
&A_2 = a_1\sin(z/a) + a_2\cos(z/a), \\
&B_1 = b_1\cos(z/a) - b_2\sin(z/a), \quad B_3 = b_3 \\
&B_2 = b_1\sin(z/a) + b_2\cos(z/a), \\
&M_{13} = m_1\cos(z/a) - m_2\sin(z/a), \quad M_{33} = m_3 \\
&M_{23} = m_1\sin(z/a) + m_2\cos(z/a), \\
&M_{31} = n_1\cos(z/a) - n_2\sin(z/a), \\
&M_{32} = n_1\sin(z/a) + n_2\cos(z/a), \\
&M_{11} = M\cos(2z/a) + N\sin(2z/a) + D_1, \\
&M_{22} = -M\cos(2z/a) - N\sin(2z/a) + D_1, \\
&M_{12} = -N\cos(2z/a) + M\sin(2z/a) + D_2, \\
&M_{21} = -N\cos(2z/a) + M\sin(2z/a) - D_2,
\end{aligned} \tag{41}$$

where all the coefficients a_i, b_i, m_i, n_i, M, N, and D_i are constants. The hermiticity conditions (22) in this case imply that b_i, n_i, D_i, M, N, V, and a_3 are real and

$$\mathrm{Im}a_1 = m_2/2a, \quad \mathrm{Im}a_2 = -m_1/2a. \tag{42}$$

J. The group E(2) ⊗ T(1): J_3, P_1, P_2, P_3

The conditions $[J_3, H] = [P_1, H] = [P_2, H] = 0$ lead to the Hamiltonian (39). The additional condition $[P_3, H] = 0$ implies that $a, \ldots, e$ in (39) are all real constants.

K. The entire Euclidean group E(3)

Requiring that the Hamiltonian (39) should commute with P_3 and be rotationally invariant (commute with say J_1), we find the Euclidean invariant Hamiltonian

$$H = -\tfrac{1}{2}\Delta + V + c(\sigma \mathbf{P}), \tag{43}$$

where V and c are real constants.

Thus, as in the case of a scalar particle, we find a one-to-one correspondence between subgroups of E(3) and symmetry breaking terms in the Hamiltonian. Again, symmetry under T(3), $\overline{E(2)}$, E(2)$\otimes$T(1) or E(3) specifies H up to some constants. All other symmetry groups allow for the presence of arbitrary functions of one or two variables. Gauge invariance could be added to our requirements to simplify the obtained results, and parity and time reversal invariance would pose further restrictions, which we do not go into.

5. WAVEFUNCTIONS FOR THE SCHRÖDINGER EQUATION WITH AN INVARIANT HAMILTONIAN

We shall now consider each of the Hamiltonians found in Sec. 3 and 4 and show how the invariance with respect to a subgroup of E(3) allows us to obtain solutions of the Schrödinger equation, or at least some properties of these solutions.

TABLE II. Solutions of Schrödinger equation.

No.	Symmetry group	Separable coordinates	Eigenfunctions $\psi(\mathbf{r})$	Additional equations
1	T(1)	Cartesian coordinate z	$\Phi(x,y)\exp(ik_3z)$	$P_3\psi = k_3\psi$
2	O(2)	Cylindrical coordinate ϕ (azimuthal angle)	$\Phi(\rho, z)e^{im\phi}$	$J_3\psi = m\psi$
3	$\overline{O(2)}$	Helical	$\Phi(\rho,v)e^{i\omega u}$	$(J_3 + aP_3)\psi = \omega\psi$
4	T(2)	Cartesian	$\Phi(z)\exp[i(k_1x + k_2y)]$	$P_1\psi = k_1\psi$, $P_2\psi = k_2\psi$ $\Phi''(z) + 2iA_3(z)\Phi'(z) - 2[V(z) - E + \frac{1}{2}(k_1^2 + k_2^2) + A_1(z)k_1 + A_2(z)k_2]\times\Phi(z) = 0$
5	O(2)$\otimes$T(1)	Cylindrical	$\Phi(\rho)e^{im\phi}\exp(ik_3z)$	$J_3\psi = m\psi$, $P_3\psi = k_3\psi$ $\Phi''(\rho) + [1/\rho + 2i\rho a(\rho)]\Phi'(\rho) - 2[V(\rho) - E + m^2/2\rho^2 + k_3^2/2 + b(\rho)k_3 + c(\rho)m]\Phi(\rho) = 0$
6	O(3)	a. Spherical	$\Phi(r)Y_{lm}(\theta,\phi)$	$J^2\psi = l(l+1)\psi$, $J_3\psi = m\psi$
		b. Spheroconical	$\Phi(r)\Lambda_{lh'}(\alpha)\Lambda_{lh}(\beta)$	$J^2\psi = l(l+1)\psi$, $(J_1^2 + r^2J_2^2)\psi = \kappa^2\psi$ $h = \kappa^2/4$, $h' = J(J+1) - \kappa^2/4$ a. and b.: $\Phi''(r) + 2[1/r + ia(r)r]\Phi'(r) - 2[V(r) - E + l(l+1)/2r^2]\Phi(r) = 0$
7	E(2)	a. Cartesian	$\Phi(z)\exp[i(k_1x + k_2y)]$	$(P_1^2 + P_2^2)\psi = \kappa^2\psi$, $P_i\psi = k_i\psi$, $i = 1,2$, $k_1^2 + k_2^2 = \kappa^2$
		b. Cylindrical	$\Phi(z)J_m(\kappa\rho)e^{im\phi}$	$(P_1^2 + P_2^2)\psi = \kappa^2\psi$, $J_3\psi = m\psi$
		c. Parabolic	$\Phi(z)D_{-i\omega^2/(2\kappa)-1/2}(i\kappa\xi)\times D_{i\omega^2/(2\kappa)-1/2}(i\kappa\eta)$	$(P_1^2 + P_2^2)\psi = \kappa^2\psi$, $(J_3P_2 + P_2J_3)\psi = \omega^2\psi$
		d. Elliptic cylinder	$\Phi(z)\begin{cases}\mathrm{Ce}_n(\alpha,q)\mathrm{ce}_n(\beta,q)\\ \mathrm{Se}_n(\alpha,q)\mathrm{se}_n(\beta,q)\end{cases}$ $q = d^2\kappa^2/4$	$(P_1^2 + P_2^2)\psi = \kappa^2\psi$, $(J_3^2 + d^2P_1^2)\psi = \omega_n^2\psi$ a., b., c. and d.: $\Phi''(z) + 2ia(z)\Phi'(z) - 2[V(z) - E + \frac{1}{2}\kappa^2]\Phi(z) = 0$
8	$\overline{E(2)}$	Cartesian	$e^{-ihz}Z(z/2a + \alpha/2)\exp[i(k_1x + k_2y)]$	$P_i\psi = k_i\psi$, $i = 1,2$ $Z''(\xi) + 4a^2(A - D + 2D\cos^2\xi)Z(\xi) = 0$ (Mathieu equation) $A = 2(E - V) - k_1^2 - k_2^2$, $D\cos\alpha = -fk_1 + gk_2$, $D\sin\alpha = fk_2 + gk_1$
9	T(3)	Cartesian	$e^{i\mathbf{kr}}$	$P_i\psi = k_i\psi$, $i = 1,2,3$ $\mathbf{k}^2 + V + \mathbf{Ak} = E$
10	E(2)$\otimes$T(1)	a. Cartesian	$e^{i\mathbf{kr}}$	$(P_1^2 + P_2^2)\psi = \kappa^2\psi$, $P_i\psi = k_i\psi$, $i = 1,2,3$, $k_1^2 + k_2^2 = \kappa^2$
		b. Cylindrical	$J_m(\kappa\rho)^{im\phi}e^{ik_3z}$	$(P_1^2 + P_2^2)\psi = \kappa^2\psi$, $J_3\psi = m\psi$, $P_3\psi = k_3\psi$
		c. Parabolic cylinder	$D_{-i\omega^2/(2\kappa)-1/2}(i\kappa\xi)\times D_{i\omega^2/(2\kappa)-1/2}(i\kappa\eta)e^{ik_3z}$	$(P_1^2 + P_2^2)\psi = \kappa^2\psi$, $(J_3P_2 + P_2J_3)\psi = \omega^2\psi$, $P_3\psi = k_3\psi$
		d. Elliptic cylinder	$\left.\begin{matrix}\mathrm{Ce}_n(\alpha,q)\mathrm{ce}_n(\beta,q)\\ \mathrm{Se}_n(\alpha,q)\mathrm{se}_n(\beta,q)\end{matrix}\right\} e^{ik_3z}$	$(P_1^2 + P_2^2)\psi = \kappa^2\psi$, $(J_3^2 + d^2P_1^2)\psi = \omega_n^2\psi$, $P_3\psi = k_3\psi$, $E = \frac{1}{2}(\kappa^2 + k_3^2) + V + ak_3$
11	E(3)	See text and articles on Helmholtz equation.		

TABLE III. Separable coordinates, diagonal operators and eigenfunctions in Table II.

Coordinates	Operators	Eigenfunctions
Cartesian: x, y, z	P_i	Exponentials
Cylindrical: $x=\rho\cos\phi$, $y=\rho\sin\phi$, z	J_3 and $P_1^2+P_2^2$ or P_3	Bessel and exp
Parabolic cylinder: $x=\frac{1}{2}(\xi^2-\eta^2)$, $y=\xi\eta$, z	$J_3P_2+P_2J_3$ and $P_1^2+P_2^2$	Parabolic cylinder[13]
Elliptic cylinder: $x=d\cosh\alpha\cos\beta$, $y=d\sinh\alpha\sin\beta$, z	$J_3^2+dP_1^2$, $P_1^2+P_2^2$	Mathieu and associate Mathieu[13]
Spherical: $x=r\sin\theta\cos\phi$, $y=r\sin\theta\sin\phi$, $z=r\cos\theta$	$\mathbf{J}^2$, J_3	Spherical harmonics
Conical: $x=r\mathrm{sn}(\alpha,k)\mathrm{dn}(\beta,k')$, $y=r\mathrm{cn}(\alpha,k)\mathrm{cn}(\beta,k')$, $z=\mathrm{dn}(\alpha,k)\mathrm{sn}(\beta,k')$ (Jacobian elliptic functions)	$\mathbf{J}^2$, $J_1^2+r^2J_2^2$	Lamé polynomials[14]
Helical: $x=\rho\cos(u-v)$, $y=\rho\sin(u-v)$, $z=a(u+v)$ (nonorthogonal; partial separation)	$\{J_3+aP_3\}$	{Exponential}

A. Scalar particles

Let us consider the Schrödinger equation

$$[-\tfrac{1}{2}\Delta + V(\mathbf{r}) + \mathbf{A}(\mathbf{r})\mathbf{P}]\psi = E\psi \tag{44}$$

and specify the interactions $V(\mathbf{r})$ and $\mathrm{A}(\mathbf{r})$, following Table I.

First of all, if (44) is invariant with respect to E(3), then $\mathbf{A}=0$, $V(\mathbf{r})=V=\text{const}$, and (44) reduces to the Helmholtz equation. It is well known that the Helmholtz equation allows the complete separation of variables in 11 orthogonal systems of coordinates.[11]

It has been shown that the separable coordinates can be characterized by the fact that the separated eigenfunctions of the Laplace operator are simultaneously eigenfunctions of a pair of second order commuting operators in the enveloping algebra of the Lie algebra of E(3). All such pairs were classified into equivalence classes under E(3) and a one-to-one correspondence between each class and a separable coordinate system was established.[12] The eigenfunctions and many of their group theoretic properties have been studied in detail elsewhere.[13]

Here we shall partially extend this group theoretical approach to Schrödinger equations with "subgroup invariant" potentials. We proceed to discuss the individual cases and summarize the situation in Table II.

The one-dimensional invariance groups generated by P_3, J_3 and J_3+aP_3, respectively, only provide one operator, that can be diagonalized. Hence they only lead to a partial separation of variables. The diagonalization of a first order operator leads to an exponential dependence on the appropriate variable. The remaining two variables (see Table II) are contained in a function Φ satisfying a partial differential equation that can be obtained directly from the Schrödinger equation.

Invariance under a two-dimensional or higher-dimensional group implies the complete separation of variables in at least one coordinate system. Invariance under T(2), O(2)⊗T(1), O(3), and E(2) does not specify the Hamiltonian completely. It does, however, completely specify the dependence of the wavefunctions on two variables. The dependence on the third variable is contained in a function Φ, satisfying an ordinary differential equation, specified in Table II. Invariance under T(3), $\overline{\mathrm{E}(2)}$, E(2)⊗T(1), or E(3) specifies H and $\psi(\mathbf{r})$ completely. The wavefunctions, etc., are given in Table II, with some explanations in Table III. For more details on eigenfunctions involving Mathieu functions, parabolic cylinder functions, Lamé polynomials, etc., we refer to the literature.[14,15]

B. Spinor particles

The equation to be considered in this case is the Pauli equation

$$[-\tfrac{1}{2}\Delta + V(\mathbf{r}) + \mathbf{A}(\mathbf{r})\sigma + \mathbf{B}(\mathbf{r})\mathbf{P} + M_{ik}(\mathbf{r})\sigma_iP_k]\psi = E\psi, \tag{45}$$

in which all the functions V, A_i, B_i, and M_{ik} are to be specified so as to correspond to the considered invariance group, as listed in Sec. 4. Separation of variables in spinor equations has not been investigated with the same amount of detail as in scalar ones. While we consider this to be a very worthwhile project, we do not go into it here and for each interaction only consider the simplest types of separated solutions.

This time we shall proceed from the higher dimensional groups to the lower ones.

1. *The group* E(3): We have

$$[-\tfrac{1}{2}\Delta + V + a(\sigma\mathbf{P})]\psi = E\psi, \tag{46}$$

with V and a real constants. Requiring that

$$P_i\psi = k_i\psi, \tag{47}$$

we find

$$\psi(\mathbf{r}) = \begin{pmatrix}\mu\\ \nu\end{pmatrix} e^{i\mathbf{k}\mathbf{r}}, \tag{48}$$

i.e., a constant spinor times an exponential. Substituting (48) into (46), we find

$$E_\pm = \tfrac{1}{2}\mathbf{k}^2 + V \pm a|k| \tag{49}$$

and

$$\nu = -[(k_3 \mp |k|)/(k_1 - ik_2)]\mu. \tag{50}$$

2. *The group* E(2)⊗T(1): We have

$$[-\tfrac{1}{2}\Delta + V + a(\sigma\mathbf{P}) + b\sigma_3 + cP_3 + d\sigma_3P_3 + e(\sigma_1P_2 - \sigma_2P_1)]\psi = E\psi \tag{51}$$

and again add the conditions (47), leading to separation in Cartesian coordinates. The wavefunction is again of the form (48); however,

$$E_{\pm}=R\pm(S^2+T^2+U^2)^{1/2} \tag{52}$$

$$\nu=-\{[S\mp(S^2+T^2+U^2)^{1/2}]/(T+iU)\}\mu, \tag{53}$$

where

$$R=\tfrac{1}{2}\mathbf{k}^2+V+ck_3,\quad T=ak_1+ek_2,$$
$$S=(a+d)k_3+b,\quad U=-ak_2+ek_1. \tag{54}$$

If we replace the conditions $P_i\psi=k_i\psi$ by

$$P_3\psi=k_3\psi,\quad J_3\psi=m\psi,\quad (P_1^2+P_2^2)\psi=\kappa^2\psi, \tag{55}$$

we are led to cylindrical coordinates and obtain

$$\psi=\begin{pmatrix}\mu\, J_{m-1/2}(\kappa\rho)\exp\{i[(m-\frac{1}{2})\phi+k_3z]\}\\ \nu\, J_{m+1/2}(\kappa\rho)\exp\{i[(m+\frac{1}{2})\phi+k_3z]\}\end{pmatrix}, \tag{56}$$

with

$$E_{1,2}=\tfrac{1}{2}(\kappa^2+k_3^2)+V+ck_3$$
$$\pm\{[(a+d)k_3+b]^2+(a^2+e^2)\kappa^2\}^{1/2}, \tag{57}$$

$$\nu=\frac{1}{(ia-e)\kappa}((a+d)k_3+b\mp\{[(a+d)k_3+b]^2$$
$$+(a^2+e^2)\kappa^2\}^{1/2}). \tag{58}$$

3. *The group* T(3): We have

$$(-\tfrac{1}{2}\Delta+V+\mathbf{A}\sigma+\mathbf{B}\mathbf{P}+M_{ik}\sigma_iP_k)\psi=E\psi. \tag{59}$$

Adding the separation conditions (47), we obtain the wavefunction ψ in the form (48) with $E_{1,2}$, ν, and μ satisfying (52) and (53). Equations (54) defining R, S, T, and U are replaced by

$$R=\tfrac{1}{2}\mathbf{k}^2+V+\mathbf{B}\mathbf{k},\quad T=A_1+M_{1j}k_j,$$
$$S=A_3+M_{3j}k_j,\quad U=-A_2-M_{2j}k_j. \tag{60}$$

4. *The group* $\overline{\mathrm{E}(2)}$: The invariant Schrödinger–Pauli equation can in this case be written as

$$\Big\{-\frac{1}{2}\Delta+A\left(\cos\frac{z}{a}\sigma_1+\sin\frac{z}{a}\sigma_2\right)$$
$$+B\left(\sin\frac{z}{a}\sigma_1-\cos\frac{z}{a}\sigma_2\right)+C\sigma_3+V$$
$$+D\left(\cos\frac{z}{a}P_1+\sin\frac{z}{a}P_2\right)+E\left(\sin\frac{z}{a}P_1-\cos\frac{z}{a}P_2\right)$$
$$+FP_3+G\left[-\sin\frac{2z}{a}(\sigma_1P_1-\sigma_2P_2)+\cos\frac{2z}{a}(\sigma_1P_2+\sigma_2P_1)\right]$$
$$+H\left[\cos\frac{2z}{a}(\sigma_1P_1-\sigma_2P_2)+\sin\frac{2z}{a}(\sigma_1P_2+\sigma_2P_1)\right]$$
$$+J(\sigma\mathbf{P})+K\sigma_3P_3+L(\sigma\times\mathbf{P})_3+M\left(\cos\frac{z}{a}\sigma_1+\sin\frac{z}{a}\sigma_2\right)P_3$$
$$+N\left(\sin\frac{z}{a}\sigma_1-\cos\frac{z}{a}\sigma_2\right)P_3+P\left(\cos\frac{z}{a}P_1+\sin\frac{z}{a}P_2\right)\sigma_3$$
$$+Q\left(\sin\frac{z}{a}P_1-\cos\frac{z}{a}P_2\right)\sigma_3\Big\}\psi=E\psi. \tag{61}$$

Adding the conditions

$$P_i\psi=k_i\psi,\quad i=1,2, \tag{62}$$

we find

$$\psi=\begin{pmatrix}\mu(z)\\ \nu(z)\end{pmatrix}\exp[i(k_1x+k_2y)]. \tag{63}$$

Substituting (63) back into (61), we obtain a pair of coupled ordinary differential equations for $\mu(z)$ and $\nu(z)$,

$$\Big[-\frac{1}{2}\frac{d^2}{dz^2}-i(F+J+K)\frac{d}{dz}-\frac{k_1^2+k_2^2}{2}$$
$$+C+V-E+(D+P)\left(k_1\cos\frac{z}{a}+k_2\sin\frac{z}{a}\right)$$
$$+(E+Q)\left(k_1\sin\frac{z}{a}-k_2\cos\frac{z}{a}\right)\Big]\mu(z)$$
$$+\Big\{-iMe^{-iz/a}\frac{d}{dz}+Ae^{-iz/a}+Be^{iz/a}$$
$$+\exp(-i2z/a)[G(k_2-ik_1)+H(k_1+ik_2)]$$
$$+J(k_1-ik_2)+L(k_1+ik_2)\Big\}\nu(z)=0, \tag{64}$$

$$\Big[-\frac{1}{2}\frac{d^2}{dz^2}-i(F-J-K)\frac{d}{dz}-\frac{k_1^2+k_2^2}{2}$$
$$-C+V-E+(D-P)\left(k_1\cos\frac{z}{a}+k_2\sin\frac{z}{a}\right)$$
$$+(E-Q)\left(k_1\sin\frac{z}{a}-k_2\cos\frac{z}{a}\right)\Big]\nu(z)$$
$$+\Big\{-iMe^{iz/a}\frac{d}{dz}+Ae^{iz/a}+Be^{-iz/a}$$
$$+e^{i2z/a}[G(k_2+ik_1)+H(k_1-ik_2)]+J(k_1+ik_2)$$
$$+L(k_1-ik_2)\Big\}\mu(z)=0.$$

Judging from the spinless case, it should be possible to express their solutions in terms of a generalization of Mathieu functions, but we have not explored this possibility. It is a simple matter to decouple the equations, at the price of introducing third derivatives of $\mu(z)$ and $\nu(z)$.

5. *The group* E(2): We have

$$[-\tfrac{1}{2}\Delta+V(z)+a(z)(\sigma\mathbf{P})+b(z)\sigma_3+c(z)P_3$$
$$+d(z)\sigma_3P_3+e(z)(\sigma_1P_2-\sigma_2P_1)]\psi=E\psi. \tag{65}$$

Adding the conditions (62), we obtain a solution of the form (63), where μ and ν satisfy the coupled equations

$$\left(-\frac{1}{2}\frac{d^2}{dz^2}-i(a+c+d)\frac{d}{dz}-\frac{k_1^2+k_2^2}{2}+V+b-E\right)\mu(z)$$
$$+[k_1(a+ie)+k_2(e-ia)\nu(z)]=0,$$
$$\left(-\frac{1}{2}\frac{d^2}{dz^2}+i(-a+c+d)\frac{d}{dz}+V-b-E\right)\nu(z) \tag{66}$$
$$+[k_1(a-ie)+k_2(a+ie)]\mu(z)=0.$$

Since $a, \ldots, e$ are arbitrary functions of z, we can proceed no further. We could equally well separate in cylindrical or possibly other coordinates.

6. *The group* O(3): We have

$$[-\tfrac{1}{2}\Delta + V(r) + a(r)(\boldsymbol{\sigma}\mathbf{P}) + b(r)(\boldsymbol{\sigma}\mathbf{r}) + c(r)(\mathbf{r}\mathbf{P}) + d(r)(\boldsymbol{\sigma}\mathbf{L})]\psi = E\psi. \tag{67}$$

In addition we require

$$\mathbf{J}^2\psi = J(J+1)\psi, \quad J_3\psi = M\psi. \tag{68}$$

The solution of (68) can be written in the form

$$\begin{aligned}\psi &\equiv \psi_{JM}(r,\theta,\phi) \\ &= \sum_{\kappa=\pm 1/2} \alpha_{J+\kappa}(r) \sum_{\mu=\pm 1/2} (J+\kappa\ \tfrac{1}{2}\ M-\mu\,\mu|JM) Y_{J+\kappa,M-\mu}(\theta,\phi)\chi_{\mu/2} \\ &\equiv \sum_{\kappa=\pm 1/2} \alpha_{J+\kappa}(r) W_{JM}^{J+\kappa}(\theta,\phi),\end{aligned} \tag{69}$$

where $(l_1 l_2 m_1 m_2 | LM)$ denotes an O(3) Clebsch–Gordan coefficient,[16] $Y_{JM}(\theta,\phi)$ a spherical harmonic, $\chi_{\mu/2}$ a constant two-component spinor, and $\alpha_{J\pm 1/2}(r)$ are arbitrary functions of r. The following relations are consequences of elementary angular momentum theory[16,17]:

$$\begin{aligned}&(\boldsymbol{\sigma}\mathbf{L})W_{JM}^{J+\kappa} = 2\kappa(J+\tfrac{1}{2}-2\kappa)W_{JM}^{J-\kappa}, \\ &(\mathbf{r}\mathbf{P})\alpha_{J+\kappa}(r)W_{JM}^{J+\kappa} = -ir\frac{d\alpha_{J+\kappa}(r)}{dr}W_{JM}^{J+\kappa}, \\ &\frac{(\boldsymbol{\sigma}\mathbf{r})}{r}\alpha_{J+\kappa}^{(r)}W_{JM}^{J+\kappa} = \alpha_{J+\kappa}^{(r)}W_{JM}^{J-\kappa}, \\ &(\boldsymbol{\sigma}\mathbf{P})\alpha_{J+\kappa}W_{JM}^{J+\kappa} = -i\frac{d\alpha_{J+\kappa}(r)}{dr}W_{JM}^{J-\kappa} \\ &\qquad -2\kappa\frac{i}{r}(J+\tfrac{1}{2}+2\kappa)\alpha_{J+\kappa}(r)W_{JM}^{J-\kappa}(r).\end{aligned} \tag{70}$$

Substituting (69) into (67) and using (70), we obtain two coupled equations for $\alpha_{J\pm 1/2}(r)$:

$$\begin{aligned}&\left\{-\frac{1}{2}\left[\frac{d^2}{dr^2}+\left(\frac{2}{r}-irc(r)\right)\frac{d}{dr}+\frac{(J+1/2)(J+3/2)}{r^2}\right]\right. \\ &\left.+V(r)-(J+\tfrac{3}{2})d(r)-E\right\}\alpha_{J+1/2}(r) \\ &+\left(-ia(r)\frac{d}{dr}-ia(r)+b(r)\right)\alpha_{J-1/2}(r)=0, \\ &\left\{-\frac{1}{2}\left[\frac{d^2}{dr^2}+\left(\frac{2}{r}-irc(r)\right)\frac{d}{dr}+\frac{(J-1/2)(J+1/2)}{r^2}\right]\right. \\ &\left.+V(r)+(J-\tfrac{1}{2})d(r)-E\right\}\alpha_{J-1/2}(r) \\ &+\left(-ia(r)\frac{d}{dr}-ia(r)+b(r)\right)\alpha_{J+1/2}(r)=0.\end{aligned} \tag{71}$$

7. *The group* O(2)⊗T(1): The Pauli equation in this case is

$$\begin{aligned}&\{-\tfrac{1}{2}\Delta + V(\rho) - E + A(\rho)(x\sigma_1 + y\sigma_2) + B(\rho)(x\sigma_2 - y\sigma_1) + C(\rho)\sigma_3 \\ &+ D(\rho)(xP_1 + yP_2) + E(\rho)L_3 + F(\rho)P_3 + G(\rho)(x\sigma_1 + y\sigma_2)P_3 \\ &+ H(\rho)(x\sigma_2 - y\sigma_1)P_3 + J(\rho)\sigma_3 P_3 + K(\rho)(xP_1 + yP_2)\sigma_3 \\ &+ L(\rho)\sigma_3 L_3 + M(\rho)[(x^2-y^2)(\sigma_1 P_1 - \sigma_2 P_2) + 2xy(\sigma_1 P_2 + \sigma_2 P_1)] \\ &+ N(\rho)[2xy(\sigma_1 P_1 - \sigma_2 P_2) - (x^2-y^2)(\sigma_1 P_2 - \sigma_2 P_1)] \\ &+ P(\rho)(\sigma_1 P_1 + \sigma_2 P_2) + Q(\rho)(\sigma_1 P_2 - \sigma_2 P_1)\}\psi(\rho,z,\phi) = 0.\end{aligned} \tag{72}$$

Adding the conditions

$$J_3\psi = (-i\frac{\partial}{\partial\phi} + \tfrac{1}{2}\sigma_3)\psi = M\psi, \quad P_3\psi = k_3\psi, \tag{73}$$

we find

$$\psi = \begin{pmatrix}\mu(\rho)\exp[i(M-1/2)\phi]\exp(ik_3 z) \\ \nu(\rho)\exp[i(M+1/2)]\exp(ik_3 z)\end{pmatrix}. \tag{74}$$

Substituting (74) into (72), we obtain a pair of coupled ordinary second order differential equations for $\mu(\rho)$ and $\nu(\rho)$.

8. *The group* T(2): We have

$$[-\tfrac{1}{2}\Delta + V(z) - E + \mathbf{A}(z)\boldsymbol{\sigma} + \mathbf{B}(z)\mathbf{P} + M_{ik}(z)\sigma_i P_k]\psi = 0. \tag{75}$$

Adding the conditions

$$P_i\psi = k_i\psi, \quad i=1,2 \tag{76}$$

we find

$$\psi = \begin{pmatrix}\mu(z) \\ \nu(z)\end{pmatrix}\exp[i(k_1 x + k_2 y)]. \tag{77}$$

Substituting (77) in (75), we obtain a pair of ordinary coupled differential equations for $\mu(z)$ and $\nu(z)$.

9. *One-dimensional groups* T(1), O(2), and $\overline{\mathrm{O(2)}}$: The Hamiltonians are given in Sec. 4. The diagonalization of the generator P_3, J_3, or $J_3 + aP_3$ leads to a partial separation of variables, as in the case of scalar particles.

Thus, $P_3\psi = k_3\psi$ implies

$$\psi = \begin{pmatrix}\mu(x,y) \\ \nu(x,y)\end{pmatrix}\exp(ik_3 z), \tag{78}$$

$J_3\psi = M\psi$ implies

$$\psi = \begin{pmatrix}\mu(\rho,z)\exp[i(M-1/2)\phi] \\ \nu(\rho,z)\exp[i(M+1/2)\phi]\end{pmatrix} \tag{79}$$

and $(J_3 + aP_3)\psi = \omega\psi$ implies (in helical coordinates)

$$\psi = \begin{pmatrix}\mu(\rho,v)\exp[i(M-1/2)u] \\ \nu(\rho,v)\exp[i(M+1/2)u]\end{pmatrix}. \tag{80}$$

The coupled partial differential equations for the functions μ and ν can be obtained in each case by substituting back into the appropriate Pauli equation.

To summarize: While the solutions are considerably more cumbersome for spinor particles than for scalar ones, the over-all picture is the same. Thus, invariance under a one-dimensional group leads to a partial separation of variables. Invariance under higher-dimensional group leads to a complete separation of variables and to the explicit form of the dependence on

at least two of the variables. Invariance under $\overline{E(2)}$, T(3), $E(2)\otimes T(3)$, or E(3) in principle completely specifies the wavefunctions.

If we wish to proceed further with a study of the wavefunctions and energy levels we must specify the problem further, i.e., impose appropriate boundary conditions, specify the arbitrary functions in the potentials, etc.

6. CONCLUSIONS

The main thesis of this article is that the complete knowledge of the subgroup structure of the invariance group of a physical system makes possible a systematic study of symmetry breaking for this system. In particular, it makes it possible to generate all related systems, the symmetry group of which is a subgroup of the original group. Invariance with respect to the subgroup can then be used to study some or all of the properties of the new more general problem.

In this article we have demonstrated the validity of the above thesis in the case of the nonrelativistic quantum mechanics of a scalar or spinor particle. In particular, in the considered case there is actually a one-to-one correspondence between subgroups of the symmetry group E(3) of the free Schrödinger (or Pauli) equation and symmetry breaking interactions of the considered type. Let us mention that if we had considered only a spinless particle in a scalar potential $V(\mathbf{r})$, the relation between potentials and subgroups would not have been one-to-one.

We have restricted ourselves, on one hand, to nonrelativistic stationary (time independent) quantum mechanics of a single particle with spin 0 or $\frac{1}{2}$ in an external field, subject to a quite specific ansatz (namely that the interaction is linear in the momentum). On the other hand, we have restricted ourselves to explicit symmetry breaking, i.e., modifications of the Hamiltonian, without considering the possibility of "spontaneous" symmetry breaking.

Clearly all the above restrictions can be abandoned and the approach of this article applied to more general problems. Trivial (conceptually) generalizations would be to consider particles of higher spin, possibly of arbitrary spin and a more general ansatz, concerning the Hamiltonian. Let us just mention several other extensions of the present work, that are currently being considered, as well as related work by other groups.

1. Symmetries of the time-dependent Schrödinger equation. This equation for a free particle is known to be invariant under a considerably larger group than the Euclidean one, namely the so-called Schrödinger group S_n (n is the number of spatial dimensions), containing the Galilei group as a subgroup.[18] Many properties of this group have been studied.[18–20] In particular, in the case of one space dimension all subgroups of S_1 have been found and the problem of finding the interactions, reducing the symmetry from S_1 to each of its subgroups has been solved.[19] The ansatz was general enough to include nonlinear interactions and indeed many nonlinear generalizations of the Schrödinger equation were obtained and at least partially solved. In this connection we would like to mention that Lie theory provides powerful tools for studying nonlinear equations. Similar work for the Schrödinger groups S_2 and S_3 is in progress.

2. The complete symmetry group for a given Hamiltonian. The invariance group of a Hamiltonian is not necessarily a subgroup of the invariance group of the free Hamiltonian. Indeed, some of the Hamiltonians found in this article may in general or in special cases have larger symmetry groups. The additional generators will not be linear combinations of P_i and J_i.

3. Groups of canonical transformations. The symmetry group considered in this article is a purely geometrical one reflecting properties of space–time rather than any special dynamics. This reflects itself in the fact that all the group generators are first order linear operators. Dynamical invariance groups, corresponding to general canonical transformations, like, e.g., O(4) for the hydrogen atom or SU(3) for the harmonic oscillator would make their appearance if we considered generators that are second or higher order operators.[12,21] The problem of symmetry breaking for groups of canonical transformations is in itself of interest, e.g., in connection with a hydrogen type atom in an external field[22] and more generally in connection with interactions removing accidental degeneracy.[23,24] In general the consideration of higher orders operators as group generators may lead away from Lie groups and to the use of, e.g., Bäcklund transformations.[25]

4. Explicit symmetry breaking in relativistic theory. One of the basic assumptions of relativistic quantum theory is that the state vectors of a free (noninteracting) elementary physical system transforms according to an irreducible unitary representation of the Poincaré group.[26] Alternatively and equivalently, the state vector obeys the Bargmann–Wigner equations.[27] Now consider a particle in a classical external field—in itself a problem of considerable interest.[28] The external field will violate Poincaré invariance and depending on its own symmetry reduce the invariance group of the system to a subgroup of the Poincaré group. Thus, similarly as in the nonrelativistic case, a classification of the subgroups of the Poincaré group will provide us with a classification of symmetry breaking external fields. The representation theory of the corresponding subgroup will then provide us with a handle for studying the corresponding relativistic equation for a particle in a field.

Related problems have been treated in the literature. The subgroups of the Poincaré group have been classified in Ref. 2 and independently by Bacry *et al.*[29] Some of the subgroups have been used in a theoretical analysis of elementary particles in an external electromagnetic field[30] and of electromagnetic fields with certain symmetries.[31] An interesting series of papers has been devoted to symmetries of electromagnetic fields making use of certain types of discrete subgroups of the Poincaré group.[32]

5. Spontaneous symmetry breaking.[33–37] This is a different mechanism for breaking a given symmetry in that the Lagrangian (or in our case the Hamiltonian)

of the system remains fully invariant with respect to the original symmetry group. The symmetry is broken by the fact that the vacuum state is not invariant under the group, but only under a certain subgroup. The relevance of a subgroup classification in this case is obvious. Spontaneous symmetry breaking plays an important role in particle physics and also in solid state theory and other fields. As a model of this type of symmetry breaking we intend to perform a systematic study of various differential equations of physics with nontrivial invariance groups. To these equations we shall add symmetry breaking boundary conditions and see how a classification of subgroups leads to a classification of boundary conditions. The heat equation (the classical diffusion equation) is presently under consideration from this point of view—it is a convenient candidate, since its symmetry group[38,39] is isomorphic to the Schrödinger group, the subgroups of which we already know.[19]

ACKNOWLEDGMENTS

The authors are much indebted to Professor C.P. Boyer, Professor W. Miller, Jr., Professor R.T. Sharp, and Professor H. Zassenhaus for helpful discussions. One of the authors (J.B.) thanks the Centre de recherches mathématiques, Université de Montréal, where the present study was performed, for its hospitality. A NATO grant making this stay possible is much appreciated.

*Service de Physique Mathématique, Université de Liège, Belgium.

†Work partially supported by research grants from NATO and National Research Council of Canada.

[1]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 15, 1378, 1932 (1974).
[2]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 16, 1597, 1615 (1975).
[3]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 17, 717 (1976).
[4]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. 17, 977, 986 (1976).
[5]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, Can. J. Phys. 54, 950 (1976).
[6]E.G. Kalnins, J. Patera, R.T. Sharp, and P. Winternitz, Phys. Rev. D 8, 2552 (1973).
[7]W. Miller Jr., *Lie Theory and Special Functions* (Academic, New York, 1968),
[8]W. Miller Jr., Comm. Pure Appl. Math. 18, 527 (1964).
[9]A.S. Wightman, Rev. Mod. Phys. 34, 845 (1962).
[10]N.Ya. Vilenkin, *Special Functions and the Theory of Group Representations* (Amer. Math. Soc., Providence, R.I., 1968).
[11]L.P. Eisenhart, Phys. Rev. 74, 87 (1948); Ph.M. Morse and H. Feshbach, *Methods of Theoretical Physics, Part* 1 (McGraw-Hill, New York, 1953); P. Moon and D.E. Spencer, *Field Theory Handbook* (Springer-Verlag, Berlin, 1961).
[12]A.A. Makarov, Ya.A. Smorodinskiĭ, Kh. Valiev, and P. Winternitz, Nuovo Cimento 52 A, 1061 (1967).
[13]C.P. Boyer, E.G. Kalnins, and W. Miller, Jr., Nagoya Math. J. 60, 35 (1976).
[14]J. Patera and P. Winternitz, J. Math. Phys. 8, 1130 (1973); J. Chem. Phys. 65, October (1976).
[15]I. Fris̆ and P. Winternitz, Yad. Fiz. 1, 889 (1965) [Sov. J. Nucl. Phys. 1, 636 (1965)].
[16]D.M. Brink and G.R. Satchler, *Angular Momentum* (Clarendon, Oxford, 1968).
[17]M.E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).
[18]U. Niederer, Helv. Phys. Acta, 46, 191 (1973); C.P. Boyer, Helv. Phys. Acta 47, 589 (1974); R.L. Anderson, S. Kumei, and C.E. Wulfman, Rev. Mex. Fis. 21, 1, 35 (1972); J. Math. Phys. 14, 1527 (1973); E.G. Kalnins and W. Miller Jr., J. Math. Phys. 15, 1728 (1974).
[19]C.P. Boyer, R.T. Sharp, and P. Winternitz, J. Math. Phys. 17, 1439 (1976).
[20]G. Burdet, M. Perrin, and P. Sorba, J. Math. Phys. 15, 1436, 2253 (1974). G. Burdet and M. Perrin, J. Math. Phys. 16, 1692, 2172 (1975); Lett. Nuovo Cimento 4, 651 (1972).
[21]P. Winternitz, Ya.A. Smorodinskiĭ, M. Uhlíř, and I. Fris̆, Yad. Fiz. 4, 625 (1966) [Sov. J. Nucl. Phys. 4, 444 (1967)].
[22]E.G. Kalnins, W. Miller Jr., and P. Winternitz, SIAM J. Appl. Math. 30, 630 (1976).
[23]J.D. Louck, M. Moshinsky, and K.B. Wolf, J. Math. Phys. 14, 692, 696 (1973).
[24]M. Moshinsky, J. Patera, and P. Winternitz, J. Math. Phys. 16, 82 (1975). M. Moshinsky and J. Patera, J. Math. Phys. 16, 1866 (1975).
[25]R.L. Anderson and N.H. Ibragimov, "Lie—Bäcklund Transformations" (Preprint 1975 and talk at International Symposium on Mathematical Physics, Mexico City, 1976); S. Kumei, J. Math. Phys. 16, 2461 (1975).
[26]E. Wigner, Ann. Math. 40, 149 (1939).
[27]V. Bargmann and E. Wigner, Proc. Nat. Acad. Sci. USA 34, 211 (1948).
[28]W.E. Thirring, *Principles of Quantum Electrodynamics* (Academic, New York, 1958); A.S. Wightman, in *Proceeding of Fifth Coral Gables Conference* on *Symmetres at High Energy, Univ. of Miami* (Gordon and Breach, New York, 1968); B. Schroer, R.Seiler, and J. A. Swieca, Phys. Rev. D 2, 2927 (1970); J. Bellissard, Commun. Math. Phys. 41, 235 (1975).
[29]H. Bacry, Ph. Combe, and P. Sorba, Rep. Math. Phys. 5, 145, 361 (1974).
[30]H. Bacry, Ph. Combe, and J.L. Richard, Nuovo Cimento A 67, 267 (1970); A 70, 289 (1970); J.L. Richard, Nuovo Cimento A 8, 485 (1972).
[31]Ph. Combe and P. Sorba, Physics A 80, 271 (1975).
[32]A. Bieri and A.Janner, Physica 50, 573 (1970); A. Janner and E. Ascher, Physica 48, 425 (1970); 54, 77 (1971); Helv. Phys. Acta 43, 296 (1970).
[33]J. Goldstone, Nuovo Cimento 19, 154 (1961); J. Goldstone, A. Salam, and S. Weinberg, Phys. Rev. 127, 965 (1962).
[34]P.W. Higgs, Phys. Lett. 12, 132 (1964); Phys. Rev. Lett. 13, 508 (1964); Phys. Rev. 145, 1156 (1966).
[35]T.W.B. Kibble, Phys. Rev. 155, 1554 (1967).
[36]P.W. Anderson, Phys. Rev. 112, 1900 (1958).
[37]J. Bernstein, Rev. Mod. Phys. 46, 7 (1974).
[38]G.W. Bluman and J.D. Cole, J. Math. Mech. 18, 1025 (1969).
[39]G.W. Bluman and J.D. Cole, *Similarity Methods for Differential Equations* (Springer, New York, 1974).

Clebsch potentials in the theory of electromagnetic fields admitting electric and magnetic charge distributions*

Hanno Rund

Department of Mathematics, University of Arizona, Tucson, Arizona 85721
(Received 12 July 1976)

Any skew-symmetric tensor field on a four-dimensional pseudo-Riemannian space V_4 admits a representation in terms of Clebsch potentials and their derivatives. Since the usual 4—potential representation of the electromagnetic field tensor of classical electrodynamics breaks down in the presence of magnetic charges, these Clebsch potentials are treated as the field variables of an invariant variational principle. The resulting Euler–Lagrange equations determine not only a useful representation of the electromagnetic field tensor (in the presence of magnetic charges), but also give rise to Maxwell-type field equations. The associated Lagrange density defines a unique energy–momentum tensor entirely on the basis of invariance consideration. A generalized variational principle is postulated, whose Euler–Lagrange equations specify the behavior of both the electromagnetic field tensor and the metric tensor of V_4. These are generalized Einstein–Maxwell equations. For the case of a spherically symmetric line-element and a static electromagnetic field an explicit solution of these equations is found which generalizes the well known Reissner–Nordström metric. In the course of the construction of this solution the magnetic charge of the central mass appears naturally as a constant of integration of the associated differential equations for the Clebsch potentials. The equations of motion of a test particle in an external electromagnetic field are also deduced from a variational principle; subject to fairly weak restrictions, the expected generalization of the classical Lorentz force emerges from this analysis.

I. INTRODUCTION

The well-known derivation of the classical Maxwell equations of electromagnetic theory from an invariant variational principle depends crucially upon the fact that the electromagnetic field tensor can be described as the curl of a 4-potential whose components are, in fact, the field variables in the associated Lagrange density. However, when the inclusion of magnetic charges is allowed in the theory, as, for instance, in the recent work of Schwinger,[1] the existence of a 4-potential in its usual form is precluded by virtue of the fact that the divergence of the magnetic field need not vanish. While most of the literature dealing with the possible existence of magnetic charges is concerned primarily with quantun mechanical consequences thereof, it is the objective of the present paper to describe, against a purely classical background, the systematic construction of suitable generalized 4-potentials, irrespective of the existence or otherwise of magnetic and electric charges.

The primary tool used for this purpose is a theorem which represents a generalization of the Clebsch transformations of hydrodynamics,[2] according to which an arbitrary differentiable vector field on a three-dimensional Euclidean space may be expressed as the sum of a gradient and a scalar multiple of another gradient. This generalization[3] entails the representation on an n-dimensional differentiable manifold ($n \geq 3$) of any totally skew-symmetric geometric object field endowed with $n-2$ indices, this representation being such that, when $n=3$, it reduces to that of Clebsch. When $n=4$, the theorem give rise to the representation of an arbitrary skew-symmetric geometric object field in terms of the so-called Clebsch potentials and their first derivatives. In the present paper, therefore, it is assumed that any electromagnetic field on a four-dimensional pseudo-Riemannian space–time V_4 is represented by a skew-symmetric type (0,2) tensor field F_{hj} and that the properties of this field are determined entirely by a single invariant variational principle whose Lagrangian depends solely on F_{hj}, the metric tensor g_{hj} of V_4, and possibly on given electric and magnetic charge distributions. In this variational principle, however, the Clebsch potentials of the representation of F_{hj} play the role which is usually assumed by the 4-potentials of classical electromagnetic theory.

An analysis of the resulting Euler—Lagrange equations indicates that effectively only two sets of Clebsch potentials survive in the representation of F_{hj}. Moreover, for the case of a flat V_4, the field equations satisfied by F_{hj} are precisely the extended Maxwell equations which had been postulated previously,[1] while the given Lagrangian L defines a unique energy—momentum tensor density whose derivation is based solely on invariance properties of L.

This state of affairs immediately suggests a generalization of the entire theory in terms of a single invariant variational principle from which the field equations for both of the tensor fields g_{hj} and F_{hj} are to be derived: In fact, it is found that thus a corresponding extension of the Einstein—Maxwell equations is obtained. An explicit solution of these equations may be constructed for the case of a spherically symmetric line element and a static electromagnetic field *in vacuo*; the integration of the associated ordinary differential equations for the Clebsch potentials appears to predict the existence of electric as well as magnetic monopoles (unless certain constants of integration are arbitrarily set equal to zero).

Because of the existence of two sets of Clebsch potentials, two distinct types of duality rotations are encountered. These may be used to construct a Lagrangian of a single integral variational problem for the determination of the equations of motion of a test particle in a combined gravitational and electromagnetic field. In this connection some difficulties are encountered in the sense that the existence of the aforementioned

Lagragian depends on the validity of certain relationships between the two sets of potentials. However, as a result of the field equations, these relationships are satisfied under fairly general conditions, particularly in the vacuum, and the resulting equations of motion reduce, in the nonrelativistic limit, precisely to those postulated by Schwinger,[4] presumably on the grounds of invariance under duality rotations.

It should be reiterated that the entire analysis presented below is of a purely classical nature, that is, nonquantum mechanical.

II. THE CLEBSCH REPRESENTATION OF THE ELECTROMAGNETIC FIELD TENSOR

Let x^j denote the local coordinates[5] of a four-dimensional differentiable manifold V_4. As shown elsewhere,[3] any class C^2 (that is, twice continuously differentiable) skew-symmetric geometric object field $X_{hj}(x^k)$ admits a local Clebsch representation of the form

$$X_{hj} = \psi_{j,h} - \psi_{h,j} + R(P_{,h}Q_{,j} - P_{,j}Q_{,h}), \tag{2.1}$$

in which the so-called Clebsch potentials are appropriate functions of the positional coordinates x^k. If, in particular, the given field X_{hj} is a type (0,2) tensor field, the potentials P,Q,R are scalars, while ψ_j is a convariant vector field. These potentials are not uniquely determined by X_{hj}; they may, in fact, be subjected to various kinds of (Clebsch) gauge transformations[3] which leave X_{hj} unchanged.

In order to introduce a tensorial dual of a skew-symmetric tensor field X_{hj} it is necessary that we endow V_4 with a prescribed Riemannian (or pseudo-Riemannian) metric tensor $g_{hj}(x^k)$, of which it is assumed that it is symmetric, of class C^2, and such that $g = |\det(g_{hj})| \neq 0$. The dual of X_{hj} is defined to be

$$X^{*hj} = \tfrac{1}{2} i g^{-1/2} \epsilon^{hjlk} X_{lk}. \tag{2.2}$$

According to (2.1) the corresponding Clebsch representation of X^{*hj} is given by

$$X^{*hj} = i g^{-1/2} \epsilon^{hjlk} (\psi_{k,l} + R P_{,l} Q_{,k}). \tag{2.3}$$

Since any skew-symmetric tensor field may be regarded as the dual of another, it follows that (2.2) represents the prototype of an equally acceptable Clebsch representation of an arbitrary skew-symmetric tensor field.

Let us now assume that a given electromagnetic field is represented by a class C^2 skew-symmetric tensor field $F_{hj}(x^k)$ on V_4. Thus the most general Clebsch representation of F_{hj} is given by a combination of the representations (2.1) and (2.2), namely

$$F_{hj} = f_{hj} + i b_{hj}, \tag{2.4}$$

where

$$f_{hj} = \psi_{j,h} - \psi_{h,j} + R(P_{,h}Q_{,j} - P_{,j}Q_{,h}), \tag{2.5}$$

and

$$b_{hj} = g_{hl} g_{jk} b^{lk}, \tag{2.6}$$

with

$$b^{hj} = g^{-1/2} \epsilon^{hjlk} (\phi_{k,l} + W U_{,l} V_{,k}). \tag{2.7}$$

This construction does not entail any restriction on the field F_{hj} at the present stage. However, the properties of the various Clebsch potentials which occur in (2.5) and (2.7) are to be determined by means of a single variational principle, in which each potential will give rise to an Euler–Lagrange equation. Thus the inclusion of a possibly superfluous potential will be compensated by the constraints implied by the corresponding Euler–Lagrange equation.

The presence of electric and magnetic charge distributions which give rise to F_{hj} is represented by electric and magnetic current density vectors, to be denoted by J^k and S^k respectively. It is now postulated that the behavior of the field F_{hj} is specified by the Euler–Lagrange equations resulting from an invariant variational principle whose Lagrange density is given by

$$L = \tfrac{1}{4} \sqrt{g}\, F^{hj} F_{hj} - \psi_h J^h + \phi_h S^h, \tag{2.8}$$

where ψ_h, ϕ_h are the vector fields which occur in the representations (2.5) and (2.7) and where

$$F^{hj} = g^{hl} g^{jk} F_{lk}, \tag{2.9}$$

it being understood that the field variables consist of all Clebsch potentials of the representation (2.4)–(2.7).

In order to derive the Euler–Lagrange equations corresponding to (2.8), we observe that, for a typical potential, temporarily denoted by z, we have, by virtue of (2.7) and the symmetry of g_{hl},

$$\frac{\partial}{\partial z} (\tfrac{1}{4} \sqrt{g}\, F^{hj} F_{hj}) = \tfrac{1}{2} \sqrt{g}\, F^{hj} \frac{\partial F_{hj}}{\partial z} = \tfrac{1}{2} \sqrt{g}\, F_{hj} \frac{\partial F^{hj}}{\partial z}.$$

Thus, for instance, it follows from (2.8), (2.4), and (2.7) that

$$\frac{\partial L}{\partial \phi_{h,j}} = \tfrac{1}{2} \sqrt{g}\, F_{lk} \frac{\partial F^{lk}}{\partial \phi_{h,j}} = \tfrac{1}{2} i \sqrt{g}\, F_{lk} \frac{\partial b^{lk}}{\partial \phi_{h,j}}$$

$$= \tfrac{1}{2} i \sqrt{g}\, F_{lk} g^{-1/2} \epsilon^{lkmp} \frac{\partial \phi_{p,m}}{\partial \phi_{h,j}}$$

$$= \sqrt{g}\, (\tfrac{1}{2} i g^{-1/2} \epsilon^{jhlk} F_{lk}) = \sqrt{g}\, F^{*jh},$$

where, in the last step, we have used (2.2). Hence the Euler–Lagrange tensor density corresponding to the potential ϕ_h is given by

$$E_{\phi_h}(L) \equiv \frac{\partial}{\partial x^j} \left(\frac{\partial L}{\partial \phi_{h,j}} \right) - \frac{\partial L}{\partial \phi_h} = \frac{\partial}{\partial x^j} (\sqrt{g}\, F^{*jh}) - S^h.$$

However, since $\sqrt{g}\, F^{*jh}$ is a skew-symmetric tensor density,[6]

$$(\sqrt{g}\, F^{*jh})_{|j} = \frac{\partial}{\partial x^j} (\sqrt{g}\, F^{*jh}) + \sqrt{g}\, F^{*lh} \Gamma^j_{lj} + \sqrt{g}\, F^{*jl} \Gamma^h_{lj}$$

$$- \sqrt{g}\, F^{*jh} \Gamma^l_{lj},$$

in which all terms involving the Christoffel symbols drop out in view of the symmetry of the latter. Accordingly, the Euler–Lagrange equations corresponding to ϕ_k may be expressed as

$$E_{\phi_k}(L) = (\sqrt{g}\, F^{*jk})_{|j} - S^k = 0. \tag{2.10}$$

By means of the same procedure the remaining set of Euler–Lagrange equations is found to assume the following form:

$$E_{\psi_k}(L) = (\sqrt{g}\, F^{jk})_{|j} + J^k = 0, \tag{2.11}$$

$$E_P(L) = (\sqrt{g}\, RF^{jk}Q_{,k})_{|j} = 0, \tag{2.12}$$

$$E_Q(L) = -(\sqrt{g}\, RF^{jk}P_{,k})_{|j} = 0, \tag{2.13}$$

$$E_R(L) = -\sqrt{g}\, F^{jk}P_{,j}Q_{,k} = 0, \tag{2.14}$$

$$E_U(L) = (\sqrt{g}\, WF^{*jk}V_{,k})_{|j} = 0, \tag{2.15}$$

$$E_V(L) = -(\sqrt{g}\, WF^{*jk}U_{,k})_{|j} = 0, \tag{2.16}$$

$$E_W(L) = -\sqrt{g}\, F^{*jk}U_{,j}V_{,k} = 0. \tag{2.17}$$

The set (2.10)—(2.17) represents the field equations satisfied by the Clebsch potentials. In order to determine the roles played by the potentials P, Q, R, we note that (2.11) and (2.12) yield

$$0 = (\sqrt{g}\, F^{jk})_{|j}RQ_{,k} + \sqrt{g}\, F^{jk}(RQ_{,k})_{|j} = -J^k RQ_{,k} + \sqrt{g}\, F^{jk}(RQ_{,k})_{|j},$$

and hence, since $(Q_{,k})_{|j}$ is symmetric in k and j, while F^{jk} is skew-symmetric in these indices,

$$\sqrt{g}\, F^{jk}R_{,j}Q_{,k} = RJ^kQ_{,k}. \tag{2.18}$$

Similarly, the combination of (2.11) with (2.13) yields

$$\sqrt{g}\, F^{jk}R_{,j}P_{,k} = RJ^kP_{,k}. \tag{2.19}$$

The constraints imposed on P, Q, R are determined entirely by (2.14), (2.18), and (2.19). The last two equations may be written in a more useful form if we introduce a vector field I_h defined by

$$I_j = 4g^{-1/2}(F\times F)^{-1}J^hF^*_{jh}, \tag{2.20}$$

where we have used the notation[7]

$$(F\times F) = F^{jk}F^*_{jk}. \tag{2.21}$$

Because of the well-known identity

$$F^{jk}F^*_{jh} = \tfrac{1}{4}\delta^k_h(F\times F), \tag{2.22}$$

we may write (2.20) as

$$J^k = \sqrt{g}\, F^{jk}I_j. \tag{2.23}$$

When this is substituted in (2.18) and (2.19), the latter reduce to the form

$$\sqrt{g}\, F^{jk}\tilde{R}_jQ_{,k} = 0, \quad \sqrt{g}\, F^{jk}\tilde{R}_jP_{,k} = 0, \tag{2.24}$$

where, for the sake of brevity, we have put

$$\tilde{R}_j = R_{,j} - RI_j. \tag{2.25}$$

Now, Eq. (2.14) is obviously satisfied if $P_{,j}$ and $Q_{,j}$ are proportional, that is, if there exist scalars λ, ν, not both zero, such that

$$\lambda P_{,j} = \nu Q_{,j}. \tag{2.26}$$

However, let us suppose to the contrary that $P_{,j}$, $Q_{,j}$ are linearly independent, in which case they span a two-dimensional plane Π_2 in the dual tangent spaces $T_4(p)$ of V_4 at each point p of V_4. From (2.14) and the identity $F^{jk}P_{,j}P_{,k} = 0$ it then follows that the vector $F^{jk}P_{,j}$ is normal to Π_2; the same inference may be drawn similarly for $F^{jk}Q_{,j}$, while (2.24) implies that $F^{jk}\tilde{R}_j$ is also normal to Π_2. Thus the three vectors $F^{jk}P_{,j}$, $F^{jk}Q_{,j}$, $F^{jk}\tilde{R}_j$ are contained in the two-dimensional orthogonal complement of Π_2, which implies their linear dependence. Accordingly there exist three scalars a, b, c, not all zero, such that

$$aP_{,j} + bQ_{,j} + c\tilde{R}_j = 0. \tag{2.27}$$

If $c = 0$, this is identical with (2.26), and accordingly we shall suppose that $c \neq 0$; for the same reason we may also suppose that $Q_{,j}$, $\tilde{R}_j$ are nonparallel and hence span Π_2. Therefore, the orthogonality of $F^{jk}P_{,j}$ to Π_2 may be expressed as

$$\sqrt{g}\, F^{hk}P_{,h} = \mu_m\epsilon^{mkhl}Q_{,h}\tilde{R}_l, \tag{2.28}$$

where μ_m is some nonvanishing vector field. [In fact, Eq. (2.28) is the solution of (2.14) and the second member of (2.24).] We now substitute for $\tilde{R}_l$ from (2.27) in (2.28), obtaining

$$\sqrt{g}\, F^{hk}P_{,h} = -ac^{-1}\mu_m\epsilon^{mkhl}Q_{,h}P_{,l}.$$

This is multiplied by F^*_{jk}, the identity (2.22) being taken into account, which yields

$$\begin{aligned}
&\tfrac{1}{4}\sqrt{g}\,(F\times F)P_{,j} \\
&\quad = -ac^{-1}\mu_m\epsilon^{mkhl}F^*_{jk}Q_{,h}P_{,l} \\
&\quad = -\tfrac{1}{2}iac^{-1}\sqrt{g}\,\mu_m\epsilon^{mkhl}\epsilon_{jkrs}F^{rs}Q_{,h}P_{,l} \\
&\quad = -\tfrac{1}{2}iac^{-1}\sqrt{g}\,\mu_m\delta^{mhl}_{jrs}F^{rs}Q_{,h}P_{,l} \\
&\quad = -iac^{-1}\sqrt{g}\,\mu_m(\delta^m_jF^{hl} + \delta^h_jF^{lm} + \delta^l_jF^{mh})Q_{,h}P_{,l},
\end{aligned}$$

or, if we use (2.14) once more,

$$[\tfrac{1}{4}(F\times F) + iac^{-1}F^{mh}\mu_mQ_{,h}]P_{,j} = iac^{-1}(F^{mh}\mu_mP_{,h})Q_{,j}. \tag{2.29}$$

Here it should be noted that Eq. (2.28) implies that the coefficient of $Q_{,j}$ on the right-hand side of (2.29) vanishes identically, which gives rise to (2.26) with $\nu = 0$. Accordingly (2.29) implies (2.26) once more, which is therefore an inevitable consequence of (2.14) and (2.24).

Moreover, since the structures of Eqs. (2.10), (2.15), (2.16), (2.17) are identical with those of (2.11)—(2.14), an analogous analysis applied to the former set yields the conclusion that there exist scalars σ, τ, not both zero, such that

$$\sigma U_{,j} = \tau V_{,j}. \tag{2.30}$$

The substitution of (2.26), (2.30) in (2.5), (2.7), respectively, reduces these representations to the simple forms[8]

$$f_{hj} = \psi_{j,h} - \psi_{h,j}, \tag{2.31}$$

$$b^{hj} = g^{-1/2}\epsilon^{hjlk}\phi_{k,l}. \tag{2.32}$$

Our conclusions may now be summarized as follows: *The field equations associated with the single invariant variational principle specified by the Lagrangian* (2.8) *uniquely determine the structure of the Clebsch representation* (2.4) *of the electromagnetic field tensor* F_{hj} *as given by* (2.31) *and* (2.32); *furthermore, the tensor* F_{hj} *satisfies the extended Maxwell equations*

$$(\sqrt{g}\, F^{hj})_{|j} = J^h, \tag{2.33}$$

$$(\sqrt{g}\,F^{*hj})_{|j} = -S^h, \tag{2.34}$$

where J^h, S^h denote the electric and magnetic current densities respectively.

In connection with these field equations the following observations may be of some relevance. From (2.32) it follows that

$$b^{hj}{}_{|j} = g^{-1/2}\epsilon^{hjlk}\phi_{k|l|j}.$$

But

$$\phi_{k|l|j} = \phi_{k|j|l} + \phi_m g^{mp} R_{pklj},$$

where R_{pklj} denotes the curvature tensor of V_4, while

$$\epsilon^{hjlk} R_{pklj} = 0$$

identically by virtue of the cyclic identities satisfied by the curvature tensor. Thus

$$b^{hj}{}_{|j} = 0 \tag{2.35}$$

identically. Also, it follows from (2.31) that

$$f^{*hj} = ig^{-1/2}\epsilon^{hjlk}\psi_{k|l},$$

so that, as before,

$$f^{*hj}{}_{|j} = 0 \tag{2.36}$$

identically. With the aid of (2.4) the field equations (2.33) and (2.34) may therefore be expressed in the form

$$(\sqrt{g}\,f^{hj})_{|j} = J^h, \tag{2.37}$$

$$i(\sqrt{g}\,b^{*hj})_{|j} = -S^h. \tag{2.38}$$

From the form of the representations (2.31) and (2.32) it is evident that (2.37) and (2.38) involve the fields ψ_h and ϕ_h separately. It would appear, therefore, that the electric and magnetic current densities give rise independently to the vector potentials ψ_h and ϕ_h respectively. Also, since the covariant derivatives on the left-hand sides of (2.37) and (2.38) may be replaced by partial derivatives, these equations imply the conservation laws

$$J^h{}_{|h} = 0, \quad S^h{}_{|h} = 0. \tag{2.39}$$

The representations (2.31) and (2.32) clearly indicate that f_{hj}, b_{hj}, and hence F_{hj}, are invariant under ordinary gauge transformations. However, F_{hj} is invariant under a far more general gauge transformation: namely

$$\psi_j \to \psi_j + \theta_j, \quad \phi_j \to \phi_j + \chi_j, \tag{2.40}$$

where θ_j, χ_j are arbitrary solutions of the system of partial differential equations

$$\theta_{j,h} - \theta_{h,j} + ig^{-1/2} g_{hl} g_{jk} \epsilon^{lkmp} \chi_{p,m} = 0. \tag{2.41}$$

Ordinary gauge transformations are obviously special cases of (2.40).

In conclusion, let us briefly glance at the explicit form of the field equations (2.33) and (2.34) in flat space–time. Relative to a coordinate system in which $x^4 = ict$, where c is the velocity of light *in vacuo*, the electromagnetic field vectors, $\mathbf{E}$, $\mathbf{H}$ are represented as usual by[9]

$$iE_\alpha = F^{4\alpha}, \quad H_\alpha = \tfrac{1}{2}\epsilon_{\alpha\beta\gamma}F^{\beta\gamma}, \tag{2.42}$$

where $\epsilon_{\alpha\beta\gamma}$ is the three-dimensional permutation symbol. Equivalently,

$$E_\alpha = \tfrac{1}{2}\epsilon_{\alpha\beta\gamma}F^{*\beta\gamma}, \quad iH_\alpha = F^{*\alpha 4}. \tag{2.43}$$

If the 4-vector representation of J^h, S^h is taken to be

$$J^h = \left(\frac{1}{c}\mathbf{j},\ i\rho\right), \quad S^h = \left(\frac{1}{c}\mathbf{s},\ i\sigma\right), \tag{2.44}$$

where ρ, σ respectively denote the electric and magnetic charge densities, it is easily verified with the aid of (2.42) and (2.43) that (2.33) is equivalent to

$$\nabla\cdot\mathbf{E} = \rho, \quad \nabla\times\mathbf{H} - \frac{1}{c}\frac{\partial\mathbf{E}}{\partial t} = \frac{1}{c}\mathbf{j}, \tag{2.45}$$

while (2.34) reduces to

$$\nabla\cdot\mathbf{H} = \sigma, \quad -\nabla\times\mathbf{E} - \frac{1}{c}\frac{\partial\mathbf{H}}{\partial t} = \frac{1}{c}\mathbf{s}. \tag{2.46}$$

These field equations are identical with those postulated by Schwinger.[1] Moreover, when (2.42) is written out in full in terms of (2.4), (2.31), and (2.32), in which we put $\psi = (\psi_1, \psi_2, \psi_3)$, $\phi = (\phi_1, \phi_2, \phi_3)$, together with

$$\psi_4 = iV, \quad \phi_4 = iU, \tag{2.47}$$

it is found that

$$\mathbf{E} = -\nabla V - \frac{1}{c}\frac{\partial\psi}{\partial t} - \nabla\times\phi, \tag{2.48}$$

and

$$\mathbf{H} = \nabla\times\psi - \nabla U - \frac{1}{c}\frac{\partial\phi}{\partial t}, \tag{2.49}$$

whose formal symmetry is quite striking. In fact, these representations are equivalent to the Helmholtz decompositions (according to which any vector on a three-dimensional Euclidean space can be expressed as the sum of a gradient and a curl) of the vectors

$$\mathbf{E} + \frac{1}{c}\frac{\partial\psi}{\partial t} \quad \text{and} \quad \mathbf{H} + \frac{1}{c}\frac{\partial\phi}{\partial t}$$

respectively.

III. THE ENERGY-MOMENTUM TENSOR

For any invariant variational principle based on a Lagrangian density whose functional dependence is given by

$$L = L(g_{hk};\ \psi_h,\ \psi_{h,k};\ \phi_h,\ \phi_{h,k}), \tag{3.1}$$

there is a simple prescription for the determination of a unique, symmetric type (2.0) tensor density, whose divergence vanishes whenever the field equations are satisfied. More precisely, writing

$$\Lambda^{hj} = \frac{\partial L}{\partial g_{hj}} = \Lambda^{jh}, \tag{3.2}$$

it may be shown, merely by virtue of the fact that L is a scalar density, that[10]

$$2\Lambda^{hj} = Lg^{hj} - g^{lj}\left[\frac{\partial L}{\partial\psi_h}\psi_l + \frac{\partial L}{\partial\psi_{h,k}}(\psi_{l,k} - \psi_{k,l}) + \frac{\partial L}{\partial\phi_h}\phi_l + \frac{\partial L}{\partial\phi_{h,k}}(\phi_{l,k} - \phi_{k,l})\right], \tag{3.3}$$

the divergence of which is a linear combination of the Euler–Lagrange expressions associated with L.

In the case of the Lagrangian

$$L_0 = \tfrac{1}{4}\sqrt{g}\, F^{lk}F_{lk}, \tag{3.4}$$

where F_{lk} is specified by (2.4), (2.31), and (2.32), we have, as before,

$$\frac{\partial L_0}{\partial \psi_{h,k}} = -\sqrt{g}\, F^{hk}, \quad \frac{\partial L_0}{\partial \phi_{h,k}} = -\sqrt{g}\, F^{*hk},$$

with

$$f_{kl} = \psi_{l|k} - \psi_{k|l}, \quad ib^*_{lk} = \phi_{l|k} - \phi_{k|l}, \tag{3.5}$$

so that (3.3) assumes the form

$$2\Lambda^{hj} = \sqrt{g}\,[g^{lj}(F^{hk}f_{kl} + iF^{*hk}b^*_{lk}) + \tfrac{1}{4}(F\cdot F)g^{hj}], \tag{3.6}$$

where we have introduced the notation

$$(F\cdot F) = F^{lk}F_{lk}. \tag{3.7}$$

The field equations corresponding to (3.4) are simply

$$F^{hj}{}_{|j} = 0, \quad F^{*hj}{}_{|j} = 0, \tag{3.8}$$

as is immediately evident from (2.33) and (2.34); according to the general theory $\Lambda^{hj}{}_{|j} = 0$ whenver (3.8) is satisfied, which, because of (3.6), implies that

$$T^j_{h|j} = 0, \tag{3.9}$$

where we have put

$$c^2T^j_h = -2\Lambda^j_h = \sqrt{g}\,[F^{jl}(\psi_{l|h} - \psi_{h|l}) + F^{*jl}(\phi_{l|h} - \phi_{h|l}) - \tfrac{1}{4}(F\cdot F)\delta^j_h]. \tag{3.10}$$

We shall regard (3.10) as the *energy–momentum tensor density of the electromagnetic field.*

Since (3.9) is crucial to the entire development, we shall briefly indicate how this assertion may be verified from first principles. A simple calculation yields

$$\tfrac{1}{4}(F\cdot F)_{|h} = F^{jl}\psi_{l|j|h} + F^{*jl}\phi_{l|j|h}, \tag{3.11}$$

and thus, if (3.8) is taken into account, it follows directly from (3.10) that

$$c^2T^j_{h|j} = \sqrt{g}\,F^{jl}(\psi_{l|h|j} - \psi_{h|l|j} - \psi_{l|j|h}) + \sqrt{g}\,F^{*jl}(\phi_{l|h|j} - \phi_{h|l|j} - \phi_{l|j|h}). \tag{3.12}$$

But the term involving F^{jl} may be written as

$$\sqrt{g}\,F^{jl}[(\psi_{l|h|j} - \psi_{l|j|h}) - \tfrac{1}{2}(\psi_{h|l|j} - \psi_{h|j|l})]$$
$$= -\sqrt{g}\,F^{jl}g^{km}\psi_k(R_{lmhj} - \tfrac{1}{2}R_{hmlj})$$
$$= \tfrac{1}{2}\sqrt{g}\,F^{jl}g^{km}\psi_k(R_{lmhj} - R_{jmhl} - R_{hmlj}),$$

which is zero by virtue of cyclic identity satisfied by the curvature tensor. Similarly, the term involving F^{*jl} in (3.12) may be shown to vanish, which establishes (3.9).

In order to facilitate the treatment of the algebraic properties of the energy–momentum tensor density (3.10), it is appropriate to use matrix instead of component notation. Let us write

$$F = (F^j{}_h) = (F^{jl}g_{lh}) = (g^{jl}F_{lh}),$$

and similarly for F^*, f, b, etc. In accordance with (3.7) and (2.21) we shall also put

$$(f\cdot b) = -f^{lk}b_{kl} = -f^l{}_k b^k{}_l = -\mathrm{tr}(fb) = (b\cdot f), \tag{3.13}$$

$$(f\times b) = -f^{*lk}b_{kl} = -f^{*l}{}_k b^k{}_l = -\mathrm{tr}(f^*b) = (b\times f). \tag{3.14}$$

Because of (2.4), we have $F = f + ib$, so that

$$(F\cdot F) = (f\cdot f) - (b\cdot b) + 2i(f\cdot b), \tag{3.15}$$

$$(F\times F) = (f\times f) - (b\times b) + 2i(f\times b). \tag{3.16}$$

By means of a direct calculation the following identities may be verified:

$$f^*b + b^*f = -\tfrac{1}{2}(f\times b)I, \quad fb^* + bf^* = -\tfrac{1}{2}(f\times b)I, \tag{3.17}$$

$$f^*b^* - bf = \tfrac{1}{2}(f\cdot b)I, \quad fb - b^*f^* = -\tfrac{1}{2}(f\cdot b)I, \tag{3.18}$$

where I denotes the unit (4×4) matrix. In particular, putting $b = f$ in (3.17) and (3.18), one obtains

$$f^*f = -\tfrac{1}{4}(f\times f)I, \quad (f^*)^2 - f^2 = \tfrac{1}{2}(f\cdot f)I. \tag{3.19}$$

It should be observed that the identities (3.17)–(3.19) hold for any pair of skew-symmetric matrices f_{jh}, b_{jh}.

With the aid of (3.18) and (3.15) we may now express (3.10) in the form

$$c^2g^{-1/2}T = -Ff + iF^*b^* - \tfrac{1}{4}(F\cdot F)I$$
$$= -f^2 - (b^*)^2 + i(f^*b - bf) - \tfrac{1}{4}[(f\cdot f) - (b\cdot b) + 2i(f\cdot b)]I$$
$$= -f^2 - (b^*)^2 - \tfrac{1}{4}[(f\cdot f) - (b\cdot b)]I. \tag{3.20}$$

Alternatively, if (3.19) is applied, it is seen that

$$-c^2g^{-1/2}T = \tfrac{1}{2}[f^2 + (f^*)^2 + b^2 + (b^*)^2]. \tag{3.21}$$

Since

$$\mathrm{tr}(f^2) = -(f\cdot f) = (f^*\cdot f^*) = -\mathrm{tr}(f^*)^2,$$

it follows directly from (3.21) that

$$\mathrm{tr}(T) = 0. \tag{3.22}$$

It should be noted, however, that the trace-free character of T is merely a result of its algebraic structure, and does not depend on the specific nature of the tensors f_{hj} and b_{hj}.

In classical electrodynamics the square of the energy–momentum tensor is a scalar multiple of the unit matrix, a phenomenon which is exploited, for instance, in the "already unified field theory" of Rainich.[11] Accordingly we shall now evaluate T^2 from (3.21) in order to determine the circumstances under which this property is preserved for the general case under discussion. To this end we note that, because of (3.19),

$$[f^2 + (f^*)^2]^2 = [f^2 - (f^*)^2]^2 + 4fff^*f^* = \tfrac{1}{4}[(f\circ f)^2 + (f\times f)^2]I,$$

while

$$[f^2 + (f^*)^2][b^2 + (b^*)^2] = (f\circ f)b^2 + (b\circ b)f^2 + 4f^2b^2 + \tfrac{1}{4}(f\circ f)(b\ b)I.$$

Thus, (3.21) gives

$$4g^{-1}c^4T^2=\tfrac{1}{4}[(f\cdot f)^2+(b\cdot b)^2+(f\times f)^2 + (b\times b)^2+2(f\cdot f)(b\cdot b)]I+R_{(1)}, \tag{3.23}$$

where

$$R_{(1)}=2(f\cdot f)b^2+2(b\cdot b)f^2+4(f^2b^2+b^2f^2) =4[f^2(b^*)^2+b^2(f^*)^2],$$

in which we have applied (3.9) once more. But by means of (3.17) it is readily established that

$$f^2(b^*)^2+b^2(f^*)^2=\tfrac{1}{4}(f\times b)^2I-fbf^*b^*-bfb^*f^*,$$

so that

$$R_{(1)}=(f\times b)^2I+R_{(2)}, \tag{3.24}$$

where

$$-R_{(2)}=4[fbf^*b^*-bfb^*f^*]. \tag{3.25}$$

Now, let us introduce the commutator

$$\lambda=fb-bf, \tag{3.26}$$

so that, by (3.18),

$$f^*b^*-b^*f^*=-\lambda. \tag{3.27}$$

With the aid of (3.19) we may then write

$$fbf^*b^*=\tfrac{1}{16}(f\times f)(b\times b)I+\lambda f^*b^*,$$

$$bfb^*f^*=\tfrac{1}{16}(f\times f)(b\times b)I-\lambda b^*f^*,$$

and hence, with the use of (3.27), it is seen that (3.25) becomes

$$R_{(2)}=-\tfrac{1}{2}(f\times f)(b\times b)I+4\lambda^2.$$

This is substituted in (3.24), by means of which (3.23) can be expressed as

$$16g^{-1}c^4T^2=\{[(f\cdot f)-(b\cdot b)]^2+[(f\times f)-(b\times b)]^2 +4(f\circ f)(b\cdot b)+4(f\times b)^2\}I+16\lambda^2. \tag{3.28}$$

Because of (3.15) and (3.16) this is equivalent to

$$16g^{-1}c^4T^2=\{(F\cdot F)^2+(F\times F)^2-4i[(F\circ F)(f\cdot b) +(F\times F)(f\times b)] +4(f\circ f)(b\cdot b)-4(f\cdot b)^2\}I+16\lambda^2. \tag{3.29}$$

Clearly T^2 is a scalar multiple of the unit matrix if and only if

$$\lambda^2=\kappa I, \tag{3.30}$$

where κ is an arbitrary scalar. This condition is equivalent to the requirement that

$$\lambda=\nu\lambda^*, \tag{3.31}$$

where

$$\nu(\lambda\times\lambda)=-4\kappa. \tag{3.32}$$

In order to analyze the implications of (3.31) completely, we have to evaluate λ^* explicitly. From (3.26) it follows that

$$\lambda^{*lk}=\tfrac{1}{2}ig^{-1/2}\epsilon^{lkhj}(f_{hm}b^m{}_j-b_{hm}f^m{}_j) =ig^{-1/2}\epsilon^{lkhj}f_{hm}b^m{}_j, \tag{3.33}$$

since $b^m{}_j=-b_j{}^m$. But

$$ig^{-1/2}\epsilon^{lkhj}f_{hm} =-ig^{-1/2}\epsilon^{lkhj}(\tfrac{1}{2}ig^{1/2}\epsilon_{hmrs}f^{*rs}) =\tfrac{1}{2}\delta^{lkj}_{mrs}f^{*rs}=[\delta^l_mf^{*kj}+\delta^k_mf^{*jh}+\delta^j_mf^{*lk}],$$

so that (3.33) reduces to

$$\lambda^{*lk}=f^{*lj}b_j{}^k-b^l{}_jf^{*jk},$$

or

$$\lambda^*=f^*b-bf^*. \tag{3.34}$$

Because of (3.17) this is equivalent to

$$\lambda^*=fb^*-b^*f. \tag{3.35}$$

Accordingly the condition (3.31) assumes the form

$$(fb-bf)=\nu(f^*b-bf^*)=\nu(fb^*-b^*f),$$

so that

$$(f-\nu f^*)b=b(b-\nu f^*),\quad f(b-\nu b^*)=(b-\nu b^*)f.$$

However, a few simple manipulations indicate that this is consistent with (3.17) and (3.18) if and only if $\nu^2=-1$ and

$$p(f\pm if^*)=q(b\pm ib^*), \tag{3.36}$$

where p and q are arbitrary nonzero scalars. Conversely, it is easily seen with the aid of (3.34) and (3.35) that (3.36) implies (3.30) with

$$4\kappa=\mp i(\lambda\times\lambda). \tag{3.37}$$

It is therefore concluded that *the square of the energy–momentum tensor density* (3.10) *is a scalar multiple of the unit matrix if and only if* f *and* b *satisfy the condition* (3.36), *in which case the scalar* κ *is related to the commutator* $\lambda=fb-bf$ *according to* (3.37).

IV. EFFECTS OF ELECTRIC AND MAGNETIC CHARGES ON GRAVITATIONAL FIELDS

Up to this stage it had been assumed that the metric tensor of the underlying manifold V_4 assumes preassigned values. It is more instructive, however, to consider the more general case when the field equations to be satisfied by *both* the tensor fields g_{hj} and F_{hj} are determined by a *single* invariant variational principle. To this end we shall consider a Lagrangian of the form

$$\hat{L}(g_{hk};g_{hk,l};g_{hk,l,m};\psi_h;\psi_{h,k};\phi_h,\phi_{h,k})=a\sqrt{g}\,R+L, \tag{4.1}$$

where L is given by (2.8), while R is the curvature scalar $R=g^{jh}R_{jh}$, defined in terms of the Ricci tensor $R_{jh}=R_j{}^k{}_{hk}$, and a is some constant to be specified presently.

The Euler–Lagrange tensor density corresponding to the metric tensor field in (4.1) is given by

$$E^{hj}(\hat{L})=-\frac{\partial^2}{\partial x^l\partial x^m}\left(\frac{\partial\hat{L}}{\partial g_{hj,l,m}}\right)+\frac{\partial}{\partial x^l}\left(\frac{\partial\hat{L}}{\partial g_{hj,l}}\right)-\frac{\partial\hat{L}}{\partial g_{hj}} =E^{hj}(a\sqrt{g}\,R)-\Lambda^{hj} \tag{4.2}$$

in terms of the notation (3.2) applied to (2.8) [it being observed that Λ^{hj} assumes precisely the same form for (2.8) and (3.4)]. It is well known that[12]

$$E^{hj}(a\sqrt{g}\,R) = a\sqrt{g}\,[R^{hj} - \tfrac{1}{2}g^{hj}R], \tag{4.3}$$

and thus it follows from (4.2) and (3.10) that the corresponding Euler–Lagrange equations are

$$a\sqrt{g}\,[R_h^j - \tfrac{1}{2}\delta_h^j R] + \tfrac{1}{2}c^2 T_h^j = 0. \tag{4.4}$$

However, according to (3.22), we have $T_j^j = 0$, so that a contraction over j and h in (4.4) yields $R = 0$. Thus, putting $a = (16\pi\kappa)^{-1}c^{-4}$, where κ is the constant of gravitation, we may express (4.4) in the form

$$\sqrt{g}\,R_{jh} = CT_{jh}, \tag{4.5}$$

where

$$C = 8\pi\kappa c^{-2}. \tag{4.6}$$

Moreover, since the term involving $a\sqrt{g}\,R$ in the Lagrangian (4.1) does not involve the fields ψ_h, ϕ_h in any way whatsoever, the analysis of Sec. II concerning the Euler–Lagrange equations involving the Clebsch potentials is applicable *verbatim*, thus yielding the field equations (2.37) and (2.38), that is,

$$(\sqrt{g}\,f^{hj})_{|j} = J^h, \quad i(\sqrt{g}\,b^{*hj})_{|j} = -S^h. \tag{4.7}$$

The system consisting of (4.5) *and* (4.7) *represents the Einstein–Maxwell equations in the presence of electric as well as magnetic charge distributions*, it being understood that the energy–momentum tensor density which appears on the right-hand side of (4.5) is given explicitly by (3.10). Again it should be emphasized that this derivation of the field equations does not presuppose a particular representation of the electromagnetic field tensor F_{hj} in terms of 4-potentials.

In order to determine the nature of the interaction between the gravitational and electromagnetic fields as implied by this system, we shall now construct an explicit solution of these equations for the case of a pair of static fields ψ_h, ϕ_h. Such fields are prescribed by the conditions

$$\psi_\alpha = 0, \quad \psi_4 = iV(x^\alpha), \quad \phi_\alpha = 0, \quad \phi_4 = iU(x^\alpha). \tag{4.8}$$

According to (3.5), we then have

$$f_{\alpha\beta} = 0, \quad f_{\alpha 4} = iV_{,\alpha}, \quad b^*_{\alpha\beta} = 0, \quad b^*_{\alpha 4} = -U_{,\alpha}. \tag{4.9}$$

From (4.8) and (4.9) it is evident that

$$\begin{aligned} &\sqrt{g}\,f^{*\alpha\beta} = -\epsilon^{\alpha\beta\gamma}V_{,\gamma}, \quad f^{*\alpha 4} = 0, \\ &i\sqrt{g}\,b^{\alpha\beta} = -\epsilon^{\alpha\beta\gamma}U_{,\gamma}, \quad b^{\alpha 4} = 0, \end{aligned} \tag{4.10}$$

which immediately implies the validity of the identities (2.35) and (2.36) [as was to be expected, of course]. However, the potentials U and V must be such that the field equation (4.7) is satisfied, where it is to be observed that the explicit structure of these equations depends crucially on the nature of the metric. The latter is now assumed to be static and spherically symmetric in the sense that, relative to a coordinate system in which $x^1 = r$, $x^2 = \theta$, $x^3 = \phi$, $x^4 = ict$, the metric tensor may be expressed in the form

$$g_{hj} = -\operatorname{diag}[e^\lambda, r^2, r^2\sin^2\theta, e^\nu], \tag{4.11}$$

where $\operatorname{diag}[a, b, c, d]$ denotes the 4×4 diagonal matrix with (ordered) diagonal entries a, b, c, d, while λ and ν are real functions of r only, these functions to be ultimately determined by the field equations. Thus

$$g = |\det(g_{hj})| = r^4\sin^2\theta\, e^{\lambda+\nu}, \tag{4.12}$$

while the inverse of (4.12) is given by

$$g^{hj} = -\operatorname{diag}[e^{-\lambda}, r^{-2}, r^{-2}\sin^{-2}\theta, e^{-\nu}]. \tag{4.13}$$

Now, for an arbitrary metric, we have from (4.9),

$$\begin{aligned} b^{*hj} &= g^{hl}g^{jm}b^*_{lm} \\ &= (g^{h4}g^{j1} - g^{h1}g^{j4})U_{,1} + (g^{h4}g^{j2} - g^{h2}g^{j4})U_{,2} \\ &\quad + (g^{h4}g^{j3} - g^{h3}g^{j4})U_{,3}. \end{aligned}$$

But since (4.13) is diagonal, it follows that, relative to the metric (4.11),

$$\begin{aligned} &b^{*12} = 0, \quad b^{*13} = 0, \quad b^{*23} = 0, \\ &b^{*14} = -g^{11}g^{44}U_{,1}, \quad b^{*24} = -g^{22}g^{44}U_{,2}, \quad b^{*34} = -g^{33}g^{44}U_{,3}, \end{aligned} \tag{4.14}$$

so that $(i\sqrt{g}\,b^{\alpha j})_{|j} = 0$ identically for $\alpha = 1, 2, 3$. Thus, according to the second member of (4.7), our original assumptions regarding the field b_{hj} can be consistent solely with a magnetic charge distribution for which $S^\alpha = 0$. But, because of (4.14), the fourth component of the second member of (4.7) yields

$$\frac{\partial}{\partial x^1}(\sqrt{g}\,g^{11}g^{44}U_{,1}) + \frac{\partial}{\partial x^2}(\sqrt{g}\,g^{22}g^{44}U_{,2}) + \frac{\partial}{\partial x^3}(\sqrt{g}\,g^{33}g^{44}U_{,3}) = iS^4,$$

and hence, if we substitute from (4.12) and (4.13),

$$\frac{\partial}{\partial r}\left(r^2\sin\theta e^{-(\lambda+\nu)/2}\frac{\partial U}{\partial r}\right) + \frac{\partial}{\partial\theta}\left(\sin\theta e^{(\lambda-\nu)/2}\frac{\partial U}{\partial\theta}\right) + \frac{\partial}{\partial\phi}\left(\frac{1}{\sin\theta}e^{(\lambda-\nu)/2}\frac{\partial U}{\partial\phi}\right) = iS^4. \tag{4.15}$$

This is the field equation to be satisfied by the magnetic potential $U(r, \theta, \phi)$. For a spherically symmetric potential $U = U(r)$ *in vacuo* (that is, in regions where $S^4 = 0$), this reduces to

$$\frac{d}{dr}[r^2 e^{-(\lambda+\nu)/2}U'] = 0,$$

where $U' = dU/dr$, which may be integrated to yield

$$U' = \gamma r^{-2}e^{(\lambda+\nu)/2}, \tag{4.16}$$

in which γ is a constant of integration. By analogy with classical magnetostatics, we shall interpret γ as the magnetic charge which gives rise to the static potential $U(r)$. [Here it should be observed that γ is real by virtue of the fact that U is real, as is evident from (2.48).]

An analogous treatment of the first member of (4.7) gives rise to the conclusion that $J^\alpha = 0$, while the potential $V(r, \theta, \phi)$ must satisfy a field equation which is identical with the equation (4.15) for U, except that the right-hand side is to be replaced by iJ^4. Again, for a spherically symmetric potential $V = V(r)$ *in vacuo* ($J^4 = 0$), this gives rise to

$$V' = \epsilon r^{-2}e^{(\lambda+\nu)/2}, \tag{4.17}$$

where the constant of integration ϵ is interpreted as the electric charge which gives rise to V.

Let us now turn to the field equation (4.5), whose treatment clearly requires the explicit evaluation of T_{jh} for the static fields defined by (4.8) and (4.9). From the second member of (4.10) it follows that the only nonvanishing elements of b^{hj} are given by $i\sqrt{g}\, b^{23} = -U' = -i\sqrt{g}\, b^{32}$. With the aid of (4.11) it is then inferred that the only nonvanishing components of $i\sqrt{g}\, b^j{}_h$ are $i\sqrt{g}\, b^2{}_3 = U'r^2\sin^2\theta$, $i\sqrt{g}\, b^3{}_2 = -U'r^2$. Accordingly the square of the matrix $(b^j{}_h)$ is given by

$$gb^2 = (U')^2 r^4 \sin^2\theta \,\mathrm{diag}[0,1,1,0].$$

Because of (4.12) this implies that

$$b^2 = (U')^2 e^{-(\lambda+\nu)} \mathrm{diag}[0,1,1,0]. \tag{4.18}$$

Similarly, since it follows from (4.9) that the only nonvanishing elements of $(b^*{}_{hj})$ are $b^*_{14} = -U' = -b^*_{41}$, it is seen that the only nonvanishing components of $b^{*h}{}_j$ are $b^{*1}{}_4 = -e^{-\lambda}U'$, $b^{*4}{}_1 = e^{-\nu}U'$, giving

$$(b^*)^2 = -(U')^2 e^{-(\lambda+\nu)} \mathrm{diag}[1,0,0,1]. \tag{4.19}$$

If we now substitute from (4.16), we may infer from (4.18) and (4.19) that

$$b^2 + (b^*)^2 = \gamma^2 r^{-4} \mathrm{diag}[-1,1,1,-1]. \tag{4.20}$$

In the same manner it is easily established with the aid of (4.10)–(4.13), and (4.17), that

$$f^2 + (f^*)^2 = \epsilon^2 r^{-4} \mathrm{diag}[1,-1,-1,1]. \tag{4.21}$$

This, together with (4.20) is substituted in (3.21) to yield the following expression for the energy–momentum tensor density:

$$T = \tfrac{1}{2}c^{-2}(\epsilon^2 - \gamma^2)\sqrt{g}\, r^{-4} \mathrm{diag}[-1,1,1,-1]. \tag{4.22}$$

Because of (4.11), this is equivalent to

$$T_{jh} = \tfrac{1}{2}c^{-2}(\epsilon^2 - \gamma^2)\sqrt{g}\, r^{-4} \begin{pmatrix} e^{\lambda} & 0 & 0 & 0 \\ 0 & -r^2 & 0 & 0 \\ 0 & 0 & -r^2\sin^2\theta & 0 \\ 0 & 0 & 0 & e^{\nu} \end{pmatrix}. \tag{4.23}$$

This expression occupies the right-hand side of the field equation (4.5). The explicit form of the components of the Ricci tensor which occur on the left-hand side, relative to the metric (4.11) is, of course, well known, being given by[13]

$$R_{11} = \tfrac{1}{2}\nu'' - \tfrac{1}{4}\lambda'\nu' + \tfrac{1}{4}(\nu')^2 - r^{-1}\lambda',$$

$$R_{22} = e^{-\lambda}[1 + \tfrac{1}{2}r\nu' - \tfrac{1}{2}r\lambda'] - 1, \quad R_{33} = R_{22}\sin^2\theta,$$

$$R_{44} = e^{\nu-\lambda}[\tfrac{1}{2}\nu'' - \tfrac{1}{4}\lambda'\nu' + \tfrac{1}{4}(\nu')^2 + r^{-1}\nu'].$$

Putting

$$\hat{C} = \tfrac{1}{2}c^{-2}(\epsilon^2 - \gamma^2)C, \tag{4.24}$$

and proceeding along the usual lines,[14] we see that Eqs. (4.5) assume the following explicit forms for $(j,h) = (1,1)$, and $(j,h) = (4,4)$:

$$\tfrac{1}{2}\nu'' - \tfrac{1}{4}\lambda'\nu' + \tfrac{1}{4}(\nu')^2 - r^{-1}\lambda' = \hat{C}r^{-4}e^{\lambda},$$

$$\tfrac{1}{2}\nu'' - \tfrac{1}{4}\lambda'\nu' + \tfrac{1}{4}(\nu')^2 + r^{-1}\nu' = e^{\lambda-\nu}\hat{C}r^{-4}e^{\nu} = \hat{C}r^{-4}e^{\lambda},$$

the combination of which yields

$$-r^{-1}(\lambda' + \nu') = 0.$$

If it is assumed that the metric is flat at spatial infinity, this implies that $\lambda = -\nu$. Equations (4.5) for $(j,h) = (2,2)$ and (3,3) are identical, and now reduce to

$$e^{\nu}(1 + r\nu') - 1 = -\hat{C}r^{-2},$$

yielding

$$e^{\nu} = 1 - 2m/r + \hat{C}/r^2, \tag{4.25}$$

where m is a constant of integration, which, by analogy with the Schwarzschild solution, is identified with the mass of the object which gives rise to the spherically symmetric gravitational field. On substituting from (4.6) and (4.24) in (4.25), we finally obtain

$$e^{\nu} = e^{-\lambda} = 1 - \frac{2m}{r} + \frac{4\pi\kappa}{c^4 r^2}(\epsilon^2 - \gamma^2). \tag{4.26}$$

The metric (4.11), in which λ, ν are given by (4.26), represents a solution of the Einstein–Maxwell equations in the presence of a pair of static fields $V(r)$, $U(r)$.[15] Moreover, as expected, the field equations (4.16) and (4.17) imply the inverse square laws by virtue of the fact that $\lambda = -\nu$.

The respective signs which appear before the terms ϵ^2 and γ^2 in (4.26) may be somewhat surprising at first sight, it being recalled that ϵ and γ must be real as a consequence of our construction. In order to trace the origin of this phenomenon, which is obviously inherent in the structure of the energy–momentum tensor, let us evaluate the component $-T_{44}$ (which is interpreted as the energy density in classical electrodynamics) by means of the representation (2.42) in flat space–time. From (3.10) and (3.5) we have

$$c^2 T^j_h = F^{jl} f_{hl} - iF^{*jl} b^*_{hl} - \tfrac{1}{4}\delta^j_h (F \cdot F). \tag{4.27}$$

In terms of the notation (2.47) it follows directly from (2.42), (2.43), and (3.5) that

$$F_{4\alpha} f_{4\alpha} = \mathbf{E} \cdot \left(\nabla V + \frac{1}{c}\frac{\partial \psi}{\partial t}\right), \quad iF^*_{4\alpha} b^*_{4\alpha} = \mathbf{H} \cdot \left(\nabla U + \frac{1}{c}\frac{\partial \phi}{\partial t}\right),$$

while

$$(F \cdot F) = 2(H^2 - E^2).$$

When these values are substituted in (4.27) with $j = h = 4$, it is found that

$$c^2 T^4_4 = \mathbf{E} \cdot \left(\nabla V + \frac{1}{c}\frac{\partial \psi}{\partial t}\right) - \mathbf{H} \cdot \left(\nabla U + \frac{1}{c}\frac{\partial \phi}{\partial t}\right) - \tfrac{1}{2}(H^2 - E^2), \tag{4.28}$$

or, after an application of (2.48) and (2.49),

$$c^2 T^4_4 = \tfrac{1}{2}(H^2 - E^2) - \mathbf{E} \cdot (\nabla \times \phi) - \mathbf{H} \cdot (\nabla \times \psi). \tag{4.29}$$

In classical electromagnetic theory, in which $\phi = 0$ and $\mathbf{H} = \nabla \times \psi$, this reduces to the well-known form

$$-c^2 T^4_4 = \tfrac{1}{2}(E^2 + H^2),$$

which is always nonnegative. However, in the present context, this can no longer be asserted. Indeed, in the case of the static field defined by (4.8), it follows from (2.42), (2.43), (4.9), and (4.10) that

$$\mathbf{E} = -\nabla V, \quad \mathbf{H} = -\nabla U,$$

so that (4.28) assumes the form

$$-c^2T_4^4=\tfrac{1}{2}(E^2-H^2), \tag{4.30}$$

which may obviously assume both positive and negative values. In particular, when the solutions (4.16) and (4.17) are substituted in (4.30), the latter assumes the value

$$-c^2T_4^4=\tfrac{1}{2}(\epsilon^2-\gamma^2)r^{-4},$$

[as is also evident directly from (4.22)], whose sign depends on the relative magnitudes of ϵ and γ.

V. DUALITY ROTATIONS

A conceptually important role in previous theories[16] based on the existence of both electric and magnetic charges is played by the concept of invariance under the so-called duality rotations.[17] It would appear that in Schwinger's theory a most significant aspect of these rotations is contained in the relations

$$\tilde{\rho}=\rho\cos\theta+\sigma\sin\theta,\quad \tilde{\sigma}=-\rho\sin\theta+\sigma\cos\theta, \tag{5.1}$$

by means of which a magnetic charge at a point may, in a sense, be "transformed away" by suitable choice of the angle θ. Clearly (5.1) is the fourth component of the relation

$$\tilde{J}^h=J^h\cos\theta+S^h\sin\theta,\quad \tilde{S}^h=-J^h\sin\theta+S^h\cos\theta, \tag{5.2}$$

which ensures the invariance of the field equations (2.33) and (2.34) under the general duality rotation

$$\tilde{F}_{hj}=F_{hj}\cos\theta-F^*_{hj}\sin\theta,\quad \tilde{F}^*_{hj}=F^*_{hj}\cos\theta+F_{hj}\sin\theta. \tag{5.3}$$

We shall see that, within the context of the present analysis, there are two fundamentally distinct approaches, both of which lead directly to (5.1)–(5.3). The first of these is based on transformations of the potentials, namely,

$$\tilde{\psi}_j=\psi_j\cos\theta+\phi_j\sin\theta,\quad \tilde{\phi}_j=-\psi_j\sin\theta+\phi_j\cos\theta, \tag{5.4}$$

these being suggested directly by (5.1). Because of (3.5) this gives rise to

$$\tilde{f}_{hj}=f_{hj}\cos\theta-ib^*_{hj}\sin\theta,\quad f^*_{hj}=f^*_{hj}\cos\theta+b_{hj}\sin\theta, \tag{5.5}$$

together with

$$\tilde{b}_{hj}=b_{hj}\cos\theta+if^*_{hj}\sin\theta,\qquad \tilde{b}^*_{hj}=b^*_{hj}\cos\theta-if_{hj}\sin\theta. \tag{5.6}$$

It is immediately evident from (2.4) that (5.3) is implied by (5.5) and (5.6), which will be called the *induced duality rotations* [having been induced by (5.4)]. On the other hand, the rotation (5.3) also results, in a trival manner, from

$$\hat{f}_{hj}=f_{hj}\cos\theta-f^*_{hj}\sin\theta,\qquad \hat{f}^*_{hj}=f^*_{hj}\cos\theta+f_{hj}\sin\theta, \tag{5.7}$$

$$\hat{b}_{hj}=b_{hj}\cos\theta-b^*_{hj}\sin\theta,\qquad \hat{b}^*_{hj}=b^*_{hj}\cos\theta+b_{hj}\sin\theta, \tag{5.8}$$

which we shall call the *standard duality rotation* [since its structure is identical with that of (5.3)]. The field equations (2.33) and (2.34) are invariant under both types of duality rotations, provided that (5.2) is associated with each, as will be assumed henceforth. For future reference we observe that, as a consequence of (5.2), the electric and magnetic charges ϵ and γ of a particle must be subjected to the transformations

$$\tilde{\epsilon}=\epsilon\cos\theta+\gamma\sin\theta,\quad \tilde{\gamma}=-\epsilon\sin\theta+\gamma\cos\theta, \tag{5.9}$$

in order to ensure the invariance of the theory under (5.3).

There are, however, significant differences between induced and standard duality rotations. For instance, the identities (2.35) and (2.36), together with the special form (2.37) and (2.38) of the field equations, are invariant under the former, but not under the latter. On the other hand, it may be easily verified that the expression (3.21) for the energy–momentum tensor density T is invariant under the standard duality rotation, while this is not true for the induced duality rotation. Thus an application of the former cannot affect the question as to whether or not T^2 is a multiple of the unit matrix. However, it may also be shown that the commutator $\lambda=fb-bf$ is invariant under induced duality rotations, and hence the latter is also incapable of affecting the condition (3.30) which determines the character of T^2. Nevertheless, it may be worth observing that the equivalent condition (3.36), which is automatically invariant under the standard duality rotation for fixed p,q, is invariant under the induced duality rotation if and only if p and q transform according to

$$\tilde{p}=p\cos\theta+q\sin\theta,\quad \tilde{q}=-p\sin\theta+q\cos\theta, \tag{5.10}$$

whose structure is identical with that of (5.9).

VI. THE EQUATIONS OF MOTION OF A TEST PARTICLE

We shall now consider the motion in V_4 of a test particle of rest mass m, endowed with electric and magnetic charges, under the influence of an external electromagnetic field F_{hj}. It is supposed that the effects of m and these charges on the tensor fields g_{hj} and F_{hj} are negligible. If the equations of motion are to be derived from a variational principle, as is assumed here, the central problem revolves about the determination of a suitable (single integral) Lagrangian $\Lambda(x^j,\dot{x}^j)$, where $\dot{x}^j=dx^j/d\tau$ for an as yet arbitrary parameter τ.

It is supposed that Λ is of the form

$$\Lambda=\Lambda_0-c^{-1}\Lambda_I, \tag{6.1}$$

where Λ_0 refers to the inertia of the particle and $c^{-1}\Lambda_I$ is the interaction Lagrangian. If it is required that the action integral $\int\Lambda\,d\tau$ be invariant under coordinate as well as parameter transformations, it is necessary to stipulate (i) that Λ is invariant under coordinate transformations, and (ii) that Λ is homogeneous[18] of the first degree in $\dot{x}^j$. This leads to the usual form for Λ_0:

$$\Lambda_0=mc(g_{hj}\dot{x}^h\dot{x}^j)^{1/2}, \tag{6.2}$$

for which (under the exclusion of null-directions[19])

$$E_j(\Lambda_0)=\frac{d}{d\tau}\left(\frac{\partial\Lambda_0}{\partial\dot{x}^j}\right)-\frac{\partial\Lambda_0}{\partial x^j}=m^2c^2\Lambda_0^{-1}\left[g_{hj}\left(\frac{D\dot{x}^h}{D\tau}\right)-\Lambda_0^{-1}\left(\frac{d\Lambda_0}{d\tau}\right)g_{hj}\dot{x}^h\right], \tag{6.3}$$

where $D\dot{x}^h/D\tau$ denotes the absolute derivative of $\dot{x}^h$ with respect to τ. If, in particular, the parameter τ is identified with the proper time σ of the particle, that is,

$$d\tau^2 = d\sigma^2 = g_{hj}dx^h dx^j, \tag{6.4}$$

then $\Lambda_0 = mc$, and (6.3) reduces to

$$E_j(\Lambda_0) = mcg_{hj}(D\dot{x}^h/D\sigma). \tag{6.5}$$

As regards the interaction Lagrangian Λ_I the situation is somewhat less obvious. In addition to the conditions (i) and (ii) it is stipulated (iii) that Λ_I does not contribute terms involving higher derivatives of $\dot{x}^j$ to $E_j(\Lambda)$; (iv) that in the absence of an external electromagnetic field Λ_I is the derivative with respect to τ of some scalar function $\sum$ [so that $E_j(\Lambda_I) = 0$ identically whenever $F_{hj} = 0$]; (v) that Λ_I is invariant under the general duality rotation (5.3) (so that the resulting equations of motions possess this property).

Condition (iii) implies that $\partial^2\Lambda_I/\partial\dot{x}^j\partial\dot{x}^h = 0$, and thus, by (ii), that

$$\Lambda_I = Z_j\dot{x}^j, \tag{6.6}$$

where $Z_j(x^h)$ is an as yet unspecified vector field. We therefore introduce the 1-form

$$\omega = Z_j dx^j = \Lambda_I d\tau, \tag{6.7}$$

whose exterior derivative is

$$d\omega = \tfrac{1}{2}(Z_{j,h} - Z_{h,j})dx^h \wedge dx^j, \tag{6.8}$$

which can be expressed as

$$d\omega = \tfrac{1}{2}E_j(\Lambda_I)d\tau \wedge dx^j. \tag{6.9}$$

According to condition (iv), we must have $\omega = d\sum$, that is, $d\omega = 0$, whenever $F_{hj} = 0$. This requirement is certainly satisfied if there exist parameters α, β such that

$$d\omega = \tfrac{1}{2}(\alpha F_{hj} - \beta F^*_{hj})dx^h \wedge dx^j. \tag{6.10}$$

If it assumed that $d\omega$ does indeed possess this form, then, by condition (v), the coefficient of $dx^h \wedge dx^j$ in $d\omega$ must be invariant under the duality rotation (5.3). Thus, if $\tilde{\alpha}, \tilde{\beta}$ denote the transforms of α and β under (5.3), it is necessary that

$$(\tilde{\alpha}\tilde{F}_{hj} - \tilde{\beta}\tilde{F}^*_{hj}) = (\tilde{\alpha}\cos\theta - \tilde{\beta}\sin\theta)F_{hj} - (\tilde{\alpha}\sin\theta + \tilde{\beta}\cos\theta)\tilde{F}^*_{hj}$$

$$= \alpha F_{hj} - \beta F^*_{hj},$$

yielding

$$\tilde{\alpha} = \alpha\cos\theta + \beta\sin\theta, \quad \tilde{\beta} = -\alpha\sin\theta + \beta\cos\theta. \tag{6.11}$$

This transformation has a structure which is identical with that of (5.9), and accordingly it would be consistent to ultimately identify α and β with constant multiples of ϵ and γ respectively, that is, to set

$$\alpha = k\epsilon, \quad \beta = k\gamma. \tag{6.12}$$

Let us write (6.10) in the form

$$d\omega = \tfrac{1}{2}\Omega \tag{6.13}$$

where we have put

$$\Omega = \tfrac{1}{2}(\alpha F_{hj} - \beta F^*_{hj})dx^h \wedge dx^j. \tag{6.14}$$

Obviously the most crucial aspect of the identification (6.13) is the resulting requirement that

$$d\Omega = 0, \tag{6.15}$$

which, in turn, is equivalent to

$$\alpha(F_{hj,k} + F_{jk,h} + F_{kh,j}) - \beta(F^*_{hj,k} + F^*_{jk,h} + F^*_{kh,j}) = 0. \tag{6.16}$$

In order to analyze this condition, we substitute from (2.4), noting at the same time that, by virtue of (3.5),

$$f_{hj,k} + f_{jk,h} + f_{kh,j} = 0, \quad b^*_{hj,k} + b^*_{jk,h} + b^*_{kh,j} = 0 \tag{6.17}$$

identically. Hence (6.16) reduces to

$$M_{hj,k} + M_{jk,h} + M_{kh,j} = 0, \tag{6.18}$$

where we have written

$$M_{hj} = i\alpha b_{hj} - \beta f^*_{hj}. \tag{6.19}$$

But (6.18) implies the existence of a vector field $W_j(x^h)$ such that

$$M_{hj} = W_{j,h} - W_{h,j}, \tag{6.20}$$

which is clearly a condition on the potential fields ψ_h and ϕ_h that is not necessarily inherent in the field equations. However, unless this requirement is satisfied, there cannot exist an interaction Lagrangian Λ_I as specified by (6.7) and (6.13).[20]

Therefore, before proceeding, let us glance at the implications of (6.20) more closely. Clearly (6.20) may be expressed as

$$\sqrt{g}\,M^{*hj} = \tfrac{1}{2}i\epsilon^{hjlk}M_{lk} = i\epsilon^{hjlk}W_{k|l}, \tag{6.21}$$

from which it follows that

$$\frac{\partial}{\partial x^j}(\sqrt{g}\,M^{*hj}) = (\sqrt{g}\,M^{*hj})_{|j} = \tfrac{1}{2}\epsilon^{hjlk}(W_{k|l|j} - W_{k|j|l}) = 0,$$

where, in the last step, the cyclic identity satisfied by the curvature tensor has been used. Conversely, if this condition is satisfied, there exists a totally skew-symmetric relative tensor field W^{hjl} such that $\sqrt{g}\,M^{*hj} = \partial W^{hjl}/\partial x^l$. However, any such field W^{hjl} may be expressed in the form $i\epsilon^{hjlk}W_k$, with $3!\,W_k = i\epsilon_{hjlk}W^{hjl}$, which guarantees the existence of a field W_k such that (6.21) holds. Thus a necessary and sufficient condition that (6.20) be satisfied is represented by

$$(\sqrt{g}\,M^{*hj})_{|j} = 0. \tag{6.22}$$

On the other hand, it follows from (6.19) that

$$\sqrt{g}\,M^{*hj} = i\alpha\sqrt{g}\,b^{*hj} + \beta\sqrt{g}\,f^{hj}, \tag{6.23}$$

so that the field equations (2.37) and (2.38) imply

$$(\sqrt{g}\,M^{*hj})_{|j} = \beta J^h - \alpha S^h. \tag{6.24}$$

Thus the crucial condition (6.22), *and hence also* (6.15), *is always satisfied in vacuo*, or more generally, when J^h, S^h are related according to

$$\beta J^h - \alpha S^h = 0. \tag{6.25}$$

Because of (5.2) and (6.11) the left-hand side of (6.25) is invariant under the duality rotation (5.3); moreover, it is remarkable that the tensor M_{hj} as defined by (6.19) is invariant under the induced duality rotations (5.5) and (5.6), while this is not true for the standard duality rotation. Also, because of (2.35) and (2.36),

$$(\sqrt{g}\,M^{hj})_{|j} = 0$$

identically.

Assuming, then, that (6.22) is satisfied, it is concluded by means of (6.8) and (6.14) that the identification (6.13) yields

$$Z_{j,h} - Z_{h,j} = \alpha F_{hj} - \beta F^*_{hj}. \tag{6.26}$$

In view of (2.4), (6.19), and (6.20) this is equivalent to

$$Z_{j,h} - Z_{h,j} = (\chi_{j,h} - \chi_{h,j}) + (W_{j,h} - W_{h,j}),$$

where we have put

$$\chi_j = \alpha \psi_j + \beta \phi_j, \tag{6.27}$$

whose invariance under the duality rotation (5.4) follows directly from (6.11). Thus

$$Z_j = \chi_j + W_j + \Phi_{,j}, \tag{6.28}$$

where Φ is an arbitrary scalar field. This vector field is to be substituted in (6.6) to yield the final explicit expression for the interaction Lagrangian Λ_I, where it should be noted that $\Phi_{,j}$ may be omitted since it does not contribute to $E_j(\Lambda_I)$.

More directly, it is evident from (6.8), (6.9), and (6.26) that

$$E_j(\Lambda_I) = (\alpha F_{hj} - \beta F^*_{hj})\dot{x}^h. \tag{6.29}$$

This holds for any choice of the parameter τ; however, if we again identify τ with σ in accordance with (6.4), the derivatives $\dot{x}^h$ will represent the components of the 4-velocity of our particle. A combination of (6.4) with (6.29) now yields the final form of the equations of motion, namely

$$E_j(\Lambda) = 0, \tag{6.30}$$

where

$$E_j(\Lambda) = mcg_{hj}(D\dot{x}^h/D\sigma) + c^{-1}(\alpha F_{jh} - \beta F^*_{jh})\dot{x}^h, \tag{6.31}$$

which, by virtue of our construction, is tensorial and invariant under the general duality rotation (5.3).

In flat space–time, with $g_{hj} = -\delta_{hj}$, it is found with the aid of the representation (2.42) that Eqs. (6.30) assume the form

$$\frac{d}{dt}\left(\frac{m\mathbf{v}}{(1 - v^2/c)^{1/2}}\right) = \alpha(\mathbf{E} + c^{-1}\mathbf{v}\times\mathbf{H}) + \beta(\mathbf{H} - c^{-1}\mathbf{v}\times\mathbf{E}) \tag{6.32}$$

for $j = 1,2,3$. Because of the homogeneity of Λ [which implies that $E_j(\Lambda)\dot{x}^j = 0$ identically], the fourth member of (6.30) is a direct consequence of the three equations (6.32): in fact, it is merely the energy conservation law:

$$\frac{d}{dt}\left(\frac{mc^2}{(1 - v^2/c^2)^{1/2}}\right) = (\alpha\mathbf{E} + \beta\mathbf{H})\cdot\mathbf{v}. \tag{6.33}$$

In view of (6.11), and the classical expression for the Lorentz force, the parameters α and β are to be identified with the electric and magnetic charge respectively of the test particle.

In conclusion let us briefly glance at the explicit form of the condition (6.18) in flat space–time for the case of a static field as defined by (4.8). From (6.19), (6.23), and (4.9) it follows that under these circumstances

$$M^*_{\mu\nu} = 0, \quad M^*_{\nu 4} = -i\alpha U_{,\nu} + i\beta V_{,\nu}.$$

Thus $M_{4\nu} = 0$, while the vector $\mathbf{M} = (M_{23}, M_{31}, M_{12})$ is given by

$$\mathbf{M} = -\nabla(\alpha U - \beta V). \tag{6.34}$$

The condition (6.18) then reduces to

$$\alpha\nabla^2 U - \beta\nabla^2 V = 0,$$

which is automatically satisfied by the solutions of (4.16) and (4.17). Thus, in this case, the validity of the equations of motion (6.32) does not depend on any additional conditions (as was to be expected from the general theory above). Indeed, it is precisely within a context of this kind that these equations of motion are applied by Schwinger,[21] who considers the motion of a test particle with electric and magnetic charges α, β in the field generated by a stationary body possessing charges ϵ and γ. It is remarkable that the conserved angular momentum vector of Schwinger may, by virtue of (6.34), by expressed (in the nonrelativistic limit) as

$$\mathbf{r}\times m\mathbf{v} + c^{-1}\mathbf{M}r^2,$$

which may shed some further light on the physical significance of the tensor M_{hj}.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. David Lovelock and Dr. Marlan O. Scully, both of the University of Arizona, for valuable discussions.

*This research was sponsored in part by NSF G P 32830.

[1] J. Schwinger, (a) Phys. Rev. D 12, 3105 (1975); (b) Science 165, 757 (1969).

[2] H. Lamb, *Hydrodynamics* (Cambrdige U. P., London and New York, 1932), 6th ed., p. 248.

[3] H. Rund, *Topics in Differential Geometry*, edited by H. Rund and W. F. Forbes (Academic, New York, 1976), p. 111.

[4] Reference 1 (b).

[5] Latin indices $j, h, k, \cdots$ range from 1 to 4; the summation convention is operative throughout with respect to these suffixes. Partial derivatives with respect to positional coordinates are denoted by commas, e.g., $\psi_{j,h} = \partial\psi_j/\partial x^h$.

[6] Here, and in the following, the vertical bar denotes covariant differentiation, the latter being defined as usual in terms of the Christoffel symbols Γ^j_{lk} associated with the metric g_{hj}.

[7] This step depends tacitly upon the assumption that $(F\times F)$, which is proportional to $[\det(F_{hj})]^{1/2}$, does not vanish. If this assumption is avoided, the resulting analysis is considerably more complicated and involves an examination of the null space of F^{hj}. See Ref. 3 for the treatment of the case $J^k = 0$, $S^k = 0$.

[8] For the case of a flat space–time this representation reduces to one suggested by N. Cabibbo and E. Ferrari [Nuovo Cimento 23, 1147 (1962)] on the grounds that the counterparts of (2.33) and (2.34) can be solved by means of such a decomposition.

[9] Here, and in the sequel, Greek indices $\alpha, \beta, \gamma, \cdots$ range from 1 to 3; the summation convention is operative in respect of these indices also.

[10] H. Rund, Abh. Math. Sem. Univ. Hamburg 29, 243 (1966); D. Lovelock and H. Rund, *Tensors, Differential Forms, and Variational Principles* (Wiley-Interscience, New York 1975), pp. 302–4. The calculations given in these references are concerned solely with the case when $L = L(g_{hk}; \psi_h; \psi_{h,k})$; however, the derivation of (3.3), when L depends on an additional

field ϕ_h and its derivatives $\phi_{h,k}$, follows precisely the same pattern, and is therefore suppressed here.

[11]G.Y. Rainich, Trans. Am. Math. Soc. **27**, 106 (1925).

[12]Reference 10, p. 313

[13]A.S. Eddington, *The Mathematical Theory of Relativity* (Cambridge U.P., Cambridge, 1952), 2nd ed., p. 85.

[14]Reference 13, p. 185—86.

[15]Clearly (4.26) represents a generalization of the well-known Reissner—Nordström solution, to which it reduces when $\gamma = 0$ (Ref. 13, p. 185.)

[16]Reference 1; also D. Zwanziger, Phys. Rev. **176**, 1489 (1968).

[17]These seem to have appeared for the first time in Ref. 11.

[18]H. Rund, *The Hamilton—Jacobi Theory in the Calculus of Variations* (Van Nostrand, London and New York, 1966; Krieger reprint, New York, 1973), pp. 143 and 180.

[19]Reference 18, p. 179.

[20]This difficulty does not arise in classical electrodynamics, for by putting $\beta = 0$ in (6.14), and disregarding b_{hj} in (6.16), the latter condition is simply reduced to the first identity in (6.17). It would appear that this phenomenon is closely related to a similar difficulty discussed in detail by F. Rohrlich [Phys. Rev. **150**, 1104 (1966)].

[21]Reference 1(b), p. 738.

On Wentzel's proof of the quantization condition for a single-well potential

Nanny Fröman and Per Olof Fröman

Institute of Theoretical Physics, University of Uppsala, Uppsala, Sweden
(Received 2 August 1976)

Wentzel's elegant method for deriving the quantal generalization of the Sommerfeld quantization condition, which has been criticized by other authors, is justified and put on an irrefutable basis. Clarifying comments on some questions related to the discussion in the present paper are also made.

1. INTRODUCTION

Wentzel (cf. pp. 519—21 in Ref. 1) has devised an elegant method for deriving, in the JWKB approximation of any order, the asymptotic quantal generalization of Sommerfeld's condition for bound states in a single-well potential. It was, however, pointed out by Kramers (cf. p. 836 in Ref. 2) that a further establishment of Wentzel's proof is needed. Later Dunham,[3] who repeated and strengthened Kramers's criticism of Wentzel's proof (cf. pp. 715 and 718 in Ref. 3), used essentially Zwaan's[4] method to justify Wentzel's result. More recently Wentzel's method of proving the quantization condition has been used by other authors (cf. p. 186 in Ref. 5, pp. 384—5 in Ref. 6, and p. 630 in Ref. 7) but without the desired clarification of the proof. The purpose of the present paper is to clarify Wentzel's proof by filling certain gaps in it. In this way one can justify Wentzel's proof completely and thus in a way reject the criticism raised against it. Unless otherwise stated, our notation follows those used by Wentzel.[1]

As preparation we discuss in Sec. 2 the representation in the complex x plane of wavefunctions, which tend to zero as the independent variable x tends, along the real axis, to one of the limits $-\infty$ or $+\infty$, but which need not correspond to eigenfunctions (i.e., which need not be zero at both $-\infty$ and $+\infty$). This discussion is related to results obtained by Birkhoff (cf. pp. 696—700 in Ref. 8), Kemble (cf. pp. 558—60 in Ref. 9 and pp. 574—5 in Ref. 10), Seifert (cf. Satz 1 on pp. 175—6 in Ref. 11), Olver (cf. p. 809 in Ref. 12 and Secs. 2 and 4 in Ref. 13), and N. Fröman and P.O. Fröman [cf. Eqs. (3.41), (4.4a), and (4.4b) in Ref. 14; cf. also Ref. 15].

In Sec. 3 we describe in passing how the single-valuedness of the wavefunction has been used by Sommerfeld (cf. pp. 160—1 and 166 in Ref. 16) and by Kemble (cf. pp. 559 and 560 in Ref. 9 and pp. 105, 107, and 574—5 in Ref. 10) for a very simple derivation of Wentzel's quantal generalization of the Sommerfeld quantization condition.

Our analysis of Wentzel's own proof of this quantization condition is then given in Sec. 4.

In Wentzel's quantization condition [cf. Eq. (12) in Ref. 1] the integrand in the left-hand member is not real in the classically allowed region of the real axis, and the right-hand member is proportional to an integer, k. For the first-order JWKB approximation, Kramers[2] wrote the quantization condition in a more convenient form, where the integrand on the left-hand side is given by a simpler expression, which is real in the classically allowed region of the real axis, and the right-hand side is proportional to a half-integer, $k+\frac{1}{2}$. Dunham (cf. pp. 719—20 in Ref. 3) transformed Wentzel's general quantization condition involving higher-order JWKB approximations into a generalization of Kramer's half-integer quantization condition. Dunham considered, however, only the first few higher-order JWKB corrections. The transformation was later made quite general by Urban[5] who considered JWKB approximations of any order (cf. pp. 186—7 in Ref. 5). Starting from Wentzel's quantization condition with all terms included, and using a transformation of the sum of the terms with odd indices in the JWKB series obtained by N. Fröman [cf. Eq. (5b) in Ref. 15], we give in Sec. 5 an alternative derivation of the half-integral quantization condition involving phase-integral approximations of arbitrary order. This quantization condition is the same as the quantization condition derived by Dunham and Urban (cf. above).

2. REPRESENTATION OF WAVEFUNCTIONS VANISHING AT $-\infty$ OR AT $+\infty$ BY JWKB FUNCTIONS OR PHASE-INTEGRAL FUNCTIONS IN CERTAIN REGIONS OF THE COMPLEX X PLANE FAR AWAY FROM THE CLASSICALLY ALLOWED REGION

Consider a single-well potential and assume that for the function $p^2(x)$ in Wentzel's[1] Eq. (1), the zeros off the real axis as well as the singularities lie far away from the classically allowed region of the real axis. Cutting the complex x plane between the two classical turning points, i.e., the two pertinent real zeros of $p^2(x)$, we choose $p(x)$ to be positive on the upper lip of the cut. Whether the bottom of the potential is approximately parabolic or not, it follows from the general behavior in the complex x plane of the level lines of the function $|\exp[i\int^x p(x)dx]|$ that, if one excludes a certain region around the classically allowed region and also a more remote region containing the further zeros and singularities of $p(x)$, which may possibly exist, and considers a certain band encircling the classically allowed region, there is in this band a region L to the left in which any point x can be joined by a path from $-\infty$ along which $\exp[-i\int^x p(x)dx]$ increases as one moves from $-\infty$ to x, and there is a region R to the right in which any point x can be joined by a path from $+\infty$ along which $\exp[-i\int^x p(x)dx]$ increases as one moves from $+\infty$ to x, and for L and R there is a common region above as well as below the real axis. See Fig. 1.

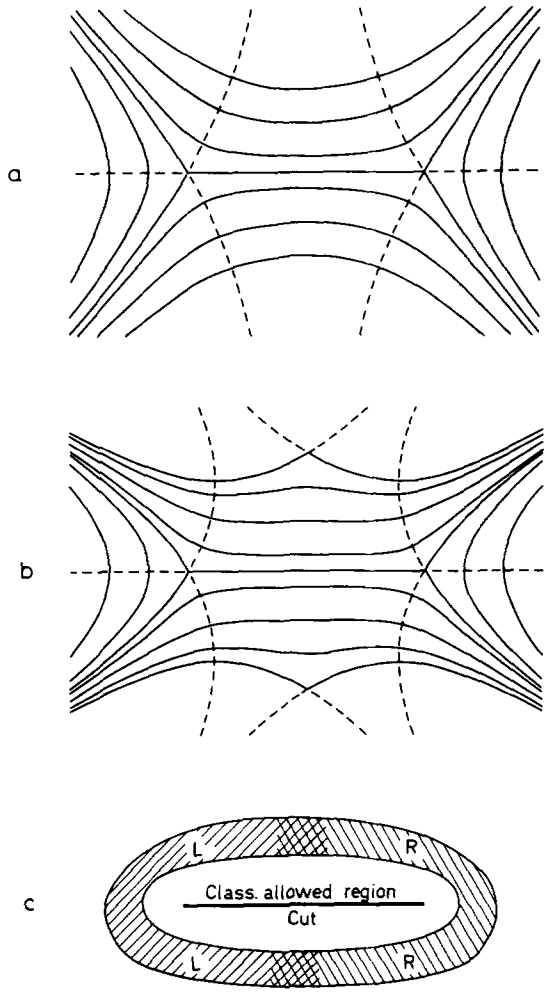


FIG. 1. The full-drawn curves in Fig. 1a (which refers to the case of a harmonic oscillator) and Fig. 1b (which refers to the case of a quartic oscillator) are anti-Stokes' lines, i.e., level lines of the function $|\exp\{i\int^x p(x)\,dx\}|$, where $p(x)$ is the function in Wentzel's[1] Eq. (1). The dashed lines are Stokes' lines. In Fig. 1c the regions L (indicated by ////) and R (indicated by ////) are shown.

Recalling results obtained by Birkhoff (cf. pp. 696–700 in Ref. 8) and by Seifert (cf. Satz 1 on pp. 175–6 in Ref. 11), one realizes that, even if the energy does not correspond to a bound state, a solution ψ_L of Wentzel's[1] Eq. (1) tending to zero as $x \to -\infty$ is, except for a constant factor, in the region L asymptotically represented by the expression

$$\exp\left[\int^x \frac{2\pi i}{h}\sum_{\nu=0}^{\infty}\left(\frac{h}{2\pi i}\right)^{\nu} y_\nu(x)\,dx\right], \tag{1}$$

provided that we choose the sign of $y_0(x)$ so that $y_0(x) = -p(x)$ [cf. Eq. (6) in Ref. 1]. Similarly, a solution ψ_R tending to zero as $x \to +\infty$ is, except for a constant factor, in the region R asymptotically represented by the expression (1) with the same choice of sign for $y_0(x)$ as above.

To make Wentzel's proof free from objections it is, as will be seen in Sec. 4, sufficient to use what has already been said above. At this point we shall, however, make some further comments related to the contents of the present section but of interest also in other contexts.

In the first-order JWKB approximation with $y_0(x) = -p(x)$, expression (1) becomes [cf. Eq. (9) in Ref. 1]

$$\exp\left[\int^x \frac{2\pi i}{h}\sum_{\nu=0}^{1}\left(\frac{h}{2\pi i}\right)^{\nu} y_\nu(x)\,dx\right]$$

$$= \text{const}\cdot p^{-1/2}(x)\exp\left(-\frac{2\pi i}{h}\int^x p(x)\,dx\right). \tag{1'}$$

From what we have said above, it is, in particular, clear that a solution ψ_L of Wentzel's[1] Eq. (1), which tends to zero as $x \to -\infty$, is approximately equal to the expression (1'), except for a constant factor, in the region L, and that a solution ψ_R which tends to zero as $x \to +\infty$ is approximately equal to the expression (1'), except for a constant factor, in the region R. With quantitative estimates of the approximations involved, this result can also be obtained from work by Olver (cf. p. 809 in Ref. 12 and Secs. 2 and 4 in Ref. 13) and by N. Fröman and P.O. Fröman [cf. Eqs. (3.41), (4.4a), and (4.4b) in Ref. 14].

The first-order JWKB approximation is the same as the lowest order of certain phase-integral approximations introduced by N. Fröman,[15] the higher orders of which are related to the higher-order JWKB approximations in the way described on pp. 453–4 in Ref. 17. For the applications to certain physical problems which often arise, for instance, in connection with the radial Schrödinger equation, it is important that there now exists a satisfactory and simple procedure for modifying these phase-integral approximations of arbitrary order (cf. Ref. 18 and Ref. 19, pp. 126–9). When these (unmodified or modified) phase-integral approximations are used, one can prove, with quantitative estimates, according to the methods developed in Ref. 14 (cf. also Ref. 15) that a wavefunction (ψ_L or ψ_R, respectively) which tends to zero as $x \to -\infty$ or as $x \to +\infty$, respectively, is, apart from a constant factor, in a region similar to the above mentioned region L or R, respectively, approximately given by the expression

$$q^{-1/2}(x)\exp\left[-i\int^x q(x)\,dx\right], \tag{1''}$$

where $q(x)$ is the truncated series (16) in Ref. 19 with the appropriate choice of sign for $q(x)$. In the case of unmodified phase-integral approximations of the order $2N+1$ the function $q(x)$ is given by the truncated series $-(2\pi i/h)\sum_{n=0}^{N}[h/(2\pi i)]^{2n} y_{2n}(x)$ containing only terms $y_\nu(x)$ with even indices ν; cf. (1). Note the differing sign conventions for the functions $y_{2n}(x)$ and $q(x)$.

When the energy is not equal to an eigenvalue, the previously defined solutions $\psi_L(x)$ and $\psi_R(x)$ are quite different and linearly independent, but yet, after having been multiplied by convenient constant factors, these solutions can both be represented by the same truncated JWKB expression (1) or the same phase-integral expression (1'') in the common parts of the regions L and R (cf. Fig. 1). Therefore, the Wronskian of $\psi_L(x)$ and $\psi_R(x)$, which is different from zero since $\psi_L(x)$ and $\psi_R(x)$ are linearly independent, is obviously not obtained correctly when calculated by means of (1) or (1''). To

explain this fact we note that, when $\psi_L(x)$ and $\psi_R(x)$ are represented by (1) or (1″) in the regions common to L and R, the relative errors are small compared to unity, whereas the absolute errors are not small compared to the values which $\psi_L(x)$ and $\psi_R(x)$, respectively, assume in the classically allowed region of the real axis, since, in the common parts of the regions L and R, the values of $|\psi_L(x)|$ and $|\psi_R(x)|$ are large compared to the corresponding values in the classically allowed region. The Wronskian of $\psi_L(x)$ and $\psi_R(x)$, which can be obtained when $\psi_L(x)$ and $\psi_R(x)$ are approximately known in the classically allowed region, can therefore not be obtained correctly when $\psi_L(x)$ and $\psi_R(x)$ are represented by (1) or (1″) in the common parts of the regions L and R. The calculation of this Wronskian is important in connection with the normalization of the wavefunction [cf. Eq. (6) in Ref. 20; also cf. Refs. 21 and 22] and in problems concerning the probability density at the origin.[23]

3. USE OF THE SINGLE-VALUEDNESS OF THE WAVEFUNCTION FOR DERIVING WENTZEL'S QUANTIZATION CONDITION IN A VERY SIMPLE WAY

The exact wavefunction ψ, corresponding to an eigenvalue of the energy, shall tend to zero when $x \to -\infty$ as well as when $x \to +\infty$. In the case of a single-well potential this eigenfunction ψ must, except for a constant factor, be asymptotically represented by the expression (1) in the whole band consisting of the regions L and R (cf. the previous section). Since the *exact* wavefunction is single-valued also around the classical turning points, its asymptotic expression (1) must be single-valued when x moves, for example, one turn around both classical turning points. This requirement was used by Sommerfeld (cf. pp. 160—1 in Ref. 16) and by Kemble (cf. pp. 558—9 in Ref. 9 and pp. 574—5 in Ref. 10) for deriving Wentzel's quantization condition, i.e., Eq. (12) in Ref. 1, where k now appears as an unspecified integer which is not *a priori* related to the number of nodes of the eigenfunction. Being based on the representation of the bound state wavefunction by expression (1) at some distance from the classically allowed region, such a proof is to some extent related to, although more straightforward than, Wentzel's proof, which will be discussed in the next section. When the quantization condition is fulfilled, expression (1) is single-valued in the whole band consisting of the regions L and R, and this simple expression represents there the bound state wavefunction $\psi(x)$. It is, in fact, only when x approaches the classically allowed region between the classical turning points that one needs to use a linear combination of both JWKB functions to represent the bound state eigenfunction $\psi(x)$.

4. JUSTIFICATION OF WENTZEL'S PROOF

Consider the *exact* wavefunction $\psi_k(x)$ for a bound state of a quantal particle of energy E_k in a one-dimensional single-well potential. It is well known (cf. p. 630 in Ref. 17) that the solutions of the one-dimensional wave equation have no zeros outside the real axis, and that the zeros on the real axis are all simple zeros situated in the classically allowed region. Furthermore, if the index k enumerating the energy eigenstates, which are all nondegenerate, is chosen such that $E_0 < E_1 < E_2 < \cdots$, the eigenfunction $\psi_k(x)$ has precisely k (≥ 0) such simple zeros, but no singularities, in a region of the complex x plane from which we exclude the remote region containing possibly existing poles of the potential. In the region thus considered the wavefunction $\psi_k(x)$ is a regular analytic function of the complex variable x with precisely k zeros, which are all simple. In the same region the quotient $\psi_k'(x)/\psi_k(x)$ is thus an analytic function of x with simple poles at the k zeros of $\psi_k(x)$, and at each one of these poles the residue is equal to 1. If C is a closed contour which lies in the region of the complex x plane under consideration and encircles the k zeros of $\psi_k(x)$ in the *positive* sense, we thus have the formula

$$\frac{1}{2\pi i}\oint_C \frac{\psi_k'(x)}{\psi_k(x)}\,dx = k, \tag{2}$$

which is *exact*, since $\psi_k(x)$ is assumed to be an *exact* eigenfunction.

We can choose the contour C to enclose the two classical turning points and to pass everywhere far away from the classically allowed region, so that we can represent the eigenfunction $\psi_k(x)$ by the asymptotic expression (1) everywhere on the contour of integration C. From (1) it then follows that the asymptotic formula

$$\frac{\psi_k'(x)}{\psi_k(x)} \sim \frac{2\pi i}{h}\sum_{\nu=0}^{\infty}\left(\frac{h}{2\pi i}\right)^{\nu} y_\nu(x) \tag{3}$$

is valid on the whole contour C. Since the series on the right-hand side of (3), when truncated, is slowly varying on the contour C, it is not dangerous to use such a truncated expression for approximately calculating the integral in the exact quantization condition (2). We can thus insert (3) into (2), getting Wentzel's asymptotic quantization condition [Eq. (12) in Ref. 1]

$$\oint_C \sum_{\nu=0}^{\infty}\left(\frac{h}{2\pi i}\right)^{\nu} y_\nu(x)\,dx = kh, \tag{4}$$

where, as previously mentioned, $y_0(x) = -p(x)$ with $p(x)$ positive on the upper lip of the cut along the classically allowed interval of the real axis, and where the path of integration C encircles the classically allowed region in the positive sense. In (4) the contour C can be deformed conveniently provided that it does not cross any branch cuts or singularities of the functions $y_\nu(x)$.

Although not as simple and direct as the proof described in the previous section, Wentzel's proof has the merit of showing clearly that the quantum number k is equal to the number of nodes of the eigenfunction.

5. TRANSFORMATION OF WENTZEL'S QUANTIZATION CONDITION INTO A SIMPLER FORM

In Wentzel's quantization condition (4) the terms y_0, y_2, $\cdots$ as well as the terms y_1, y_3, $\cdots$ appear in the integrand in the left-hand member. Because of the presence of the latter terms, the integrand is not real in the classically allowed region of the real axis. As we have already mentioned in the Introduction, Dunham[3] and Urban[5] have transformed Wentzel's quantization condition into the now well-known form, in which the integrand in the left-hand member contains only the terms

$y_0, y_2, \cdots$ and is real in the classically allowed region of the real axis, and the right-hand member is proportional to the half-integer quantum number $k+\frac{1}{2}$. The argumentation used by Dunham and Urban can be explained as follows. From Wentzel's[1] Eqs. (9) and (6) it follows that $y_1(x) = -\frac{1}{4}d[\ln p^2(x)]/dx$, and recalling that $p^2(x)$ has precisely two simple zeros inside the contour C defined in the previous section, one thus realizes that $\oint_C y_1(x)\,dx = -\pi i$. Since, furthermore, it can be shown that for any odd value of $\nu \geq 3$ the function $y_\nu(x)$ can be written as the derivative with respect to x of an analytic function which is single-valued on the contour C, one gets the formula

$$\oint_C \sum_{n=0}^{\infty} \left(\frac{h}{2\pi i}\right)^{2n+1} y_{2n+1}(x)\,dx = -\tfrac{1}{2}h. \tag{5}$$

Using this formula, one can write the quantization condition (4) in the form

$$\oint_C \sum_{n=0}^{\infty} \left(\frac{h}{2\pi i}\right)^{2n} y_{2n}(x)\,dx = (k+\tfrac{1}{2})h, \tag{6}$$

where the asymptotic series in the integrand on the left-hand side is real in the classically allowed region of the real axis. By reversing the sense of integration along the path C and at the same time changing the signs of the functions $y_{2n}(x)$ in (6), one obtains this quantization condition in a form corresponding to Eqs. (10.22) in Ref. 14 and (13) in Ref. 15.

We shall now demonstrate another way of transforming Wentzel's quantization condition, i.e., Eq. (4) in the present paper, into the form (6). Following N. Fröman (cf. pp. 542–3 in Ref. 15), we realize that

$$\sum_{\nu=0}^{\infty}\left(\frac{h}{2\pi i}\right)^{\nu} y_\nu(x) = \sum_{n=0}^{\infty}\left(\frac{h}{2\pi i}\right)^{2n} y_{2n}(x) + \frac{h}{2\pi i}\frac{d}{dx}\left\{\ln\left[\sum_{n=0}^{\infty}\left(\frac{h}{2\pi i}\right)^{2n} y_{2n}(x)\right]^{-1/2}\right\}. \tag{7}$$

According to Eq. (6) in Ref. 1 we have $y_0^2(x) = p^2(x)$. The function $p^2(x)$ is single-valued and regular in the region of the complex x plane under consideration, and in this region the only zeros of $p^2(x)$ are the classical turning points, which are simple zeros of $p^2(x)$. The truncated series $\sum_{n=0}^{N}[h/(2\pi i)]^{2n} y_{2n}(x)/y_0(x)$, which has no branch points in the region of the complex x plane under consideration, has a certain number of simple zeros in the neighborhood of each classical turning point, and at the two classical turning points it has poles, the order of each pole being equal to the number of zeros in the neighborhood of the classical turning point in question (cf. pp. 456–9 in Ref. 17). Hence this truncated series is single-valued in the region in question of the complex x plane, and it is, furthermore, close to unity far away from the classical turning points. For the reasons mentioned we get [cf. (7)]

$$\begin{aligned}\oint_C \frac{h}{2\pi i}\frac{d}{dx}\left\{\ln\left[\sum_{n=0}^{N}\left(\frac{h}{2\pi i}\right)^{2n} y_{2n}(x)\right]^{-1/2}\right\}dx &= \frac{h}{2\pi i}\oint_C \frac{d}{dx}\left[\ln y_0^{-1/2}(x)\right]dx \\ &= -\frac{h}{8\pi i}\oint_C \frac{d}{dx}[\ln p^2(x)]\,dx = -\tfrac{1}{2}h. \end{aligned} \tag{8}$$

Inserting (7) into (4), and using (8) with $N=\infty$, we perform, in an alternative way, the transformation of Wentzel's quantization condition into the form (6).

Using the symmetric phase-integral approximations mentioned in Sec. 2, N. Fröman[15] has in another way directly derived the quantization condition in a form equivalent to (6).

[1] G. Wentzel, Z. Phys. **38**, 518–29 (1926).
[2] H. A. Kramers, Z. Phys. **39**, 828–40 (1926).
[3] J. L. Dunham, Phys. Rev. **41**, 713–20 (1932).
[4] A. Zwaan, *Intensitäten im Ca-Funkenspektrum*, *Academisch Proefschrift* (Joh. Enschedé en Zonen, Haarlem, 1929). Also published in Arch. Néer. Sci. Exactes, Ser. IIIA **XII**, 1–76 (1929).
[5] P. Urban, Acta Phys. Austriaca **6**, 181–94 (1952).
[6] T. Tietz, Acta Phys. Hungary **XIV**, 383–5 (1962).
[7] L. Bertocchi, S. Fubini, and G. Furlan, Nuovo Cimento **35**, 599–632 (1965).
[8] G. D. Birkhoff, Bull. Am. Math. Soc. **39**, 681–700 (1933).
[9] E. C. Kemble, Phys. Rev. **48**, 549–61 (1935).
[10] E. C. Kemble, *The Fundamental Principles of Quantum Mechanics* (McGraw-Hill, New York, 1937); unabridged and corrected republication (Dover, New York, 1958).
[11] H. Seifert, Math. Z. **48**, 173–92 (1942–43).
[12] F. W. J. Olver, Proc. Cambridge Philos. Soc. **57**, 790–810 (1961).
[13] F. W. J. Olver, J. Res. Nat. Bur. Stand. Sec. B **69**, 271–90 (1965).
[14] N. Fröman and P. O. Fröman, *JWKB-Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965).
[15] N. Fröman, Ark. Fys. **32**, 541–8 (1966).
[16] A. Sommerfeld, *Atombau und Spektrallinien. Wellenmechanischer Ergänzungsband* (Vieweg, Braunschweig, 1929).
[17] N. Fröman, Ann. Phys. (N.Y.) **61**, 451–64 (1970).
[18] N. Fröman and P. O. Fröman, Ann. Phys. (N.Y.) **83**, 103–7 (1974).
[19] N. Fröman and P. O. Fröman, Nuovo Cimento B **20**, 121–32 (1974).
[20] S. Yngve, J. Math. Phys. **13**, 324–31 (1972).
[21] S. Yngve, J. Math. Phys. **12**, 114–7 (1971).
[22] P. O. Fröman, Ann. Phys. (N.Y.) **88**, 621–30 (1974).
[23] N. Fröman and P. O. Fröman, Phys. Rev. A **6**, 2064–7 (1972).

Symmetry and exact dyon solutions for classical Yang–Mills field equations*

J. P. Hsu and E. Mac

Center for Particle Theory, University of Texas at Austin, Austin, Texas 78712
(Received 19 September 1975; final revised manuscript received 19 March 1976)

We show that the generalized electromagnetic field tensor $\bar{F}_{\mu\nu}$ and the magnetic and electric charges in non-Abelian gauge theories have little to do with the Higgs scalars and/or the dynamics of the Lagrangian. They are consequences of a symmetry in the theory. We present several exact static dyon solutions to the nonlinear classical field equations in both massless and massive Yang–Mills theories, which possess both electric and magnetic charges. The implications of $\bar{F}_{\mu\nu}$ are also discussed.

1. INTRODUCTION

Yang and Mills emphasized that local isospin gauge symmetry leads to gauge fields B^a_μ and completely determines the interaction dynamics of the gauge fields.[1] Several particular solutions to the classical Yang–Mills field equations have been discussed.[2–5] Recently, it was pointed out that because of the *local* isospin gauge symmetry, one can introduce a unit isovector, e.g., $v^a = r^a/r$, to connect the gauge field B^a_μ and the Abelian electromagnetic field tensor $\bar{F}_{\mu\nu}$ associated with the magnetic monopole in such a way that (a) $\bar{F}_{\mu\nu}$ is isospin gauge invariant and (b) $\bar{F}_{\mu\nu}$ reduces to the usual electromagnetic field tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ if $v^a = (0, 0, 1)$.[6] Because $v^a(x)$ is not a dynamical quantity, the theory consists of kinematics and dynamics. The kinematics possesses the symmetry in question, the dynamics does not. The symmetry referred to and used was that of the kinematics throughout.

In this paper, we consider the static solutions with $B^a_0 \neq 0$ and we obtain the dyon solution, having both magnetic and electric charges. We also find an exact static solution B^a_μ for which the total energy of the system is finite, but the solutions B^a_μ are nonpointlike and the dyon carries an imaginary electric charge. It is shown that the conserved magnetic and electric charges in non-Abelian theories have nothing to do with the Higgs scalars and/or the symmetry structure of the Lagrangian. Rather, they are a consequence of the symmetry of the kinematics part of the theory. A massive Yang–Mills theory is also considered; it is found that in this case the dyon solution exists too.

2. SPECIAL STATIC SPHERICALLY SYMMETRIC SOLUTIONS

The classical field equations for the Yang–Mills field B^a_μ are[1,7]

$$B^a_{\mu\nu} = \partial_\mu B^a_\nu - \partial_\nu B^a_\mu + e\epsilon^{abc} B^b_\mu B^c_\nu, \quad B^\mu_3 \equiv A^\mu, \tag{1}$$

$$\partial^\nu B^a_{\mu\nu} + e\epsilon^{abc} B^\nu_b B^c_{\mu\nu} = 0, \tag{2}$$

$$\partial^\mu B^a_\mu = 0. \tag{3}$$

We look for the static spherically symmetric solution of the form,[3]

$$B^a_i = \epsilon_{iab} v^b(\mathbf{r}) B(r), \quad i, a, b = 1, 2, 3, \tag{4}$$

$$B^a_0 = v^a(\mathbf{r}) G(r), \quad v^a(\mathbf{r}) = r^a/r. \tag{5}$$

The constraint equation (3) is automatically satisfied. Equation (2) can be written as

$$\partial_i\partial_i B - \frac{2B}{r^2} - \frac{3eB^2}{r} - e^2B^3 + \frac{eG^2}{r}(1 + erB) = 0, \tag{6}$$

$$\partial_i\partial_i G - \frac{2G}{r^2} - \frac{4eBG}{r} - 2e^2B^2G = 0. \tag{7}$$

Introducing the variable $C(r) = 1 + erB(r)$, we have

$$r^2\frac{d^2C}{dr^2} = C(C^2 - 1) - e^2r^2G^2C, \tag{8}$$

$$\frac{d}{dr}\left(r^2\frac{dG}{dr}\right) = 2GC^2. \tag{9}$$

The trivial solution $C = 0$ to (8) leads to the nontrivial special solution,

$$B(r) = -1/(er), \quad G(r) = G_0/r + G_1, \quad G_0 = e, \tag{10}$$

where G_0 and G_1 are constants of integration and $G_0 = e$ because of the requirement $\bar{F}_{\mu\nu} = F_{\mu\nu}$ when $v^a = (0, 0, 1)$.

We try to find a solution B which is regular at $r = 0$. By a stroke of luck, we find the following solution to (8) and (9) (note that here C and G can be either positive or negative),

$$\begin{aligned} C(r) &= \alpha r/\sin\alpha r, \\ G(r) &= i(\alpha r\cos\alpha r - \sin\alpha r)/(er\sin\alpha r), \end{aligned} \tag{11}$$

where α is an arbitrary constant. Unfortunately, these solutions are undefined as $r \to \infty$ if α is real. However, if α is imaginary, i.e., $\alpha = i\beta$ with β real, we have an exact solution to Eqs. (8) and (9),

$$\begin{aligned} C(r) &= \frac{\beta r}{\sinh\beta r} \quad \text{or} \quad B(r) = \frac{\beta r - \sinh\beta r}{er\sinh\beta r}, \\ G(r) &= \frac{i}{er\sinh\beta r}(\beta r\cosh\beta r - \sinh\beta r), \end{aligned} \tag{12}$$

which can be easily verified. These solutions are regular at both $r = 0$ and $r = \infty$.[8] We note that such a finite solution for arbitrary r is possible because B^a_0 [or $G(r)$] is nonvanishing. If B^a_0 is set equal to zero, Eq. (8) becomes $r^2d^2C/dr^2 = C(C^2 - 1) \approx C^3$ in the region $C^2 \gg 1$, say $r \approx r_f$, where we have the asymptotic solution $C \approx \sqrt{2}\, r_f/(r_f - r)$, $r \approx r_f$. In this case, the nontrivial solution $C(r)$ diverges at some finite $r = r_f$.[5]

On the other hand, if one looks for the static solution of the form

$$B_i^a = r^i f^a(r)/r, \quad B_0^a = g^a(r), \tag{13}$$

one has the Ikeda—Miyachi solution[2]

$f^a(r) = A^a/r^2$, A^a = integration constants,

$$g^a(r) = \left(1 - \frac{k}{r}\right)\left[a'^a + b'^a \cos\left(\frac{eA}{r}\right) + c'^a \sin\left(\frac{eA}{r}\right)\right],$$

$$k = \text{const}, \tag{14}$$

where $A \equiv |\mathbf{A}|$, $a'^a = \text{const} A^a$, $|\mathbf{b}'| = |\mathbf{c}'|$, and $\mathbf{a}'$, $\mathbf{b}'$, and $\mathbf{c}'$ form a right-handed orthogonal system. It is interesting to note that if $\mathbf{A} \times \mathbf{g} \neq 0$ the vector $\mathbf{g}$ appears to rotate around the $\mathbf{A}$ axis, with the angle between $\mathbf{A}$ and $\mathbf{g}$ constant everywhere, when we travel from one pointe to another point in space.

3. LOCAL GAUGE SYMMETRY AND THE DYON

What is the "physical" meaning of these classical solutions? One may regard the 6-vector $B_{\mu\nu}^a$ as composed by isovector "electric" and isovector "magnetic" fields[3]:

$$E_j^a = B_{j0}^a, \quad H_k^a = \tfrac{1}{2}\epsilon_{kij} B_{ij}^a, \tag{15}$$

which do not correspond to physical reality in nature. It would be more interesting and meaningful if $B_{\mu\nu}^a$ could be related to the usual electromagnetic fields. Indeed, this is possible because of the local gauge symmetry.[6]

We introduce a local unit vector $v^a(x)$ $(v^a v^a = 1)$, which transforms like an isovector under the local SU(2) transformation and we construct a generalized electromagnetic field tensor $\bar{F}_{\mu\nu}$ more closely related to $F_{\mu\nu}$, so that we may actually split the 6-vector $\bar{F}_{\mu\nu}$ into the usual electric and magnetic field associated with interesting objects such as the magnetic monopoles or the dyons. Their existence has been speculated in relation to the fundamental problem of charge quantization and discussed by many authors.[9] To accomplish this in the Yang—Mills theory with $B_3^\mu \equiv A^\mu$, it is natural that $\bar{F}_{\mu\nu}$ satisfies the requirements: (i) $\bar{F}_{\mu\nu}$ is gauge invariant and (ii) $\bar{F}_{\mu\nu}$ can be reduced to the usual $F_{\mu\nu}$ when $v^a = (0, 0, 1)$. We define[6]

$$\begin{aligned} &\bar{F}_{\mu\nu} = v^a B_{\mu\nu}^a - e^{-1}\epsilon^{abc} v^a (D_\mu v^b)(D_\nu v^c), \quad v^a = v^a(x), \\ &v^a v^a = 1, \quad (D_\mu v^b) = \partial_\mu v^b + e\epsilon^{bcd} B_\mu^c v^d, \end{aligned} \tag{16}$$

which is indeed gauge invariant, and $\bar{F}_{\mu\nu} = F_{\mu\nu}$ when $v^a = (0, 0, 1)$. Now, the electric and the magnetic fields, E_j and H_k, are related to $\bar{F}_{\mu\nu}$ by

$$E_j = \bar{F}_{j0}, \quad H_k = \tfrac{1}{2}\epsilon_{kij}\bar{F}_{ij}. \tag{17}$$

Definition (16) is interesting because E_j and H_k in (17) can be interpreted as the fields produced by the magnetic monopole or the dyons. Since $v^a v^a = 1$, we may write (16) as

$$\bar{F}_{\mu\nu} = \partial_\mu (v^a B_\nu^a) - \partial_\nu (v^a B_\mu^a) - e^{-1}\epsilon^{abc} v^a (\partial_\mu v^b)(\partial_\nu v^c). \tag{18}$$

For completeness one might mention that the components of B_μ^a perpendicular to v^a in isospace form a charged vector meson relative to the components of B_μ^a along v^a as electromagnetic field.

If the theory involves a scalar field ϕ^a, one usually defines[10] $\bar{F}_{\mu\nu}$ in (16) with v^a replaced by $\phi^a/|\phi|$, as proposed by 't Hooft.[9] We stress that the ratio $\phi^a/|\phi|$ has nothing to do with dynamics, although ϕ^a is determined by dynamics. The reason is that one looks for solutions of the form $\phi^a = r^a \xi(r)$, where $\xi(r)$ is the only quantity determined dynamically. Therefore, $\phi^a/|\phi|$ has nothing to do with $\xi(r)$ or dynamics, just like the unit isovector v^a. Note that the form $\phi^a = r^a \xi(r)$ is determined by symmetry consideration alone.

To see the sources which generate the field $\bar{F}_{\mu\nu}$, let us define the magnetic current[10] k_λ and the electric current j_μ:

$$k_\lambda = \tfrac{1}{2}\epsilon_{\lambda\rho\mu\nu}\partial^\rho \bar{F}^{\mu\nu}, \tag{19}$$

$$j_\mu = \partial^\nu \bar{F}_{\mu\nu}. \tag{20}$$

They are both obviously conserved currents,

$$\partial^\lambda k_\lambda = 0, \quad \partial^\mu j_\mu = 0, \tag{21}$$

and hence the magnetic charge M and the electric charge Q are conserved,

$$\begin{aligned} &\frac{d}{dt} M = \frac{d}{dt}\int k_0 \, d^3r/(4\pi) = 0, \\ &\frac{d}{dt} Q = \frac{d}{dt}\int j_0 \, d^3r/(4\pi) = 0. \end{aligned} \tag{22}$$

When the fields B_μ^a are free from singularity lines, we have the identity

$$\epsilon_{\lambda\rho\mu\nu}\partial^\rho [\partial_\mu (v^a B_\nu^a) - \partial_\nu (v^a B_\mu^a)] = 0 \tag{23}$$

and the magnetic current k_λ can be written as[9]

$$k_\lambda = -(1/2e)\epsilon_{\lambda\rho\mu\nu}\epsilon^{abc}\partial^\rho v^a \partial^\mu v^b \partial^\nu v^c. \tag{24}$$

The relations (18)—(24) *hold for arbitrary local unit isovector* $v^a(\mathbf{r}, t)$, $v^a v^a = 1$. It is striking that these results are the consequence of the symmetry embedded in (16) and have nothing to do with the interaction dynamics of the system. We note that, in fact, symmetry is the most basic concept in gauge theories. The concept of symmetry leads to the gauge fields B_μ^a and completely determines the interaction dynamics of B_μ^a.[1,3] In the same sense, we may regard the dyon as the consequence of local isospin gauge symmetry.

In general, the magnetic charge M is given by

$$M = \frac{1}{4\pi}\int k_0 \, d^3r = -\frac{1}{8\pi e}\int \epsilon_{ijk}\epsilon_{abc}\partial_i [v^a (\partial_j v^b)(\partial_k v^c)] \, d^3r. \tag{25}$$

We require that $v^a v^a = 1$ and that $v^a(x)$ is single valued. We can write M as

$$M = -\frac{3!}{8\pi e}\int \frac{\partial(v_1, v_2, v_3)}{\partial(r_1, r_2, r_3)} d^3r. \tag{26}$$

This equals the integration over a unit sphere in isospace multiplied by a wrapping number n,

$$M = -(3/4\pi e)\int_{v^2 \leq 1} n \, d^3v = -n/e, \tag{27}$$

where n is an integer. In particular, for fixed constant v^a we have $M = 0$. For the case[10]

$$\begin{aligned} &v_1 = ar_1/A, \quad v_2 = r_2 r_3/A, \quad v_3 = [(r_3)^2 - a^2]/A, \\ &A \equiv [(ar_1)^2 + (r_2 r_3)^2 + ((r_3)^2 - a^2)^2]^{1/2}, \end{aligned} \tag{28}$$

we have the total magnetic charge $M=-2/e$. In this context, it is worth pointing out that the magnetic charge density k_0 in (25) vanishes everywhere except at the monopole position where the space derivatives of v^a have δ type singularities. This position corresponds to the zeros of the Higgs field $\phi^a(x)$, as discussed by 't Hooft.[9] We emphasize that the form $\phi^a(x)=R^a(x)\xi(x)$ with $R^a=\delta_{a3}$ or $R^a=r^a/r$ is specified before the field equation for ϕ^a is solved. Usually we determine the explicit form of $R^a(x)$ by some sort of symmetry considerations. We may remark that although (25) does not involve gauge fields one should not regard gauge fields as having nothing to do with the magnetic charge.[10] (See Sec. 7.)

4. THE DYONS

For the special type of solution given by (4) and (5), the definition (16) with $v^a=r^a/r$ leads to

$$\mathbf{E}=\nabla G(r), \tag{29}$$

$$\mathbf{H}=-\mathbf{r}/(er^3). \tag{30}$$

The fields E_j and H_k related to solution (10) are

$$\mathbf{E}=-e\mathbf{r}/r^3, \quad \mathbf{H}=-\mathbf{r}/(er^3). \tag{31}$$

This shows the presence of a stable dyon[11] at $\mathbf{r}=0$ with an electric charge e and a magnetic charge $g=1/e$, satisfying the Schwinger condition $eg=1$.[9]

One may wonder why such a dyon with a point source is possible. The answer is that the solution (10) is singular at $r=0$ and, therefore, the solution (10) has a δ function type source[3] S^a_μ defined by

$$S^a_\mu \equiv \partial^\nu B^a_{\mu\nu}+e\epsilon^{abc}B^\nu_b B^c_{\mu\nu} \propto \begin{cases}\epsilon_{\mu ab}v^b\delta^3(\mathbf{r}), & \mu=1,2,3, \quad v^b=r^b/r,\\ v^a\delta^3(\mathbf{r}), & \mu=0.\end{cases}$$

However, the magnetic charge g in (31) has nothing directly to do with this source S^a_μ. (See Sec. 3.)

The solution (12) leads to the following electric and magnetic fields:

$$\mathbf{E}=\frac{i}{e}\frac{\mathbf{r}}{r^3}\left(1-\frac{\beta^2r^2}{(\sinh\beta r)^2}\right), \quad \mathbf{H}=\frac{-\mathbf{r}}{er^3} \tag{32}$$

Unfortunately, this corresponds to a dyon with a point magnetic charge surrounded by a cloud of imaginary electric charge without a pointlike core. Thus, the particular solution (12) is, in contrast to the solutions (10) and (14), not physically meaningful.[8]

The Ikeda–Miyachi solution (14) gives

$$\mathbf{E}=\mathbf{r}k(\mathbf{r}\cdot\mathbf{g})/[r^4(1-k/r)]+\mathbf{g}/r-\mathbf{r}(\mathbf{r}\cdot\mathbf{g})/r^3 +\mathbf{r}(1-k/r)eA[\mathbf{r}\cdot\mathbf{b}'\sin(eA/r)-\mathbf{r}\cdot\mathbf{c}'\cos(eA/r)]/r^4, \tag{33}$$

$$\mathbf{H}=-\mathbf{r}/(er^3)+\mathbf{A}\times\mathbf{r}/r^4. \tag{34}$$

The magnetic flux is

$$\int \mathbf{H}\cdot d\mathbf{s}=-4\pi/e \tag{35}$$

because the second term in (34) does not originate from the magnetic charge. The electric field E is a rapidly oscillating function of r as $r\to 0$, and there is no simple picture for this solution.

5. STATIC ENERGY

If one defines the static energy E_s of the system by $E_s=-L_{\mathrm{YM}}$ (where L_{YM} is the Yang–Mills Lagrangian[1]), as suggested by 't Hooft,[9] one has

$$E_s=\int d^3r\left[\tfrac{1}{4}B^a_{ij}B^a_{ij}-\tfrac{1}{2}B^a_{0i}B^a_{0i}\right] =\frac{4\pi}{e^2}\int_0^\infty dr\left[\left(\frac{dC}{dr}\right)^2+\frac{(C^2-1)^2}{2r^2}-e^2G^2C^2-\frac{e^2}{2}\left(\frac{dG}{dr}\right)^2r^2\right] \tag{36}$$

which is divergent for the solutions (10) and (11). The Ikeda–Miyachi solution (13) and (14) also leads to a divergent energy. The remarkable feature of the solution (12) is that it gives a positive and finite energy

$$E_s=\frac{4\pi}{e^2}|\beta|, \quad \left(\frac{d}{dr}\left[Gr^2\frac{dG}{dr}\right]=r^2\left(\frac{dG}{dr}\right)^2+2G^2C^2\right), \tag{37}$$

where β is real. We note that direct calculation of the Hamiltonian[1] H_s of the system shows that $H_s=0$ for the solution (12). Also, in general β in (12) may be complex with $\mathrm{Re}\beta\neq 0$. In this case, we have $E_s=4\pi\beta/e^2$ for $\mathrm{Re}\beta>0$, $E_s=-4\pi\beta/e^2$ for $\mathrm{Re}\beta<0$ and $H_s=0$.

6. MASSIVE YANG-MILLS FIELDS

From the experimental viewpoint, if the dyon exists, it is probably massive. We wish to point out that the above discussions also hold for a theory involving massive Yang–Mills fields $f^a_\mu(x)$. Let us consider the Lagrangian[12]

$$\begin{aligned}\mathcal{L}=&-\tfrac{1}{4}\mathbf{f}_{\mu\nu}\cdot\mathbf{f}^{\mu\nu}+\tfrac{1}{2}M_f^2\mathbf{f}^2_\mu+\tfrac{1}{2}\partial_\mu U\partial^\mu U+\tfrac{1}{2}\partial_\mu\boldsymbol{\phi}\cdot(\partial^\mu+e\mathbf{f}^\mu\times)\boldsymbol{\phi}\\ &-\tfrac{1}{2}e\mathbf{f}_\mu\cdot(U\partial^\mu\boldsymbol{\phi}-\boldsymbol{\phi}\partial^\mu U)+\tfrac{1}{8}e^2\mathbf{f}^2_\mu(\boldsymbol{\phi}^2+U^2)+\tfrac{1}{2}eM_f\mathbf{f}^2_\mu U\\ &+M_f\boldsymbol{\phi}\cdot\partial_\mu\mathbf{f}^\mu-\tfrac{1}{2}\xi(\partial_\mu\mathbf{f}^\mu+M_f\boldsymbol{\phi}/\xi)^2,\end{aligned} \tag{38}$$

$$\mathbf{f}_{\mu\nu}=\partial_\mu\mathbf{f}_\nu-\partial_\nu\mathbf{f}_\mu+e\mathbf{f}_\mu\times\mathbf{f}_\nu,$$

which is, except the last gauge-fixing term with the parameter ξ, invariant under distorted local SU(2) gauge transformation.[13] The Lagrangian (38) leads to field equations for $f^a(x)$, $\phi^a(x)$, and $U(x)$, and $\phi^a(x)=0$ is a trivial solution. In analogy with (4) and (5), we set $f^a_i=\epsilon_{iab}v^b\bar{B}(r)$, $f^a_0=v^a\bar{G}(r)$, $v^a=r^a/r$. We find that

$$\nabla^2\bar{B}-\frac{2}{r^2}\bar{B}-\frac{3e}{r}\bar{B}^2-e^2\bar{B}^3+\frac{e\bar{G}^2}{r}(1+er\bar{B})-\frac{e^2}{4}\bar{B}\left(U+\frac{2M_f}{e}\right)^2=0, \tag{39}$$

$$\nabla^2\bar{G}-\frac{2}{r^2}\bar{G}-\frac{4e}{r}\bar{B}\bar{G}-2e^2\bar{B}^2\bar{G}-\frac{e^2}{4}\bar{G}\left(U+\frac{2M_f}{e}\right)^2=0, \tag{40}$$

$$\nabla^2U+\frac{e^2}{4}(\bar{G}^2-2\bar{B}^2)(U+2M_f/e)=0. \tag{41}$$

By inspection, we see that $U=-2M_f/e$ is a trivial solution to Eq. (41), and Eqs. (39) and (40) reduce to Eqs. (6) and (7) respectively. Therefore, all subsequent discussions hold equally well for the Lagrangian (38) involving massive Yang–Mills fields.

7. REMARKS AND CONCLUSIONS

In the solutions (10) and (14), B_0^a may be zero and the nonvanishing B_i^a leads to the magnetic monopole.[6] When $B_0^a = 0$, the static particlelike solution with positive and finite energy such as (12) has not been found.[5,6] We conjecture that the nonvanishing B_0^a is necessary for a positive and finite energy E_s.

We note that $\bar{F}_{\mu\nu}$ in (16) is invariant under isospin rotation. However, when one identifies the local unit isovector v^a with r^a/r, (16) must be understood to be invariant under the *combined isospin and space rotation*.[6] Also, although (16) leads to zero magnetic charge (i.e., reduces to the usual $F_{\mu\nu}$) if one chooses a *fixed constant* isovector $v^a = \delta_{a3}$; yet when one chooses $v^a = r^a/r$ and rotates it to $v^a = \delta_{a3}$ one will get the same magnetic charge e^{-1} with the type of solution given by (4) and (5). This comes about because the solution of Yang–Mills fields will also be rotated and become singular as discussed in Ref. 10.

The proposition of the definition (16) for the electromagnetic field tensor is not a direct dynamical consequence of the Lagrangian in non-Abelian gauge theory. The isovector defining the gauge invariant field tensor (16) need not be a Higgs field but might have some other origin without having any dynamical nature. The local unit isovector $v^a(x)$ is not a dynamical field because it does not obey any field equation. In Ref. 10 the conserved magnetic charge in non-Abelian gauge theories is regarded as the consequence of the topological structure of three Higgs scalar fields in a three-dimensional space. However, from the above discussions, both the existence and the conservation of the electric charge Q and the magnetic charge M, as shown in (22) and (27), have nothing to do with the presence of Higgs scalars. Rather, they are related to the basic concept of local gauge symmetry. We note that $v^a(x)$ is an absolute element because it is not a dynamical variable, even though it transforms. Thus, the local SU(2) symmetry of the present theory is broken due to the presence of $v^a(x)$ as an absolute element.[14] This is related to the fact that all the quantities in a theory having transformation laws under a group does not mean the theory is symmetric under the group.[15] On the other hand, $v^a(x)$ could be treated as a dynamical field by putting appropriate terms, i.e., $\frac{1}{2}(\partial^\nu v^a + e\epsilon^{abc}B_b^\nu v^c)^2$ in the Lagrangian. This term combined with a constraint $v^a(x)v^a(x) = \text{const}$ will lead to a mass for the gauge field B_μ^a and the constraint can be treated by the usual method of Lagrange multipliers.

*Work supported in part by the U.S. Atomic Energy Commision.

[1]C.N. Yang and R.L. Mills, Phys. Rev. 96, 191 (1954).
[2]M. Ikeda and Y. Miyachi, Progr. Theor. Phys. 29, 474 (1962); H.G. Loos, Nucl. Phys. 72, 677 (1965).
[3]T.T. Wu and C.N. Yang, in *Properties of Matter under Unusual Conditions*, edited by H. Mark and S. Fernbac (Interscience, New York, 1969), pp. 349–54.
[4]R.P. Treat, Nuovo Cimento 50, 871 (1967).
[5]G. Rosen, J. Math. Phys. 13, 595 (1972); 9, 996, 999 (1968).
[6]J.P. Hsu, Lett. Nuovo Cimento 14, 189 (1975).
[7]Equation (3) is the gauge condition (i.e., the Landau gauge) necessary for quantizing the Yang–Mills theory and it need not be gauge invariant. Of course, one may choose the gauge condition such as $\partial^\mu B_\mu^a + e\epsilon^{abc}B_b^\mu B_\mu^c = 0$ instead of (3); it is still satisfied by solutions of the form (4) and (5).
[8]J.P. Hsu, Phys. Rev. Lett. 36, 646 (1976). The solution is very much similar to that obtained by M.K. Prasad and C.M. Sommerfield, Phys. Rev. Lett. 35, 760 (1975). The complex solution (12) in SU(2) is not meaningful for it leads to imaginary total isospin of the system. However, it is interesting because it can help one find real solutions by analytic continuation or enlarging the gauge group as discussed by T.T. Wu and C.N. Yang, Phys. Rev. D 12, 3843 (1975). Prof. Yang informed one of us (J.P.) that, based on the solution (12), a real analytic solution of gauge field for the group SL(2,C) has been found and discussed by T.T. Wu and C.N. Yang, "A Static Sourceless Gauge Field," Preprint ITP-SB-76-2.
[9]P.A.M. Dirac, Proc. R. Soc. London A 133, 60 (1931); Phys. Rev. 74, 817 (1948); J. Schwinger, Phys. Rev. 144, 1087 (1966); C.N. Yang, Phys. Rev. D 1, 2360 (1970); G. 't Hooft, Nucl. Phys. B 79, 276 (1974); D.M. Stevens, "Magnetic Monopoles: An Updated Bibliography," Virginia Polytechnic Institute and State University Preprint VPI-EPP-73-5; A.M. Polyakev, Zh. Eksp. Teor. Fiz. Pis. Red. 20, 430 (1974) [JETP Lett. 20, 194 (1974)].
[10]J. Arafune, P.G.O. Freund, and C.J. Goebel, J. Math. Phys. 16, 433 (1975).
[11]J. Schwinger, Science 165, 757 (1969); B. Julia and A. Zee, Phys. Rev. D 11, 2227 (1975).
[12]J.P. Hsu and J.A. Underwood, Phys. Rev. D 12, 620 (1975); J.P. Hsu and E.C.G. Sudarshan, Nucl. Phys. B 91, 477 (1975).
[13]J.P. Hsu, Lett. Nuovo Cimento 11, 525 (1974).
[14]J.L. Anderson, *Principles of Relativity Physics* (Academic, New York, 1967), Chapter 4.
[15]For example, Newtonian mechanics can be written in arbitrary curvilinear coordinates using tensor analysis, but its symmetry group includes the Galilean group and not the Einstein group. The Einstein symmetry is broken by the presence of a certain connection and metric tensor as absolute elements of the theory.

Maxwell and Wien processes as special cases of the generalized Feller diffusion process

Siegfried H. Lehnigk

US Army Missile Command, Redstone Arsenal, Alabama 35809
(Received 28 July 1976; revised manuscript received 28 August 1976)

It is shown that Maxwell and Wien type processes are special cases of the generalized Feller diffusion process. In particular, both are obtained for specific parameter values from the delta function initial condition solution of the generalized Feller equation. For specific values of the independent time variable, one obtains the well-known distribution laws of Maxwell and Wien of statistical physics.

1. THE GENERALIZED FELLER EQUATION

In a recent paper,[1] this author discussed delta function initial condition solutions of the generalized (one-dimensional, autonomous, parabolic) Feller equation

$$l(u)=A(z)u_{zz}+B(z)u_z+C(z)u-u_t=0, \quad u=u(z,t), \tag{1.1a}$$

on the domain $z>0$, $t>0$, with coefficients

$$\begin{aligned} &A(z)=\alpha z^{\lambda+1}, \quad \alpha>0,\\ &B(z)=\beta_1 z^{\lambda}+\beta_2 z, \quad \beta_{1,2}\epsilon(-\infty,\infty),\\ &C(z)=\rho z^{\lambda-1}+\beta_2, \quad \rho=\lambda[\beta_1-\alpha(1+\lambda)]. \end{aligned} \tag{1.1b}$$

We assume in this paper that $\lambda<1$. The variable z will be interpreted as a length variable, t as a time variable. Since in the applications in Secs. 2 and 3 the parameter β_2 will not appear, we set from now on $\beta_2=0$ in (1.1b).

Because of the choice of ρ in the coefficient $C(z)$, Eq. (1.1) is of Fokker–Planck type, i.e., it can be written in the form ($\beta_2=0$)

$$l(u)=[A(z)u]_{zz}-[\tilde{B}(z)u]_z-u_t=0 \tag{1.2}$$

with

$$\tilde{B}(z)=2A'(z)-B(z)=[2\alpha(1+\lambda)-\beta_1]z^{\lambda}.$$

The functions $A(z)$ and $\tilde{B}(z)$ are called the diffusion and drift coefficients, respectively.

Equation (1.2) and its adjoint have the basic solution

$$V(z,t;y,s)=b^{\lambda}\xi^{-\lambda-\nu(1-\lambda)/2}\eta^{1+\lambda+\nu(1-\lambda)/2}Z_{\nu}(2i\zeta^{(1-\lambda)/2})\times\exp(-\xi^{1-\lambda}-\eta^{1-\lambda}),$$

in which

$$\begin{aligned} &\zeta=\xi\eta, \quad \xi=zb^{-1}, \quad \eta=yb^{-1},\\ &b=b(t,s)=[\alpha(1-\lambda)^2(t-s)]^{(1-\lambda)^{-1}},\\ &t>s\geq 0, \quad \nu=(1-\lambda)^{-1}(\alpha^{-1}\beta_1-1-2\lambda), \end{aligned}$$

and where $Z_{\nu}(x)$ is a cylinder function, i.e., a solution of the Bessel equation

$$u''+x^{-1}u'+(1-\nu^2x^{-2})u=0.$$

By saying that V is a basic solution we mean that, as a function of z and t, V is a solution of $l(u)=0$ and, as a function of y and s, V is a solution of the adjoint equation.

Using V, we form the function

$$\begin{aligned} V^*(z,t;y)&=(1-\lambda)b_0^{-1-\lambda}\eta_0^{-1-\lambda}V(z,t;y,0)\\ &=(1-\lambda)b_0^{-1}\xi_0^{-\lambda-\nu(1-\lambda)/2}\eta_0^{\nu(1-\lambda)/2}\mathcal{Z}_{\nu}(r_0)\\ &\quad\times\exp(-\xi_0^{1-\lambda}-\eta_0^{1-\lambda}), \end{aligned} \tag{1.3}$$

$$r_0=2\xi_0^{(1-\lambda)/2},$$

in which the subscript 0 signifies that $s=0$. By writing $\mathcal{Z}_{\nu}$ in (1.3) instead of the original Z_{ν}, we want to indicate that we have chosen particular cylinder functions, namely,

$$\mathcal{Z}_{\nu}=\begin{cases} I_{-\nu}(r_0) & (\nu<1),\\ I_{\nu}(r_0) & (\nu>0), \end{cases} \tag{1.4}$$

where $I_{\pm\nu}$ is a modified Bessel function of the first kind. In going from V to V^*, we have disregarded the constant factor $\exp(-\nu\pi i/2)$ which occurs if we specify $Z_{\nu}(ir_0)$ to be the Bessel function $J_{\nu}(ir_0)$.

The function $V^*(z,t;y)$ is a solution of $l(u)=0$ as a function of z and t. Details and a more general mathematical exposition of the properties of the generalized Feller equation will be presented elsewhere.

With $\mathcal{Z}_{\nu}$ given by (1.4) one can show[2] (formula 6.643.2) that

$$\int_0^{\infty} V^*(z,t;y)\,dy$$

converges since $\lambda<1$ by assumption.

Let $\delta(z-z_0)$ be the delta function applied at $z_0>0$. We consider the delta function initial condition solution

$$u_{\delta,0}(z,t)=\int_0^{\infty} V^*(z,t;y)\delta(y-z_0)\,dy=V^*(z,t;z_0) \tag{1.5}$$

of $l(u)=0$.

We concentrate our attention now on the case that V^* in (1.5) is taken for $\mathcal{Z}_{\nu}(r_0)=I_{-\nu}(r_0)$ with $\nu<1$. Then we may let $z_0\downarrow 0$ in (1.5). Using the asymptotic formula for $I_{-\nu}$ for small argument and observing (1.3), we obtain the positive solution

$$\begin{aligned} u_{\delta}(z,t)&=\lim_{z_0\to 0}u_{\delta,0}(z,t)\\ &=\frac{1-\lambda}{\Gamma(1-\nu)}b_0^{-1}\xi_0^{-\lambda-\nu(1-\lambda)}\exp(-\xi_0^{1-\lambda}) \end{aligned} \tag{1.6}$$

of $l(u)=0$, which corresponds to the delta function initial condition applied at $z=0$. This means that

$$u_{\delta}(z,t)\downarrow 0 \quad \text{as } t\downarrow 0, \ z>0,$$

independent of the parameters $\lambda < 1$ and $\nu < 1$. In other words, the solution $u_\delta(z,t)$ of $l(u)=0$ corresponds to a completely concentrated initial state at $z=0$.

Furthermore, $u_\delta(z,t)$ has the following properties:

(1) $\int_0^\infty u_\delta(z,t)\,dz \equiv 1, \quad \nu < 1 \ (\lambda < 1)$,

i.e., $u_\delta(z,t)$ is conservative and a source solution of $l(u)=0$ if $\nu < 1$ $(\lambda < 1)$.

(2) $u_\delta(z,t) \downarrow 0$ as $z \downarrow 0$, $t > 0$, $\lambda + \nu(1-\lambda) < 0$ $(\lambda < 1)$,

i.e., $u_\delta(z,t)$ is singular if $\lambda + \nu(1-\lambda) < 0$ $(\lambda < 1)$. If it is singular, it is also conservative $(\lambda < 1)$.

(3) If $u_\delta(z,t)$ is singular, then, for fixed $t > 0$ and as a function of z, $u_\delta(z,t)$ has exactly one maximum which is located at

$$z_m(t) = \{-\alpha[\lambda + \nu(1-\lambda)](1-\lambda)t\}^{(1-\lambda)^{-1}} \quad (\beta_2 = 0). \tag{1.7}$$

$z_m(t)$ and the velocity $\dot{z}_m(t)$ with which the maximum propagates are functions of the parameters λ and ν.

2. THE MAXWELL DISTRIBUTION LAW

In the coefficients of $l(u)=0$ we set $\lambda = -1$, $\alpha = t_0 kT/2m$, $\beta_1 = -2\alpha$ $(\beta_2 = 0)$, where

k = Boltzmann constant,

T = temperature,

m = mass of molecule,

t_0 = some characteristic time.

We observe that t_0 may be expressed as $t_0 = z_0 v_0^{-1}$ in terms of some characteristic length z_0 and some characteristic velocity v_0. Then (1.2) takes the form of a heat equation with constant diffusion coefficient and nonzero z-dependent drift coefficient,

$$[t_0 kT/2m]u_{zz} - [(t_0 kT/m)z^{-1}u]_z - u_t = 0.$$

Since $\nu = -\frac{1}{2}$, the function (1.6) multiplied by t_0 becomes

$$u_\delta(z,t) = \frac{4t_0}{\sqrt{\pi}} b_0^{-1}(t)\xi_0^2 \exp(-\xi_0^2),$$
$$b_0(t) = b(t,0) = \left(\frac{2t_0 kT}{m}t\right)^{1/2}, \quad \xi_0 = z b_0^{-1}(t). \tag{2.1}$$

It is of Maxwell type for every $t > 0$ as a function of z. It is singular and conservative, and

$$\int_0^\infty u_\delta(z,t)\,dz = t_0.$$

According to (1.7), the abscissa $z_m(t)$ of the maximum of the Maxwell type function (2.1) is

$$z_m(t) = \left(\frac{2t_0 kT}{m}\right)^{1/2} t^{1/2} = b_0(t), \tag{2.2}$$

and the propagation velocity $\dot{z}_m(t)$ of the maximum is

$$\dot{z}_m(t) = \left(\frac{t_0 kT}{2m}\right)^{1/2} t^{-1/2} > 0.$$

If we introduce the velocity variable $v = t_0^{-1}z$, where v denotes the magnitude of the velocity vector, we obtain from (2.1) for $t = t_0$ the well-known Maxwell distribution law[3]

$$P(v) = u_\delta(t_0 v, t_0) = \frac{4}{\sqrt{\pi}}\left(\frac{m}{2kT}\right)^{3/2} v^2 \exp\left(-\frac{m}{2kT}\right)v^2,$$

in which $P(v)$ has the physical unit probability/velocity interval dv.

Using (2.2) with $t = t_0$, we obtain the also well-known formula for the most probable velocity v_p,

$$v_p = t_0^{-1} z_m(t_0) = (2kT/m)^{1/2}.$$

The indicated transformation of variables may be directly applied by setting $u(z,t) = U(v,t)$, $z = t_0 v$, and leads to the equation

$$\frac{kT}{2mt_0}U_{vv} - \left[\frac{kT}{mt_0}v^{-1}U\right]_v - U_t = 0.$$

3. THE WIEN DISTRIBUTION LAW

In the coefficients of $l(u)=0$ we set now $\lambda = 0$, $\alpha = 2\pi ckT/\omega_0\hbar$, $\beta_1 = -2\alpha$ $(\beta_2 = 0)$, where the meaning of k and T is the same as in Sec. 2 and where

c = velocity of light,

$\hbar$ = Planck's constant,

ω_0 = some characteristic frequency.

Then (1.2) reduces to the special Feller equation

$$\left[\frac{2\pi ckT}{\omega_0\hbar}zu\right]_{zz} - \frac{8\pi ckT}{\omega_0\hbar}u_z - u_t = 0.$$

Here $\nu = -3$, and (1.6) multiplied by $12(kT)^4/\pi\omega_0^2 c^2\hbar^3$ becomes

$$u_\delta(z,t) = [2(kT)^4/\pi\omega_0^2 c^2\hbar^3]b_0^{-1}(t)\xi_0^3\exp(-\xi_0),$$
$$b_0(t) = b(t,0) = (2\pi ckT/\omega_0\hbar)t, \quad \xi_0 = z b_0^{-1}(t). \tag{3.1}$$

This function is of Wien type for every $t > 0$ as a function of z. It is singular and conservative, and

$$\int_0^\infty u_\delta(z,t)\,dz = 12(kT)^4/\pi\omega_0^2 c^2\hbar^3.$$

From (1.7) follows the abscissa of the maximum of the Wien type function (3.1) as

$$z_m(t) = (6\pi ckT/\omega_0\hbar)t = 3b_0(t), \tag{3.2}$$

which implies that the propagation velocity of the maximum is identically constant,

$$\dot{z}_m(t) = 6\pi ckT/\omega_0\hbar > 0.$$

We introduce now a frequency variable ω by setting $z = 2\pi c\omega_0^{-2}\omega$. Then, for $t = \omega_0^{-1}$, we obtain from (3.1) the well-known Wien distribution law (limiting case of the Planck law for high frequencies) (Ref. 3, p. 173),

$$Q(\omega) = u_\delta(2\pi c\omega_0^{-2}\omega, \omega_0^{-1}) = (\hbar/\pi^2 c^3)\omega^3\exp(-\hbar/kT)\omega,$$

in which $Q(\omega)$ has the physical unit energy density/frequency interval $d\omega$.

By means of (3.2), the most probable frequency ω_p is obtained as

$$\omega_p = (\omega_0^2/2\pi c)z_m(\omega_0^{-1}) = 3kT/\hbar.$$

Carrying out the change from z to ω in the original equation, we obtain

$$\omega_0 kT\hbar^{-1}\omega U_{\omega\omega} - 2\omega_0 kT\hbar^{-1}U_\omega - U_t = 0.$$

[1] S.H. Lehnigk, J. Math. Phys. 17, 973 (1976).

[2] I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1965), 4th ed., formula 6.643.2.

[3] L.D. Landau and E.M. Lifshitz, *Statistical Physics, Course of Theoretical Physics, Vol.* 5 (Pergamon, London, 1958), p. 83.

Nonexistence theorem for spherically asymmetric solutions of the Fermi–Thomas model of atoms and ions

David Eimerl

Physical Dynamics, Inc., P. O. Box 556, La Jolla, California 92038
(Received 5 August 1976)

It is proved that there are no spherically asymmetric solutions of the Fermi–Thomas equation $\nabla^2(u/r) = (u/r)^{3/2}\xi$, where ξ is unity inside the surface $u = 0$, and zero otherwise. This is proved by showing that the boundary conditions on u at the origin and on the surface at infinity necessarily lead to unphysical discontinuities in u at the surface $u = 0$, except for the case where u is spherically symmetric. A simple, only approximately statistical model is presented, which may replace the Fermi–Thomas model for asymmetric systems.

I. INTRODUCTION

The statistical model of Fermi[1] and Thomas[2] is a nonempirical method of determining atomic properties, which has been applied to several problems in atomic physics. It has been used with some success to calculate interatomic and interionic potentials, and also to calculate atomic energy levels and wavefunctions.[3,4] It is particularly useful for obtaining approximate wavefunctions for moderately excited single electron terms, as a complement to the Bates–Damgaard[5,6] method. It has also been used to calculate wavefunctions for any electron in a many-electron atom or ion, by providing an effective potential for the electron in question. The main advantage of the Thomas–Fermi method is that it is rapid and may be extended over the whole range of energies available to a bound electron. The self-consistent field methods of Hartree or Hartree–Fock are practicable only for the lower lying states, and the long numerical iteration schemes may be justified only by the increased accuracy of the results, compared to the Thomas–Fermi (TF) method.

In all these applications, the TF potential is chosen to be spherically symmetric. This has the obvious advantage of simplicity, and of giving a potential with no parameters to adjust (or at most one).[4] For single electron levels, this choice corresponds with the expected state of the core on which the tower of single electron terms is built. On the other hand, if it is known that the core for such a tower of states is asymmetric, the use of asymmetric TF potentials suggests itself, especially for comparing corresponding states in towers built on different cores. In fact, many of the tabulated atomic states[7] form towers of single electron levels built on asymmetric cores. A typical example is neutral oxygen, where all but a very few of the tabulated levels belong to one of three towers:

$$|O_1\rangle = |^4S^\circ_{3/2},\ nl\rangle, \tag{1.1a}$$

$$|O_2\rangle = |^2D^\circ_{3/2},\ nl\rangle, \tag{1.1b}$$

$$|O_3\rangle = |^2P^\circ_{3/2},\ nl\rangle. \tag{1.1c}$$

The cores for these towers correspond very closely with the ground states of nitrogen and O II.

Atomic cores may be asymmetric for a number of reasons, most obviously angular momentum specification and configuration interaction. The asymmetry is intimately related to the shell structure of the atom, stemming mainly from the imbalance in electron configuration in unfilled shells, and generally decreases with the number of electrons. Now the TF equation contains only a statistical assumption, and has in it no information on shell structure. (This is its main advantage). Consequently, it is probably a poorer approximation to the core, the larger the asymmetry is. On the other hand, it is not unreasonable to expect that if the asymmetry is not too large, the TF equation may still be a sufficiently good approximation to give useful quantitative results. Asymmetric TF potentials would be of physical interest in this context because of the simplicity and rapidity of their application.

Here the results of an investigation of asymmetric TF potentials are presented. An analytic theorem is proved, that the only solutions which satisfy the physical boundary conditions are those which are spherically symmetric. This somewhat surprising theorem is remarkable because such proofs for nonlinear differential equations are quite rare. The theorem shows that the lack of physical information in the TF equation is fatal; there are *no* asymmetric solutions. The potential of an asymmetric core must be obtained otherwise. Whereas this potential certainly does not satisfy the TF equation, it may satisfy an equation which differs from it by terms which are small, in some sense. Such asymmetric potentials would be only approximately statistical.

The next section reviews the TF model and proves the theorem, and the last section suggests a simple approximately statistical model of asymmetric cores.

II. THE THOMAS-FERMI MODEL

The TF model of atoms and ions is well known[8] and will only be reviewed here. In terms of the rescaled coordinate x,

$$\mathbf{x} = \mathbf{r}/\sigma a_0, \tag{2.1}$$

where

$$\sigma = (3\pi/4)^{2/3}(2Z^{1/3})^{-1} \tag{2.2}$$

the potential is

$$V = \frac{Ze}{4\pi\epsilon_0\sigma a_0}\left(\frac{u(x)}{x} + c\right), \tag{2.3}$$

where u satisfies the TF equation

$$\nabla^2(u/x) = (u/x)^{3/2} \equiv \rho. \tag{2.4}$$

The constant c represents the maximum energy that electrons in the ion may have; for $c=0$, there is no net charge, and the model describes an atom. The charge density, ρ, is normalized such that

$$Z\int \rho\, d^3x = N \leq Z, \tag{2.5}$$

where N is the number of electrons in the atom. The boundary condition at $x=0$ is

$$u(0)=1. \tag{2.6}$$

The boundary of the atom is that surface where the charge density vanishes. [The existence of this surface is actually an assumption based on physical considerations. However, its existence is shown below to follow from the condition in (2.5).] It is defined by

$$u(x_s)=0. \tag{2.7}$$

Clearly (2.4) requires u to be positive inside this surface, and therefore the gradient of u is negative at the boundary x_s.

$$\mathbf{n}\cdot\nabla u\big|_s \leq 0. \tag{2.8}$$

Outside this surface the potential satisfies Laplace's equation

$$\nabla^2(u/x)=0, \tag{2.9}$$

subject to the boundary condition that at large distances the potential tends to zero. Namely

$$\lim_{x\to\infty} u = -c|x| + q + O(1/|x|), \tag{2.10}$$

where q is the total charge on the ion,

$$q = 1 - N/Z. \tag{2.11}$$

Across the surface $u=0$, the potential and its gradient are continuous. The possibility that there may be a surface density of charge is rejected as unphysical, since at the boundary there is no phase space available, and ρ is regarded as a smooth function away from $x=0$.

The model is investigated here by considering two particular curves which extend out from the origin to infinity, along which u satisfies a differential equation with certain useful properties. The positivity of u and ρ is crucial to the argument, which shows that the boundary conditions on the derivative of u are not satisfied simultaneously at the origin, boundary, and at infinity, unless the solutions have spherical symmetry.

Consider the region where (2.4) applies, and assume that it contains no absolute maxima or minima in u, with the exception of the origin itself. Since u is continuous and single-valued, the surfaces

$$u=\text{const}=M \tag{2.12}$$

do not intersect each other. These surfaces cannot extend to infinity since then the total charge carried by the electrons diverges. The surface $u=0$ is a possible exception to this and may be at infinity. Consequently, the surfaces (2.12) are all closed, and therefore must contain an absolute extremum of u. By assumption, the origin is the only such point, and therefore it is enclosed by all these surfaces.

If the surfaces are ordered according to their minimum distance from the origin, $R_{\min}$, then a sequence of the M values associated with each surface is defined. This sequence must be monotonic, since otherwise at least one extremum of M ($R_{\min}$) exists, and on that surface u must be an absolute extremum, violating our assumption. In addition, M must be monotonically decreasing. For, if it were increasing, it would start from unity at the origin and increase, always remaining greater than unity. Again the charge would diverge. Thus M must decrease monotonically from unity at the origin, remaining positive everywhere. Thus on each surface $u=M$, the gradient of u is negative.

$$\mathbf{n}\cdot\nabla u\big|_M \leq 0. \tag{2.13}$$

Equation (2.8) is a special case of this formula.

On each surface there are at least two points where the distance from the origin is an extremum. The set of all these points form at least two lines extending from the origin out at least as far as the surface $u=M_0$. If, due to a symmetry of u, there is a degeneracy, it may be removed by choosing one of the lines of minima and one of the lines of maxima. By defining, in general

$$f = x\nabla^2 u, \tag{2.14}$$

it is shown in Appendix A that along the line of maxima, where u is denoted w, w satisfies the following equation:

$$d^2w/dr^2 = f(w) - p_1\, dw/dr, \tag{2.15}$$

where p_1 is a positive definite function, which is zero only if the surface $u=M_1$, whose maximum distance $R_{\max}$ is r, is spherical, and r is the distance of the point on the curve from the origin (not the distance along the curve). Similarly, along the line of minima, where u is denoted v, v satisfies

$$d^2v/dr^2 = f(v) + p_2\, dv/dr, \tag{2.16}$$

where p_2 is positive definite, and is zero only if the surface $u=M_2$, whose minimum distance $R_{\min}$ is r, is spherical. It is important that f for the TF equation is positive definite. By solving the TF equation for $x<x_0$, where x_0 is small but finite, Appendix B derives the boundary conditions on these functions:

$$v(0)=w(0)=1, \tag{2.17}$$

$$\frac{dv}{dr}(0)=\frac{dw}{dr}(0)=-g. \tag{2.18}$$

From the discussion preceding (2.13), g is a positive quantity, and the derivatives of w and v are either negative or zero.

This system of equations gives the function u on two curves. At a given value of r it is clear that

$$d^2w/dr^2 \geq d^2v/dr^2 \tag{2.19}$$

and since these functions satisfy identical boundary conditions at $r=0$, it follows that

$$dw/dr \geq dv/dr \tag{2.20}$$

and

$$w(r) \geq v(r). \tag{2.21}$$

The equality holds if and only if $p_1=p_2=0$, which is the spherically symmetric case.

The possibility that u has absolute maxima and minima may be discussed using these equations. It is proved

here that such extrema do not exist, so that the initial assumption of their absence is exhaustive, covering all cases of physical interest. An extremum may take one of two forms; either lying on a closed surface, or lying on an open surface, a line or at a point. In the former case the gradient of u vanishes on the closed surface and from (2.15) and (2.16) it follows that u must be a minimum on the surface, $u = u_{\min}$. Outside the surface $u > u_{\min}$ and the charge (2.5) diverges unless there is a maximum beyond the surface, so that u turns over and may tend to zero. In the latter case, the set of points where u is extremal are enclosed in closed surfaces, $u = M$. (If $M < 1$ then the surfaces $u = M$ may consist of two or more disjoint closed surfaces, one enclosing the origin, the other the extremal points of u.) There is an extremal line associated with these surfaces which satisfies (2.15) before the extremum is reached and (2.16) beyond the extremum. At the extremum the gradients vanish, and hence beyond the extremum (2.15) requires u to increase. Therefore, all such extrema are minima. However, if $w > 0$ and is increasing with r, then the charge diverges, unless a maximum is reached so that u may turn over and tend to zero. Thus, in all cases, whatever the form of the extremum, only minima exist, and the finiteness of the total charge requires that there are maxima beyond each minima. (Without such maxima u never reaches zero and remains finite as $x \to \infty$.) Therefore there are no extrema of u at all, other than at the origin.

The finiteness of the charge of the electrons requires that u go to zero at some finite surface or that u goes to zero faster than $x^{-3/2}$ at infinity. Thus the existence of the surface $u = 0$ is established.

In the case that the boundary surface $u = 0$, is not at infinity, the previous discussion may be applied to the region outside this surface, where $f = 0$. The equations corresponding to (2.15) and (2.16) may then be integrated once to obtain the derivatives of w and v:

$$dw/dr = dw/dr|_{R_{\max}} \exp\left(- \int_{R_{\max}}^{r} p_1 \, dr \right), \tag{2.22}$$

$$dv/dr = dv/dr|_{R_{\min}} \exp\left(\int_{R_{\min}}^{r} p_2 \, dr \right). \tag{2.23}$$

where $R_{\max}$ and $R_{\min}$ refer to the boundary surface. On letting $r \to \infty$ and using (2.10), one finds

$$dw/dr|_{R_{\max}} \leq -c \tag{2.24}$$

and

$$dv/dr|_{R_{\min}} \geq -c. \tag{2.25}$$

The equalities apply, as usual, only if u is spherically symmetric, outside the surface.

The final step is to match the boundary conditions at the boundary. From (2.20) and the positivity of the second derivative of w, it follows that

$$dw/dr|_{R_{\max}} \geq dw/dr|_{R_{\min}} \geq dv/dr|_{R_{\min}}. \tag{2.26}$$

But from (2.24) and (2.25),

$$dw/dr|_{R_{\max}} \leq dv/dr|_{R_{\min}}. \tag{2.27}$$

These equations are inconsistent unless the equalities apply, namely only for the spherically symmetric case.

If the boundary surface is at infinity, then the finiteness of the charge requires that the derivatives vanish faster than $x^{-5/2}$ at infinity. Then

$$\lim_{R_{\min} \to \infty} dw/dr|_{R_{\max}} - dv/dr|_{R_{\min}} = O(R_{\min})^{-5/2}, \tag{2.28}$$

where these variables refer to the surface $u = M$ and the limit is equivalent to $M \to 0$. But from (2.19) and (2.26) the left hand side is necessarily an increasing function of $R_{\min}$. Again the boundary conditions cannot be satisfied except for the case of spherical symmetry.

The only possibility not covered by these two cases is that where the boundary surface has a finite $R_{\min}$, but infinite $R_{\max}$. This case is highly unphysical, due to the enormous asymmetry, and one does not expect the TF model to apply. However, mathematically, it may be considered as a limiting case of a finite boundary, in which case the above analysis applies and shows that such solutions do not exist.

This completes the proof of the absence of spherically asymmetric solutions of the TF model.

III. DISCUSSION

The arguments presented here may be adapted trivially to the more general case

$$\nabla^2 \phi = \phi^{\nu} \tag{3.1}$$

where ν is any noninteger number. The crux of the proof is the positivity of the right-hand side and the positivity of ϕ. It does not apply to complex solutions of (3.1). Because the potential is coulombic around its only singularities, it must be spherically symmetric everywhere. This phenomenon is well known for linear differential equations, and has now been extended to the system (3.1) also.

The problem of determining the potential of an asymmetric atomic core cannot be solved using the TF model. However, a simple procedure for obtaining an approximately statistical potential is to perform an angle dependent scale transformation on the spherically symmetric solution $u_s(r)$. Then

$$u_p(x, \theta, \phi) = u_s\left(x \sum_{lm} \eta_{lm} Y_{lm}(\theta, \phi) \right) \tag{3.2}$$

in a simple generalization of the technique of Ref. 4, where only η_{00} is included to obtain a one-parameter semiempirical model. This procedure has the attractive feature that in the limit where the asymmetry vanishes, the potential satisfies a model which is remarkably accurate.[4]

APPENDIX A

In spherical polar coordinate (x, ξ_i) the gradient of a scalar function u is

$$\begin{aligned} \nabla u &= \mathbf{n} \left\{ u_x^2 + \sum_i (h_i^{-1} u_{\xi_i})^2 \right\}^{1/2} \\ &= \mathbf{n} u_n, \end{aligned} \tag{A1}$$

where $\mathbf{n}$ is the unit normal to the surface $u = \text{const}$. Similarly

$$\begin{aligned} x \nabla^2 (u/x) &= \nabla^2 u - (2/x) u_x \\ &= (\kappa + \mathbf{n} \cdot \nabla) u_x - (2/x) u_x, \end{aligned} \tag{A2}$$

where

$$\kappa = \mathrm{div}\mathbf{n}. \tag{A3}$$

Along the surface $u = \text{const}$, the extrema of x are given by $u_{\xi_i} = 0$, and at such an extremum

$$x\nabla^2(u/x) = u_{xx} + (\kappa - 2/x)u_x. \tag{A4}$$

If $l(\xi_i)$ is a solution of the equation $u = \text{const}$, then the extremal value of l is obtained by varying ξ_i:

$$\begin{aligned} 0 &= \delta u(l, \xi_i) \\ &= \delta l u_x + \sum_i \delta\xi_i u_{\xi_i} + \tfrac{1}{2}\delta l^2 u_{xx} + 2\delta l \sum_i \delta\xi_i u_{l\xi_i} \\ &\quad + \sum_{i,j} \delta\xi_i \delta\xi_j u_{\xi_i\xi_j} . \end{aligned} \tag{A5}$$

Therefore at the extremal

$$u_{\xi_i}(l, \xi_i) = 0 \tag{A6}$$

and

$$\lambda \equiv \sum_{i,j} \delta\xi_i \delta\xi_j u_{\xi_i\xi_j} = \eta u_x, \tag{A7}$$

where η is positive at a maximum of l and negative at a minimum. Writing the angular coordinates of the extremum as $\xi_i(R)$, where R is the extremum, gives the identity

$$u_{\xi_i}(R, \xi_i(R)) = 0. \tag{A8}$$

Differentiating with respect to R, we have

$$u_{l,\xi_i} + u_{\xi_i\xi_j} d\xi_j/dR = 0 \tag{A9}$$

at all points along the curve defined by $\xi_i(R)$. These are the extremal curves of the text. If $U(R)$ is defined by

$$U(R) = u(R, \xi_i(R)), \tag{A10}$$

one obtains

$$U_{RR} = f - (\kappa - 2/R + \eta)U_R \tag{A11}$$

along the maximal curve, where (A9) and (A4) have been used, and f is defined in the text (2.14). Now κ is related to the curvatures of the surface $u = \text{const}$ at the extremum, namely

$$\kappa = 2/r_c \tag{A12}$$

where r_c is a mean radius of curvature.[9] At a maximum $r_c < r$ and at a minimum $r_c > r$. Then the function

$$\rho(R) = \kappa - 2/R + \eta \tag{A13}$$

is positive definite at a maximum and negative definite at a minimum. Thus along a curve of maxima

$$w_{RR} = f - \rho_1(R)w_R \tag{A14}$$

and along a curve of minima

$$V_{RR} = f + p_2(R)V_R. \tag{A15}$$

If $p_1(R)$ or $p_2(R)$ vanishes, then the surface is locally spherical up to order $(\delta\xi)^3$, as may be seen from (A5). However, $l(\xi)$ satisfies a second order differential equation on the surface $u = \text{const}$, and if both first and second derivatives vanish at some point, then $l(\xi)$ is a constant. Therefore the condition for sphericity of the surface $u = \text{const}$ is

$$p_1(R) = p_2(R) = 0. \tag{A16}$$

APPENDIX B

Near the origin write

$$u = u_0 + x\sum_{l=1}^{\infty} f_{lm}x^l Y_{lm}, \tag{B1}$$

where u_0 satisfies the symmetrical equation

$$u_0'' = x^{-1/2}u_0^{3/2} \tag{B2}$$

and expand the left-hand side of (2.4) to obtain the linearized equations

$$f_{lm}'' + [2(l+1)/x]f_{lm}' = \tfrac{3}{2}(u_0/x)^{1/2}f_{lm}. \tag{B3}$$

The radial functions are expanded in a power series in $x^{1/2}$, and have the form:

$$u_0 = 1 - gx + \tfrac{4}{3}x^{3/2} + O(x^{5/2}), \tag{B4}$$

$$f_{lm} = 1 + x^{3/2}(2l + 5/2)^{-1} + O(x^{5/2}). \tag{B5}$$

Then, near the origin

$$u_x = -g + O(x^{1/2}) \tag{B6}$$

and is spherically symmetric. Therefore,

$$\frac{d\omega}{dr}(0) = \frac{dV}{dr}(0) = -g. \tag{B7}$$

[1] E. Fermi, Z. Physik **48**, 73 (1928).
[2] L. H. Thomas, Proc. Cam. Phil. Soc. **23**, 542 (1972).
[3] R. Latter, Phys. Rev. **99**, 510 (1955).
[4] J. C. Stuart and M. Rotenberg, Phys. Rev. **140**, A1508 (1965).
[5] D. R. Bates and A. Damgaard, Phil. Trans. Roy. Soc. London A **242**, 101 (1949).
[6] H. R. Griem, *Plasma Spectroscopy* (McGraw-Hill, New York, 1964), Chap. 3.
[7] W. L. Weise, M. W. Smith, and B. M. Glennon, *Atomic Transition Probabilities*, NSRDS NBS 4 (Natl. Bur. Stds., Washington, D.C., 1966).
[8] P. Gombas, *Die Statistische Theorie des Atoms* (Springer-Verlag, Vienna, 1949).
[9] *Mathematical Handbook for Scientists and Engineers*, edited by Korn and Korn (McGraw-Hill, New York, 1961). The mean curvature is the average of the two principal curvatures.

A classical perturbation theory*

Charles Schwartz

Department of Physics, University of California, Berkeley, California 94720
(Received 27 August 1975; resubmitted manuscript received 5 January 1976)

A compact formula is found for the perturbation expansion of a general one-dimensional Hamiltonian system in classical mechanics. The technique is also applied to the mathematical problem of functional inversion.

PERTURBATION THEORY FOR CLASSICAL MECHANICS

We consider a system with one degree of freedom, with the Hamiltonian $H=H(p,q)$ being independent of the time t. We assume that we have bounded, periodic motion at some value E of the energy. That is, the equation $E=H(p,q)$ describes a single closed curve (the orbit) in the $p-q$ plane (phase space). Now consider the integral

$$T=\int\int dp\,dq\,\delta(E-H(p,q)), \tag{1}$$

involving the Dirac delta function; the domain of the integral is to include the orbit. We shall first show that this integral T is equal to the period of motion at energy E.

Doing the integral over p, we get

$$T=\int dq \sum_m \left(\left|\frac{\partial H(p,q)}{\partial p}\right|_{p=p_m}\right)^{-1}, \tag{2}$$

where p_m are all points satisfying $E=H(p_m,q)$ for fixed q. But we have from Hamilton's equation of motion

$$\dot{q}=\frac{\partial H}{\partial p}, \tag{3}$$

and we then see that the expression T is just

$$\oint_E \frac{dq}{\dot{q}}=\oint_E dt=T(E), \tag{4}$$

where the integral goes once around the orbit.

We can thus express the time average of any function F of the dynamical variables taken over the orbit at energy E, as

$$\langle F;E\rangle=\frac{1}{T(E)}\int\int dp\,dq\,\delta(E-H)F(p,q). \tag{5}$$

Now, for the perturbation theory, suppose that we are given $H=H_0+\lambda H_1$ and we seek an expansion in powers of λ. The basic step is to regard E as an independent variable and then write the Taylor series expansion,

$$\delta(E-H_0-\lambda H_1)=\sum_{n=0}^{\infty}\frac{1}{n!}\left(-\lambda H_1\frac{d}{dE}\right)^n\delta(E-H_0), \tag{6}$$

$$T(E)=\sum_{n=0}^{\infty}\frac{(-\lambda)^n}{n!}\left(\frac{d}{dE}\right)^n\int\int dp\,dq(H_1(p,q))^n\delta(E-H_0)$$

$$=\sum_{n=0}^{\infty}\frac{(-\lambda)^n}{n!}\left(\frac{d}{dE}\right)^n T_0(E)\langle H_1^n;E\rangle_0. \tag{7}$$

Here, the subscript "0" means that the averages are performed over the orbit of the zeroth-order Hamiltonian H_0. This formula is very compact; its evaluation involves only the operation of integration over the unperturbed orbits, followed by differentiation with respect to the energy. For comparison, one may look at the formulas obtained in "canonical perturbation theory" (see, for example, Saletan and Cromer[1]). That analysis is based upon the action-angle formalism (our result can be reexpressed in terms of action-angle variables but there is no particular advantage in doing so) and appears as an expansion for the energy E, thought of as a dependent variable; the expansion formula is there worked out only to second order in λ, and the form is quite messy in appearance.

We also get the formula

$$T(E)\langle F;E\rangle=\sum_{n=0}^{\infty}\frac{(-\lambda)^n}{n!}\left(\frac{d}{dE}\right)^n T_0(E)\langle H_1^n F;E\rangle_0. \tag{8}$$

We will compute some examples based upon the harmonic oscillator,

$$H_0=\frac{p^2}{2m}+\frac{k}{2}q^2, \tag{9}$$

which has the solutions (at energy E) given by

$$q=\sqrt{2E/k}\sin(\omega_0 t+\phi),\quad \omega_0=\sqrt{k/m},$$
$$p=\sqrt{2Em}\cos(\omega_0 t+\phi), \tag{10}$$
$$T_0(E)=2\pi/\omega_0\ \text{(independent of } E).$$

1. Consider the perturbation Hamiltonian $H_1=|q|^\sigma$

$$\langle H_1^n;E\rangle=\frac{1}{2\pi}\int_0^{2\pi}d\theta\left(\frac{2E}{k}\right)^{n\sigma/2}|\sin\theta|^{n\sigma}. \tag{11}$$

This integral may be evaluated and the derivatives with respect to E are likewise easily evaluated; the result is

$$T(E)=\frac{2\pi}{\omega_0}\sum_{n=0}^{\infty}\frac{(n\sigma/2-\frac{1}{2})!}{(-\frac{1}{2})!}\frac{1}{n!(n\sigma/2-n)!}\times\left[\frac{-2\lambda}{k}\left(\frac{2E}{k}\right)^{\sigma/2-1}\right]^n. \tag{12}$$

For $\sigma=2$ we get the familiar result,

$$T(E)=\frac{2\pi}{\omega}\sum_{n=0}^{\infty}\frac{(n-\frac{1}{2})!}{(-\frac{1}{2})!\,n!}\left(\frac{-2\lambda}{k}\right)^n=\frac{2\pi}{\omega_0}\left(1+\frac{2\lambda}{k}\right)^{-1/2}$$
$$=2\pi\left(\frac{m}{k+2\lambda}\right)^{1/2}.$$

Having this series explicitly given, we can ask about its radius of convergence. The ratio of successive terms, for large n, is

$$R_n\sim\frac{-\lambda\sigma}{k}\left(\frac{2E}{k}\,\frac{\sigma}{\sigma-2}\right)^{(\sigma-2)/2} \tag{13}$$

and thus we will have convergence up to that value of the energy E for which this ratio is 1. We now ask for the

significance of this critical value of the energy,

$$E^* = \left| \frac{k}{2}\left(\frac{\sigma-2}{\sigma}\right)\left(\frac{-\lambda\sigma}{k}\right)^{2/2-\sigma} \right|. \tag{14}$$

It can be readily shown that at this energy the orbit reaches an amplitude at which the total potential energy has a zero slope and the motion thereafter is qualitatively different. Thus we conclude, at least for this example, that the perturbation series will converge for all energies for which the period is a finite and continuous function of the energy. (In this we must allow for changing the sign of λ.)

2. Consider the perturbation

$$H_1 = \frac{a}{3}q^3 + \frac{b}{4}q^4 + \frac{c}{5}q^5 + \frac{d}{6}q^6 + \cdots. \tag{15}$$

We calculate

$$T(E)\frac{\omega_0}{2\pi} = 1 + \frac{E}{k^2}\left(-\frac{3}{4}b + \frac{5}{6}\frac{a^2}{k}\right) + \frac{E^2}{k^3}\left(-\frac{5}{4}d + \frac{105}{64}\frac{b^2}{k} + \frac{7}{2}\frac{ac}{k} - \frac{105}{16}\frac{a^2b}{k^2} + \frac{385}{144}\frac{a^4}{k^3}\right) + O(E^3), \tag{16}$$

which gives the leading energy dependent corrections to the period of a general nonlinear oscillator.

APPENDIX

The expansion technique used above finds application to some problems removed from Hamiltonian mechanics. Consider a given function F whose inverse is sought:

$$y = F(x), \quad x = F^{-1}(y).$$

We assume that F is a monotonic function, so that this inverse is unique. Now suppose we have $F = F_0 + \lambda F_1$, where λ is again a small parameter. We would expect to find an expansion

$$F^{-1}(y) = F_0^{-1}(y) + \sum_{n=1}^{\infty} \lambda^n G_n(y),$$

where the terms G_n could be found by a lenghy process of Taylor expansions. What is somewhat surprising is that we can find a compact formula for the general term in this series.

Again, starting from the integral of a delta function, we have

$$\int dx\, \delta(y - F(x)) = \frac{1}{F'(x)}\bigg|_{x=F^{-1}(y)} = \frac{dx}{dy} = \frac{d}{dy}F^{-1}(y).$$

Substituting $F = F_0 + \lambda F_1$, we make the Taylor series expansion of this same integral to get

$$\sum_{n=0}^{\infty} \frac{(-\lambda)^n}{n!}\left(\frac{d}{dy}\right)^n \int dx\, \delta(y - F_0(x))F_1^n(x).$$

Equating these two expressions, and then performing one integral in y, we get

$$F^{-1}(y) = F_0^{-1}(y) + \sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n!}\left(\frac{d}{dy}\right)^{n-1} \frac{(F_1(x))^n}{F_0'(x)}\bigg|_{x=F_0^{-1}(y)}.$$

Again, the trick in finding this compact formula was to regard y, and not x, as the independent variable.

For one simple example of application of this formula, consider

$$y = x^\alpha + \lambda x^\beta;$$

we find the inversion:

$$x = y^{1/\alpha} + \sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n!}\left(\frac{d}{dy}\right)^{n-1} \frac{x^{n\beta}}{\alpha x^{\alpha-1}}\bigg|_{x=y^{1/\alpha}}$$
$$= y^{1/\alpha}\sum_{n=0}^{\infty} (-\lambda y^{(\beta-\alpha)/\alpha})^n \frac{(n\beta/\alpha + 1/\alpha - 1)!}{n!\,\alpha\{n[(\beta-\alpha)/\alpha] + 1/\alpha\}!}.$$

This series is readily seen to be convergent up to that point at which dy/dx becomes zero for either sign of λ.

Now consider extending this technique to functions of several variables:

$$y_i = F_{0_i}(\mathbf{x}) + \lambda F_{1_i}(\mathbf{x}), \quad \mathbf{x} = (x_1, x_2, \ldots, x_N), \quad i = 1, N,$$

where we wish to solve for x_i in terms of the y_j. For simplicity we take the function F_0 to be the identity function:

$$y_i = x_i + \lambda\phi_i(\mathbf{x})$$

[later, one can set $x_i = F_{0_i}(\mathbf{z})$ to recover the more general case]. Now consider the following integral:

$$\int\cdots\int d\xi_1 \cdots d\xi_N\, \delta(y_1 - \xi_1 - \lambda\phi_1(\xi)) \cdots \delta(y_N - \xi_N - \lambda\phi_N(\xi))$$
$$\cdot \det\left|\delta_{ij} + \lambda\frac{\partial\phi_i(\xi)}{\partial\xi_j}\right| \cdot G(\xi).$$

by changing integration variables from the ξ_i's to

$$\eta_i = \xi_i + \lambda\phi_i(\xi),$$

we see that this integral has just the value $G(\mathbf{x})$ where x_i are related to y_i by the equations given above. (The determinant is the Jacobian which is needed to make this transformation work out simply.)

Now we use the Taylor series expansion, as before, using the variables y_i to expand the arguments of the delta functions in power series in λ:

$$G(\mathbf{x}) = \int\cdots\int d\xi_1 \cdots d\xi_N \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{n!}\left(\sum_{i=1}^{N} \frac{\partial}{\partial y_i}\phi_i(\xi)\right)^n$$
$$\times\delta(y_1 - \xi_1)\cdots\delta(y_N - \xi_N)\det\left|\delta_{ij} + \lambda\frac{\partial\phi_i(\xi)}{\partial\xi_j}\right| \cdot G(\xi)$$
$$= \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{n!} \sum_{i_1, i_2, \ldots, i_n=1}^{N} \frac{\partial}{\partial y_{i_1}} \cdots \frac{\partial}{\partial y_{i_n}}$$
$$\times\phi_{i_1}(\mathbf{y})\cdots\phi_{i_n}(\mathbf{y})\det\left|\delta_{ij} + \lambda\frac{\partial\phi_i(\mathbf{y})}{\partial y_j}\right| G(\mathbf{y}).$$

This general formula is not exactly in the form of a power series in λ because the determinant is an Nth degree polynomial in λ; but it is the most compact form of the functional inversion problem for several variables. Using the equation that results from setting $G(\mathbf{x}) = 1$, we can rewrite the general formula as

$$G(\mathbf{x}) = G(\mathbf{y}) + \sum_{n=1}^{\infty} \frac{(-\lambda)^n}{n!} \sum_{i_1 \cdots i_n=1}^{N}$$
$$\times\left[\frac{\partial}{\partial y_{i_1}} \cdots \frac{\partial}{\partial y_{i_n}}, G(\mathbf{y})\right]\phi_{i_1}(\mathbf{y})\cdots\phi_{i_n}(\mathbf{y})$$

$$\times \det\left|\delta_{ij}+\lambda\frac{\partial\phi_i(\mathbf{y})}{\partial y_j}\right|,$$

involving the commutator of G with the derivative operators.

For the case $N=1$ we have

$$G(x)=G(y)+\sum_{n=1}^{\infty}\frac{(-\lambda)^n}{n!}\left[\left(\frac{\partial}{\partial y}\right)^n,G(y)\right]\phi^n(y)\times\left[1+\lambda\frac{\partial\phi(y)}{\partial y}\right],\quad y=x+\lambda\phi(x),$$

and this can be rearranged into a strict power series in λ, yielding

$$G(x)=G(y)+\sum_{n=1}^{\infty}\frac{(-\lambda)^n}{n!}\left(\frac{\partial}{\partial y}\right)^{n-1}\phi^n(y)\frac{\partial G(y)}{\partial y}.$$

This formula is equivalent to our earlier result on function inversion with one variable. This formula was first published by Lagrange in 1770 (see Whittaker and Watson[2]; the derivation given there does not use delta functions and has the added virtue that one can more readily see what the radius of convergence of the series will be). The formula we have given above for several variables is, as far as we know, new. For $N=2$ the series can be rearranged and, with some care, we obtain

$$y_1=x_1+\lambda\phi_1(x_1,x_2),\quad y_2=x_2+\lambda\phi_2(x_1,x_2),$$

$$G(x_1,x_2)=G(y_1,y_2)+\sum_{n=1}^{\infty}(-\lambda)^n\sum_{l_1+l_2=n}\frac{1}{l_1!}\frac{1}{l_2!}\left(\frac{\partial}{\partial y_1}\right)^{l_1-1}\left(\frac{\partial}{\partial y_2}\right)^{l_2-1}\times\left[\frac{\partial^2G(y_1,y_2)}{\partial y_1\,\partial y_2}+\frac{\partial G(y_1,y_2)}{\partial y_1}\,l_1\frac{\partial\ln\phi_1(y_1,y_2)}{\partial y_2}+\frac{\partial G(y_1,y_2)}{\partial y_2}\,l_2\frac{\partial\ln\phi_2(y_1,y_2)}{\partial y_1}\right]\phi_1^{l_1}(y_1,y_2)\phi_2^{l_2}(y_1,y_2)$$

of which a special case is

$$x_1=y_1+\sum_{n=1}^{\infty}(-\lambda)^n\sum_{l}\frac{1}{l!\,(n-l)!}\left(\frac{\partial}{\partial y_1}\right)^{n-l-1}\left(\frac{\partial}{\partial y_2}\right)^{l-1}\phi_2^{l}\frac{\partial\phi_1^{n-l}}{\partial y_2}.$$

For another special case consider the linear forms

$$\phi_i(\mathbf{x})=\sum_{j=1}^{N}A_{ij}x_j\quad\text{for any } N.$$

Then, taking $G=1$, we find the formula

$$\frac{1}{\det(1+\lambda A)}=\sum_{n=0}^{\infty}\frac{(-\lambda)^n}{n!}\sum_{l_1,l_2,\ldots,l_n=1}^{N}\sum_{k_1,k_2,\ldots,k_n}\times A_{l_1k_1},A_{l_2k_2},\ldots,A_{l_nk_n},$$

where the set of labels $(k_1,k_2,\ldots,k_n)$ goes over each permutation of the set $(l_1,l_2,\ldots,l_n)$.

*This work was supported by the National Science Foundation under the Grant Number MPS 74 08175 A01.

[1]E. J. Saletin and A. H. Cromer, *Theoretical Mechanics* (Wiley, New York, 1971), p. 241.

[2]E. T. Whittaker and G. N. Watson, *A Course in Modern Analysis* (Cambridge U. P., Cambridge, 1958), p. 133.

Gleason measures on infinite tensor products of Hilbert spaces

Mark J. Christensen

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332
(Received 9 August 1976)

Nowak [Bull. Acad. Pol. Sci. Ser. Sci. Math. Astron. Phys. **22**, 393–5 (1974)] has given an example of a consistent (in the sense of Kolmogorov) family of Gleason measures [A. M. Gleason, J. Math. Mech. **6**, 885–94 (1957)] $\{m_n\}$ defined over $\otimes_{i=1}^{n} H_i$ which do not extend to a Gleason measure on $\otimes_{i=1}^{\infty} {}^{\Phi}H_i$ for a given construction of the infinite tensor product. In this paper we show: (1) In the example of Nowak it is not necessary to assume, as is done, that the H_i are infinite dimensional. (2) That every consistent family developed from pure states, which is the type considered by Nowak, extends over the complete infinite tensor product of von Neumann [Compositio. Math. **6**, 1–77 (1938)]. (3) Even if each H_i is two-dimensional and the complete infinite tensor product of von Neumann is used, it is possible to give a simple counterexample to the conjecture that every consistent family of Gleason measures extends by the use of nonpure states.

1. INTRODUCTION

A *Gleason measure* is a function m from the collection $\mathcal{E}$ of closed subspaces of a Hilbert space H into the unit interval, countably additive on orthogonal decompositions, with $m(\emptyset)=0<m(H)<\infty$. Gleason[1] has shown that every Gleason measure m can be represented as $m(E)=\mathrm{Tr}(MP_E)$, where Tr denotes trace, P_E is the projection operator of the closed subspace E, and M is a bonded, positive, self-adjoint, trace class operator on H. It is evident that every such operator with trace equal to 1 will likewise yield a Gleason measure.

The most elementary case of a Gleason measure occurs when M is the projection onto a one-dimensional subspace of H. In that case m is referred to as a "pure state," this terminology arising from the quantum mechanical origins of the subject.

Jajte has asked if there is for Gleason measures on tensor products of Hilbert spaces a theorem analogous to the well-known Kolmogorov extension theorem for probability measures on cartesian products. Nowak[2] has recently given an example to show that a consistent family of Gleason measures need not extend on a given construction of the infinite tensor product. In Sec. 3 of this note we show that for the case of products of pure states, such as considered by Nowak, there is always as extension Gleason measure on the complete infinite tensor product of von Neumann which, considering the quantum mechanical origins of the subject of Gleason measures, is the natural setting for the consideration of such questions. In Sec. 4, with an eye towards more general Gleason measures, the properties of operators on infinite tensor products are given. Finally, in Sec. 5, we show that a completely general family of product Gleason measures will extend over the complete infinite tensor product of von Neumann if and only if a condition of "asymptotic purity" is obeyed. It therefore follows that, even if each H_n is two-dimensional, it is possible to give a counterexample to the conjecture that every such family of Gleason measures has an extension.

2. TENSOR PRODUCTS OF HILBERT SPACES

Let N denote the positive integers, and let $\{H_n\}_{n\in N}$ be a family of complex Hilbert spaces, with $\langle\cdot,\cdot\rangle_n$ and $\|\cdot\|_n$ denoting the inner product and norm of H_n, respectively. The finite tensor products $\hat{\otimes}_{n=1}^{p} H_n$ are defined in the usual way, as being the completion with respect to the inner product

$$\left[\sum_k \overset{p}{\underset{n=1}{\otimes}} y_n^k, \sum_h \overset{p}{\underset{n=1}{\otimes}} z_n^h\right]_p = \sum_{k,h} \prod_{n=1}^{p} \langle y_n^k, z_n^h\rangle_n$$

of the linear span of the functionals

$$\left(\overset{p}{\underset{n=1}{\otimes}} y_n\right)(x) = \prod_{n=1}^{p} \langle x_n, y_n\rangle_n,$$

where $x=(x_1,\ldots,x_p)$.

To define von Neumann's complete infinite tensor product (cf. Ref. 3), we proceed as follows. Denote by Δ the set of all $x\in\times_{n\in N}H_n$ such that $\sum_{n=1}^{\infty}|\,\|x_n\|_n-1|<\infty$. By $\otimes_{n=1}^{\infty}y_n$, with $\{y_n\}\in\Delta$, we shall mean the bounded functional on Δ defined by

$$\left(\overset{\infty}{\underset{n=1}{\otimes}} y_n\right)(x) = \prod_{n=1}^{\infty}\langle x_n, y_n\rangle_n,$$

where the infinite product is quasiconvergent in the sense of von Neumann.

Definition 1: $\hat{\otimes}_{n=1}^{\infty} H_n$ is the completion of the linear span of all bounded linear functionals of the form $\otimes_{n=1}^{\infty} y_n$ with respect to the inner product

$$\left[\sum_k \overset{\infty}{\underset{n=1}{\otimes}} y_n^k, \sum \overset{\infty}{\underset{n=1}{\otimes}} z_n^h\right] = \sum_{k,h} \prod_{n=1}^{\infty} \langle y_n^k, z_n^h\rangle_n.$$

This definition is, however, only one of many. The definition currently receiving the greatest usage is that of the algebraic tensor product.

Definition 2: Let $\Phi=\{e_n^0\}$ be a sequence of unit vectors, with $e_n^0\in H_n$. Let $\Delta_\Phi=\{x\in\times_{n=1}^{\infty}H_n : x_n\neq e_n^0$ for only finitely many $n\}$. The Hilbert space obtained by completing the linear span of the linear functionals $\otimes_{n=1}^{\infty}y_n$ with $y\in\Delta_\Phi$ with respect to the inner product $[\cdot,\cdot]$ is called the infinite tensor product of the H_n constructed about Φ and is denoted by

$$\hat{\underset{n=1}{\overset{\infty}{\otimes}}} {}^{\Phi}H_n.$$

Each

$$\hat{\underset{n=1}{\overset{\infty}{\otimes}}} {}^{\Phi}H_n$$

is a subspace of $\hat{\otimes}_{n=1}^{\infty} H_n$. Furthermore, von Neumann[3] has shown that these subspaces decompose $\hat{\otimes}_{n=1}^{\infty} H_n$ into mutually orthogonal subspaces.

3. A KOLMOGOROV EXTENSION THEOREM FOR PURE STATES

Now let $\{H_n\}$ be a countable family of Hilbert spaces as before. For each n let m_n be a pure state on H_n. Let e_n be the unit vector of H_n which represents the state m_n. Then, for every p, $\otimes_{n=1}^{p} e_n$ is a unit vector of $\hat{\otimes}_{n=1}^{p} H_n$ and hence defines a Gleason measure $\otimes_{n=1}^{p} m_n$ over that p-fold tensor product.

The proofs of Propositions 1, 2, and 3 are given in the Appendix.

Proposition 1 (1 of Ref. 2): The family $\{\otimes_{n=1}^{p} m_n, \hat{\otimes}_{n=1}^{p} H_n\}_p$ is consistent in the sense that if E is a closed subspace of $\hat{\otimes}_{n=1}^{p} H_n$, then

$$\left(\overset{p}{\underset{n=1}{\otimes}} m_n\right)(E) = \left(\overset{p+1}{\underset{n=1}{\otimes}} m_n\right)(E \otimes H_{p+1}).$$

We remark that Proposition 1 is true for arbitrary Gleason measures. The following proposition is then the Kolmogorov extension theorem for pure states.

Proposition 2: The family $\{\otimes_{n=1}^{p} m_n, \hat{\otimes}_{n=1}^{p} H_n\}_p$ has an extension to $\hat{\otimes}_{n=1}^{\infty} H_n$, namely that Gleason measure represented by the unit vector $\hat{\otimes}_{n=1}^{\infty} e_n$. That is, there exists a Gleason measure m on $\hat{\otimes}_{n=1}^{\infty} H_n$ satisfying

$$m\left(E \otimes \left(\overset{\infty}{\underset{n=p}{\hat{\otimes}}} H_n\right)\right) = \overset{p-1}{\underset{n=1}{\otimes}} m_n(E)$$

for every p and closed subspace E of $\hat{\otimes}_{n=1}^{p-1} H_n$.

The next proposition gives the criterion for the extension to lie in any particular construction of the infinite tensor product.

Proposition 3: Let $\Phi = \{e_n^0\}_{n \in N}$ be a sequence of unit vector with $e_n^0 \in H_n$ and let

$$\overset{\infty}{\underset{n=1}{\hat{\otimes}}}{}^{\Phi} H_n$$

be the algebraic tensor product constructed about Φ. Then the consistent family $\{\otimes_{n=1}^{p} m_n, \hat{\otimes}_{n=1}^{p} H_n\}_{p \in N}$ has an extension to

$$\overset{\infty}{\underset{n=1}{\hat{\otimes}}}{}^{\Phi} H_n$$

if and only if

$$0 < \prod_{n=L}^{\infty} |\langle e_n, e_n^0 \rangle_n|$$

for some L.

We remark that in the example of Nowak[2] $\langle e_n, e_n^0 \rangle_n = 0$ for all $n > 1$.

4. TENSOR PRODUCTS OF OPERATORS ON HILBERT SPACES

Let $\{H_n\}_{n \in N}$ denotes a countable family of separable Hilbert spaces over the complex numbers Z. Let $\hat{\otimes}_{n=1}^{p} H_n$ be the tensor product of $H_1, \ldots, H_n$ defined in Sec. 2. Further let $\{A_n\}_{n \in N}$ be a family of bounded linear operators, each A_n mapping H_n into itself. Then we may define[3] an operator $\otimes_{n=1}^{p} A_n$ from $\hat{\otimes}_{n=1}^{p} H_n$ into itself by the formula

$$\left(\overset{p}{\underset{n=1}{\otimes}} A_n\right)\left(\overset{p}{\underset{n=1}{\otimes}} v_n\right) = \overset{p}{\underset{n=1}{\otimes}} (A_n v_n).$$

This operator may then be extended by linearity to the completion $\hat{\otimes}_{n=1}^{p} H_n$. The operator $\otimes_{n=1}^{p} A_n$ will be bounded over $\hat{\otimes}_{n=1}^{p} H_n$ with norm

$$\left\| \overset{p}{\underset{n=1}{\otimes}} A_n \right\| = \prod_{n=1}^{p} \|A_n\|.$$

The trace of this operator is likewise given by formula

$$\mathrm{Tr}\left(\overset{p}{\underset{n=1}{\otimes}} A_n\right) = \prod_{n=1}^{\infty} \mathrm{Tr}\,(A_n).$$

Therefore, the operator $\hat{\otimes}_{n=1}^{\infty} A_n$ exists as a bounded, linear, nontrivial operator if and only if the infinite product

$$\prod_{n=1}^{\infty} \|A_n\|$$

converges to a nonzero quantity.

5. THE KOLMOGOROV EXTENSION THEOREM FOR GLEASON MEASURES

Let $\{m_n\}$ be a family of Gleason measures over the family of Hilbert spaces $\{H_n\}$. Furthermore, let $\{M_n\}$ be the corresponding family of operators. Then we have the following proposition.

Proposition 4: The family $\{\otimes_{n=1}^{p} m_n, \hat{\otimes}_{n=1}^{p} H_n\}_p$ has an extension to $\hat{\otimes}_{n=1}^{\infty} H_n$ if and only if $0 < \prod_{n=1}^{\infty} \|M_n\|$. That is, there exists a Gleason measure m on $\hat{\otimes}_{n=1}^{\infty} H_n$ satisfying

$$m\left(E \otimes \left(\overset{\infty}{\underset{n=p}{\hat{\otimes}}} H_n\right)\right) = \left(\overset{p-1}{\underset{n=1}{\otimes}}\right) m_n \ (E),$$

for every closed subspace E of $\hat{\otimes}_{n=1}^{p-1} H_n$ for every p if and only if $0 < \prod_{n=1}^{\infty} \|M_n\|$.

Thus we see that if $H_n = R^2$ for each n and $M_n = \frac{1}{2} I$, where I is the identity operator of R^2, then M_n will represent a uniform distribution on H_n and yet there will be no extension on $\hat{\otimes}_{n=1}^{\infty} H_n$.

The condition $0 < \prod_{n=1}^{\infty} \|M_n\|$ implies that $\|M_n\| \to 1$ and since $\mathrm{Tr}(M_n) = 1$, we see that M_n tends to a projection on a one-dimensional subspace if both are to be true. Hence the states $\{m_n\}_{n \in N}$ must become asymptotically pure.

ACKNOWLEDGMENT

The authors would like to thank Professor A. T. Bharucha-Reid for bringing this problem to his problem.

APPENDIX

Note: The proofs of the propositions given here utilize the results of Sec. 4. This yields both simplier proofs and results of greater generality.

Proof of Proposition 1

Let H_n be the given countable family of Hilbert spaces with m_n a given Gleason measure defined on H_n and M_n the positive, trace-class operator representing m_n. Then for each p we define the operator $\otimes_{n=1}^{p} M_n$. Thus we

have

$$\left(\overset{p}{\underset{n=1}{\otimes}} m_n\right)(E) = \mathrm{Tr}\left[\left(\overset{p}{\underset{n=1}{\otimes}} M_n\right) P_E\right]$$
$$= \mathrm{Tr}\left[\left(\overset{p+1}{\underset{n=1}{\otimes}} M_n\right)\left(P_E \otimes I_{p+1}\right)\right]$$
$$= \left(\overset{p+1}{\underset{n=1}{\otimes}} m_n\right)\left(E \otimes H_{p+1}\right),$$

where I_{p+1} is the identity operator on H_{p+1}.

Proof of Proposition 2

Let m_n be the given family of pure states and suppose e_n is a unit vector in the range of M_n. Then let $e = \hat{\otimes}_{n=1}^{\infty} e_n$ and let M be the orthogonal projection onto the subspace spanned by e. Then, if we define

$$E' = E \otimes \left(\overset{\infty}{\underset{n=p+1}{\hat{\otimes}}} H_n\right)$$

for every closed subspace E of $\hat{\otimes}_{n=1}^{p} H_n$ we shall have

$$\mathrm{Tr}(MP_{E'}) = \mathrm{Tr}\left(\left(\overset{p-1}{\underset{n=1}{\otimes}} M_n\right) P_E\right) \mathrm{Tr}\left(\left(\overset{\infty}{\underset{n=p}{\otimes}} M_n\right) \overset{\infty}{\underset{n=p}{\otimes}} I_n\right)$$
$$= \mathrm{Tr}\left(\left(\overset{p-1}{\underset{n=1}{\otimes}} M_n\right) P_E\right) = \left(\overset{p-1}{\underset{n=1}{\otimes}} m_n\right)(E).$$

Thus M extends the M_n. That this extension is non-trivial follows from the fact that $\|M\| = \Pi_{n=1}^{\infty} \|M_n\| = 1 > 0$.

Proof of Proposition 3

Let $\Phi = \{e_n^0\}$ be the sequence of unit vectors about which

$$\overset{\infty}{\underset{n=1}{\hat{\otimes}}} {}^{\Phi}H_n$$

is to be constructed and let e_n be the sequence of unit vectors lying in the range of the M_n's. We must show that the operator $M = \otimes_{n=1}^{\infty} M_n$ is not the null operator on

$$\overset{\infty}{\underset{n=1}{\hat{\otimes}}} {}^{\Phi}H_n.$$

This will insure that the extension exists on this construction of the tensor product.

First we observe that

$$1 = \|M\| \geq \|M\|_{\Phi}$$
$$= \sup\left\{[Mx, Mx] : x \in \overset{\infty}{\underset{n=1}{\hat{\otimes}}} {}^{\Phi}H_n, \|x\| = 1\right\}$$

and since the elements of the form $x = \otimes_{n=1}^{\infty} y_n$, $y_n = e_n^0$ for all but finitely many n generate the tensor product under consideration we see that $\|M\|_{\Phi} = 0$ if and only if $Mx = 0$ for every such generator.

However, if we set

$$x = \left(\overset{L-1}{\underset{n=1}{\otimes}} y_n\right) \otimes \left(\overset{\infty}{\underset{n=L}{\otimes}} e_n^0\right),$$

then

$$[Mx, Mx] = \prod_{n=1}^{L-1} |\langle y_n, e_n\rangle_n|^2 \prod_{n=L}^{\infty} |\langle e_n^0, e_n\rangle_n|^2$$

for every choice of L and $y_1, \ldots, y_{L-1}$. Thus, if $\|M\|_{\Phi} = 0$, it follows that

$$\prod_{n=L}^{\infty} |\langle e_n^0, e_n\rangle_n| = 0$$

for every value of L. On the other hand, if $\|M\|_{\Phi} > 0$, then for some choice of L and $y_1, \ldots, y_{L-1}$, $Mx \neq 0$ and hence for that choice of L

$$\prod_{n=L}^{\infty} |\langle e_n^0, e_n\rangle_n| > 0.$$

[1] A. M. Gleason, J. Math. Mech. 6, 885–94 (1957).
[2] B. Nowak, Bull. Acad. Polon, Sci. Ser. Sci. Math. Astronom. Phys. 22, 393–95 (1974).
[3] J. von Neumann, Compositio Math. 6, 1–77 (1938).

Exact results for second-neighbor surface magnons in an fcc lattice

J. Dobson

Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853
(Received 1 March 1976)

An exact dispersion equation $f(\Omega)=0$ is obtained for (001) surface magnons in an fcc lattice with nearest and second nearest neighbor Heisenberg coupling. On the Brillouin zone boundary a closed expression is given for the frequency Ω, while a simple root search yields Ω for wave vectors inside the zone. These exact frequencies confirm the results obtained by Trullinger from a Gottlieb polynomial expansion; a discussion of the latter method is given.

Recently Trullinger[1] demonstrated a new approximation scheme for solving discrete one-sided linear equations by expanding in orthonormal functions of a discrete variable, the Gottlieb functions. As an example he solved a surface magnon problem with coupling through second nearest neighbors. The purpose of the present paper is to exhibit an essentially exact solution of this latter problem, by means of which several features of the Gottlieb polynomial method are clarified.

Let $S_m^-(x,y,t)$ be the left-hand circularly polarized component of spin on the particle at position (x,y) in the mth layer of an fcc lattice [$m=0,1,2,\ldots$, measured inwards from the (001) surface]. With the ansatz

$$S_m^-(x,y,t)=S(m)\exp i(k_x x+k_y y-\omega t), \tag{1}$$

the equation of motion, previously given in equation (3.1) of Trullinger's[1] paper, can be written in a convenient form:

$$\Omega S(m)=L_2(S(m-2)+S(m+2))+L_1(S(m-1)+S(m+1))+L_0 S(m),\quad m\geq 2, \tag{2}$$

$$\Omega S(1)=L_2 S(3)+L_1(S(2)+S(0))+L_{01}S(1), \tag{3}$$

$$\Omega S(0)=L_2 S(2)+L_1 S(1)+L_{00}S(0). \tag{4}$$

Here

$$\Omega=\omega/SJ_1, \tag{5a}$$

$$L_2=-J_2/J_1=-\gamma, \tag{5b}$$

$$L_1=-[1+\cos k_x a+\cos k_y a+\cos(k_x+k_y)a], \tag{5c}$$

$$L_0=12-2(\cos k_x a+\cos k_y a)+\gamma[6-2\cos(k_x+k_y)a-2\cos(k_x-k_x)a], \tag{5d}$$

$$L_{01}=L_0-\gamma, \tag{5e}$$

$$L_{00}=L_{01}-4. \tag{5f}$$

J_1 and J_2 are exchange integrals, and a is the lattice spacing.

Equation (1) is a finite difference equation for the $\{S(m)\}$, with constant coefficients L_2, L_1, L_0. As such it has basic solutions of the form

$$S_i(m)=\exp(-\lambda_i m)=t_i^m, \tag{6}$$

where direct substitution in (1) yields an equation for the $\{t_i\}$,

$$(t+t^{-1})^2+(L_1/L_2)(t+t^{-1})+L_0/L_2-2-\Omega/L_2=0. \tag{7}$$

The four solutions are

$$t=\tfrac{1}{2}[W_\pm+\sqrt{W_\pm-4}\,],\quad \tfrac{1}{2}[W_\pm-\sqrt{W_\pm-4}], \tag{8}$$

where

$$W_\pm=\tfrac{1}{2}\{-(L_1/L_2)\pm[(L_1/L_2)^2+(4/L_2)(\Omega+2L_2-L_0)]^{1/2}\}. \tag{9}$$

In general these roots occur in reciprocal pairs, say t_1, t_1^{-1}, t_2, t_2^{-1}, where

$$|t_1|<1,\quad |t_2|<1,\quad |t_1^{-1}|>1,\quad |t_2^{-1}|>1. \tag{10}$$

The general *surface* magnon solution, which vanishes as $m\to\infty$, must thus be of the form

$$S(m)=At_1^m+Bt_2^m. \tag{11}$$

Equation (11) automatically satisfies (2) but must be subjected specifically to the edge conditions (3) and (4). This gives two linear equations for A and B, namely

$$X_{11}A+X_{12}B=0, \tag{12a}$$

$$X_{21}A+X_{22}B=0, \tag{12b}$$

where

$$X_{11}=\Omega t_1-L_2t_1^3-L_1(t_1^2+1)-L_{01}t_1, \tag{13a}$$

$$X_{12}=\Omega t_2-L_2t_2^3-L_1(t_2^2+1)-L_{01}t_2, \tag{13b}$$

$$X_{21}=\Omega-L_2t_1^2-L_1t_1-L_{00} \tag{13c}$$

$$X_{22}=\Omega-L_2t_2^2-L_1t_2-L_{00}. \tag{13d}$$

The condition that (12) have a nontrivial solution is the vanishing of a determinant,

$$f(\Omega)=X_{11}X_{22}-X_{12}X_{21}=0. \tag{14}$$

Equation (14), with the definitions (13), (10), (9), (8), and (5) is an *exact* secular equation giving the surface magnon frequency Ω. Its terms contain only trigonometric and algebraic functions and so it is trivially programmed for a numerical root search. For wavevectors $(k_x,\Pi/a)$ on the zone boundary or these on the zone diagonal $(k_x,\Pi/a-k_x)$, Eq. (5c) shows that $L_1=0$, so that neighboring layers are not coupled in the equation of motion. A trivial exact solution is then obtained with the ansatz

$$S(2n)=p^n,\qquad S(2n+1)=0,\quad n=0,1,\cdots. \tag{15}$$

Direct substitution in (2) and (4) then gives

$$p = r/(r+4),$$

$$\Omega(k_x, \Pi/a) = 10 + 5r - 2(1-2r)\cos k_x a - r^2/(r+4), \tag{16}$$

$$\Omega(k_x, \Pi/a - k_x) = 8 + 7r + 2r\cos(2k_x a) - r^2/(r+4). \tag{17}$$

For example, if $r = 0.2$ then (16) gives

$$\Omega(\Pi/a, \Pi/a) = 12.190476190477,$$

which agrees with Table I of Trullinger, up to 11 significant figures. This amply verifies the excellent convergence of the Gottlieb polynomial method, at least at the zone corner. At other points covered by (16) and (17) similar agreement was found with the results obtained by expanding in 10—15 Gottlieb functions. Indeed, from a root search on (14) good agreement was obtained everywhere in the zone.

Several points now emerge concerning the Gottlieb function method:

(i) Excellent results can be obtained for the eigenfrequencies, as tested against an exact solution.

(ii) The Gottlieb solutions $S(m)$ need not have the correct behavior as $m \to \infty$. In the present case the exact solutions are sums of exponentials $\sum_i A_i \exp(-\lambda_i m)$, whereas the Gottlieb approximations of nth degree have the simpler asymptotic form $m^n \exp(-\lambda m)$. Since considerable freedom is available in the choice of λ, the asymptotic behavior is not necessarily well given by the Gottlieb expansion. Even the λ_{opt} corresponding to optimum Gottlieb convergence appears to be only roughly correlated with the exact λ_i's. In general it is not surprising that the eigenfunctions $S(m)$ are less accurately determined than the eigenfrequencies Ω; for the problem can be formulated as a variational calculation in which the frequency is stationary with respect to changes in the eigenfunction. It should be noted that, for many surface problems where the surface mode decays rapidly, the asymptotic form of the excitations is largely irrelevant.

(iii) The method presented in this paper is readily generalized to give essentially exact results for linear equations of the form

$$S(m) = \sum_{m' \geq 0} K(|m - m'|)\, S(m'), \tag{18}$$

where K has a finite range, say n. The method requires a root search of an nth degree polynomial and of a specific derived function $f(\Omega)$, but is otherwise exact. It is therefore probably preferable to the Gottlieb expansion method, which is approximate and requires diagonalization of matrices whose size increases with the desired accuracy.

However, for a large class of nontranslationally-invariant problems (e.g., the equation

$$S(m) = \alpha(m)\, S(m) + \sum_{m' \geq 0} K(|m - m'|)\, S(m'),$$

where $\alpha(m)$ is *not* constant), the Gottlieb function expansion continues to be available while the method of the present paper fails.

ACKNOWLEDGMENTS

The author is grateful for useful discussions with S. E. Trullinger and W. Lamb.

[1]S. E. Trullinger, J. Math. Phys. 17, 1884 (1976).

Hamiltonians with x^{-2}-like singularity*

Štěpán Pick

The J. Heyrovský Institute of Physical Chemistry and Electrochemistry, The Czechoslovak Academy of Sciences, Prague, Czechoslovakia †

(Received 13 February 1976)

A class of singular Hamiltonians with deficiency indices (1,1) is studied. By using asymptotic behavior of functions from their domain, all self-adjoint extensions are described. The Friedrichs extension is shown to be (in a certain sense) a limit of nonsingular Hamiltonians.

I. BASIC PROPERTIES OF SINGULAR HAMILTONIANS

During the past few years, a special type of the singular Hamiltonian—quantal oscillator—has been investigated by several authors from various points of view.[1] The aim of the present paper is to explain some properties of a more general class of singular Hamiltonians. Let us consider the linear differential operator

$$L_0 = -\frac{d^2}{dx^2} + \frac{s^2 - 1/4}{x^2} + v(x), \quad 0 \leq s < 1, \tag{1}$$

as the linear operator in the complex Hilbert space $L_2(0,\infty)$. For simplicity, $v(x)$ is supposed to be a real function, continuous and bounded from below on $[0,\infty)$. L_0 is defined on the domain $D(L_0) = C_0^\infty$ of all infinitely differentiable functions with the compact support in $(0,\infty)$ (finite functions). Clearly, L_0 is a symmetric operator. At the point $x=0$ the case of the limit circle occurs and at $x=\infty$ that of the limit point occurs.[2] From the Glazman theorem we conclude, that deficiency indices of L_0 are $(1,1)$.[3] Therefore, one-parameter family of self-adjoint extensions (s. e.) exists.[4] It is important to mention another property of L_0[5]:

$$(u, L_0 u) \geq \inf v(x)\, \|u\|^2, \quad u \in C_0^\infty. \tag{2}$$

In other words, L_0 is bounded from below. Now, the following can be stated.[6] Let L be any s. e. of L_0. The spectrum of L in the interval $(-\infty, \inf v(x))$ consists either of the only simple eigenvalue or it is empty. Vice versa, any point of this interval is the eigenvalue of some s. e. of L_0. To obtain further results, it is useful to consider as an example the quantal oscillator[1]:

$$M_0 = -\frac{d^2}{dx^2} + \frac{s^2 - 1/4}{x^2} + x^2, \quad 0 \leq s < 1, \tag{3}$$

with the domain $D(M_0) = C_0^\infty$. Let us also consider the adjoint operator M_0^*. The spectrum of M_0^* is the entire complex plane $\mathbb{C}$. Indeed, the eigenfunction, corresponding to the eigenvalue $-4p + 2 + 2s$, $p \in \mathbb{C}$, is

$$B_{p,s}(x) = x^{s+1/2} \exp -\left(\frac{x^2}{2}\right) \Psi(p, s+1, x^2). \tag{4}$$

In (4), $\Psi(a,c,x)$ is the confluent hypergeometric function of the second kind.[7] It is also possible to write

$$B_{p,s}(x) = x^{-1/2} W_{k,m}(x^2), \quad p = 1/2 - k + m, \quad s = 2m, \tag{5}$$

where $W_{k,m}$ is the Whittaker function. It is easy to understand why M_0^* is not self-adjoint. Integrating by parts the expression

$$Q(u,v) = (u, M_0^* v) - (M_0^* u, v), \quad u, v \in D(M_0^*),$$

we obtain

$$Q(u,v) = \lim_{x \to 0^+} (\bar{u}' v - \bar{u} v')(x). \tag{6}$$

$Q(u,v)$ is generally different from zero on $D(M_0^*) \times D(M_0^*)$. The behavior of functions (4) for $x \to 0^+$ is

$$0 < s < 1: f \sim x^{1/2}(x^{-s} + \sigma x^s) + O(x^{5/2-s}),$$

$$\sigma = \Gamma(p-s)/\Gamma(p) \in \mathbb{C}, \tag{7a}$$

$$f \sim x^{1/2+s} + O(x^{5/2-s}), \quad \sigma = +\infty, \tag{7b}$$

$$s = 0: f \sim x^{1/2}(\ln x - \sigma) + O(x^{5/2} \ln x),$$

$$\sigma = -(1/2\psi(p) + \gamma) \in \mathbb{C}, \tag{7c}$$

$$f \sim x^{1/2} + O(x^{5/2}), \quad \sigma = +\infty. \tag{7d}$$

In (7c), $\psi(p)$ is the logarithmic derivative of the Γ function and γ is the Euler constant.[7]

Due to the continuity of $v(x)$, all functions from the domain of L_0^* have the same behavior at $x=0$ as in the case of M_0^*. Further, on the domain of any s. e. of L_0 the form (6) has to vanish. [There is no term with $x \to \infty$ in (6) for $v(x)$ considered.]

This and the general theory[4] proves the following theorem.

Theorem 1: For any fixed $\sigma \in \mathbb{R} \cup \{+\infty\}$, let L^σ be the differential operator in Eq. (1) with the domain $D(L^\sigma)$, defined as follows: $u \in D(L^\sigma)$, if

(1) $u \in L_2(0,\infty)$ is continuous in $(0,\infty)$;

(2) u' exists, locally absolutely continuous in $(0,\infty)$;

(3) $L^\sigma u \in L_2(0,\infty)$;

(4) In some right neighborhood of $x=0$, u is a linear combination of the function with behavior (7) (for chosen σ) and a function $g = o(x^{1/2+s})$.

L^σ is a s. e. of L_0 and any s. e. of L_0 can be obtained in this way. $L^\infty \equiv L_F$ is the Friedrichs extension (F. e.)[8] of L_0 and has the same lower bound $b \geq \inf v(x)$ as L_0.

Let us consider again M_0(3). M^σ has a simple, purely point spectrum $\lambda_0 < \lambda_1 < \cdots$. For $\sigma \to -\infty$, $\lambda_0 \to -\infty$, $\lambda_k \to 4k - 2 + 2s$, $k = 1, 2, \cdots$. For $\sigma \to +\infty$, $\lambda_k \to 4k + 2 + 2s$, $k = 0, 1, \cdots$. For $\sigma = +\infty$, we obtain the well-known case[1] with the eigenfunctions $x^{1/2+s} \exp(-x^2/2) L_s^k(x^2)$; $L_s^k(x)$ are the Laguerre polynomials.[7]

It is worthwhile to note the following fact: Due to the vanishing of Eq. (6), the flux of probability is nonsingular on $D(L^\sigma)$ in the one-dimensional case. Analogi-

cally, in the three-dimensional case ($\varphi(r) = u(r)r^{-1}$) the flux of probability $\sim j_r r^2$ is nonsingular.

II. REGULARIZATION

In the monograph of Landau and Lifshitz,[8] a simple method of regularization for singular Hamiltonians is used. On the interval $[0, \epsilon]$, $\epsilon > 0$, the constant value $(s^2 - 1/4)/\epsilon^2$ is substituted instead of $(s^2 - 1/4)/x^2$. Adding the boundary condition $u(0) = 0$, we define uniquely a self-adjoint operator L_ϵ.[9] In the limit $\epsilon \to 0$, we get the behavior (7b) for eigenfunctions for the case $s > 0$.[8] It is easy, however, to obtain an analogical result for $s = 0$.

We can associate a closed quadratic form $1_\epsilon(u) = (u, L_\epsilon u)$ with any L_ϵ.[9] From the results given by Kato,[10] Theorem 2 follows immediately.

Theorem 2: For any $z \in \mathbb{C}$, $\mathrm{Im} z \neq 0$, the resolvent $(L_\epsilon - z)^{-1}$ converges strongly to the resolvent $(L_F - z)^{-1}$ of the F. e. of L_0 (as $\epsilon \to 0$). Let $dE(\lambda)$ be the spectral measure of L_F ($L_F = \int \lambda\, dE(\lambda)$), and let λ_0 be a continuity point of $dE(\lambda)$. The spectral projector $E(\lambda_0) = \int_{-\infty}^{\lambda_0} dE(\lambda)$ is the strong limit of the corresponding spectral projectors of L_ϵ.

This theorem suggests, in fact, the possibility of numerical treatment of problems with singular operators. Other approximative methods, using functions with correct behavior at $x = 0$, have been proposed recently.[11] (In this paper, singular Hamiltonians are used to describe physically real systems—quasilinear molecules.)

ACKNOWLEDGMENTS

The author is indebted to Dr. D. Papoušek, V. Danielis, and their colleagues from the Department of Molecular Spectroscopy, at the J. Heyrovský Institute of Physical Chemistry and Electrochemistry, for valuable discussions and the encouragement to write this paper.

*The paper is based on the author's doctoral thesis, Charles University, Prague, 1975.

†Máchova 7, 121 38 Praha 2, Czechoslovakia.

[1]F. Calogero, J. Math. Phys. 10, 2191 (1969); 12, 419 (1971); B. Sutherland, J. Math. Phys. 12, 246 (1971); N.I. Allilujev, Zh. Eksp. Teor. Fiz. 61, 15 (1971); F. Zirilli, J. Math. Phys. 15, 1202 (1974); L. Lathouwers, J. Math. Phys. 16, 1393 (1975).

[2]Compare E.A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955), Chap. IX.

[3]M.A. Naijmark, *Linear Differential Operators* (in Russian) (Nauka, Moscow, 1969), 2nd ed., Sec. 17.5.

[4]For the theory of s.e. of dif. operators see Ref. 3, especially Secs. 14—18.

[5]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966), Chap. VI, Sec. 4.

[6]Ref. 3, Sec. 14, Theorems 13 and 16.

[7]A. Erdélyi, Ed., *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I, Chap. VI.

[8]L.D. Landau and E.M. Lifshitz, *Quantum Mechanics* (Nauka, Moscow, 1974) (Eng. transl., Pergamon, New York, 1958), 3rd ed., Sec. 35.

[9]Ref. 5, Chap. VI.

[10]Ref. 5, Chap. VIII, Theorems 3.11, 3.13, and 1.15.

[11]I. Bohâček and V. Danielis, D. Papoušek, and V. Špirko, J. Mol. Struct. (in preparation).

Hamiltonian dynamics in symplectic phase space and Nambu's formulation of mechanics*

Colber G. Oliveira†

Physics Department, Syracuse University, Syracuse, New York 13210
(Received 2 January 1976, revised manuscript received 1 March 1976)

Starting from the formulation of the Hamiltonian dynamics in phase spaces with a symplectic structure, we show that it is possible to formulate the mechanics recently proposed by Nambu by means of a mapping from the even-dimensional Hamiltonian phase space to a "phase space" which displays all properties of Nambu's theory. The difficulties concerning the process of quantization of Nambu's theory can be presently discussed at classical level by means of a suitable interpretation of the meaning of the several Hamiltonians of the theory. Some examples are given.

I. INTRODUCTION

Recently Nambu[1] proposed a new formulation of the analytical mechanics for phase spaces which may have even or odd number of dimensions. The physical motivation for constructing such formulation is based on the Liouville theorem. Nambu shows that is possible to formulate this theorem in phase spaces of even or odd dimensionality by means of a suitable increase in the number of the Hamiltonians. For an n-dimensional phase space his theory has $(n-1)$ Hamiltonians, all of them being constants of motion. The proof that the Liouville theorem follows from this method is a direct consequence of the result that the field of velocities is divergenceless. One of the properties of the usual Hamiltonian theory is also of this nature, namely, the fact that the field of velocities $v \to (\dot{q}_i, \dot{p}_i)$ is divergenceless allows for a proof of the Liouville theorem.[2] By introducing a matrix representation for Hamilton's theory it is possible to show that the proof of Liouville's theorem follows similarly to the case of Nambu's theory, the only difference, of course, being the dimensionality of both phase spaces. This fact together with the result that the quantized version of Nambu's theory is the usual Heisenberg representation of quantum mechanics seems to indicate that there exists a connection, at classical level, between these two formulations. In this paper we look for such a relationship between the usual canonical dynamics and the new version proposed by Nambu. The method which will be used is the matrix representation of the Hamiltonian mechanics. This formulation is a well known mathematical method which has made infrequent appearances in the physics literature.

In Sec. 1 we give a short summary of the matrix formulation of the Hamiltonian dynamics, under the several forms that it may be exhibited. In Sec. 2 we show how it is possible to transform to another representation which possesses all properties of Nambu's theory. At the last section of this paper we present some examples of our method.

1. FORMULATION OF THE HAMILTONIAN DYNAMICS IN 2s-DIMENSIONAL SYMPLECTIC PHASE SPACES

We denote by (q_i, p_i), $i = 1, \ldots, 2s$, the variables of a canonical phase space. Let V_{2s} be the space of the variables y^a, $a = 1, \ldots, 2s$, defined as a column matrix:

$$y^a = \begin{pmatrix} q_i \\ p_i \end{pmatrix}.$$

Define the antisymmetric $2s \times 2s$ matrix η as[3]

$$\eta^{a\, b+s} = \delta^{ab}, \quad \eta^{a+s\, b} = -\delta^{ab}, \tag{1.1}$$

and $\eta^{ab} = 0$ otherwise. From now on we shall use the first letters of the Latin alphabet to indicate degrees of freedom ranging from 1 to 2s if they belong to the y variables, or respectively from 1 to s and from $s+1$ to 2s if they refer to the q and p variables. Thus,

$$y^a = q_a, \quad a = 1, \ldots, s,$$

$$y^a = p_{a-s}, \quad a = s+1, \ldots, 2s.$$

The Hamiltonian equations assume the form

$$\frac{dy^a}{dt} = \eta^{ab} \frac{\partial H}{\partial y^b}. \tag{1.2}$$

The matrix η^{ab} satisfies the condition

$$\eta^{ab} = [y^a, y^b], \tag{1.3}$$

where $[A, B]$ denotes the Poisson bracket (PB) of the dynamical functions A and B. The relations (1.2) and (1.3) show that we can write the Hamiltonian equations under the usual notation of Poisson brackets. The Poisson bracket is here defined by

$$[A, B] = \eta^{ab} \frac{\partial A}{\partial y^a} \frac{\partial B}{\partial y^b} \tag{1.4}$$

and the equations of motion for closed dynamical systems may be written as

$$\frac{dF}{dt} = [F, H] = \eta^{ab} \frac{\partial F}{\partial y^a} \frac{\partial H}{\partial y^b}. \tag{1.5}$$

The field of velocities, $v^a = \dot{y}^a$, is divergenceless,

$$\frac{\partial v^a}{\partial y^a} = \eta^{ab} H_{,ab} = 0 \tag{1.6}$$

Equations (1.5) and (1.6) for $F = \rho(y)$, with $\rho(y)$ the statistical distribution function allow us to prove that for closed dynamical systems

$$\frac{d\rho}{dt} = \frac{\partial}{\partial y^a}(\rho v^a),$$

and the rhs vanishes as a consequence of the continuity

equation for steady flow in V_{2s}. This proves the Liouville theorem in similar steps as those suggested by Nambu. However, we still may have a difference in dimensionality for both phase spaces. What we intend to show here is just an analogy to the mathematical steps followed in the proof of this theorem. Of course this has nothing to do with any deeper analogy between the two formalisms, but suggests that such an analogy may exist.

A transformation of V_{2s} on itself is a canonical transformation if the fundamental PB relations are left invariant,

$$y \to \bar{y}, \quad [\bar{y}^a, y^b] = \eta^{ab}.$$

Thus, canonical transformations leave invariant the tensor η^{ab},

$$\bar{\eta}^{ab} = \eta^{ab}. \tag{1.7}$$

From (1.4) we have

$$[\bar{y}^a, \bar{y}^b] = \eta^{rs} \frac{\partial \bar{y}^a}{\partial y^r} \frac{\partial \bar{y}^b}{\partial y^s} = \eta^{ab}. \tag{1.8}$$

This relation represents an usual tensor law of transformation, but now we have the constraint (1.7). Taking determinants on both sides of (1.8), we find (use that $|\eta| \neq 0$, indeed, $|\eta| = +1$)

$$\left| \frac{\partial \bar{y}}{\partial y} \right| = \pm 1.$$

The connected components of a canonical transformation are those with Jacobian equal to (+1). A special class of canonical transformations are the linear transformations in V_{2s},

$$\bar{y}^a = L^a{}_b\, y^b.$$

The connected components satisfy $|L| = 1$. From (1.8) we have in this case

$$L \eta L^T = \eta. \tag{1.9}$$

For infinitesimal linear canonical transformations we have

$$\bar{y}^a = y^a + \xi^a(y), \quad \xi^a = \epsilon^a{}_b\, y^b.$$

From (1.9) we get

$$\eta^{ac} \epsilon^b{}_c + \eta^{cb} \epsilon^a{}_c = 0 \tag{1.10}$$

The quantity playing the role of a metric here is the tensor η^{ab}. However since this tensor is antisymmetric we have to define properly the process of raising and lowering of indices. First we note that the inverse of the matrix η is the matrix $\eta^{-1} = (\eta_{ab}) = -(\eta^{ab})$. Raising and lowering of indices will be defined as

$$\chi^a = \eta^{ab} \chi_b = -\chi_b \eta^{ba},$$

$$\chi_a = \chi^b \eta_{ba} \eta_{ab} \chi^b.$$

In general, for objects with several indices we have

$$\chi^{a \cdots d \cdots c} = \eta^{db} \chi^{a \cdots}{}_{b}{}^{\cdots c},$$

and so on. Using these results in Eq. (1.10) we find

$$\epsilon^{ba} = \epsilon^{ab}.$$

Thus, infinitesimal linear canonical transformations correspond to $2s \times 2s$ symmetric matrices when all indices are written as contravariant indices. The condition $|L| = 1$ is then identically verified by matrices of the form $L^a{}_b = \delta^a{}_b + \epsilon^a{}_b$, if $\epsilon^{ab} = \epsilon^{ba}$. A possible generator for infinitesimal linear canonical transformations is the velocity field for quadratic Hamiltonians,

$$H = \tfrac{1}{2} b_{ac}\, y^a y^c, \quad b^T = b,$$

since then

$$\xi^a = \epsilon v^a = \epsilon \eta^{ac} b_{cs}\, y^s = \epsilon\, b^a{}_s\, y^s.$$

We can easily verify that $b^{ac} = b^{ca}$ (use that $b_{ac} = b_{ca}$ by definition). Of course not all infinitesimal canonical transformations need be generated in this way. We can also present this transformation in the PB notation

$$\delta y^a = \xi^a = \epsilon [y^a, H].$$

For our present purposes the representation which we have discussed up to now (we will call it the y representation from now on) is not much useful. We introduce another representation by means of a mapping on the elements of the y representation as

$$y^a \to w^a, \quad w^a = \tau^a{}_b\, y^b$$

in such way that the doublet structure of the Hamiltonian phase space is brought out explicitly,

$$w^{2a-1} = q_a, \quad a = 1, 2, \ldots, s,$$

$$w^{2a} = p_a, \quad a = 1, 2, \ldots, s.$$

This mapping is clearly not canonical for the y representation since (1.9) is not verified for $L = \tau$. It corresponds to a relabeling of variables y^a defining a new column matrix w. Under this mapping the metric η^{ab} of the y representation transforms to a new metric ξ^{ab} according to

$$\xi = \tau \eta \tau^T, \quad \xi^T = -\xi \tag{1.11}$$

(note that τ is a $2s \times 2s$ orthogonal matrix). From (1.3) and using the property that the $\tau^a{}_b$ are numbers, we easily see that

$$\xi^{ab} = (\tau \eta^T \tau)^{ab} = [w^a, w^b]. \tag{1.12}$$

From (1.2) we have

$$\dot{w}^a = \tau^a{}_b\, \eta^{bc} \frac{\partial H}{\partial w^d} \frac{\partial w^d}{\partial y^c} = (\tau \eta^T \tau)^{ad} \frac{\partial H}{\partial w^d}.$$

Thus, the equation of motion takes the canonical form

$$\dot{w}^a = \xi^{ab} \frac{\partial H}{\partial w^b} = [w^a, H]. \tag{1.13}$$

We denote this representation as the w representation. All mathematical formulas holding for the y representation will also hold here. The PB is presently defined as

$$[A, B] = \xi^{ab} \frac{\partial A}{\partial w^a} \frac{\partial B}{\partial w^b}. \tag{1.14}$$

For any dynamical function $F(w)$ we have

$$\frac{dF}{dt} = \frac{\partial F}{\partial w^a} \dot{w}^a = \xi^{ab} \frac{\partial F}{\partial w^a} \frac{\partial H}{\partial w^b} = [F, H] \tag{1.15}$$

(as before when we considered closed systems). The

inverse of $\xi=(\xi^{ab})$ is the matrix $\xi^{-1}=(\xi_{ab})$, with $\xi_{ab}=-\xi^{ab}$.

We may define a symplectic form on V_{2s} as

$$\langle x, w\rangle=\xi_{ab}x^a w^b=\xi^{ab}x_a w_b$$

for two elements x, w in V_{2s}. But since V_{2s} is the space of the vectors $\omega=(w^a)$ it follows that we must have $x=f(w)$, since otherwise this inner product will not be fully defined. The PB is an example of a inner product of this type,

$$[A,B]=\left\langle\frac{\partial A}{\partial w},\frac{\partial B}{\partial w}\right\rangle. \tag{1.16}$$

Thus, the Poisson bracket between the dynamical functions $A(w)$ and $B(w)$ may be given by the symplectic inner product of the normals to the two hypersurfaces $A(w)=\text{const}$ and $B(w)=\text{const}$ in V_{2s}. The property $[A,A]=0$ is then a direct consequence of the symplectic structure of V_{2s}.

The process of raising and lowering of indices is similar to that used previously for the y representation, with ξ^{ab} and ξ_{ab} taking the role of a metric. Canonical transformations in the w representation are defined as

$$w^a\to\bar{w}^a,\quad[\bar{w}^a,\bar{w}^b]=\xi^{ab},$$

which imply $\bar{\xi}=\xi$. From (1.14) we have for these transformations

$$\xi^{cd}\frac{\partial\bar{w}^a}{\partial w^c}\frac{\partial\bar{w}^b}{\partial w^d}=\xi^{ab}. \tag{1.17}$$

Similarly as before the Jacobian of these transformations has value ±1.

Linear canonical transformations are presently defined as

$$u^a=L^a{}_b\,w^b,$$

and from (1.17) we have

$$\xi=L\xi L^T.$$

From the equation of motion (1.13) we see that the closed dynamical system moves along a sheet of the family of hypersurfaces $H(w)=\text{const}$, where H is the Hamiltonian

$$\frac{dH}{dt}=\frac{\partial H}{\partial w^a}v^a=\xi^{ab}\frac{\partial H}{\partial w^a}\frac{\partial H}{\partial w^b}=[H,H]=0.$$

From the geometrical point of view this means that conservation of energy is associated to the symplectic structure of V_{2s}.

From now on we will restrict the discussion to the particular case of $s=3$. The equation of motion (1.13) can be rewritten under the form

$$\begin{pmatrix}\dot{\omega}^1\\ \dot{\omega}^2\\ \dot{\omega}^3\end{pmatrix}=\begin{pmatrix}\epsilon&0&0\\0&\epsilon&0\\0&0&\epsilon\end{pmatrix}\begin{pmatrix}\partial H/\partial\omega^1\\ \partial H/\partial\omega^2\\ \partial H/\partial\omega^3\end{pmatrix} \tag{1.18}$$

for

$$\omega^i=\begin{pmatrix}q_i\\p_i\end{pmatrix}=\begin{pmatrix}\omega^{i(1)}\\ \omega^{i(2)}\end{pmatrix},\quad i=1,2,3,$$

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

$$\nabla_i\equiv\frac{\partial}{\partial\omega^i}=\begin{pmatrix}\partial/\partial q_i\\ \partial/\partial p_i\end{pmatrix}.$$

Thus, the equations of motion (1.18) may also be presented as

$$\dot{\omega}^i=\epsilon\frac{\partial H}{\partial\omega^i}. \tag{1.19}$$

(These results may, of course, be easily generalized for arbitrary values of s.) We denote this form of rewriting Hamilton's equations as the ω representation. It brings out explicitly the doublet structure of phase space. The transpose of the operator ∇_i is the row matrix operator

$$\overset{T}{\nabla}_i=\left(\frac{\partial}{\partial q_i}\ \frac{\partial}{\partial p_i}\right).$$

Operating on the lhs of (1.19) with the operator $\overset{T}{\nabla}_i$, we get again the result that the field of velocities is divergenceless,

$$\overset{T}{\nabla}_i\dot{\omega}^i=\overset{T}{\nabla}_i\epsilon\nabla_i H=\epsilon^{(a)(b)}\frac{\partial}{\partial\omega^{i(a)}}\frac{\partial H}{\partial\omega^{i(b)}}=0.$$

In this expression as well as in all the others which follow we are summing over all repeated indices. The fundamental PB relations here take the form

$$[\omega^{i(a)},\omega^{k(b)}]=\epsilon^{(a)(b)}\delta^{ik}. \tag{1.20}$$

They clearly show that elements corresponding to different doublets have vanishing Poisson brackets. From (1.20) we have for the PB of the dynamical functions $A(\omega)$ and $B(\omega)$

$$[A,B]=\epsilon^{(a)(b)}\frac{\partial A}{\partial\omega^{i(a)}}\frac{\partial B}{\partial\omega^{i(b)}}. \tag{1.21}$$

Thus, from (1.19) and (1.20) we have

$$\dot{\omega}^{i(a)}=[\omega^{i(a)},H].$$

In this representation $\epsilon=(\epsilon^{(a)(b)})$ and $\epsilon^{-1}=(\epsilon_{(a)(b)})$ correspond to the metric. Raising and lowering indices which label the components on each doublet is carried out similarly as before,

$$\chi^{(a)}=\epsilon^{(a)(b)}\chi_{(b)}=-\chi_{(b)}\epsilon^{(b)(a)},$$

$$\chi_{(a)}=\chi^{(b)}\epsilon_{(b)(a)}=-\epsilon_{(a)(b)}\chi^{(b)}.$$

The inner product of the two elements $x_i^{(a)}=f_i^{(a)}(\omega)$ and $\omega_i^{(a)}$ in V_{2s} is given by the symplectic form

$$\langle\mathbf{x},\boldsymbol{\omega}\rangle=\epsilon^{(a)(b)}x_{i(a)}\omega_{i(b)}=\epsilon^{(a)(b)}\mathbf{x}_{(a)}\cdot\boldsymbol{\omega}_{(b)}$$

(note that the indices labeling the several doublets, that is, the indices $i,j,\cdots$, may be written in any position since they correspond to an Euclidean structure). In this notation the PB takes the form

$$[A,B]=\left\langle\frac{\partial A}{\partial\boldsymbol{\omega}},\frac{\partial B}{\partial\boldsymbol{\omega}}\right\rangle.$$

Canonical transformations in this representation are

mappings which preserve the form of the metric,

$$\omega^{i(a)} \to \bar{\omega}^{i(a)}, \quad [\bar{\omega}^{i(a)}, \bar{\omega}^{k(b)}] = \epsilon^{(a)(b)}\delta^{ik}.$$

Therefore, from (1.21) we have for canonical transformations

$$\epsilon^{(c)(d)} \frac{\partial \bar{\omega}^{i(a)}}{\partial \omega^{j(c)}} \frac{\partial \bar{\omega}^{k(b)}}{\partial \omega^{j(d)}} = \delta^{ik}\epsilon^{(a)(b)}. \tag{1.22}$$

2. REPRESENTATION OF THE ANALYTICAL MECHANICS IN TERMS OF FUNCTIONS OF DOUBLETS

The ω representation studied at the end of the last section has two basic properties: It has an Euclidean structure on the indices $i, j, \cdots$ labeling the several doublets (for phase spaces of dimension $2s$ these indices go from $\underline{1}$ to $\underline{s}$) and has a symplectic structure for the degrees of freedom associated to each doublet. In this section we consider the case for $s=3$, the reason for this choice will be made clear in the exposition which follows. Presently we want to introduce a pure triplet representation of analytical mechanics. In order to obtain such a representation we have first of all to define a mathematical process for going from the previous set of three doublets to a pure state of three variables defining the triplet state. For doing this we consider a mapping from the space of the quantities $\omega^{i(a)}$ to a space of variables z^i, $z^i = z^i(\omega^{j(a)})$ in such way that in the new representation the indices corresponding to internal degrees of freedom for each doublet are no longer present. In other words, in the mapping functions we are summing over all indices of the type (a), (b), etc. This transformation will not preserve the canonical symmetry of the ω representation. Therefore, we now have a formula similar to (1.22) but in the rhs we have to write a new metric field. We denote this new metric by ρ^{ik},

$$\rho^{ik}(z) = [z^i, z^k] = \epsilon^{(a)(b)} \frac{\partial z^i}{\partial \omega^{j(a)}} \frac{\partial z^k}{\partial \omega^{j(b)}}. \tag{2.1}$$

In general this new metric becomes a function of the variables z^i, a property which was not present in the previous representations. We call this new representation the z representation.

The equations of motion become

$$\frac{dz^i}{dt} = \frac{\partial z^i}{\partial \omega^{j(a)}} \dot{\omega}^{j(a)} = \frac{\partial z^i}{\partial \omega^{j(a)}} \epsilon^{(a)(b)} \frac{\partial H}{\partial \omega^{j(b)}}.$$

Considering that H may be written as a function of the z^i we have

$$\frac{dz^i}{dt} = \rho^{ik} \frac{\partial H}{\partial z^k} = [z^i, H]. \tag{2.2}$$

The PB is here defined as

$$[A, B] = \frac{\partial A}{\partial z^i} [z^i, z^k] \frac{\partial B}{\partial z^k} = \rho^{ik} \frac{\partial A}{\partial z^i} \frac{\partial B}{\partial z^k}. \tag{2.3}$$

The indices $i, j, \cdots$ are written as contravariant indices, formally we may regard the matrix ρ^{ik} as a symplectic metric; however, here this interpretation cannot be taken in the sense used before (defining an inner product) since the choice that we will use for the ρ^{ik} will imply that this matrix is singular. Thus, we keep all indices which refer to the z variables as contravariant indices and do not use any definition for lowering such indices.

We define the operation

$$\{A, B, C\} = \epsilon^{ijk} \frac{\partial A}{\partial z^i} \frac{\partial B}{\partial z^j} \frac{\partial C}{\partial z^k}, \tag{2.4}$$

which corresponds to the volume contained by the three vectors ∇A, ∇B, and ∇C. The symbol ϵ^{ijk} is the Levi-Civita symbol with $\epsilon^{123} = +1$. This operation may be formally interpreted as a generalized "Poisson bracket" derivable from the "fundamental generalized PB relations,"

$$\{z^i, z^j, z^k\} = \epsilon^{ijk}. \tag{2.5}$$

The relations (2.4) and (2.5) are Nambu's definition of the PB.

We now particularize the general mapping (2.1) in such a form that the antisymmetric tensor ρ^{ik} takes the form

$$\rho^{ik} = \{z^i, G, z^k\} = \{z^i, z^j, z^k\} \frac{\partial G}{\partial z^j} = \epsilon^{ijk} \frac{\partial G}{\partial z^j} \tag{2.6}$$

[this is obtained from (2.4) for $A = z^i$, $B = G$, $C = z^k$], for $G(z)$ a given dynamical function. Then,

$$\rho^{ik} = [z^i, z^k] = \epsilon^{ijk} \frac{\partial G}{\partial z^j}. \tag{2.7}$$

For the choice (2.7) the equation of motion (2.2), for the z^i, takes the form postulated by Nambu,

$$\frac{dz^i}{dt} = \epsilon^{ijk} \frac{\partial G}{\partial z^j} \frac{\partial H}{\partial z^k}. \tag{2.8}$$

Equivalently we may write

$$\frac{dz^i}{dt} = [z^i, H] = \{z^i, z^j, z^k\} \frac{\partial G}{\partial z^j} \frac{\partial H}{\partial z^k} = \{z^i, G, H\}. \tag{2.9}$$

Therefore the pair of dynamical functions G and H which Nambu calls Hamiltonians are here represented by $G(z)$ associated to the weighting factor ρ^{ik} through Eq. (2.7) and by $H(z)$ which plays the role of the Hamiltonian in the usual sense (note that H appears through the terms $\dot{\omega}^{j(a)}$).

In general we have, for any two given dynamical functions $A(z)$ and $B(z)$,

$$[A, B] = \rho^{ik} \frac{\partial A}{\partial z^i} \frac{\partial B}{\partial z^k} = \{z^i, G, z^k\} \frac{\partial A}{\partial z^i} \frac{\partial B}{\partial z^k}.$$

Then,

$$[A, B] = \{A, G, B\}. \tag{2.10}$$

The Eq. (2.10) relating the two types of PB is consistent with the alternation and derivation laws. The first law is obviously verified by both types of PB. The derivation law is verified for $A \to A_1 A_2$, or for $B \to B_1 B_2$, for fixed G. From (2.10) we have for any dynamical function $A(z)$,

$$\frac{dA}{dt} = \frac{\partial A}{\partial z^i} \dot{z}^i = [A, H] = \{A, G, H\}. \tag{2.11}$$

In particular, for $A \to G$ we get

$$\frac{dG}{dt} = [G, H] = \{G, G, H\} = 0. \tag{2.12}$$

Therefore, the function $G(z)$ associated with the "metric" of this phase space is a constant of the motion. Again from (2.10) we obtain, for $B \to G$,

$$[A, G] = \{A, G, G\} = 0.$$

Thus, the function $G(z)$ has vanishing PB with any dynamical function $A(z)$. This property is a direct consequence of the definition of PB, with the ρ^{ik} as a weighting factor,

$$[A, G] = \rho^{ik} \frac{\partial A}{\partial z^i} \frac{\partial G}{\partial z^k} = \epsilon^{ijk} \frac{\partial G}{\partial z^j} \frac{\partial A}{\partial z^i} \frac{\partial G}{\partial z^k} = 0.$$

The function $G(z)$ is here the analog of a c-number (in the z representation) in the quantum theory. This explains the difficulties concerning the problem of quantization of this formulation, and the result that the quantized theory is equivalent to the usual Heisenberg representation of quantum mechanics.

As an example, for the free rigid rotator where $z^i = L^i$ (angular momentum, which presently is written in function of the doublets as $z^i = \frac{1}{2}\epsilon^{ijk}\omega_j^T \epsilon \omega_k$) the function $G(z)$ is proportional to the square of the angular momentum, and therefore has vanishing PB with any function of the angular momentum.

3. EXAMPLES

Besides the example of the rigid rotator given by Nambu, some other examples of his theory are known in the literature.[4] Presently we consider briefly three examples which are of interest. These three examples will have similar symmetry properties under rotations [taking Cartesian coordinates ($q_j = x_j$), we consider the class of Hamiltonians which have null PB with the generator of infinitesimal rotations, $\mathcal{G} = -\frac{1}{2}\alpha_{rs}\epsilon_{rsk}L_k$, where α_{rs} is an antisymmetric matrix (first order infinitesimals) related to an orthogonal rotation matrix by $R_{rs} = \delta_{rs} + \alpha_{rs}$]. This property will allow us to use the same set of constants of motion for these three examples: The total energy and the square of the angular momentum. Since the Lie algebra of the phase space variables z^i is related to the form of the constant of motion $G(z)$ through its definition [Eq. (2.7)], it follows that we can use the same set of variables z^i for these three examples (since we require the existence of the same constant of motion for all three cases). A convenient set of functions $z^i(\omega^{r(a)})$ for these examples is

$$z^i = \omega_j^{(a)} \beta^i_{(a)(b)} \omega_j^{(b)} = \omega_j^T \beta^i \omega_j,$$

for

$$\beta^1 = \tfrac{1}{2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \beta^2 = \tfrac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \beta^3 = \tfrac{1}{2}\begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}.$$

Thus,

$$z^1 = \tfrac{1}{2} p_i p_i, \quad z^2 = \tfrac{1}{2} x_i x_i, \quad z^3 = x_i p_i \tag{3.1}$$

(presently we use Cartesian coordinates, $q_i = x_i$). The Lie algebra of the PB of the variables z^i has the form

$$[z^1, z^2] = -z^3, \quad [z^1, z^3] = -2z^1, \quad [z^2, z^3] = 2z^2 \tag{3.2}$$

(which satisfies Jacobi's identity). From (2.7) and (3.2) we have

$$\frac{\partial G}{\partial z^3} = z^3, \quad \frac{\partial G}{\partial z^2} = -2z^1, \quad \frac{\partial G}{\partial z^1} = -2z^2$$

which has as solution

$$G(z) = \tfrac{1}{2}(z^3)^2 - 2z^1 z^2, \tag{3.3}$$

From (3.1) and (3.3) it follows directly that $G(z)$ is proportional to the square of the angular momentum

$$G(z) = -\tfrac{1}{2}\mathbf{L}^2, \tag{3.4}$$

i. e., it has the dimension of the square of an angular momentum. We have

$$[A(z), G(z)] = 0. \tag{3.5}$$

This condition holds for every function $A(z)$ (as was seen before), and implies that

$$[z^i, G(z)] = 0. \tag{3.6}$$

These conditions can be easily understood in the present case by the following. From (3.1), (3.4), and (3.6) we see that the square of the angular momentum has null PB with:

(a) z^1 = kinetic energy times mass (z^1/m has the dimension of a Hamiltonian for a free particle).

(b) z^2 = function of r^2, that is $z^2 = \frac{1}{2}r^2$, which corresponds to the potential energy of a particle in a central field of forces for a harmonic oscillator, where $z^2 = (m\omega^2)^{-1}V$.

(c) z^2 = scalar product of the vectors **x** and **p**.

These three results are a direct consequence of the invariance of the quantities z^1, z^2, and z^3 under rotations in the space of the variables **x** and **p**. The corresponding conditions for the case of a rigid rotator are simply the conditions $[L^i, \mathbf{L}^2] = 0$, which say that $\mathbf{L}^2$ is invariant under rotations.

The examples that will be considered are:

(1) *Free particle*: For this case the Hamiltonian becomes equal to the kinetic energy which we denote by the symbol H_0:

$$H = H_0 = (1/2m)\, p_i p_i = z^1/m. \tag{3.7}$$

Then, the two constants of motion $G(z) = -\frac{1}{2}\mathbf{L}^2$ and H_0 generate the motion according to Nambu's equation (2.8), and to (3.3) and (3.7) as

$$\dot{z}^1 = 0, \quad \dot{z}^2 = z^3/m, \quad \dot{z}^3 = 2z^1/m.$$

(2) *Isotropic harmonic oscillator*: Here the Hamiltonian has the form

$$H = H_0 + \tfrac{1}{2}k\mathbf{x}^2, \quad k = m\omega^2, \tag{3.8}$$

and is written in the function of the variables z^i as

$$H = z^1/m + kz^2. \tag{3.9}$$

The equations of motion in Nambu's representation follow from (2.8), (3.3), and (3.9) as

$$\dot{z}^1 = -kz^3, \quad \dot{z}^2 = (1/m)\, z^3, \quad \dot{z}^3 = (2/m)\, z^1 - 2kz^2,$$

which may easily be seen to correspond to the usual

equations of motion for the isotropic harmonic oscillator.

(3) *Motion of a pointlike mass in a central field of forces*: In this case we have the Hamiltonian

$$H = H_0 + V(r),$$

where $V(r)$ is the potential energy of the particle of mass m in this central field of forces. This Hamiltonian is written in term of the variables z^i as

$$H = z^1/m + V((2z^2)^{1/2}). \tag{3.10}$$

The equations of motion in the Nambu representation then follow from (2.8), (3.3), and (3.10) as

$$\dot{z}^1 = -(z^3/r) \cdot V'(r), \quad \dot{z}^2 = z^3/m,$$
$$\dot{z}^3 = (-2z^2/r) \cdot V'(r) + (2z^1/m),$$

where $V'(r) = dV/dr$.

ACKNOWLEDGMENTS

The major part of this work was completed when the author was on a leave of absence from the Department of Physics of the University of Brazilia, Brazil. The author acknowledges the Department of Physics of Syracuse University for the warm hospitality extended to him during this period. This work was supported by the National Council for Scientific and Technological Development (CNPq) of Brazil, under Grant No. 3800/73.

*Research supported by a fellowship of the CNPq Brazil.

†Present address: Department of Physics, University of Brazilia, 70.000 Brazilia, DF, Brazil.

[1]Y. Nambu, Phys. Rev. D 7, 2405 (1973).
[2]L. Landau and E. M. Lifschitz, *Statistical Physics*, Vol. 5 (Pergamon, New York, 1958), p. 10.
[3]C. W. Kilmister, *Hamiltonian Dynamics* (Wiley, New York, 1964), p. 85.
[4]M. Razavy and F. J. Kennedy, Can. J. Phys. 52, 1532 (1974).

Convergent low-frequency expansions for penetrable scatterers

George Dassios*

Department of Mathematics, University of Patras, Patras, Greece
(Received 28 October 1975; revised manuscript received 18 February 1976)

We develop convergent series solutions in powers of the wavenumber ($k = 2\pi/\lambda$) for the field (ψ) and the normalized (dimensionless) scattering amplitude (g) for scattering by lossless penetrable obstacles whose physical properties are specified by two real parameters. The first two terms of ψ are solutions of Laplace's equation and the term of order k^n, $n \geq 2$, satisfies a two-parameter Poisson equation whose inhomogeneous term is proportional to the k^{n-2} term. The leading term of g is of order k^3 (as obtained originally, by Rayleigh); the k^4 term is zero for shapes that have inversion symmetry, and vanishes in the forward direction for all shapes; the k^n terms, $n \geq 3$, are expressed as volume integrals of functions involving the terms of ψ up to order k^{n-2}. Equivalent expressions in terms of surface integrals are included. For a plane wave of arbitrary direction of incidence and a triaxial ellipsoid, we obtain explicitly the first four nonvanishing terms of ψ (to order k^3) and the first two nonvanishing terms of Img (to order k^5) and Reg (to order k^8). Corresponding results for spheroids, needle, disc, sphere, and for the one-parameter problems are obtained as special cases. The necessary transformation of the ellipsoidal harmonics are also provided.

INTRODUCTION

We consider the three-dimensional problem of Helmholtz's equation corresponding to the scattering of a wave with suppressed time factor ($\exp(-i\omega t)$) by a lossless, penetrable, smooth obstacle with all dimensions small compared to wavelength (λ). The physical properties of the scatterer are implicit in the transition conditions at its surface and in the interior value of the wavelength: the field (ψ) is continuous and its normal derivative is discontinuous at the surface, and the interior wavelength (λ_{in}) differs from λ. Thus the problem is specified by two physical parameters, the magnitude of the discontinuity of $\partial_n\psi$ and the ratio of external to internal wavelengths ($\eta = \lambda/\lambda_{in}$).

In linearized (small amplitude) acoustics, this two-parameter problem corresponds to the scattering of a pressure wave (Φ) by an obstacle whose density (ρ_-) and compressibility (C_-) are different from those (ρ_+, C_+) of the surrounding infinite medium; the associated velocity field is proportional to $\nabla\psi$. The two parameters of the problem correspond to the ratio of outer to inner densities ($\beta = \rho_+/\rho_-$) and the ratio of outer to inner compressibilities ($\mathbb{C} = C_+/C_-$) or equivalently to β and the index of refraction $\eta = (\mathbb{C}/\beta)^{1/2}$. For $\beta \to 0$ and $\mathbb{C} \to 0$ the results reduce to those for a rigid scatterer (Neumann boundary condition).

Rayleigh[1,2] was the first to consider the problem of approximating solutions of the wave equation for $k = 2\pi/\lambda \approx 0$ in terms of solutions of Laplace's equation. Using potential theory results, he obtained approximations for the far field (essentially, the scattering amplitude g) of the sphere[1] and of the triaxial ellipsoid[2] for the one-parameter problems of either equal densities or equal compressibilities, and constructed the corresponding approximation for the two-parameter problem by superposition. (We show that this superposition procedure is valid only for leading term approximation.)

Stevenson[3] developed a systematic procedure for obtaining low-frequency approximations for the corresponding electromagnetic problem. He applied the development to the ellipsoid[4] and obtained the first three terms of the expansion for the near field, and the first two terms for the far field in terms of elliptic integrals. Stevenson[4] also considers the special cases of the spheroid and sphere, and cites additional papers[3,4] which use alternative approaches.

The systematic development for the scalar problem of the soft scatterer (Dirichlet condition) of arbitrary shape is given by Morse and Feshbach.[5] For the special case of the sphere, they give the first three terms of the field and of the scattering amplitude explicitly. The exact solution for the sphere was derived originally by Rayleigh[1] as a Legendre–Hankel series; he obtained appropriate low-frequency approximations for various cases of physical interest.[1,2] The most complete low-frequency results available for spheroids are those of Burke,[6–8] who used appropriate approximations of the exact solutions in terms of spheroidal wave functions. For the soft,[6] hard,[7] and penetrable[8] cases he obtained ψ to order k^2 and g to order k^6. Existing explicit approximations for the soft and rigid sphere, the soft and rigid prolate spheroid, as well as for the soft and rigid oblate spheroid, are presented in Ref. 9, which contains many additional citations to the original literature. A survey of Rayleigh's work which also discusses his contribution to two-dimensional problems has been given by Twersky.[10] Other aspects of low-frequency scattering are surveyed by Kleinman,[11] and other approaches are considered in Refs. 12–15.

For the triaxial ellipsoid, Rayleigh's approximation of order k^3 for the scattering amplitude g corresponds to the leading term of Img. The leading term of Reg, or order k^6, was obtained by Twersky[12] by direct application of the general scattering theorem (generalized cross-section theorem) to Rayleigh's approximation. By extending Twersky's procedure, we obtain Reg to order k^8 by using our new results to order k^5 for Img.

STATEMENT OF THE PROBLEM

We seek a solution ψ of Helmholtz's equation corresponding to scattering of a plane wave ϕ by a penetrable obstacle with volume v and smooth surface s. (We take the origin of coordinates at the center of the smallest circumscribing sphere S_a, of radius a). In the exterior V the solution $\psi=\psi^+$ satisfies

$$(\nabla^2+k^2)\psi^+=0, \quad \mathbf{k}=k\hat{\mathbf{k}}, \quad k=2\pi/\lambda,$$
$$\psi^+=\phi+u, \quad \phi=\exp(i\mathbf{k}\cdot\mathbf{r}), \tag{1}$$

where the scattered wave u satisfies the radiation condition

$$\lim_{r\to\infty}\int_{S_r}|\partial_r u(\mathbf{r})-iku(\mathbf{r})|^2\,dS(\mathbf{r})=0. \tag{2}$$

In the interior, the solution $\psi=\psi^-$ is a nonsingular solution of

$$(\nabla^2+\kappa^2)\psi^-=0, \quad \kappa=\eta k, \tag{3}$$

where η is the relative index of refraction. On the boundary the field is continuous and the normal derivative is discontinuous, i. e. ,

$$\psi^+=\psi^-, \quad \partial_n\psi^+=\beta\partial_n\psi^-, \tag{4}$$

where $\partial_n=\hat{\mathbf{n}}\cdot\nabla$ and $\hat{\mathbf{n}}$ is the outer unit normal on s.

The two parameters we use are β and $C=\beta\eta^2$, or equivalently $B=\beta-1$ and $C=C-1$. If $B=C=0$, then the regions V and v are filled with the same material. In this case the incident wave meets no discontinuities in the medium of propagation and no scattering occurs (i. e. , $\psi=\phi$ and $u=0$).

In V the solution of (1)—(4) may be represented in terms of an integral over v as

$$\psi^+(\mathbf{r})=\exp(i\mathbf{k}\cdot\mathbf{r})+\frac{k}{4\pi i}\int_v[B\nabla h(k|\mathbf{r}-\mathbf{r}'|)\cdot\nabla\psi^-(\mathbf{r}')$$
$$-Ck^2h(k|\mathbf{r}-\mathbf{r}'|)\psi^-(\mathbf{r}')]\,dv(\mathbf{r}'), \tag{5}$$

where $h(x)=h_0^{(1)}(x)=\exp(ix)/ix$ is the normalized free-space Green's function for Helmholtz's equation in three dimensions. Using Gauss' theorems, we can transform (5) to a surface integral over s

$$\psi^+(\mathbf{r})=\exp(i\mathbf{k}\cdot\mathbf{r})+\frac{k}{4\pi i}\int_s[(B+1)h(k|\mathbf{r}-\mathbf{r}'|)\partial_n\psi^-(\mathbf{r}')$$
$$-\psi^-(\mathbf{r}')\partial_n h(k|\mathbf{r}-\mathbf{r}'|)]\,ds(\mathbf{r}'). \tag{6}$$

If we substitute the asymptotic form for $kr\sim\infty$ in (5) or (6), we obtain for $r\gg a$

$$u(\mathbf{r})\sim g(\mathbf{r},\hat{\mathbf{k}})h(kr), \tag{7}$$

where the normalized scattering amplitude g is given by

$$g(\hat{\mathbf{r}},\hat{\mathbf{k}})=\frac{ik^2}{4\pi}\int_v[iB\hat{\mathbf{r}}\cdot\nabla\psi^-(\mathbf{r}')+kC\psi^-(\mathbf{r}')]$$
$$\times\exp(-ik\hat{\mathbf{r}}\cdot\mathbf{r}')\,dv(\mathbf{r}') \tag{8}$$

or by

$$g(\hat{\mathbf{r}},\hat{\mathbf{k}})=\frac{k}{4\pi i}\int_s[(B+1)\nabla\psi^-(\mathbf{r}')+ik\psi^-(\mathbf{r}')\hat{\mathbf{r}}]\cdot\hat{\mathbf{n}}$$
$$\times\exp(-ik\hat{\mathbf{r}}\cdot\mathbf{r}')\,ds(\mathbf{r}'). \tag{9}$$

The proof of the well-posedness of this boundary value problem, based on proving existence and uniqueness for the solution of the corresponding integral equation formulation of the problem, is discussed in Refs. 16—18.

We restrict consideration to small ka, where a is the radius of the smallest sphere that circumscribes the scatterer, and obtain series solutions in powers of k. The incident plane wave $\phi=\exp(i\mathbf{k}\cdot\mathbf{r})$ is analytic at $k=0$, and we assume that the fields ψ^- and u are also analytic at $k=0$. Therefore, convergent Taylor series for ψ^+ and ψ^- exist in powers of k. The procedure we follow reduces the boundary value problem for Helmholtz's equation to a sequence of boundary value problems for Laplace's and Poisson's equations. Both the differential and integral equation formulation are used throughout. The integral equation which describes the problem provides the nonvanishing terms of the far field as well as the particular solutions of the Poisson's equations.

DERIVATION OF THE APPROXIMATIONS

A. The field

Since both ψ^+ and ψ^- are assumed analytic at $k=0$, the series

$$\psi^\pm(\mathbf{r})=\sum_{n=0}^{\infty}\frac{(ik)^n}{n!}\phi_n^\pm(\mathbf{r}) \tag{10}$$

converge in their regions of analyticity and can be differentiated and integrated term by term.

The leading term in each of these expansions is independent of k and is identical to the solution of the corresponding potential theory problem ($k=0$):

$$\nabla^2\phi_0^\pm=0, \quad \phi_0^+=\phi_0^-, \quad \partial_n\phi_0^+=(B+1)\partial_n\phi_0^-, \quad \phi_0^+\sim 1. \tag{11}$$

A solution to this problem is given by $\phi_0^\pm=1$ everywhere for any admissible surface s. Since the problem is well posed[16,17] and ψ^+, ψ^- are taken to be analytic at $k=0$, it follows that this solution is unique. Substituting (10) into (1), (3), and (4), equating coefficients of equal powers of k, we obtain

$$\nabla^2\phi_n^\pm=n(n-1)\eta_\pm^2\phi_{n-2}^\pm, \quad \eta_+=1, \quad \eta_-=\eta,$$
$$\phi_n^+=\phi_n^-, \quad \partial_n\phi_n^+=(B+1)\partial_n\phi_n^-. \tag{12}$$

Similarly, substituting (10), and the entire functions

$$\phi(\mathbf{r})=\exp(i\mathbf{k}\cdot\mathbf{r})=\sum_{n=0}^{\infty}\frac{(ik)^n}{n!}(\hat{\mathbf{k}}\cdot\mathbf{r})^n. \tag{13}$$

and

$$kh(k|\mathbf{r}-\mathbf{r}'|)=\frac{\exp(ik|\mathbf{r}-\mathbf{r}'|)}{i|\mathbf{r}-\mathbf{r}'|}=\sum_{n=0}^{\infty}\frac{(ik)^n}{in!}|\mathbf{r}-\mathbf{r}'|^{n-1} \tag{14}$$

into (5), we obtain

$$\phi_n^+(\mathbf{r}) = (\hat{\mathbf{k}}\cdot\mathbf{r})^n + \frac{B}{4\pi}\sum_{\rho=0}^{n-1}\binom{n}{\rho}(\rho-1) \times \int_v |\mathbf{r}-\mathbf{r}'|^{\rho-3}(\mathbf{r}-\mathbf{r}')\cdot\nabla\phi_{n-\rho}^-(\mathbf{r}')\,dv(\mathbf{r}') - \frac{n(n-1)C}{4\pi}\sum_{\rho=0}^{n-2}\binom{n-2}{\rho} \times \int_v |\mathbf{r}-\mathbf{r}'|^{\rho-1}\phi_{n-2-\rho}^-(\mathbf{r}')\,dv(\mathbf{r}'). \quad (15)$$

The expression (15) for ϕ_n^+ satisfies the first equation in (12). The term $(\hat{\mathbf{k}}\cdot\mathbf{r})^n$ of (15) is the contribution of the incident wave to the nth coefficient. The remaining terms are proportional to either B or C; those in B involve the velocity field (i. e., $\nabla\phi^-$) and those in C the pressure field. The leading term ϕ_0^+ is unity, and ϕ_1^+ depends only on B. Equation (15) expresses ϕ_n^+ at any point in V in terms of ϕ_i^-, $0\le i\le n$ (the interior coefficients of order less than or equal to the order of ϕ_n^+).

In order to determine the nonvanishing terms of ϕ_n^+ at infinity, we use the asymptotic form

$$\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|} = (\mathbf{r}-\mathbf{r}')\left[\frac{1}{r}+O\left(\frac{1}{r^2}\right)\right]\sim\hat{\mathbf{r}} \quad (16)$$

then (15) in the far field becomes

$$\phi_n^+(\mathbf{r}) \sim (\hat{\mathbf{k}}\cdot\mathbf{r})^n + \frac{B}{4\pi}\sum_{\rho=2}^{n-1}\binom{n}{\rho}(\rho-1)\hat{\mathbf{r}}\cdot\int_v \nabla\phi_{n-\rho}^-(\mathbf{r}) \times |\mathbf{r}-\mathbf{r}'|^{\rho-2}\,dv(\mathbf{r}') - \frac{n(n-1)C}{4\pi}\sum_{\rho=1}^{n-2}\binom{n-2}{\rho} \times \int_v \phi_{n-2-\rho}^-(\mathbf{r}')|\mathbf{r}-\mathbf{r}'|^{\rho-1}\,dv(\mathbf{r}'). \quad (17)$$

The contribution of the incident wave to the far field is of order r^n which is three orders of magnitude higher than the contribution of the terms that depend on the parameters B and C. This is in accord with the fact that the incident wave becomes progressively more prominent while the scattered field dies out as we recede from the scatterer. The far field form of ϕ_n^+ is expressed in terms of ϕ_i^-, $0\le i\le n-2$, which are specified functions.

For each $n=1,2,\cdots$ Eqs. (12) and (17) define a potential problem whose solution provides the coefficients ϕ_n^+ and ϕ_n^- of the expansions (10), i. e., the wave theory problem has been reduced to a sequence of potential problems which can be solved iteratively. For $n=1,2,\cdots$ we have

$$\phi_1^+ = \hat{\mathbf{k}}\cdot\mathbf{r} - \frac{B}{4\pi}\int_v \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3}\cdot\nabla\phi_1^-(\mathbf{r}')\,dv(\mathbf{r}') \sim \hat{\mathbf{k}}\cdot\mathbf{r}, \quad (18)$$

$$\phi_2^+ = (\hat{\mathbf{k}}\cdot\mathbf{r})^2 - \frac{B}{4\pi}\int_v \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3}\cdot\nabla\phi_2^-(\mathbf{r}')\,dv(\mathbf{r}') - \frac{C}{2\pi}\int_v \frac{dv(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \sim (\hat{\mathbf{k}}\cdot\mathbf{r})^2, \quad (19)$$

$$\phi_3^+ = (\hat{\mathbf{k}}\cdot\mathbf{r})^3 - \frac{B}{4\pi}\int_v \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3}\cdot\nabla\phi_3^-(\mathbf{r}')\,dv(\mathbf{r}') + \frac{3B}{4\pi}\int_v \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\cdot\nabla\phi_1^-(\mathbf{r}')\,dv(\mathbf{r}') - \frac{3C}{2\pi}\int_v \frac{\phi_1^-(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,dv(\mathbf{r}') - \frac{3Cv}{2\pi} \sim (\hat{\mathbf{k}}\cdot\mathbf{r})^3 + \frac{3B}{4\pi}\hat{\mathbf{r}}\cdot\int_v \nabla\phi_1^-(\mathbf{r}')\,dv(\mathbf{r}') - \frac{3Cv}{2\pi}. \quad (20)$$

From (12) and (17) we see that the equations defining the first coefficient ϕ_1^+ depends only on B while all higher coefficients depend on both B and C. In the far field, ϕ_1^+ and ϕ_2^+ are independent of parameters and the higher coefficients depend on both B and C.

From (17) we obtain $\phi_n^+ = O(r^n)$ as $r\to\infty$ and, therefore, the convergence of (10) becomes poorer as r increases and fails to converge as $r\to\infty$. We may overcome this difficulty by using (15) to express ϕ_n on $S_{a+\epsilon}$ (a sphere centered at the origin with radius $a+\epsilon$), where $\epsilon>0$ and a is the radius of the smallest sphere that circumscribes s; then, since $|\mathbf{r}| = a+\epsilon<\infty$ for $\mathbf{r}\in S_{a+\epsilon}$ the series (10) converges on $S_{a+\epsilon}$ and provides ψ on $S_{a+\epsilon}$. Thus by (4), Eq. (6) transform to

$$\psi^+ = \exp(i\mathbf{k}\cdot\mathbf{r}) + \frac{k}{4\pi i} \times \int_{S_{a+\epsilon}} [h(k|\mathbf{r}-\mathbf{r}'|)\partial_{r'}\psi^+(\mathbf{r}') - \psi^+(\mathbf{r}')\partial_{r'}h(k|\mathbf{r}-\mathbf{r}'|)] \times dS(\mathbf{r}'). \quad (21)$$

Equation (21) specifies $\psi^+(\mathbf{r})$ for $|\mathbf{r}|\ge a+\epsilon$ in terms of ψ^+ and $\partial_r\psi^+$ on $S_{a+\epsilon}$.

Equation (12) is Laplace's equation for $n=1$ and Poisson's equation for $n\ge 2$. From (15) we see solutions of Poisson's equation are obtained by quadratures: All ϕ_i^-, $0\le i\le n-1$ are known (from the evaluation of the successive coefficients), and the unknown term

$$-\frac{B}{4\pi}\int_v \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\cdot\nabla\phi_n^-(\mathbf{r}')\,dv(\mathbf{r}')$$

can be omitted because it is a harmonic function. The term

$$-\frac{n(n-1)}{4\pi}C\int_v \frac{\phi_{n-2}^-(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,dv(\mathbf{r}')$$

is also harmonic and can be eliminated. Thus the required particular solution of (12) is provided by the nonharmonic terms of (15).

B. The scattering amplitude

Substitution of (10) and of

$$\exp(-ik\hat{\mathbf{r}}\cdot\mathbf{r}') = \sum_{n=0}^{\infty}\frac{(-ik)^n}{n!}(\hat{\mathbf{r}}\cdot\mathbf{r}')^n \quad (22)$$

into (8) or (9) yields

$$g(\hat{\mathbf{r}},\hat{\mathbf{k}}) = \frac{k^2}{4\pi}\sum_{n=1}^{\infty}(ik)^n\sum_{\rho=0}^{n-1}\frac{(-1)^\rho}{\rho!(n-1-\rho)!}\left(-\frac{B}{n-\rho}\hat{\mathbf{r}}\cdot\int_v \nabla\phi_{n-\rho}^-(\mathbf{r}')(\hat{\mathbf{r}}\cdot\mathbf{r}')^\rho\,dv(\mathbf{r}') + C\int_v \phi_{n-1-\rho}^-(\mathbf{r}')(\hat{\mathbf{r}}\cdot\mathbf{r}')^\rho\,dv(\mathbf{r}')\right) \quad (23)$$

or, equivalently,

$$g(\hat{\mathbf{r}},\hat{\mathbf{k}}) = \frac{k^2}{4\pi}\sum_{n=1}^{\infty}(ik)^n\sum_{\rho=0}^{n}\frac{(-1)^\rho}{\rho!(n-\rho)!}\left(\frac{B+1}{n-\rho+1}\int_s (\hat{\mathbf{r}}\cdot\mathbf{r}')^\rho\hat{\mathbf{n}}\cdot\nabla\phi_{n+1-\rho}^-(\mathbf{r}')\,ds(\mathbf{r}') + \int_s \phi_{n-\rho}^-(\mathbf{r}')(\hat{\mathbf{r}}\cdot\mathbf{r}')^\rho(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}})\,ds(\mathbf{r}')\right), \quad (24)$$

where we used

$$\nabla\phi_0^- = 0, \quad \hat{\mathbf{r}} \cdot \int_s \hat{\mathbf{n}}\, ds(\mathbf{r}) = 0, \quad \int_s \hat{\mathbf{n}} \cdot \nabla\phi_1^-(\mathbf{r}')\, ds(\mathbf{r}'). \tag{25}$$

Applying Gauss's theorem to (12), we obtain

$$\int_v \nabla^2\phi_n^-(\mathbf{r}')\, dv(\mathbf{r}') = \int_s \hat{\mathbf{n}} \cdot \nabla\phi_n^-(\mathbf{r}')\, ds(\mathbf{r}') = n(n-1)\left[(C+1)/(B+1)\right] \int_v \phi_{n-2}^-(\mathbf{r}')\, dv(\mathbf{r}'), \tag{26}$$

i.e., the flux through s of the nth coefficient of the interior velocity field is proportional to the volume integral over v of the $(n-2)$th coefficient of the interior pressure field. The last equation in (25) is (26) for $n=1$.

The coefficient ϕ_n^- is a harmonic function for $n=1$ and a solution of Poisson's equation for $n \geq 2$ and, consequently, represents a potential function of known density distribution over v. The gradient of this potential $\nabla\phi_n^-$ is proportional to the velocity field. Furthermore,

$$(\hat{\mathbf{r}} \cdot \mathbf{r}')^\rho = \left(\sum_{i=1}^{3} o_i x_i'\right)^\rho = \sum_{\kappa=0}^{\rho} \sum_{\sigma=0}^{\kappa} \binom{\rho}{\kappa}\binom{\kappa}{\sigma} o_1^\sigma o_2^{\kappa-\sigma} o_3^{\rho-\kappa} x_1'^\sigma x_2'^{\kappa-\sigma} x_3'^{\rho-\kappa}. \tag{27}$$

Hence, the integrals

$$\int_v \phi_{n-1-\rho}^-(\mathbf{r}')(\hat{\mathbf{r}} \cdot \mathbf{r}')^\rho\, dv(\mathbf{r}') \tag{28}$$

are moments of the interior pressure fields, and

$$\hat{\mathbf{r}} \cdot \int_v \nabla\phi_{n-\rho}^-(\mathbf{r}')(\hat{\mathbf{r}} \cdot \mathbf{r}')^\rho\, d\rho(\mathbf{r}') \tag{29}$$

are moments of the interior velocity fields projected on the direction of observation $\hat{\mathbf{r}}$. The integrals in (24) are moments of the flux through s of the same physical quantities that appear in (28) and (29).

The first term of g is of order k^3; all odd-order terms are imaginary and all even-order terms are real. The evaluation of g to order k^n requires knowledge of ϕ_i^-, $0 \leq i \leq n-2$ if we use the volume integral representation (23), and $0 \leq i \leq n-1$ if we use the surface integral representation (24); it is therefore more advantageous to work with the form (23).

The first approximation

$$g(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = \frac{ik^3}{4\pi}\left[Cv - B\hat{\mathbf{r}} \cdot \int_v \nabla\phi_1^-(\mathbf{r}')\, dv(\mathbf{r}')\right] + O(k^4) \tag{30}$$

consists of essentially two terms, one of which depends only on B and the other only on C; here v stands for the volume. This result was obtained originally by Rayleigh,[1,2] who assumed from the start that the small k approximation consisted of two terms, one proportional to B and one proportional to C. Rayleigh obtained the two particular approximations corresponding to $B=0$ and $C=0$ in terms of the potential theory solutions and superposed the results to obtain essentially (30). We show that for $n \geq 2$ the ϕ_n^- terms depend on both B and C and g does not separate into two such sets of terms.

Equation (23) gives g as a sum of two sets of terms: the one that involves the moments of the velocity projected on $\hat{\mathbf{r}}$ has B as a factor, and the other that involves the moments of the pressure fields has C as a factor. The form suggests that if either the compressibility or density of the scatterer is equal to that of the surrounding medium, then the expression for g would be devoid of the pressure or velocity fields, respectively. If the densities are equal then $B=0$ and, therefore, both the field and its normal derivative on s are continuous: $\eta^2 = C+1$. If the compressibilities are equal, then $C=0$ and $\eta^2 = 1/(B+1)$ as holds for most gases. If the inner density and the inner compressibility tend to infinity, B and C tend to zero. The limit $B \to -1$ and $C \to -1$ corresponds to the boundary condition $\partial_n\psi^+ = 0$ on s (Neumann condition), and to $\psi^-(\mathbf{r}) = 0$ as is appropriate for the rigid scatterer.

The normalized scattering amplitude g satisfies the reciprocity theorem[19]

$$g(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = g(-\hat{\mathbf{k}}, -\hat{\mathbf{r}}). \tag{31}$$

In addition, g satisfies the scattering theorem[19]

$$-g(\mathbf{r}, \mathbf{k}) - g^*(\mathbf{k}, \mathbf{r}) = \frac{1}{2\pi}\int g(\hat{\mathbf{p}}, \hat{\mathbf{k}}) g^*(\hat{\mathbf{p}}, \hat{\mathbf{r}})\, d\Omega(\hat{\mathbf{p}}), \tag{32}$$

where * denotes the complex conjugate and the integration is over the surface of the unit sphere. If the scatterer has inversion symmetry ($\mathbf{r} \in s$ implies $-\mathbf{r} \in s$), then

$$g(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = g(\hat{\mathbf{k}}, \hat{\mathbf{r}})$$

and the scattering theorem reduces to

$$-\operatorname{Re} g(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = \frac{1}{4\pi}\int g(\hat{\mathbf{p}}, \hat{\mathbf{k}}) g(\hat{\mathbf{p}}, \hat{\mathbf{r}})\, d\Omega(\hat{\mathbf{p}}). \tag{33}$$

In terms of real functions $A_n(\hat{r}, \hat{k})$, we assume

$$g(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = ik^3 A_3 + k^4 A_4 + ik^5 A_5 + k^6 A_6 + ik^7 A_7 + k^8 A_8 + O(k^9). \tag{34}$$

Then (33) gives

$$A_4(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = 0, \tag{35}$$

$$A_6(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = -\frac{1}{4\pi}\int A_3(\hat{\mathbf{p}}, \hat{\mathbf{r}}) A_3(\hat{\mathbf{p}}, \hat{\mathbf{k}})\, d\Omega(\hat{\mathbf{p}}), \tag{36}$$

$$A_8(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = -\frac{1}{4\pi}\int \left[A_3(\hat{\mathbf{p}}, \hat{\mathbf{r}}) A_5(\hat{\mathbf{p}}, \hat{\mathbf{k}}) + A_5(\hat{\mathbf{p}}, \hat{\mathbf{r}}) A_3(\hat{\mathbf{p}}, \hat{\mathbf{k}})\right] d\Omega(\hat{\mathbf{p}}). \tag{37}$$

Therefore, for scatterers with inversion symmetry the first approximation to Reg is of order k^6.

Equation (36) gives the leading real part of g in terms of the leading imaginary part of g, and (37) gives the second term of Reg when the first two terms of Img are known. This method of evaluating the successive terms of the Reg from Img was originally described by Twersky,[12] who applied the theorem to Rayleigh's result A_3 for ellipsoids to obtain the corresponding A_6. This method reduces the calculations significantly: In order to evaluate the k^6 term from (23), we require ϕ_i^-, $1 \leq i \leq 4$, and for the k^8 term ϕ_i^-, $1 \leq i \leq 6$; however, by using (36) and (37) only ϕ_1^- is needed for the k^6 term and only ϕ_i^-, $1 \leq i \leq 3$, for the k^8 term.

More generally for unsymmetrical scatterers, from (32) in terms of (34) we obtain

$$A_3(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = A_3(\hat{\mathbf{k}}, \hat{\mathbf{r}}), \quad A_4(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = -A_4(\hat{\mathbf{k}}, \hat{\mathbf{r}}), \quad A_5(\hat{\mathbf{r}}, \hat{\mathbf{k}}) = A_5(\hat{\mathbf{k}}, \hat{\mathbf{r}}). \tag{38}$$

Consequently, if $\hat{\mathbf{r}} = \hat{\mathbf{k}}$, then

$$A_4(\hat{\mathbf{k}}, \hat{\mathbf{k}}) = 0. \tag{39}$$

The scattering theorem applied for $\hat{\mathbf{k}} = \hat{\mathbf{r}}$ becomes

$$-\mathrm{Reg}(\mathbf{k}, \mathbf{k}) = \frac{1}{4\pi} \int |g(\hat{\mathbf{p}}, \hat{\mathbf{k}})|^2 \, d\Omega(\hat{\mathbf{p}}) = \frac{k^2}{4\pi} \sigma_s, \tag{40}$$

where σ_S is the scattering cross section.

Using (34) and (39) in the first form of (40), we obtain

$$\sigma_S = -4\pi k^6 A_6(\hat{\mathbf{k}}, \hat{\mathbf{k}}) - 4\pi k^8 A_8(\hat{\mathbf{k}}, \hat{\mathbf{k}}) + O(k^8), \tag{41}$$

where

$$-A_6(\hat{\mathbf{k}}, \hat{\mathbf{k}}) = |A_6(\hat{\mathbf{k}}, \hat{\mathbf{k}})| \tag{42}$$

follows from (36) for $\hat{\mathbf{r}} = \hat{\mathbf{k}}$. The first approximation for the total energy scattered is proportional to the inverse fourth power of the wavelength[1] (Rayleigh scattering).

APPLICATIONS

A. Triaxial ellipsoid

We apply the iterative procedure (12), (17) to a triaxial ellipsoid given by

$$\sum_{i=1}^{3} \frac{x_i^2}{a_i^2} = 1, \quad a_1 > a_2 > a_3 > 0 \tag{43}$$

In ellipsoidal coordinates ρ, μ, ν we have

$$x_i^2 = \frac{(-1)^{i-1} h_i^2}{h_1^2 h_2^2 h_3^2} (\rho^2 - a_1^2 + a_i^2)(\mu^2 - a_1^2 + a_i^2)(\nu^2 - a_1^2 + a_i^2), \quad (i = 1, 2, 3), \tag{44}$$

with $-h_3 \leq \nu \leq h_3 \leq \mu \leq h_2 < \rho < +\infty$, where $2h_1$, $2h_2$, $2h_3$ are the interfocal distances of the main ellipses, i. e.,

$$h_1^2 = a_2^2 - a_3^2, \quad h_2^2 = a_1^2 - a_3^2, \quad h_3^2 = a_1^2 - a_2^2. \tag{45}$$

The value $\rho = a_1$ specifies the surface s of the scatterer, $h_2 \leq \rho < a_1$ the region v, and $a_1 < \rho < +\infty$ the region V. The outward normal derivative corresponds to ∂_ρ, and in the far field we have $r \sim \rho$. The directions of incidence and observation are given by $\hat{\mathbf{k}} = \sum_{n=1}^{3} i_n \hat{\mathbf{x}}_n$ and $\hat{\mathbf{r}} = \sum_{n=1}^{3} o_n \mathbf{x}_n$ respectively, where $\hat{\mathbf{x}}_n$ $(n = 1, 2, 3)$ are the Cartesian unit vectors.

Expressions for $(\hat{\mathbf{k}} \cdot \mathbf{r})^n$ $(n = 1, 2, 3)$ in terms of surface ellipsoidal harmonics,[10–29] as well as transformations between ellipsoidal and Cartesian coordinates are given in the Appendix.

Omitting long computations,[29] we give the solutions of the potential problems that determine the coefficients to the order k^3, as well as the two leading nonvanishing terms of Img and Reg in terms of elliptic integrals.

As shown for the general case (and as can be verified for the special case of the ellipsoid at hand) $\phi_0^* = 1$.

The coefficients of k are determined by (12) for $n = 1$ and (18). Solving this problem, we obtain

$$\phi_1^+ = \sum_{n=1}^{3} \left(1 - BV \frac{I_n^1(\rho)}{H_n^1}\right) i_n x_n, \tag{46}$$

$$\phi_1^- = \sum_{n=1}^{3} \frac{i_n x_n}{H_n^1}, \tag{47}$$

where

$$V = a_1 a_2 a_3, \quad I_i^1(\rho) = \int_\rho^\infty \frac{du}{(u^2 - a_1^2 + a_i^2) S(u)}, \quad I_i^1 = I_i^1(a_1), \quad S(u) = (u^2 - h_2^2)^{1/2} (u^2 - h_3^2)^{1/2}, \quad H_i^1 = 1 + BVI_i^1, \quad (i = 1, 2, 3). \tag{48}$$

The coefficients $\phi_2^\pm$ are determined by (12) for $n = 2$ and (19). Particular solutions of Poisson's equation are provided by

$$\phi_2^{P+} = (\hat{\mathbf{k}} \cdot \mathbf{r})^2, \quad \phi_2^{P-} = [(C+1)/(B+1)](\hat{\mathbf{k}} \cdot \mathbf{r})^2. \tag{49}$$

The solution is

$$\phi_2^+ = \phi_2^{P+} - \tfrac{2}{3} CVI^0(\rho) - 2i_1 i_2 i_3 x_1 x_2 x_3 \sum_{n=1}^{3} \frac{\Gamma_n I_n^2(\rho)}{i_n x_n} - 2VM \frac{I_\Lambda^2(\rho)}{H_n^2} \left[\sum_{n=1}^{3} i_n^2 \beta_n \mu_n^+\right] \left[\sum_{k=1}^{3} \frac{x_k^2}{\Lambda - a_k^2} + 1\right] + (\Lambda \rightleftharpoons \Lambda'), \tag{50}$$

$$\phi_2^- = \phi_2^{P-} - \tfrac{2}{3} CVI^0 - \frac{1}{3}\left(\frac{C+1}{B+1} - 1\right) \sum_{n=1}^{3} i_n^2 a_n^2 - 2i_1 i_2 i_3 x_1 x_2 x_3 \sum_{n=1}^{3} \frac{1}{i_n x_n} \left(\frac{C+1}{B+1} - \frac{1}{H_n^2}\right) - \frac{M}{H_n^2} \left[\sum_{n=1}^{3} i_n^2 \beta_n \mu_n^-\right] \left[\sum_{k=1}^{3} \frac{x_k^2}{\Lambda - a_k^2} + 1\right] + (\Lambda \rightleftharpoons \Lambda'). \tag{51}$$

The symbol $(\Lambda \rightleftharpoons \Lambda')$ represents a repetition of the last term with Λ and Λ' interchanged, where

$$\left.\begin{matrix} \Lambda \\ \Lambda' \end{matrix}\right\} = \frac{1}{3}\left\{\pm\left[\sum_{n=1}^{3}\left(a_n^4 - \frac{V^2}{a_n^2}\right)\right]^{1/2} + \sum_{n=1}^{3} a_n^2\right\}. \tag{52}$$

The functions that appear in (50) and (51) are

$$\Gamma_n=\frac{BV}{H_n^2}\left(\sum_{k=1}^{3}a_k^2-a_n^2\right),\quad \beta_n=(-1)^n\frac{h_n^2}{h_1^2h_2^2h_3^2}\frac{\Lambda'-a_n^2}{\Lambda-\Lambda'},\quad M=(\Lambda-a_1^2)(\Lambda-a_2^2)(\Lambda-a_3^2), \tag{53}$$

$$\mu_n^+=C(\Lambda-a_n^2)+Ba_n^2,\quad \mu_n^-=2C\Lambda VI_\Lambda^2+\frac{C-B}{B+1}\left(\frac{1}{\Lambda}-2VI_\Lambda^2\right)a_n^2. \tag{54}$$

The solution is expressed in terms of the elliptic integrals

$$I^0(\rho)=\int_\rho^\infty\frac{du}{S(u)},\quad I^0=I^0(a_1),\quad I_\Lambda^2(\rho)=\int_\rho^\infty\frac{du}{(u^2-a_1^2+\Lambda)^2S(u)},\quad I_\Lambda^2=I_\Lambda^2(a_1),$$

$$I_\alpha^2(\rho)=\int_\rho^\infty\frac{du}{(u^2-a_1^2+a_\beta^2)(u^2-a_1^2+a_\gamma^2)S(u)},\quad (\alpha,\beta,\gamma)=(1,2,3),\ (2,3,1),\ (3,1,2), \tag{55}$$

$$H_\Lambda^2=1+2BV\Lambda I_\Lambda^2,\quad H_\alpha^2=1+BV(a_\beta^2+a_\gamma^2)I_\alpha^2.$$

The coefficients ϕ_3^* are determined by the potential problem (12) for $n=3$ and (20). The particular solutions of Poisson's equations are

$$\phi_3^{P+}=(\hat{\mathbf{k}}\cdot\mathbf{r})^3+\frac{3BV}{2}\left[I^0(\rho)-\sum_{n=1}^{3}I_n^1(\rho)x_n^2\right]\sum_{\sigma=1}^{3}\frac{i_\sigma x_\sigma}{H_\sigma^1}, \tag{56}$$

$$\phi_3^{P-}=\frac{C+1}{B+1}\sum_{n=1}^{3}\frac{i_nx_n^3}{H_n^1}. \tag{57}$$

The solution is

$$\begin{aligned}\phi_3^+=\phi_3^{P+}&+3V\sum_{k=1}^{3}i_kx_k\frac{I_k^1(\rho)}{H_k^1}\left[\tau_1^+-\left(\Delta_k^1-\tfrac{1}{5}\sum_{n=1}^{3}i_n^2a_n^2\right)-\tau_2^+\left(\Delta_k^2-\tfrac{1}{5}\sum_{n=1}^{3}i_n^2(2a_k^2+a_n^2)\right)+Vh_1h_2h_3\sum_{k=1}^{3}i_kx_k\frac{M_k}{h_k}\frac{I_k^3(\rho)}{H_k^3}\right.\\&\left.\times\left[\tau_3^+\left(\Delta_k^3+\sum_{n=1}^{3}i_n^2\Delta_{kn}^5\right)-\tau_4^+\left(\Delta_k^4+\sum_{n=1}^{3}i_n^2\Delta_{kn}^6\right)\right]\left[\sum_{\sigma=1}^{3}\frac{x_\sigma^2}{\Lambda_k-a_\sigma^2}+1\right]+(\Lambda_k\rightarrow\Lambda_k')\right.\\&+6V^3\frac{x_1x_2x_3}{h_1^2h_2^2h_3^2}\frac{I_{123}(\rho)}{H_{123}}\left[\sum_{n=1}^{3}\frac{i_n}{a_n^2h_n^2}-(B+1)\frac{i_1i_2i_3}{h_1^2h_2^2h_3^2}\sum_{n=1}^{3}\frac{1}{a_n^2}\right],\end{aligned} \tag{58}$$

$$\begin{aligned}\phi_3^-=\phi_3^{P-}&+3V\sum_{k=1}^{3}i_kx_k\frac{1}{H_k^1}\left[\tau_1^-\left(\Delta_k^1-\tfrac{1}{5}\sum_{n=1}^{3}i_n^2a_n^2\right)-\tau_2^-\left(\Delta_k^2-\tfrac{1}{5}\sum_{n=1}^{3}i_n^2(2a_k^2+a_n^2)\right)\right]\\&+Vh_1h_2h_3\sum_{k=1}^{3}i_kh_k\frac{1}{H_k^3}\frac{M_k}{h_k}\left[\tau_3^-\left(\Delta_k^3+\sum_{n=1}^{3}i_n^2\Delta_{kn}^5\right)-\tau_4^-\left(\Delta_k^4+\sum_{n=1}^{3}i_n^2\Delta_{kn}^6\right)\right]\left[\sum_{\sigma=1}^{3}\frac{x_\sigma^2}{\Lambda_k-a_\sigma^2}+1\right]+(\Lambda_k\rightarrow\Lambda_k')\\&+6V^3\frac{x_1x_2x_3}{h_1^2h_2^2h_3^2}\frac{1}{H_{123}}\left[I_{123}\sum_{n=1}^{3}\frac{i_n}{a_n^2h_n^2}+\frac{i_1i_2i_3}{h_1^2h_2^2h_3^2}\left(\frac{1}{V^3}-I_{123}\sum_{n=1}^{3}\frac{1}{a_n^2}\right)\right].\end{aligned} \tag{59}$$

The symbol $(\Lambda_i\rightarrow\Lambda_i')$ represents a repetition of the last term with Λ_i and Λ_i' interchanged, where

$$\left.\begin{matrix}\Lambda_i\\ \Lambda_i'\end{matrix}\right\}=\frac{1}{5}\left\{\pm\left[4\sum_{n=1}^{3}a_n^4-3a_i^4-V^2\left(\sum_{n=1}^{3}\frac{1}{a_n^2}+\frac{6}{a_i^2}\right)\right]^{1/2}+2\sum_{n=1}^{3}a_n^2-a_i^2\right\},\quad (i=1,2,3). \tag{60}$$

The Δ's (functions of a_i, B, and C, i.e., of the geometry of the scatterer and its physical parameters) are

$$\begin{aligned}&\Delta_k^1=\frac{a_k^2}{5H_k^1}\left(H_k^1+\frac{C+1}{B+1}-1\right)-\frac{2BVI^0}{5H_k^1},\quad \Delta_k^2=\frac{1}{H_k^1}\left[\tfrac{3}{5}(C+1)a_k^2+\frac{B}{2}a_k^2-\frac{BVI^0}{2}\right]+\frac{BV}{10H_k^1}\sum_{n=1}^{3}\left[I_n^1(2a_k^2+a_n^2)-\frac{a_k^2}{V}\right](1+2\delta_{kn}),\\&\Delta_k^3=\frac{3BV}{2}\frac{1}{H_k^1}\sum_{n=1}^{3}\zeta_{kn}a_n^2I_n^1+\frac{C+1}{B+1}\frac{1}{H_k^1}\zeta_{kk}a_k^2,\\&\Delta_k^4=\frac{3BV}{2}\frac{1}{H_k^1}\sum_{n=1}^{3}\left[I_n^1(2a_k^2+a_n^2)-\frac{a_k^2}{V}\right]\zeta_{kn}+\frac{3(C+1)}{H_k^1}\zeta_{kk}a_k^2,\\&\Delta_{kn}^5=\frac{3}{1+2\delta_{kn}}\zeta_{kn}a_n^2,\quad \Delta_{kn}^6=\frac{3}{1+2\delta_{kn}}\zeta_{kn}(2a_k^2+a_n^2),\end{aligned} \tag{61}$$

where

$$\tau_1^+=B+1,\quad \tau_2^+=1,\quad \tau_3^+=(B+1)(\Lambda_k+2a_k^2),\quad \tau_4^+=\Lambda_k,$$

$$\tau_1^-=I_k^1-\frac{1}{V},\quad \tau_2^-=I_k^1,\quad \tau_3^-=\frac{H_k^3-(B+1)}{BV\Lambda_k},\quad \tau_4^-=\Lambda_kI_k^3, \tag{62}$$

and

$$\zeta_{kn}=(-1)^n\frac{h_k h_n^2}{h_1^3 h_2^3 h_3^3}\frac{\Lambda_k'-a_n^2}{\Lambda'-\Lambda_k'},\quad \delta_{kn}=\begin{cases}1, & k=n,\\ 0, & k\neq n,\end{cases} \tag{63}$$

$$M_\sigma=(\Lambda_\sigma-a_1^2)(\Lambda_\sigma-a_2^2)(\Lambda_\sigma-a_3^2),\quad I_n^3(\rho)=\int_\rho^\infty\frac{du}{(u^2-a_1^2+a_n^2)(u^2-a_1^2+\Lambda_n)^2S(u)},\quad I_n^3=I_n^3(a_1),$$
$$H_n^3=1+BV\Lambda_n(\Lambda_n+2a_n^2)I_n^3\quad(n=1,2,3),\quad I_{123}(\rho)=\int_\rho^\infty\frac{du}{u^2S^3(u)},\quad I_{123}=I_{123}(a_1),\quad H_{123}+1+BV^3I_{123}\Big(\sum_{i=1}^3\frac{1}{a_i^2}\Big). \tag{64}$$

The corresponding primed quantities are obtained from the above relations by replacing Λ_i by Λ_i'. In evaluating the coefficients $\phi_3^\pm$ we used

$$V\sum_{i=1}^3 I_i^1=1,\quad \sum_{i=1}^3 a_i^2 I_i^1=I^0,\quad I_1^2=\frac{I_3^1-I_2^1}{a_2^2-a_3^2},\quad I_2^2=\frac{I_3^1-I_1^1}{a_1^2-a_3^2},\quad I_3^2=\frac{I_2^1-I_1^1}{a_1^2-a_2^2}. \tag{65}$$

Having obtained the fields to order k^3, we can determine the scattering amplitude g to order k^5 by substitution in (23). Because of the ellipsoid's inversion symmetry the coefficient of order k^4 of g is zero. Since the $\phi_i^-(\mathbf{r}')$ are polynomials of degree i in x_n' $(n=1,2,3)$ for $i=1,2,3$ we see that the $\nabla\phi_i^-(\mathbf{r}')$ are polynomials of degree $i-1$. Similarly for $(\hat{\mathbf{r}}\cdot\mathbf{r}')^n$ $(n=1,2)$. Therefore, all integrands are polynomials in x_n $(n=1,2,3)$ of degree 0, 1, and 2. As a consequence of the symmetry of the ellipsoid, the only nonzero integrals are those whose integrands are either a constant, or x_n^2 $(n=1,2,3)$. Thus, the only integrals that have to be evaluated in order to obtain g to the order k^5 are the three principal moments of inertia of the ellipsoid. These are given by

$$\int_v x_n^2\,dv=\frac{4\pi V}{15}a_n^2\quad(n=1,2,3). \tag{66}$$

The term k^3 of g is proportional to the zeroth moment of the ellipsoid, i. e., the volume $4\pi V/3$. The term k^4 is proportional to the first order moments which vanish by symmetry. (The first order moments give the coordinates of the centroid which in our case coincides with the origin.) The term k^5 is proportional to the second order moments (moments of inertia). The form of the higher ellipsoidal harmonics implies that the term k^n is expressed in terms of the even moments of the ellipsoid up to the $(n-3)$th order (integrals which depend solely on the geometry of the ellipsoid).

Substituting (47), (51), and (59) into (23) and exploiting the symmetry and the form (66), we obtain the following multipole expansion in terms of the directions $\hat{\mathbf{r}}$ and $\hat{\mathbf{k}}$:

$$\mathrm{Im}g(\hat{\mathbf{r}},\hat{\mathbf{k}})\approx\frac{ik^3V}{3}\left[C-B\left(\frac{Y_1^{e0}(\hat{\mathbf{r}})Y_1^{e0}(\hat{\mathbf{k}})}{H_1^1}+\frac{Y_1^{e1}(\hat{\mathbf{r}})Y_1^{e1}(\hat{\mathbf{k}})}{H_2^1}+\frac{Y_1^{o1}(\hat{\mathbf{r}})Y_1^{o1}(\hat{\mathbf{k}})}{H_3^1}\right)\right]$$
$$+\frac{ik^5V}{30}\sum_{n,m}\sum_{n',m'}(\alpha_{nn'}^{mm'}Y_n^{em}(\hat{\mathbf{r}})Y_{n'}^{em'}(\hat{\mathbf{k}})+\beta_{nn'}^{mm'}Y_n^{om}(\hat{\mathbf{r}})Y_{n'}^{om'}(\hat{\mathbf{k}})) \tag{67}$$

where

$$\sum_{n,m}=\sum_{n=0}^3\sum_{m=0}^n,\quad Y_n^{em}(\hat{\mathbf{r}})=P_n^m(\cos\theta)\cos(m\phi),\quad Y_n^{om}(\hat{\mathbf{r}})=P_n^m(\cos\theta)\sin(m\phi)$$

and similarly for $Y_n^{em}(\hat{\mathbf{k}})$ and $Y_n^{om}(\hat{\mathbf{k}})$. The nonvanishing coefficients of the multipole expansion are

$$\alpha_{00}^{00}=\tfrac{1}{9}\sum_{n=1}^3\Big[\omega_n^3+\sum_{k=1}^3\omega_{kn}^4+3(\omega_n^6+\omega_n^7)\Big]+\omega^9,\qquad \alpha_{02}^{02}=\alpha_{20}^{20}=\tfrac{1}{18}\sum_{n=2}^3(-1)^n\Big[\omega_n^3+\sum_{k=1}^3\omega_{kn}^4+3\omega_n^7\Big].$$

$$\alpha_{02}^{00}=\alpha_{20}^{00}=\tfrac{1}{9}\sum_{n=1}^3(3\delta_{1n}-1)\Big[\omega_n^3+\sum_{k=1}^3\omega_{kn}^4+3\omega_n^7\Big],$$

$$\alpha_{22}^{02}=\alpha_{22}^{20}=-\tfrac{1}{18}\sum_{n=2}^3(-1)^n\omega_n^3+\tfrac{1}{18}\sum_{n=1}^3\sum_{k=2}^3(-1)^k(3\delta_{1n}-1)\omega_{kn}^4,$$

$$\alpha_{13}^{02}\big|_{\sigma=1}=\alpha_{31}^{20}\big|_{\sigma=1}=2\alpha_{13}^{13}\big|_{\sigma=2}=2\alpha_{31}^{31}\big|_{\sigma=2}=2\beta_{13}^{13}\big|_{\sigma=3}=2\beta_{31}^{31}\big|_{\sigma=3}=\tfrac{1}{30}(\omega_{\sigma 2}^8-\omega_{\sigma 3}^8),$$

$$\alpha_{13}^{00}=\alpha_{31}^{00}=\tfrac{1}{5}\sum_{n=1}^3(3\delta_{1n}-1)\omega_{1n}^8, \tag{*}$$

$$\alpha_{13}^{11}=\alpha_{31}^{11}=\tfrac{1}{30}\sum_{n=1}^3(4\delta_{1n}-3\delta_{2n}-\delta_{3n})\omega_{2n}^8,$$

$$\beta_{13}^{11}=\beta_{31}^{11}=\tfrac{1}{30}\sum_{n=1}^3(4\delta_{1n}-\delta_{2n}-3\delta_{3n})\omega_{3n}^8,$$

$$\alpha_{11}^{00}\big|_{\sigma=1}=\alpha_{11}^{11}\big|_{\sigma=2}=\beta_{11}^{11}\big|_{\sigma=3}=\omega_\sigma^1+\tfrac{1}{5}\sum_{n=1}^3(2\delta_{\sigma n}+1)(\omega_{\sigma n}^2+\omega_{\sigma n}^8),$$

$$\alpha_{22}^{00}=\tfrac{1}{9}\sum_{n=1}^3(3\delta_{1n}+1)\omega_n^3+\tfrac{1}{9}\sum_{n=1}^3\sum_{k=1}^3(3\delta_{1n}-1)(3\delta_{1k}-1)\omega_{kn}^4,$$

$$\alpha_{22}^{22}=\tfrac{1}{36}\sum_{n=2}^{3}[\omega_n^3+(-1)^n(\omega_{2n}^4-\omega_{3n}^4)],\quad 4\beta_{22}^{22}|_{\sigma=1}=\beta_{22}^{11}|_{\sigma=2}=\alpha_{22}^{11}|_{\sigma=3}=\tfrac{1}{9}\left(\sum_{n=1}^{3}\omega_n^3-\omega_\sigma^3+\omega_\sigma^5\right),$$

where

$$\omega_n^1=\frac{Ba_n^2}{H_n^1}\left(\frac{C+1}{B+1}+2\frac{C}{B}\right)+\frac{B(VI_n^1-1)}{(H_n^1)^2}\left[\left(H_n^1+\frac{C+1}{B+1}-1\right)a_n^2-2BVI^0\right]-\frac{BVI_n^1}{(H_n^1)^2}[3(C+1)a_n^2+B(VI_n^1+1)a_n^2+2BV(a_n^2I_n^1-I^0)],$$

$$\omega_{kn}^2=\frac{B}{H_k^1}(2Va_k^2I_k^1+a_n^2),\quad \omega_n^3=-2B\frac{C+1}{B+1}a_n^2,\quad \omega_{kn}^4=\frac{2BM}{H_\Lambda^2}\frac{a_k^2}{\Lambda-a_k^2}\beta_n\mu_n^-+(\Lambda\rightleftharpoons\Lambda'),$$

$$\omega_n^5=2B\left(\frac{C+1}{B+1}-\frac{1}{H_n^2}\right)\left(\sum_{k=1}^{3}a_k^2-a_n^2\right),\quad \omega_n^6=\frac{C}{3}\left(2\frac{C+1}{B+1}-5\right)a_n^2+\frac{2CM}{H_\Lambda^2}\beta_n\mu_n^-+(\Lambda\rightleftharpoons\Lambda'),$$

$$\omega_n^7=-Ca_n^2,\quad \omega_{kn}^8=\frac{Ba_n^2}{H_k^1},\quad \omega^9=\tfrac{10}{3}C^2VI^0,$$

where $\alpha_{nn'}^{mm'}|_{\sigma=1}$ means that $\alpha_{nn'}^{mm'}$ equals the given form with $\sigma=1$, etc. As we expect from the reciprocity theorem we have $\alpha_{nn'}^{mm'}=\alpha_{n'n}^{m'm}$ and $\beta_{nn'}^{mm'}=\beta_{n'n}^{m'm}$.

The real part of g can be evaluated from the scattering theorem (33). Equations (36) and (37) give

$$\begin{aligned}\mathrm{Re}g(\hat{\mathbf{r}},\hat{\mathbf{k}})=&-\frac{k^6V^2}{9}\left[C^2+\frac{B^2}{3}\left(\frac{Y_1^{e0}(\hat{\mathbf{r}})Y_1^{e0}(\hat{\mathbf{k}})}{(H_1^1)^2}+\frac{Y_1^{e1}(\hat{\mathbf{r}})Y_1^{e1}(\hat{\mathbf{k}})}{(H_2^1)^2}+\frac{Y_1^{o1}(\hat{\mathbf{r}})Y_1^{o1}(\hat{\mathbf{k}})}{(H_3^1)^2}\right)\right]-\frac{k^8V^2}{90}\left[2C\alpha_{00}^{00}-\frac{B}{3H_1^1}Y_1^{e0}(\hat{\mathbf{r}})\right.\\&\left.\times\sum_{n',m'}(\alpha_{1n'}^{0m'}Y_{n'}^{em'}(\hat{\mathbf{k}})+\beta_{1n'}^{0m'}Y_{n'}^{om'}(\hat{\mathbf{k}}))+(\hat{\mathbf{k}}\rightleftharpoons\hat{\mathbf{r}})-\frac{B}{3}\left(\frac{Y_1^{e1}(\hat{\mathbf{r}})}{H_2^1}+\frac{Y_1^{o1}(\hat{\mathbf{r}})}{H_2^1}\right)\sum_{n',m'}(\alpha_{1n'}^{1m'}Y_{n'}^{em'}(\hat{\mathbf{k}})+\beta_{1n'}^{1m'}Y_{n'}^{om'}(\hat{\mathbf{k}}))+(\hat{\mathbf{k}}\rightleftharpoons\hat{\mathbf{r}})\right]\\&+O(k^{10}),\end{aligned}\tag{68}$$

where the symbol $(\hat{\mathbf{k}}\rightleftharpoons\hat{\mathbf{r}})$ represents a repetition of the double sum with $\hat{\mathbf{k}}$ and $\hat{\mathbf{r}}$ interchanged. The orthogonality of the surface spherical harmonics has been used as well as

$$\int[Y_1^\rho(\hat{\mathbf{p}})]^2d\Omega(\hat{\mathbf{p}})=4\pi/3,\quad \rho=e0,e1,o1.\tag{69}$$

The form of the multipole coefficients shows the existence of terms that depend on both parameters B and C. In the limit as $B\to-1$ and $C\to-1$ we obtain the corresponding results for the rigid ellipsoid.

The scattering cross section is obtained from (40) by using (68) with $\hat{\mathbf{r}}=\hat{\mathbf{k}}$. For the special case $\hat{\mathbf{k}}=\hat{\mathbf{r}}=\hat{\mathbf{x}}_1$, we have

$$Y_n^{em}(\hat{\mathbf{x}}_1)=\delta_{0m},\quad Y_n^{om}(\hat{\mathbf{x}}_1)=0,$$

and

$$\mathrm{Im}g(\hat{\mathbf{x}}_1,\hat{\mathbf{x}}_1)\approx\frac{ik^3V}{3}\left(C-\frac{B}{H_1^1}\right)+\frac{ik^5V}{30}(\alpha_{00}^{00}+\alpha_{11}^{00}+\alpha_{22}^{00}+\alpha_{02}^{00}+\alpha_{20}^{00}+\alpha_{13}^{00}+\alpha_{31}^{00}),\tag{70}$$

$$\mathrm{Re}g(\hat{\mathbf{x}}_1,\hat{\mathbf{x}}_1)\approx-\frac{k^6V^2}{9}\left(C^2+\frac{B^2}{3H_1^2}\right)-\frac{k^8V^2}{45}\left[C\alpha_{00}^{00}-\frac{B}{3}\frac{\alpha_{11}^{00}+\alpha_{13}^{00}}{H_1^1}\right].\tag{71}$$

B. Spheroids

We obtain the corresponding results for spheroids by setting $a_2=a_3$, with $a_1>a_2$ for the prolate, and $a_1<a_2$ for the oblate. We use the spheroidal coordinates

$$x_1=\rho\cos\theta,\quad\begin{Bmatrix}x_2\\x_3\end{Bmatrix}=(\rho^2-h_3^2)^{1/2}\sin\theta\begin{Bmatrix}\cos\phi\\\sin\phi\end{Bmatrix},\quad 0\le\theta\le\pi,\ 0\le\phi<2\pi,\tag{72}$$

$$\rho=\begin{cases}h_3\cosh\omega, & a_1>a_2,\\ ih_3\sinh\omega, & a_1<a_2,\end{cases}\quad 0\le\omega<+\infty,\tag{73}$$

$$h_3=(a_1^2-a_2^2)^{1/2}=\begin{Bmatrix}1\\i\end{Bmatrix}|h_3|,\quad a_1\gtrless a_2,\tag{74}$$

and the definitions

$$\sigma=a_1/a_2\gtrless1,\quad a_1\gtrless a_2,\tag{75}$$

$$\tau=\begin{cases}(\sigma^2-1)^{-1/2}\ln[\sigma+(\sigma^2-1)^{1/2}], & a_1>a_2,\\ (1-\sigma^2)^{-1/2}\tan^{-1}[(1-\sigma^2)^{1/2}/\sigma], & a_1<a_2.\end{cases}\tag{76}$$

The elliptic integral appearing in the solution of the triaxial ellipsoid can be evaluated exactly for spheroids:

$$I^0(\rho)=\begin{cases}(1/2h_3)\ln[(\rho+h_3)/(\rho-h_3)], & a_1>a_2,\\ (1/ih_3)\tan^{-1}(ih_3/\rho), & a_1<a_2,\end{cases}\quad I^0=\tau/a_2,\quad I_1^1(\rho)=(1/h_3^2)[I^0(\rho)-1/\rho],$$

$$I_2^1(\rho)=I_3^1(\rho)=-\frac{1}{2h_3^2}\left(I^0(\rho)-\frac{\rho}{\rho^2-h_3^2}\right),\quad I_\Lambda^2(\rho)=\frac{9}{4h_3^4}\left(I^0(\rho)-\frac{3\rho}{3\rho^2-h_3^2}\right),$$

$$I^2_{\Lambda'}=I^2_1(\rho)=\frac{3}{8h_3^5}\left[I^0(\rho)-\frac{\rho(3\rho^2-5h_3^2)}{3(\rho^2-h_3^2)^2}\right],\quad I^2_2(\rho)=I^2_3(\rho)=-\frac{3}{2h_3^4}\left(I^0(\rho)-\frac{3\rho^2-2h_3^2}{3\rho(\rho^2-h_3^2)}\right),$$

$$I^3_1(\rho)=\frac{25}{4h_3^6}\left[I^0(\rho)-\frac{15\rho^2-4h_3^2}{3\rho(5\rho^2-3h_3^2)}\right],\quad I^3_2(\rho)=I^3_3(\rho)=-\frac{75}{16h_3^6}\left(I^0(\rho)-\frac{\rho}{6(\rho^2-h_3^2)}-\frac{25\rho}{6(5\rho^2-h_3^2)}\right),$$

$$I^{3'}_1(\rho)=I_{123}(\rho)=\frac{15}{8h_3^6}\left(I^0(\rho)+\frac{2\rho h_3^2}{15(\rho^2-h_3^2)^2}-\frac{7\rho}{15(\rho^2-h_3^2)}-\frac{8}{15\rho}\right),$$

$$I^{3'}_2(\rho)=I^{3'}_3(\rho)=-\frac{5}{16h_3^6}\left(I^0(\rho)-\frac{8\rho h_3^4}{15(\rho^2-h_3^2)^3}-\frac{\rho(3\rho^2-5h_3^2)}{3(\rho^2-h_3^2)^2}\right). \tag{77}$$

Also

$$\begin{aligned}
&H^1_1=1+B(\sigma\tau-1)/(\sigma^2-1),\quad H^1_2=H^1_3=1+B[(\sigma^2-1)-(\sigma\tau-1)]/2(\sigma^2-1),\\
&H^2_\Lambda=1+B\{3[2\sigma^2+1)(\sigma\tau-1)-(\sigma^2-1)]\}/2(\sigma^2-1)^2,\quad H^2_{\Lambda'}=H^2_1=1+B[3(\sigma\tau-1)+(2\sigma^2-3)(\sigma^2-1)]/4(\sigma^2-1)^2,\\
&H^2_2=H^2_3=1+B(\sigma^2+1)[(\sigma^2-1)-3(\sigma\tau-1)]/2(\sigma^2-1)^2,\quad H^3_1=1+B(4\sigma^2+1)[3(2\sigma^2+3)(\sigma\tau-1)-5(\sigma^2-1)]/4(\sigma^2-1)^3\\
&H^{3'}_1=H_{123}=1+B(2\sigma^2+1)[(2\sigma^2-7)(\sigma^2-1)+15(\sigma\tau-1)]/8(\sigma^2-1)^3,\\
&H^3_2=H^3_3=1+B(4\sigma^2+11)[(2\sigma^2+3)(\sigma^2-1)-3(4\sigma^2+1)(\sigma\tau-1)]/16(\sigma^2-1)^3,\\
&H^{3'}_2=H^{3'}_3=1+B[(\sigma^2-1)(8\sigma^4-18\sigma^2+15)-15(\sigma\tau-1)]/16(\sigma^2-1)^3.
\end{aligned} \tag{78}$$

If we substitude the above values of the elliptic integrals, as well as $a_2=a_3$, $h_1=0$,

$$\begin{aligned}
&-h_2^2=-h_3^2=3(\Lambda-a_1^2)=\Lambda'-a_1^2=\tfrac{5}{3}(\Lambda_1-a_1^2)=5(\Lambda_2-a_1^2)=5(\Lambda_3-a_1^2)=\Lambda'_1-a_1^2=\Lambda'_2-a_1^2=\Lambda'_3-a_1^2=-\tfrac{3}{2}(\Lambda-\Lambda')\\
&\quad=-\tfrac{5}{2}(\Lambda_1-\Lambda'_1)=-\tfrac{5}{4}(\Lambda_2-\Lambda'_2)=-\tfrac{5}{4}(\Lambda_3-\Lambda'_3)
\end{aligned} \tag{79}$$

in the solutions of the triaxial ellipsoid we obtain the corresponding solutions for the spheroids.

The scattering amplitude g for

$\hat{\mathbf{k}}=(\cos\alpha,\sin\alpha,0)$, $\hat{\mathbf{r}}=(\cos\theta,\sin\theta\cos\phi,\sin\theta\sin\phi)$

is given by

$$\mathrm{Im}g(\hat{\mathbf{r}},\hat{\mathbf{k}})\approx\frac{ik^3a_1a_2^2}{3}\left[C-B\sum_{i=1}^{2}\frac{Y_1^{i-1}(\hat{\mathbf{r}})Y_1^{i-1}(\hat{\mathbf{k}})}{H_i^1}\right]+\frac{ik^5a_1a_2^4}{30}\sum_{n,m}\sum_{n',m'}\gamma_{nn'}^{mm'}Y_n^m(\hat{\mathbf{r}})Y_{n'}^{m'}(\hat{\mathbf{k}}), \tag{80}$$

$$\mathrm{Re}g(\hat{\mathbf{r}},\hat{\mathbf{k}})\approx-\frac{k^6a_1^2a_2^4}{9}\left(C^2+\frac{B^2}{3}\sum_{i=1}^{2}\frac{Y_1^{i-1}(\hat{\mathbf{r}})Y_1^{i-1}(\hat{\mathbf{k}})}{(H_i^1)^2}\right)-\frac{k^8a_1^2a_2^6}{90}\left[2C\gamma_{00}^{00}-\frac{B}{3}\sum_{i=1}^{2}\frac{Y_1^{i-1}(\hat{\mathbf{r}})}{H_i^1}\left(\sum_{n',m'}\gamma_{1n'}^{(i-1)m'}Y_{n'}^{m'}(\hat{\mathbf{k}})\right)\right]-(\hat{\mathbf{k}}\leftrightarrow\hat{\mathbf{r}}). \tag{81}$$

The nonvanishing coefficients are

$$\gamma_{00}^{00}=\frac{10}{3}C^2\sigma\tau+\frac{2}{9}(\sigma^2+2)\left[(C-B)\frac{C+1}{B+1}-4C\right]-\frac{4(C-B)^2(\sigma^2-1)^2}{9B(2\sigma^2+1)}\left(\frac{1}{H_\Lambda^2}-\frac{1}{B+1}\right),$$

$$\gamma_{13}^{00}\big|_{\rho=1}=\gamma_{31}^{00}\big|_{\rho=1}=3\gamma_{13}^{11}\big|_{\rho=2}=3\gamma_{31}^{11}\big|_{\rho=2}=\frac{2B}{5H_\rho^1}(\sigma^2-1),$$

$$\gamma_{20}^{00}=\gamma_{02}^{00}=\frac{2}{9}(2-3C)(\sigma^2-1)-\frac{2}{9}\frac{C+1}{B+1}\left[B(2\sigma^2-1)+2(\sigma^2-1)\right]+\frac{4(\sigma^2-1)(C-B)}{9H_\Lambda^2},$$

$$\gamma_{22}^{00}=\frac{2B}{9}\frac{C+1}{B+1}-\frac{4B(2\sigma^2+1)}{9H_\Lambda^2},\quad \gamma_{22}^{11}=-\frac{2B}{9}\frac{\sigma^2+1}{H_3^2}, \tag{**}$$

$$\gamma_{11}^{00}=\frac{2}{(H_1^1)^2}[B^2\sigma\tau+\sigma^2(B+C)]-\frac{2B}{5H_1^1}(2\sigma^2+3)+\frac{4\sigma^2(\sigma^2-1)^2(C+1)(H_1^3-1)^2}{(4\sigma^2+1)(2\sigma^2+3)H_1^1H_1^3},$$

$$\gamma_{11}^{11}=\frac{2}{(H_2^1)^2}[B^2\sigma\tau+B+C]+\frac{2B}{5H_2^1}(\sigma^2-6)+\frac{8(\sigma^2-1)^2(C+1)(H_2^3-1)^2}{(4\sigma^2+1)(4\sigma^2+11)H_2^1H_2^3},$$

$$\gamma_{22}^{20}=\gamma_{22}^{02}=-\gamma_{20}^{20}=-\gamma_{02}^{02}=-2\gamma_{22}^{22}=\frac{B}{9}\frac{C+1}{B+1}.$$

When $\hat{\mathbf{k}}=\hat{\mathbf{r}}=\hat{\mathbf{x}}_1$, we obtain

$$\mathrm{Im}g(\hat{\mathbf{x}}_1,\hat{\mathbf{x}}_1)\approx\frac{ik^3a_1a_2^2}{3}\left(C-\frac{B}{H_1^1}\right)+\frac{ik^5a_1a_2^4}{30}\sum_{n,n'=0}^{3}\gamma_{nn'}^{00} \tag{82}$$

$$\mathrm{Re}g(\hat{\mathbf{x}}_1,\hat{\mathbf{x}}_1)\approx-\frac{k^6a_1^2a_2^4}{9}\left(C^2+\frac{B^2}{3(H_1^1)^2}\right)-\frac{k^8a_1^2a_2^6}{45}\left(C\gamma_{00}^{00}-\frac{B}{3H_1^1}\sum_{n'=0}^{3}\gamma_{1n'}^{00}\right) \tag{83}$$

For the prolate near-sphere we have

$$\tau\sigma-1=e^2[\tfrac{1}{3}+\tfrac{1}{5}e^2+\tfrac{1}{7}e^4]+O(e^8), \tag{84}$$

where $e^2=(\sigma^2-1)/\sigma^2$ is small and $\sigma\gtrsim 1$; for the oblate near-sphere we have

$$\tau\sigma-1=(e^2/\sigma^2)[-\tfrac{1}{3}+\tfrac{1}{5}e^2/\sigma^2-\tfrac{1}{7}e^4/\sigma^4]+O(e^8), \tag{85}$$

where $e^2=1-\sigma^2$ with $\sigma\lesssim 1$. Using either (84) or (85) we obtain

$$\begin{aligned} &H_i^1=1+B/3\sigma^2+O(e^2)\quad (i=1,2,3),\qquad H_\Lambda^2=H_{\Lambda'}^2=H_1^2=H_2^2=H_3^2=1+2B/5\sigma^4+O(e^2),\\ &H_1^3=H_1^{3'}=H_2^3=H_2^{3'}=H_3^3=H_3^{3'}=H_{123}=1+3B/7\sigma^2+O(e^2). \end{aligned} \tag{86}$$

These reduce to the corresponding values for the sphere as $\sigma\to 1+$ or $\sigma\to 1-$.

For the case of the needle, $a_1\gg a_2$,

$$\tau=(2/\sigma)\ln\sigma+O(1/\sigma)\to 0,\quad \sigma\to+\infty,\qquad (\sigma\tau-1)/(\sigma^2-1)=(2\ln\sigma-1)/(\sigma^2-1)+O(1/\sigma^2)\to 0,\quad \sigma\to+\infty, \tag{87}$$

we obtain

$$H_1^1, H_\Lambda^2, H_1^3\to 1,\qquad H_2^1, H_3^1, H_{\Lambda'}^2, H_1^2, H_2^2, H_3^2, H_1^{3'}, H_2^3, H_2^{3'}, H_3^3, H_3^{3'}, H_{123}\to 1+B/2. \tag{88}$$

For the case of the disc, $a_1\ll a_2$,

$$\tau=\pi/2+O(\sigma)\to\pi/2+O(\sigma)\to\pi/2,\quad \sigma\to 0+,\qquad \tau\sigma-1\to -1,\quad \sigma\to 0+, \tag{89}$$

we obtain

$$H_2^1, H_3^1, H_\Lambda^2, H_{\Lambda'}^2, H_1^2, H_2^2, H_3^2, H_2^{3'}, H_3^{3'}\to 1,\qquad H_1^1, H_2^2, H_3^2, H_1^3, H_1^{3'}, H_{123}\to 1+B. \tag{90}$$

For both cases we have

$$\sum_{i=1}^3 H_i^1=3+B,\qquad H_\Lambda^2+H_{\Lambda'}^2+\sum_{i=1}^3 H_i^2=5+2B,\qquad \sum_{i=1}^3 (H_i^3+H_i^{3'})+H_{123}=7+3B. \tag{91}$$

The appropriate form of g for the needle or disc is obtained from (80) and (81) by substituting (88) or (90) into the coefficients $\gamma_{nn'}^{mm'}$.

C. Sphere

The sphere corresponds to the degenerate ellipsoid $a_i=a$, $h_i^2=0$, $i=1,2,3$; we have $0\le\rho<\infty$, $\mu=\nu=0$, with $r=\rho$. The elliptic integrals reduce to

$$I^0(\rho)=1/\rho,\quad I_i^1(\rho)=1/3\rho^3,\quad I_\Lambda^2(\rho)=I_{\Lambda'}^2(\rho)=I_i^2(\rho)=1/5\rho^5,\quad I_i^3(\rho)=I_i^{3'}(\rho)=I_{123}(\rho)=1/7\rho^7, \tag{92}$$

$$H_i^1=1+B/3,\quad H_\Lambda^2=H_{\Lambda'}^2=H_i^2=1+\tfrac{2}{5}B,\quad H_i^3=H_i^{3'}=H_{123}=1+\tfrac{3}{7}B. \tag{93}$$

Specializing the solutions for the ellipsoid, we obtain in terms of the Legendre polynomials $P_n(\hat{\mathbf{k}},\hat{\mathbf{r}})=P_n(\cos\theta)$,

$$\phi_0^+=\phi_0^-=1, \tag{94}$$

$$\phi_1^+=\left(1-\frac{a^3}{r^3}\frac{B}{B+3}\right)rP_1(\cos\theta),\quad \phi_1^-=\frac{3}{B+3}\,rP_1(\cos\theta), \tag{95}$$

$$\phi_2^+=\frac{r^2}{3}-\frac{2a^3}{3r}C+\left(\frac{2}{3}-\frac{4a^5B}{3r^5(2B+5)}\right)r^2P_2(\cos\theta),\quad \phi_2^-=\left(\frac{r^2-a^2}{3}\right)\frac{C+1}{B+1}+\frac{a^2(1-2C)}{3}+\frac{10r^2}{3(2B+5)}\,P_2(\cos\theta), \tag{96}$$

$$\begin{aligned} \phi_3^+&=-2a^3C+3\left(\frac{a^3B}{B+3}+\frac{r^3}{5}-\frac{6a^5(B^2+B+C)}{5r^2(B+3)^2}\right)P_1(\cos\theta)+\frac{2}{5}\left(r^3-\frac{3a^7B}{r^4(3B+7)}\right)P_3(\cos\theta),\\ \phi_3^-&=-2a^3C+\frac{9}{5(B+3)}\left(\frac{a^2(5B-3C+2)}{B+3}-\frac{2a^2(C+1)}{(B+3)(B+1)}+\frac{r^2(C+1)}{B+1}\right)rP_1(\cos\theta)+\frac{14r^3}{5(3B+7)}\,P_3(\cos\theta), \end{aligned} \tag{97}$$

$$\mathrm{Im}g(\hat{\mathbf{r}},\hat{\mathbf{k}})\approx ik^3a^3\left(\frac{C}{3}-\frac{B}{B+3}P_1(\cos\theta)\right)+ik^5a^5\left(\frac{5C^2-4C}{45}+\frac{(C-B)(C+1)}{45(B+1)}+\frac{3(B^2+B+C)}{5(B+3)^2}P_1(\cos\theta)-\frac{2B}{9(2B+5)}P_2(\cos\theta)\right), \tag{98}$$

$$\mathrm{Re}g(\hat{\mathbf{r}},\hat{\mathbf{k}})\approx -k^6a^6\left(\frac{C^2}{9}+\frac{B^2}{3(B+3)^2}P_1(\cos\theta)\right)-k^8a^8\left[\frac{2C}{135}\left(5C^2-4C+\frac{(C-B)(C+1)}{B+1}\right)-\frac{2B(B^2+B+C)}{5(B+3)^2}P_1(\cos\theta)\right].$$

If we let $B\to -1$ and $C\to -1$, we obtain the solutions for a rigid scatterer:

$$\begin{aligned} &\phi_0^+=1,\quad \phi_1^+=\left(1+\frac{a^3}{2r^3}\right)rP_1(\cos\theta),\quad \phi_2^+=\left(\frac{r^2}{3}+\frac{2a^3}{3r}\right)+\left(\frac{2}{3}+\frac{4a^5}{9r^5}\right)r^2P_2(\cos\theta),\\ &\phi_3^+=2a^3+3\left(-\frac{a^3}{2}+\frac{r^3}{5}+\frac{3a^5}{10r^2}\right)P_1(\cos\theta)+\frac{2}{5}\left(1+\frac{3a^7}{4r^7}\right)r^3P_3(\cos\theta). \end{aligned} \tag{99}$$

The corresponding scattering amplitude is

$$\mathrm{Im}g(\hat{\mathbf{r}},\hat{\mathbf{k}}) \approx ik^3a^3[-\tfrac{1}{3}+\tfrac{1}{2}P_1(\cos\theta)]+ik^5a^5[\tfrac{1}{5}-\tfrac{3}{20}P_1(\cos\theta)+\tfrac{2}{27}P_2(\cos\theta)],$$
$$\mathrm{Re}g(\hat{\mathbf{r}},\hat{\mathbf{k}}) \approx -k^6a^6[\tfrac{1}{9}+\tfrac{1}{12}P_1(\cos\theta)]+k^8a^8[\tfrac{2}{15}+\tfrac{1}{10}P_1(\cos\theta)]. \tag{100}$$

The above result for the rigid sphere coincide with those given in Ref. 9, p. 376 (where $\hat{\mathbf{k}}=-\hat{\mathbf{x}}_3$).

D. Physical considerations

In this section we discuss the effects of the shape as well as the orientation of the scatterer on the total energy scattered. We consider only the leading term of the scattering cross section, i.e., the k^4 term. For the general ellipsoid we have

$$\sigma_s=\frac{4\pi V^2k^4}{9}\left(C^2+\frac{B^2}{3}\sum_{n=1}^{3}\frac{i_n^2}{(H_n^1)^2}\right), \tag{101}$$

where $H_n^1=1+BQ_n$, $Q_n=VI_n^1$, $n=1,2,3$ and $\sum_{n=1}^3 Q_n=1$. Note that $a_1>a_2>a_3$ implies that $0<Q_1<Q_2<Q_3<1$ and, therefore,

$$1/(B+1)^2<1/(H_3^1)^2<1/(H_2^1)^2<1/(H_1^1)^2<1. \tag{102}$$

Hence, the maximum σ_s occurs for incidence $\hat{\mathbf{k}}=\hat{\mathbf{x}}_1$ and the minimum occurs for incidence $\hat{\mathbf{k}}=\hat{\mathbf{x}}_3$: The larger the projection of the ellipsoid on the line of incidence (the larger the path length through the scatterer), the larger the energy scattered. If we average σ_S over orientations, we obtain

$$\langle\sigma_s\rangle\approx\frac{4\pi V^2k^4}{9}\left(C^2+\frac{B^2}{9}\sum_{n=1}^{3}\frac{1}{(H_n^1)^2}\right). \tag{103}$$

For spheroids we have

$$\sigma_s\approx\frac{4\pi V^2k^4}{9}\left[C^2+\frac{B^2}{3}\left(\frac{\cos\alpha}{H_1^1}\right)^2+\frac{B^2}{3}\left(\frac{\sin\alpha}{H_2^1}\right)^2\right]. \tag{104}$$

We give Table I for the values of $Q_1=(\sigma\tau-1)/(\sigma^2-1)$ and $Q_2=\sigma(\sigma-\tau)/2(\sigma^2-1)$, where σ and I are given in (75) and (76).

We have $(1+BQ_1)\lesseqgtr(1+BQ_2)$ for $a_1\gtreqless a_2$. Therefore, σ_s becomes maximum at $\alpha=0$ for prolate, and at $\alpha=\pi/2$ for oblate spheroids.

If we average σ_s over orientations, we obtain

$$\langle\sigma_s\rangle\approx\frac{4\pi a_1^2a_2^4k^4}{9}\left(C^2+\frac{B^2}{9(1+BQ_1)^2}+\frac{2B^2}{9(1+BQ_2)^2}\right). \tag{105}$$

As the oblate spheroid approaches the disc or the prolate spheroid approaches the needle, the maximum of σ_s increases, and the minimum decreases. Also, as the oblate or prolate spheroids approach the sphere, the maximum of σ_s decreases, and the minimum increases; both tend to the scattering cross section for the sphere,

$$\sigma_s\approx\frac{4\pi a^6k^4}{9}\left(C^2+\frac{3B^2}{(B+3)^2}\right). \tag{106}$$

The radius of the sphere that scatters the same energy (to the order k^4) as that obtained form the ellipsoid averaged over orientation is given by

$$V^{1/3}\left(C^2+\frac{B^2}{9}\sum_{n=1}^{3}\left(\frac{1}{H_n^1}\right)^2\right)\left(C^2+\frac{3B^2}{(B+3)^2}\right)^{-1}. \tag{107}$$

For the special case of equal densities, $B=0$, (81) becomes

$$\mathrm{Re}g(\hat{\mathbf{r}},\hat{\mathbf{k}})\approx-\frac{a_1^2a_2^4C^2}{9}k^6-\frac{2a_1^2a_2^6C^2(6C\sigma\tau-\sigma^2-2)}{135}k^8. \tag{108}$$

The terms of order k^6 and k^8 for $\mathrm{Re}g(\hat{\mathbf{r}},\hat{\mathbf{k}})$ are monopole contributions, i.e., if the densities are equal, then the first two nonvanishing terms of the scattering cross section are independent of the directions of incidence and observation.

APPENDIX

If we denote by E_n^m and F_n^m the Lamé functions of the first and the second kind, and by

$$\mathbb{E}_n^m(\rho,\mu,\nu)=E_n^m(\rho)E_n^m(\mu)E_n^m(\nu),$$

and

$$\mathbb{F}_n^m(\rho,\mu,\nu)=F_n^m(\rho)E_n^m(\mu)E_n^m(\nu)$$
$$=(2n+1)\mathbb{E}_n^m\int_\rho^\infty\frac{du}{(E_n^m(u))^2S(u)}$$

the interior and exterior ellipsoidal harmonics, we have the following expressions of the ellipsoidal harmonics in terms of Cartesian coordinates:

$$\mathbb{E}_0=1,\quad \mathbb{E}_1^i=h_1h_2h_3\frac{x_i}{h_i},\quad i=1,2,3,$$

$$\mathbb{E}_2^1=(\Lambda-a_1^2)(\Lambda-a_2^2)(\Lambda-a_3^2)\left(\sum_{n=1}^{3}\frac{x_n^2}{\Lambda-a_n^2}+1\right)=\mathbb{E}_2^1(\Lambda),$$

$$\mathbb{E}_2^2=\mathbb{E}_2^1(\Lambda'),$$

$$\mathbb{E}_2^{6-i}=h_1h_2h_3x_1x_2x_3\frac{h_i}{x_i},$$

$$\mathbb{E}_3^{2i-1}=h_1h_2h_3(\Lambda_i-a_1^2)(\Lambda_i-a_2^2)(\Lambda_i-a_3^2)\frac{x_i}{h_i}\left(\sum_{n=1}^{3}\frac{x_n^2}{\Lambda_i-a_n^2}+1\right)$$
$$=\mathbb{E}_3^{2i-1}(\Lambda_i),$$

$$\mathbb{E}_3^{2i}=\mathbb{E}_3^{2i-1}(\Lambda_i'),$$

$$\mathbb{E}_3^7=h_1^2h_2^2h_3^2x_1x_2x_3.$$

TABLE I.

Prolate			Oblate		
σ	Q_1	Q_2	σ	Q_1	Q_2
1.001	.32483	.33483	0.001	.99843	.00078
1.01	.32985	.33482	0.01	.98448	.00775
1.1	.30823	.34585	0.1	.86079	.06959
1.5	.23296	.38351	0.5	.52720	.23638
2	.17356	.41322	0.7	.43205	.28396
10	.02028	.48985	0.9	.36184	.31905
10^3	.00000	.49999	0.99	.33618	.33165
10^6	.00000	.50000	0.999	.33516	.33016

The Cartesian monomials are expressed in terms of surface ellipsoidal harmonics as follows:

$$x_i = \frac{h_i}{h_1h_2h_3}\,\mathbb{E}_1^i, \quad i=1,2,3,$$

$$x_i^2 = \frac{(-1)^i h_i^2(\rho_1^2+a_i^2)}{h_1^2h_2^2h_3^2} \times\left(\frac{(\Lambda'-a_i^2)E_2^1(\mu)E_2^1(\nu)-(\Lambda-a_i^2)E_2^2(\mu)E_2^2(\nu)}{\Lambda-\Lambda'} - (\Lambda-a_i^2)(\Lambda'-a_i^2)\right),$$

$$\frac{x_1x_2x_3}{x_i} = \frac{\mathbb{E}^{6-i}_2}{h_ih_1h_2h_3},$$

$$x_ix_n^2 = \frac{(\rho_1^2+a_i^2)^{1/2}(\rho_1^2+a_n^2)h_i}{h_1h_2h_3}\left\{\frac{1+2\delta_{in}}{5}E_1^i(\mu)E_1^i(\nu) + \frac{(-1)^n h_n^2}{h_1^2h_2^2h_3^2(\Lambda_i-\Lambda_i')}\left[(\Lambda_i'-a_n^2)E_3^{2i-1}(\mu)E_3^{2i-1}(\nu) - (\Lambda_i-a_n^2)E_3^{2i}(\mu)E_3^{2i}(\nu)\right]\right\}, \quad i,n=1,2,3,$$

$$x_1x_2x_3 = \prod_{n=1}^{3}\frac{(\rho_1^2+a_n^2)^{1/2}}{h_n^2}E_3^7(\mu)E_3^7(\nu),$$

where $\rho_1^2=\rho^2-a_1^2$.

Using the above relations, we obtain the following expressions for $(\hat{\mathbf{k}}\cdot\mathbf{r})^n$ in terms of surface ellipsoidal harmonics:

$$\hat{\mathbf{k}}\cdot\mathbf{r} = \frac{1}{h_1h_2h_3}\sum_{n=1}^{3} i_nh_n\mathbb{E}_1^n(\rho,\mu,\nu),$$

$$(\hat{\mathbf{k}}\cdot\mathbf{r})^2 = \frac{\rho^2-a_1^2}{3} + \tfrac{1}{3}\sum_{n=1}^{3} i_n^2a_n^2 + 2\frac{i_1i_2i_3}{h_1h_2h_3}\sum_{n=1}^{3}\frac{\mathbb{E}_2^{6-n}(\rho,\mu,\nu)}{i_nh_n} + \frac{E_2^1(\mu)E_2^1(\nu)}{h_1^2h_2^2h_3^2(\Lambda-\Lambda')}\sum_{n=1}^{3}(-1)^n i_n^2h_n^2(\rho_1^2+a_n^2)(\Lambda'-a_n^2) + (\Lambda\leftrightarrow\Lambda'),$$

$$(\hat{\mathbf{k}}\cdot\mathbf{r})^3 = 6E_3^7(\mu)E_3^7(\nu)\prod_{n=1}^{3}(\rho_1^2+a_n^2)^{1/2}\frac{i_n}{h_n^2} + \frac{3}{5h_1h_2h_3}\sum_{n=1}^{3}\sum_{k=1}^{3} i_ki_n^2h_k(\rho_1^2+a_n^2)\mathbb{E}_1^k(\rho,\mu,\nu) + \frac{3}{h_1^3h_2^3h_3^3}\sum_{n=1}^{3}\sum_{k=1}^{3}(-1)^n i_ki_n^2h_kh_n^2\frac{(\rho_1^2+a_k^2)^{1/2}(\rho_1^2+a_n^2)}{(1+2\delta_{kn})(\Lambda_k-\Lambda_k')} \cdot\left[(\Lambda_k'-a_n^2)E_3^{2k-1}(\mu)E_3^{2k-1}(\nu)-(\Lambda_k-a_n^2)E_3^{2k}(\mu)E_3^{2k}(\nu)\right]. \qquad (***)$$

*This work is part of the author's Ph.D. dissertation, in the Mathematics Department of the University of Illinois at Chicago Circle under the guidance of Professor Victor Twersky. The work was supported in part by the N.S.F. Grant GP 33368X.

[1]Lord Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. II, pp. 149, 272, 282.
[2]Lord Rayleigh, Phil. Mag. **XLIV**, 28–52 (1897).
[3]A.F. Stevenson, J. Appl. Phys. **24**, 1134–42 (1953).
[4]A.F. Stevenson, J. Appl. Phys. **24**, 1143–51 (1953).
[5]P. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. II, pp. 1085–88.
[6]J.E. Burke, J. Acoust. Soc. Am. **39**, 826–31 (1966).
[7]J.E. Burke, J. Acoust. Soc. Am. **40**, 325–30 (1966).
[8]J.E. Burke, J. Acoust. Soc. Am. **43**, 871–75 (1968).
[9]J. Bowman, T. Senior, and P. Uslenghi, *Acoustical Scattering by Simple Shapes* (North-Holland, Amsterdam, 1969), pp. 361, 376, 403, 446, 508, 515.
[10]V. Twersky, Appl. Opt. **3**, 1150–62 (1964).
[11]R. Kleinman, Proc. IEEE **53**, 848–56 (1965).
[12]V. Twersky, J. Acoust. Soc. Am. **36**, 1314–29 (1964).
[13]R. Kleinman, Appl. Sci. Res. 18, 1–8 (1967).
[14]B. Noble, *Integral Equations Perturbation Methods in Low-Frequency Diffraction*, edited by R.E. Langer (Univ. of Wisconsin Press, Madison, 1962), pp. 323–60.
[15]D. Darling and T. Senior, J. Acoust. Soc. Am. **37**, 228–34 (1965).
[16]S. Mikhlin, *Linear Equations of Mathematical Physics* (Holt, Rinehart and Winston, New York, 1967), p. 220.
[17]C. Muller, *Foundations of the Mathematical Theory of Electromagnetic Waves* (Springer-Verlag, Berlin, 1969), p. 297.
[18]C. Wilcox, Am. Math. Soc. 7, 271–76 (1956).
[19]V. Twersky, J. Appl. Phys. 25, 859–62 (1954).
[20]E. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Chelsea, New York, 1955).
[21]E. Whittaker and G. Watson, *Modern Analysis* (Cambridge U.P., Cambridge, 1962).
[22]W. MacMillan, *The Theory of the Potential* (Dover, New York, 1958), p. 56.
[23]O. Kellogg, *Foundations of Potential Theory* (Dover, New York, 1954).
[24]P. Moon and D. Spencer, *Field Theory Handbook* (Springer-Verlag, Berlin, 1961).
[25]S. Chandrasekhar, *Ellipsoidal Figures of Equilibrium* (Yale U.P., New Haven, 1969).
[26]A. Dixon, Proc. Lond. Math. Soc. **35**(i) 162–96 (11902).
[27]J. Sharma, Bull. Calcutta. Math. Soc. **XXIV**, 3 (1932).
[28]A. Erdelyi, Proc. Royal Soc. Edinburg **LXII** 247–67 (1948).
[29]G. Dassios, "Convergent Low-Frequency expansions for penetrable scatterers," Ph.D. thesis (University of Illinois, Chicago, 1975).

Macrocausal aspects of the S matrix related to the LSZ asymptotic condition

H. D. Doebner and W. Lücke

Institut für Theoretische Physik der Technischen Universität Clausthal, Clausthal-Zellerfeld, Germany
(Received 8 March 1976)

By use of the LSZ asymptotic condition in Hepp's form it is shown that one may substitute functions $s(x)$, which converge sufficiently rapidly to 1 in positive timelike direction, for the step function $\theta(x^0)$ in the usual LSZ reduction formulas. This result, partly related to earlier work by Toll, does not depend on microcausality of the interpolating field. Choosing $s(x)$ such that $s(x) = 0$ outside the forward lightcone, a complete reduction to vacuum expectation values of T products of the interpolating field thus shows that it is sufficient to know the n-point Wightman distributions at pairwise timelike arguments $x_1,\ldots,x_n$ in order to calculate the S matrix completely. It is argued that this fact indicates some macrocausal property of the S matrix introduced by the LSZ asymptotic condition.

1. INTRODUCTION

In Wightman quantum field theory the interplay between the axiom of microcausality (in short, *locality*) and properties of the S matrix has not yet been fully understood. It appears that some results of axiomatic quantum field theory usually derived by locality still remain valid when locality is replaced by some physically better motivated and mathematically weaker assumption. For example, Haag's asymptotic conditions may be directly derived from the cluster property.[1] The cluster property is physically more fundamental (for short-range forces) and mathematically weaker than locality: Rapid decrease of the field commutator in spacelike direction is sufficient for a derivation of the cluster property by translation invariance and the spectrum condition.[2] Rapid spacelike decrease of the field commutator is also sufficient for a derivation of simple dispersion relations.[3] We suppose that connected with the LSZ asymptotic condition and the hyperbolic propagation character of Klein–Gordon solutions is some mechanism introducing certain macrocausal properties into the S matrix whether the interpolating field fulfils locality or not. It is the aim of the present paper to indicate such a mechanism. We derive x-space representations of the S matrix *formally* indifferent to microcausality (see Sec. 3, Part B).

For 2–2 scattering such a representation was *heuristically* given already by Toll.[4] He modified the original LSZ asymptotic condition[5] by introducing a "causal" factor $s(x)$ with

$$\operatorname{supp} s \subset V_+, \quad s(x) \to 1 \quad \text{for } x^2 \to +\infty,\ x^0 > 0$$

(V_+ the interior of the forward lightcone), i.e., he *postulated* (Toll's asymptotic conditon)

$$\langle \Phi | A^f_{\substack{out \\ (in)}} | \Psi \rangle = i \lim_{x^0 \to (\pm)\infty} \int d\mathbf{x} f^*(x) \times \overleftrightarrow{\partial}_0 [s(\pm(x-y)) \langle \Phi | A(x) | \Psi \rangle]$$

in addition to the original LSZ asymptotic condition (in which s is dropped). Corresponding modification of the original LSZ reduction technique led him to some x-space representation of the 2–2 scattering amplitude with a primitive domain of analyticity (in p-space) even larger than that derived by locality. However, Toll's approach met two difficulties:

(i) The relation between Toll's asymptotic condition and the LSZ asymptotic condition was not exactly known.

(ii) Introduction of the factor $s(x-y)$ which implies the p-space analyticity properties unfortunately destroys the support properties in p-space which follow from the spectrum condition in the conventional approach. Therefore the usual methods of enlarging the domain of analyticity are not applicable.

While the first difficulty is overcome by the present paper, a solution to the second problem is still outstanding.

The paper is organized as follows. In Sec. 2 we introduce a rigorous version of Toll's general reduction formalism as a consequence of the LSZ asymptotic condition and some "causality" property of the test space. In Sec. 3 we first derive Toll's reduction formula and then completely reduce the whole S matrix to vacuum expectation values of modified T products which differ from the usual ones in that the step function $\theta(x^0 - y^0)$ is replaced by a causal multiplier $z(x,y)$ [$= s(x-y)$ for example]. In Sec. 4 we prove that at least the Schwartz space $\mathcal{S}(R^8)$ has the "causality" property required for the rigorous version of Toll's reduction formalism. Section 5, finally, is devoted to a short discussion of our results.

2. ASYMPTOTIC CONDITION FOR FIELDS OVER CAUSAL TEST SPACES

A. LSZ asymptotic condition and causal test spaces

Let $\mathcal{C}(R^1) \subset \mathcal{S}(R^1)$ be some standard test space.[6] Denote by $\mathcal{C}(R^n)$ its complete n-fold π tensor product,[7] by $\hat{\mathcal{C}}(R^n)$ the Fourier transformed space, and by $\mathcal{C}'(R^n)$ the topological dual of $\mathcal{C}(R^n)$. Let $A(x)$ be a scalar Hermitian field over $\mathcal{C}(R^4)$ describing particles with mass $m > 0$, and let D be its invariant dense domain in the Hilbert space $\mathcal{H}$; hence

$$\langle \Psi | A(x) | \Phi \rangle \in \mathcal{C}'(R^4) \text{ for } \Psi, \Phi \in D.$$

Without loss of generality[8] we assume D such that[9]

$$\int d\hat{x}\, A(x_1)\cdots A(x_n)h(\hat{x})\, D \subset D \tag{I}$$

for all integers $n > 0$ and for all $h \in \mathcal{C}(R^{4n})$. The vacuum is written $\Omega \in D$. Denote by $(A^f_{ex})^*$ the creation operator for an ingoing ($t \to -\infty$), resp. outgoing, ($t \to +\infty$) particle with wavefunction

$$f(x) = (2\pi)^{-3/2} \int_{p^0=\omega_{\mathbf{p}}} \frac{d\mathbf{p}}{2\omega_{\mathbf{p}}} \hat{f}(\mathbf{p}) \exp(-ipx), \tag{1}$$

$$\hat{f}(\mathbf{p}) = \tilde{\varphi}(\omega_{\mathbf{p}}, \mathbf{p}), \quad \tilde{\varphi}(p) \in \tilde{\mathcal{C}}(R^4),$$

and denote by $\mathcal{L}_{ex} \subset \mathcal{H}$ the linear span of all ingoing resp. outgoing r-particle states ($r = 0, 1, 2, \cdots$) with p-space wavefunctions in $\mathcal{C}(R^{3r})$. Then the LSZ asymptotic condition[10] demands $\mathcal{L}_{ex} \subset D$ and

$$\lim_{t\to\pm\infty} \int dx \langle \Psi | A(x) | \Phi \rangle \varphi(x; t) = \langle \Psi | (A^f_{ex})^* | \Phi \rangle \tag{AI}$$

$$\text{for } \Psi \in D, \quad \phi \in \mathcal{L}_{ex}, \quad \tilde{\varphi} \in \tilde{\mathcal{C}}(R^4),$$

where

$$\varphi(x; t) \equiv (2\pi)^{-5/2} \int dp\, \tilde{\varphi}(p) \frac{p^0 + \omega_{\mathbf{p}}}{2\omega_{\mathbf{p}}} \exp[i(p^0 - \omega_{\mathbf{p}})t] \exp(-ipx). \tag{2}$$

For the annihilation operators A^f_{ex} corresponding conditions are imposed.

Because of Eq. (I), condition (AI) implies asymptotic relations of the form

$$\lim_{t\to\pm\infty} \int d\hat{x}\, dy\, dx \langle \Psi | A(x_1)\cdots A(x_n)A(y)A(x) | \Phi \rangle h(\hat{x}, y)\varphi(x; t)$$
$$= \int d\hat{x}\, dy \langle \Psi | A(x_1)\cdots A(x_n)A(y)(A^f_{ex})^* | \Phi \rangle h(\hat{x}, y), \tag{AII}$$

with

$$h(\hat{x}, y) \in \mathcal{C}(R^{4(n+1)}), \quad \Psi \in D, \quad \Phi \in \mathcal{L}_{ex}, \quad \tilde{\varphi} \in \tilde{\mathcal{C}}(R^4),$$

which are necessary for the usual reduction of the S matrix. It is worthwhile to study the asymptotic ($t \to \pm\infty$) properties of $\varphi(x; t)$ in order to get further relations of this type.

Consider the special case

$$\tilde{\varphi}(p) = (2\pi)^{1/2} \hat{f}(\mathbf{p}) \tilde{h}(p^0 - \omega_{\mathbf{p}}), \quad h \in \mathcal{D}(R^1), \tag{3}$$

in which we have

$$\varphi(x; t) = 2ih(x^0 - t)\partial_0 f(x) + ih'(x^0 - t)f(x).$$

For $t \to +\infty$, $\varphi(\cdot; t)$ will rapidly decrease outside the forward lightcone with vertex at an arbitrary fixed space–time point[11] y. The same holds for the general case [see inequality (10) in Sec. 4, Part A]. Hence $h(\hat{x}, y)\varphi(x; t)$ with $h \in \mathcal{C}(R^{4(n+1)})$ should rapidly decrease for $x - y \notin V_+$ and even for (x, y) outside the region

$$G(\mu) \equiv \{(x, y) \in R^8 : (x - y)^2 > \mu, \ x^0 - y^0 > 0\}, \quad \mu \text{ fixed}.$$

So we expect the $t \to +\infty$ limit of the left-hand integral in (AII) to be independent of the form of the generalized function $\langle \Psi | A(x_1)\cdots A(x_n)A(y)A(x) | \Phi \rangle$ outside the region $(x, y) \in G(\mu)$, e.g., the matrix element can be multiplied with a multiplier $z(x, y)$ in $\mathcal{C}(R^8)$ without changing the limit of the integral, provided $z(x, y)$ rapidly tends to 1 within the region $(x, y) \in G(\mu)$ for $\mu \to +\infty$. This indicates a large class of multipliers $z(x, y)$ in $\mathcal{C}(R^8)$, at least for $\mathcal{C} = \mathcal{S}$, fulfilling[12]

$$\lim_{t\to+\infty} (1 - z(x, y))h(\hat{x}, y)\varphi^{\#}(x; t) = 0 \tag{4}$$

in $\mathcal{C}(R^{4(n+2)})$ for arbitrary $h \in \mathcal{C}(R^{4(n+1)})$ and $\tilde{\varphi} \in \tilde{\mathcal{C}}(R^4)$.

Such multipliers can be used instead of smooth θ functions in the reduction of outgoing particles if, in addition,

$$\lim_{t\to-\infty} z(x, y)h(\hat{x}, y)\varphi^{\#}(x; t) = 0 \tag{5}$$

in $\mathcal{C}(R^{4(n+2)})$ holds. The same $z(x, y)$ can also be used for the reduction of ingoing particles if (4) and (5) hold for $\check{z}(x, y) \equiv (1 - z(y, x))$ as well. Hence an interesting question is whether $\mathcal{C}(R^1)$ is "causal."

Definition 1: $\mathcal{C}(R^1)$ is called *causal* iff there is a *causal multiplier* $z(x, y)$ in $\mathcal{C}(R^8)$, i.e., iff there is a multiplier $z(x, y)$ in $\mathcal{C}(R^8)$ with $\operatorname{supp} z \subset \bar{V}_+$ fulfilling $(z_-(y, x) \equiv z_+(x, y) \equiv z(x, y))$:

(i) $\lim_{t\to\pm\infty}(1 - z_\pm(x, y))h(\hat{x}, y)\varphi^{\#}(x; t) = 0$ in $\mathcal{C}(R^{4(n+2)})$,

(ii) $\lim_{t\to\pm\infty} z_\mp(x, y)h(\hat{x}, y)\varphi^{\#}(x; t) = 0$ in $\mathcal{C}(R^{4(n+2)})$,

for arbitrary $h \in \mathcal{C}(R^{4(n+1)})$ and $\tilde{\varphi} \in \tilde{\mathcal{C}}(R^4)$.

In Sec. 4 we explicitly construct a causal multiplier in $\mathcal{S}(R^8)$ which is also Poincaré invariant. Causality of Jaffe spaces will be proved in a separate paper.

We summarize the main results of our discussion by the following lemma.

Lemma 1: Let $P(\hat{x}, y)$ be a derivative of $A(x_1)\cdots \times A(x_j)A(y)A(x_{j+1})\cdots A(x_n)$, $j < n$. If $\mathcal{C}(R^1)$ is causal then there are multipliers $z_+(x, y) = z_-(y, x)$ in $\mathcal{C}(R^8)$ such that for all $\Psi^{(\prime)} \in D$, $\Phi \in \mathcal{L}_{ex}$, $h \in \mathcal{C}(R^{4(n+1)})$, $\tilde{\varphi} \in \tilde{\mathcal{C}}(R^4)$ and for all multipliers $k(\hat{x}, y)$ in $\mathcal{C}(R^{4(n+1)})$,

$$\lim_{t\to\pm\infty} \int d\hat{x}\, dy\, dx \langle \Psi | P(\hat{x}, y) | \Psi' \rangle$$
$$\times z_\mp(y, x_r)h(\hat{x}, 0)\varphi^{\#}(y; t)k(\hat{x}, y) = 0 \quad \text{for } r = 1, \ldots, n \tag{AIII}$$

holds and the LSZ asymptotic conditions imply

$$\lim_{t\to\pm\infty} \int d\hat{x}\, dy\, dx \langle \Psi | P(\hat{x}, y)A(x) | \Phi \rangle z^{\#}_\pm(x, y)h(\hat{x}, y)\varphi(x; t)$$
$$= \int d\hat{x}\, dy \langle \Psi | P(\hat{x}, y)(A^f_{ex})^* | \Phi \rangle h(\hat{x}, y), \tag{AIV}$$

$$\lim_{t\to\pm\infty} \int d\hat{x}\, dy\, dx \langle \Psi | P(\hat{x}, y)A(x) | \Phi \rangle z^{\#}_\pm(x, y)h(\hat{x}, y)\varphi^*(x; t)$$
$$= \int d\hat{x}\, dy \langle \Psi | P(\hat{x}, y)A^f_{ex} | \Phi \rangle h(\hat{x}, y). \tag{AV}$$

B. Toll's asymptotic condition

As mentioned in the introduction, Toll proposed[4] on a heuristic level an asymptotic condition containing a factor $s(x - y)$. Now, a rigorous mathematical meaning of Toll's ansatz can be given if $s(x - y)$ is a multiplier in $\mathcal{C}(R^8)$,

$$\lim_{t\to\pm\infty} \int dx \langle \Psi | A(x) | \Phi \rangle s(\pm(x - y))\varphi(x, t)$$
$$= \langle \Psi | (A^f_{ex})^* | \Phi \rangle, \quad \Psi \in D, \quad \Phi \in \mathcal{L}_{ex}, \quad \tilde{\varphi} \in \tilde{\mathcal{C}}(R^4). \tag{AVI}$$

For causal $z(x, y) = s(x - y)$ this relation can even be derived from (AI) since, e.g.,

$$\lim_{t \to +\infty} (1 - z(x, y))h(y)\varphi^{\#}(x + a; t) = 0$$

in the topology of $C(R^8)$ implies $(h(0) \neq 0)$,

$$\lim_{t \to +\infty} (1 - s(x - a))\varphi^{\#}(x; t) = 0$$

in the topology of $C(R^4)$.

However, an asymptotic field theory based only on Toll's condition (AVI) meets serious difficulties. A derivation of Toll's reduction formula[4] requires asymptotic relations of the type (AIII)—(AV) which can be obtained only formally from (AVI); so one has to postulate these relations or one has to add an additional assumption to get them. A very weak postulate of this kind, valid for a sufficiently large class of standard test spaces, is causality of the test space. But in causal test spaces the asymptotic relations (AIII)—(AV) can be *directly* derived from the LSZ asymptotic condition and Toll's condition has no technical advantage as compared with the LSZ condition.

3. REDUCTION OF THE S MATRIX FOR CAUSAL TEST SPACES

A. 2-particle scattering amplitude

Consider the 2-particle matrix element

$$(S - 1)_{22} = \langle \Omega | (A^{f_4}_{\text{out}} A^{f_3}_{\text{out}} - A^{f_4}_{\text{in}} A^{f_3}_{\text{in}})(A^{f_2}_{\text{in}})^* (A^{f_1}_{\text{in}})^* | \Omega \rangle$$

with

$$\hat{f}_j(\mathbf{p}) = \tilde{\varphi}_j(\omega_{\mathbf{p}}, \mathbf{p}), \quad \varphi_j \in C(R^4) \quad \text{for } j = 1, 2, 3, 4$$

and insert for $A^{f_3}_{ex}$ the LSZ asymptotic condition.

Then $(K_x \equiv \partial_0^2 - \sum_{j=1}^3 \partial_j^2 + m^2)$,

$$(S - 1)_{22} = \lim_{a,b \to +\infty} i \int dy (K_y \langle \Omega | A^{f_4}_{\text{out}} A(y) (A^{f_2}_{\text{in}})^* \times (A^{f_1}_{\text{in}})^* | \Omega \rangle) \varphi_3^*(y; a, b)$$

holds,[13] where

$$\varphi_j(x; a, b) \equiv (2\pi)^{-3/2} \int dp\, \delta_{a,b}(p) \tilde{\varphi}_j(p) \exp(-ipx),$$

$$\delta_{a,b}(p) \equiv \frac{\exp(i(p^0 - \omega_{\mathbf{p}})a) - \exp(i(p^0 - \omega_{\mathbf{p}})b)}{2\pi i (p^0 - \omega_{\mathbf{p}}) 2\omega_{\mathbf{p}}}.$$

Here the identity

$$K_x \varphi_j(x; a, b) = i(\varphi_j(x; a) - \varphi_j(x; -b)) \tag{6}$$

has been used. Choose a causal multiplier $z(x, y)$ in $C(R^8)$ and apply to $(A^{f_2}_{\text{in}})^*$ the asymptotic relations (AIII) and (AIV). The result is the following lemma.

Lemma 2: Let $\varphi_j \in C(R^4)$ and define $\hat{f}_j(\mathbf{p}) = \tilde{\varphi}_j(\omega_{\mathbf{p}}, \mathbf{p})$. If $z(x, y)$ is a causal multiplier in $C(R^8)$ then

$$\begin{aligned}
&\langle \Omega | (A^{f_4}_{\text{out}} A^{f_3}_{\text{out}} - A^{f_4}_{\text{in}} A^{f_3}_{\text{in}})(A^{f_2}_{\text{in}})^* (A^{f_1}_{\text{in}})^* | \Omega \rangle \\
&\quad = -\lim_{a,b \to +\infty} \lim_{c,d \to +\infty} \int dx\, dy \\
&\qquad \cdot (K_x z(y, x) K_y \langle \Omega | A^{f_4}_{\text{out}} A(y) A(x) (A^{f_1}_{\text{in}})^* | \Omega \rangle) \\
&\qquad \times \varphi_3^*(y; a, b) \varphi_2(x; c, d) = -\lim_{a,b \to +\infty} \lim_{c,d \to +\infty} \int dx\, dy \\
&\qquad \cdot (K_x K_y z(y, x) \langle \Omega | A^{f_4}_{\text{out}} A(y) A(x) (A^{f_1}_{\text{in}})^* | \Omega \rangle) \\
&\qquad \times \varphi_3^*(y; a, b) \varphi_2(x; c, d)
\end{aligned}$$

holds.[14]

This lemma shows that the θ function in the corresponding heuristic LSZ reduction formula can be replaced by any causal multiplier to get a rigorous result. The same holds for the well-known reduction formulas with T or R products. Here the modified products

$$T_z(x, y) \equiv z(x, y) A(x) A(y) + z(y, x) A(y) A(x),$$

$$R_z(x, y) \equiv -iz(x, y)[A(x), A(y)]_-$$

have to be used.

A complete reduction to vacuum expectation values needs two further similar steps and gives, e.g.,

$$\begin{aligned}
(S - 1)_{22} = &\lim_{a_3,b_3 \to -\infty} \lim_{a_2,b_2 \to -\infty} \lim_{a_4,b_4 \to -\infty} \lim_{a_1,b_1 \to -\infty} \\
&\times \int dx_1 \cdots dx_4 (K_{x_4} K_{x_2} K_{x_1} z(x_4, x_3) z(x_3, x_2) \\
&\times z(x_2, x_1) K_{x_3} \langle \Omega | A(x_4) \cdots A(x_1) | \Omega \rangle) \\
&\cdot \varphi_4^*(x_4; a_4, b_4) \varphi_3^*(x_3; a_3, b_3) \varphi_2(x_2; a_2, b_2) \varphi_1(x_1; a_1, b_1).
\end{aligned}$$

Assume[15] $z(x, y) = s(x - y)$. Then, by translation invariance of the theory, this may be written in p-space [as an identity for generalized functions over $\tilde{C}(R^{16})$]:

$$\begin{aligned}
&{}_{\text{out}}\langle p', k' | p, k \rangle_{\text{in}} - {}_{\text{in}}\langle p', k' | p, k \rangle_{\text{in}} \\
&\quad = \lim_{a_3,b_3 \to -\infty} \lim_{a_2,b_2 \to -\infty} \lim_{a_4,b_4 \to -\infty} \lim_{a_1,b_1 \to -\infty} \delta^*_{a_4,b_4}(p') \\
&\qquad \cdot \delta^*_{a_3,b_3}(k') \delta_{a_2,b_2}(-k) \delta_{a_1,b_1}(-p) (2\pi)^{5/2} \\
&\qquad \times \delta(p' + k' - p - k) T_{22}(p', k'; p, k),
\end{aligned}$$

where

$$|p, k\rangle_{ex} \equiv (2\pi)^{-1} \tilde{A}_{ex}(-p) \tilde{A}_{ex}(-k) | \Omega \rangle,$$

$$\begin{aligned}
&T_{22}(-k_4, -k_3; k_1, k_2) \\
&\quad \equiv -(k_1^2 - m^2)(k_2^2 - m^2)(k_4^2 - m^2) \int \frac{d\xi_1\, d\xi_2\, d\xi_3}{(2\pi)^{9/2}} \\
&\qquad \cdot \exp[i(\xi_1 k_1 + \xi_2(k_2 + k_1) + \xi_3(k_3 + k_2 + k_1))] s(\xi_3) s(\xi_2) \\
&\qquad \cdot s(\xi_1) \langle \Omega | A(\xi_3) j(0) A(-\xi_2) A(-\xi_2 - \xi_1) | \Omega \rangle
\end{aligned}$$

$(\xi_j = x_{j+1} - x_j,\ y = x_3)$. The generalized function

$$\langle \Omega | A(\xi_3) j(0) A(-\xi_2) A(-\xi_2 - \xi_1) | \Omega \rangle$$

is implicitly defined by

$$\begin{aligned}
&\int d\xi_1\, d\xi_2\, d\xi_3 \langle \Omega | A(\xi_3) j(0) A(-\xi_2) A(-\xi_2 - \xi_1) | \Omega \rangle \\
&\qquad \times \left(\int dy\, h(y, \xi_1, \xi_2, \xi_3) \right) \\
&\quad = \int dx_1 \cdots dx_4 (K_{x_3} \langle \Omega | A(x_4) \cdots A(x_1) | \Omega \rangle) \\
&\qquad \times h(x_3, x_2 - x_1, x_3 - x_2, x_4 - x_3) \quad \text{for all } h \in C(R^{16}).
\end{aligned}$$

This result, here derived from a LSZ field theory over a causal test space, is an exact formulation of Toll's reduction formula (Ref. 4, Sec. 3), for which the primitive domain of analyticity was extensively discussed in Ref. 4.

B. Complete reduction of the S matrix to vacuum expectation values of causal T products

Define "causal T products" by substituting a causal multiplier $z(x, y)$ in $\mathcal{C}(R^8)$ for the step function $\theta(x^0 - y^0)$ in the formal definition of the ordinary T products[16] [S_r being the set of all permutations of $(1, \ldots, r)$]:

$$T_z(X) \equiv \begin{cases} 1 \text{ for } X = \phi, \\ A(x) \text{ for } X = \{x\}, \\ \sum_{\sigma \in S_r} \left[\prod_{\rho=1}^{r-1} z(x_{\sigma(\rho)}, x_{\sigma(\rho+1)}) \right] A(x_{\sigma(1)}) \cdots A(x_{\sigma(r)}) \\ \text{for } X = \{x_1, \ldots, x_r\}. \end{cases}$$

Then a complete reduction of the S matrix to vacuum expectation values of T_z products may be based on the following simple reduction formula, well known for ordinary T products.

Lemma 3: Let $\tilde{\varphi} \in \tilde{\mathcal{C}}(R^4)$ and define $\hat{f}(\mathbf{p}) = \tilde{\varphi}(\omega_{\mathbf{p}}, \mathbf{p})$. Moreover, let X be a (possibly empty) finite set of variables $x_j \in R^4$. If $z(x, y)$ is a causal multiplier in $\mathcal{C}(R^8)$ then, in the weak topology of $\mathcal{C}'$, the relations

$$\langle \Phi | (A^f_{\text{out}})^* T_z(X) - T_z(X)(A^f_{\text{in}})^* | \Phi' \rangle = -\lim_{a,b \to \infty} i \int dx (K_x \langle \Phi | T_z(X \cup \{x\}) | \Phi' \rangle) \varphi(x; a, b) \tag{R1}$$

and

$$\langle \Phi | A^f_{\text{out}} T_z(X) - T_z(X) A^f_{\text{in}} | \Phi' \rangle = \lim_{a,b \to \infty} i \int dx (K_x \langle \Phi | T_z(X \cup \{x\}) | \Phi' \rangle) \varphi^*(x; a, b) \tag{R2}$$

hold for arbitrary $\Phi \in \mathcal{L}_{\text{out}}$, $\Phi' \in \mathcal{L}_{\text{in}}$.

Proof: For $X = \phi$ relation (R1) resp. (R2) is a direct consequence of the LSZ asymptotic condition and (6). Now let X be nonempty, say $X = \{x_1, \ldots, x_r\}$. Then by (AIV) and definition of T_z we get

$$\langle \Phi | T_z(X)(A^f_{\text{in}})^* | \Phi' \rangle = \sum_{\sigma \in S_r} \lim_{b \to +\infty} \int dx \left(\prod_{\rho=1}^{r-1} z(x_{\sigma(\rho)}, x_{\sigma(\rho+1)}) \right) z(x_{\sigma(r)}, x_{r+1}) \times \langle \Phi | A(x_{\sigma(1)}) \cdots A(x_{\sigma(r)}) A(x_{r+1}) | \Phi' \rangle \varphi(x_{r+1}; -b).$$

Moreover, if $\sigma' \in S_{r+1}$ is such that $\sigma'(r+1) \neq r+1$, (AIII), implies

$$\lim_{b \to \infty} \int dx_{r+1} \prod_{\rho=1}^{r} z(x_{\sigma'(\rho)}, x_{\sigma'(\rho+1)}) \cdot \langle \Phi | A(x_{\sigma'(1)}) \cdots A(x_{\sigma'(r+1)}) | \Phi' \rangle \varphi(x_{r+1}; -b) = 0.$$

Summation over all equations, obtained so far, yields

$$\langle \Phi | T_z(X)(A^f_{\text{in}})^* | \Phi' \rangle = \lim_{b \to \infty} \int dx \langle \Phi | T_z(X \cup \{x\}) | \Phi' \rangle \varphi(x_{r+1}; -b).$$

Similarly, using (AV) for $t \to +\infty$ instead of (AIV) for $t \to -\infty$, we conclude with

$$\langle \Phi' | (T_z(X))^* A^f_{\text{out}} | \Phi \rangle = \lim_{a \to \infty} \int dx \langle \Phi' | (T_z(X \cup \{x\}))^* | \Phi \rangle \varphi^*(x; a).$$

Thus (R1) is a direct consequence of both equations (take the complex conjugate of the latter) and identity (6), again. The proof for (R2) is analogous. ∎

By application of the simple identities[17]

$$A^f_{ex} \prod_{\rho=1}^{r} (A^{f_\rho}_{ex})^* | \Omega \rangle = \sum_{\rho'=1}^{r} \langle \Omega | A^f_{ex} (A^{f_{\rho'}}_{ex})^* | \Omega \rangle \prod_{\rho \neq \rho'} (A^{f_\rho}_{ex})^* | \Omega \rangle,$$

$$\langle \Omega | A^f_{ex} (A^{f_{\rho'}}_{ex})^* | \Omega \rangle = \langle \Omega | A^f_{\text{out}} (A^{f_{\rho'}}_{\text{in}})^* | \Omega \rangle = (\langle \Omega | A^{f_{\rho'}}_{\text{out}} (A^f_{\text{in}})^* | \Omega \rangle)^*,$$

we get the following relations:

$$\langle \Omega | \prod_{\rho=1}^{k} A^{f_\rho}_{\text{out}} T_z(X) (A^f_{\text{in}})^* | \Phi' \rangle = \lim_{a,b \to +\infty} i \int dx \left(K_x \langle \Omega | \prod_{\rho=1}^{k} A^{f_\rho}_{\text{out}} T_z(X \cup \{x\}) | \Phi' \rangle \right) \times \varphi(x; a, b) + \sum_{\rho'=1}^{k} \langle \Omega | A^{f_{\rho'}}_{\text{out}} (A^f_{\text{in}})^* | \Omega \rangle \times \langle \Omega | \prod_{\rho \neq \rho'} A^{f_\rho}_{\text{out}} T_z(X) | \Phi' \rangle, \tag{7}$$

$$\langle \Phi | A^f_{\text{out}} T_z(X) \prod_{\rho=1}^{k} (A^{f_\rho}_{\text{in}})^* | \Omega \rangle = \lim_{a,b \to +\infty} i \int dx \left(K_x \langle \Phi | T_z(X \cup \{x\}) \prod_{\rho=1}^{k} (A^{f_\rho}_{\text{in}})^* | \Omega \rangle \right) \varphi^*(x; a, b) + \sum_{\rho'=1}^{k} \langle \Omega | A^f_{\text{out}} (A^{f_{\rho'}}_{\text{in}})^* | \Omega \rangle \langle \Phi | T_z(X) \prod_{\rho \neq \rho'} (A^{f_\rho}_{\text{in}})^* | \Omega \rangle. \tag{8}$$

These relations already show that complete reduction to vacuum expectation values of T_z products is indeed possible. It is merely a matter of economical notation to get a neat reduction formula.

Following Ruelle[18] we introduce the "convolution product"

$$(F * G)(M) \equiv \sum_{L \subset M} (F(L) G(M \setminus L))$$

for complex valued set functions F, G defined for finite subsets M of $Z' \equiv \{1, -1, 2, -2, \cdots\}$. This product is commutative and associative, hence corresponding brackets for multiple $*$ products are not necessary. Recursively define

$$\overset{0}{\prod_{\nu=1}}{}^* G_\nu \equiv 1, \quad \overset{n+1}{\prod_{\nu=1}}{}^* G_\nu \equiv G_{n+1} * \overset{n}{\prod_{\nu=1}}{}^* G_\nu$$

for $n = 0, 1, 2, \cdots$. Note that

$$1(M) \equiv \begin{cases} 1 \text{ for } M = \phi, \\ 0 \text{ otherwise,} \end{cases}$$

in order to ensure $1 * G = G$. Analytic functions of G can be defined via power series expansions preserving their particular properties. For example, the definitions of ln and exp,

$$\ln^* G = \sum_{\nu=1}^{\infty} \frac{(-1)^{\nu-1}}{\nu} \overset{\nu}{\prod_{j=1}}{}^* (G - 1) \quad \text{for } 0 < G(\phi) < 2,$$

$$\exp^* G \equiv \sum_{\nu=0}^{\infty} \frac{1}{\nu!} \overset{\nu}{\prod_{j=1}}{}^* G,$$

preserve the relations

$$\exp^*(\ln^* G) = G \quad \text{for } 0 < G(\phi) < 2$$

and

$$\ln^*(\exp^* F) = F \quad \text{for } 0 < (\exp^* F)(\phi) < 2.$$

In order to apply the * product formalism to the reduction of the S matrix, choose a sequence of test functions $\tilde{\varphi}_1$, $\tilde{\varphi}_{-1}$, $\tilde{\varphi}_2$, $\tilde{\varphi}_{-2}$, $\cdots$ from $\tilde{C}(R^4)$ and define

$$S(M) \equiv \left\langle \Omega \middle| \prod_{\substack{j \in M \\ j>0}} A_{\text{out}}^{f_j} \prod_{\substack{k \in M \\ k<0}} (A_{\text{in}}^{f_k})^* \middle| \Omega \right\rangle \quad \text{for } M \subset Z'$$

[$f_j(x)$ defined according to (1)]. Moreover, choose an arrangement $r_1, r_2, \cdots$ of the set Z' and define the sequence of functions

$$F_0 \equiv 1,$$

$$F_j(M) \equiv \begin{cases} S(M) & \text{if } M = \{r_j, r_k\} \text{ for some } k < j, \\ 0 & \text{otherwise,} \end{cases}$$

$(j = 1, 2, \cdots)$. Finally, define

$$S_n(M) \equiv \lim_{a_1, b_1 \to \infty} \cdots \lim_{a_n, b_n \to \infty} \int \prod_{j \in M \cap R_n} dx_j \cdot \prod_{k \in M \cap R_n} (h_k(x_k; a_k, b_k) K_{x_k}) \left\langle \Omega \middle| \prod_{l \in M \cap N_+ \setminus R_n} A_{\text{out}}^{f_l} \times T_z\left(\prod_{j' \in M \cap R_n} \{x_{j'}\} \right) \prod_{l' \in M \cap N_- \setminus R_n} (A_{\text{in}}^{f_{l'}})^* \middle| \Omega \right\rangle$$

for $n = 0, 1, 2, \ldots, \infty$, where

$$h_r(x; a, b) \equiv \begin{cases} i\varphi_r(x; a, b) & \text{for } r \in N_-, \\ i(\varphi_r(x; a, b))^* & \text{for } r \in N_+, \end{cases}$$

$$R_0 \equiv \phi, \quad R_n \equiv \{r_1, \ldots, r_n\},$$

$$R_\infty \equiv Z', \quad N_\pm \equiv \{j \in Z' : j \gtrless 0\}.$$

Then (7) and (8) may be used in the compact form

$$S_n = S_{n+1} + F_{n+1} * S_{n+1}.$$

This yields complete reduction of S to vacuum expectation values of T_z products,

$$S = \sum_{J \subset N_+} \prod\nolimits^* F_j * S_\infty.$$

More explicitly this reads

$$S(M) = \sum_{P \in \mathcal{P}_M} \prod_{(j,k) \in P} \langle f_j | f_k \rangle S_\infty(M \setminus K_P),$$

where

$$\mathcal{P}_M \equiv \{P \subset (M \cap N_+) \times (M \cap N_-) : \{j, k\} \cap \{j', k'\} = \phi \text{ for different } (j, k), (j', k') \in P\},$$

$$K_P \equiv \bigcup_{(j,k) \in P} (\{j\} \cup \{k\}).$$

Note that, in the * product formalism, the usual implicit definition of the *connected parts* S_C of the S matrix is given by

$$S = \exp^* S_C.$$

Since

$$\sum_{J \subset N_+} \prod\nolimits^* F_j = \exp^* \left(\sum_{j \in N_+} F_j \right),$$

we have

$$S_C = \ln^* S = \ln^* S_\infty + \sum_{j \in N_+} F_j.$$

Thus we finally obtain[19] the following theorem.

Theorem 1: Let $z(x, y)$ be a causal multiplier in $\tilde{C}(R^8)$. Let n, n' be positive integers with $n + n' \geq 2$. Finally, let $\tilde{\varphi}_1, \ldots, \tilde{\varphi}_n, \tilde{\varphi}_{-1}, \ldots, \tilde{\varphi}_{-n'} \in \tilde{C}(R^4)$. Then the reduction formula

$$S_C(f_1, \ldots, f_n; f_{-1}, \ldots, f_{-n'}) = i^{n+n'} \lim_{a_1, b_1 \to +\infty} \cdots \lim_{a_{n+n'}, b_{n+n'} \to +\infty} \int dx_1 \cdots dx_{n+n'} \cdot (K_{x_1} \cdots K_{x_{n+n'}} \langle \Omega | T_z(x_1, \ldots, x_{n+n'}) | \Omega \rangle^T) \times \varphi_1^*(x_1; a_1, b_1) \cdots \varphi_n^*(x_n; a_n, b_n) \varphi_{-1}(x_{n+1}; a_{n+1}, b_{n+1}) \times \cdots \varphi_{-n'}(x_{n+n'}; a_{n+n'}, b_{n+n'})$$

holds, the order of the iterated limits being arbitrary.

Remark: As usual, the truncated vacuum expectation values of the T_z products are defined by

$$\langle \Omega | T_z(x_1, \ldots, x_n) | \Omega \rangle^T = \sum_{l=1}^{n} (-1)^{l-1} (l-1)! \sum_{M \in P_l(n)} \prod_{J \in M} \langle \Omega | T_z(x_{j_1}, \ldots, x_{j_{k_J}}) | \Omega \rangle,$$

where $M = \{J_1, \ldots, J_l\} \in P_l(n)$ iff it is a partition of $\{1, \ldots, n\}$ into l (nonempty) disjoint ordered subsets J_r with the ordering in each $J = (j_1, \ldots, j_{k_J})$ being the natural relative ordering of integers.

C. LSZ asymptotic condition restricted to essentially nonoverlapping asymptotic states

Up to now, within the Wightman framework,[8] the LSZ asymptotic condition can only be proved in the form (AI) if $\mathcal{L}_{ex}$ is replaced by its subset of essentially nonoverlapping states[20]:

Definition 2: The state $\Phi \in \mathcal{L}_{ex}$ is called *essentially nonoverlapping* iff it is a (finite) linear combination of ingoing resp. outgoing r-particle states $(r = 0, 1, 2, \cdots)$ with p-space wavefunctions $\hat{f}(\mathbf{p}_1, \ldots, \mathbf{p}_r)$ vanishing together with all their derivatives at points where at least two of the arguments $\mathbf{p}_1, \ldots, \mathbf{p}_r$ coincide.

But we immediately realize that all considerations and results of Sec. 3, Part B still apply to the case of this restricted form of the asymptotic condition provided that all states corresponding to p-space wavefunctions of the form

$$\hat{f}(\mathbf{p}_1, \ldots, \mathbf{p}_n) = \tilde{\varphi}_{\hat{r}_1}(\omega_{\mathbf{p}_1}, \mathbf{p}_1) \cdots \tilde{\varphi}_{\hat{r}_n}(\omega_{\mathbf{p}_n}, \mathbf{p}_n), \quad n = 2, 3, \cdots$$

are essentially nonoverlapping for every arrangement $\hat{r}_1, \hat{r}_2, \cdots$ of $\{1, 2, 3, \cdots\}$ as well as of $\{-1, -2, -3, \cdots\}$. The results of Sec. 3, Part A remain valid even without any additional restriction on the asymptotic states.

4. CAUSALITY OF THE SCHWARTZ SPACE $\mathcal{S}(R^1)$

A. Characterization of causal multipliers in $\mathcal{S}(R^8)$

As discussed in Sec. 2, Part A, causality of the test space is related to the asymptotic ($t \to \pm\infty$) behavior of functions of type (2) stemming from their intimate connection with smooth Klein–Gordon solutions. For general $\varphi(x;t)$ this behavior may be derived from the representation[21]

$$\tilde{\varphi}(p) = (2\pi)^2 \tilde{\psi}(p) \hat{f}(\mathbf{p}), \quad \tilde{\psi} \in \tilde{\mathcal{S}}(R^4), \quad \hat{f} \in \tilde{\mathcal{S}}(R^3),$$

for which we have

$$\varphi(x;t) = i \int_{x'^0 = t} dx' \psi(x - x') \overleftrightarrow{\partial}'_0 f(x'). \tag{9}$$

Since for smooth Klein–Gordon solutions we have the physically plausible inequality[22]

$$\|y\|^N |f(x)| < A(1 + \|x - y\|^N) \quad \text{for } x \in R^4, \quad y \in R^4 \backslash V,$$

(9) implies[23]

$$\begin{aligned} \|y\|^N |\varphi(x;t)| &\leq \int_{x'^0 = t} dx' \|y\|^N |\psi(x - x') \overleftrightarrow{\partial}'_0 f(x')| \\ &\leq A' \int_{x'^0 = t} dx' (1 + \|y - x'\|^N) \\ &\quad \times \max\{|\psi(x - x')|, |\partial'_0 \psi(x - x')|\} \\ &\leq A''(1 + \|(t - y^0, \mathbf{x} - \mathbf{y})\|^N) \\ &\quad \text{for } x \in R^4, \quad y \in R^4 \backslash V. \end{aligned} \tag{10}$$

By simple multiple partial integration with respect to p in (2) we also have the inequality[24]

$$|x^0 - t|^n \|x\|^r |D_x^\alpha \varphi(x;t)| < A_{n,r,\alpha}(1 + \|t\|^r), \tag{11}$$

valid for arbitrary $x \in R^4$, $t \in R^1$, $\alpha \in Z_+^4$, and $n, r \in Z_+$. Here we adopt the usual notation

$$Z_+ \equiv N_+ \cup \{0\}, \quad Z_+^4 \equiv Z_+ \times Z_+ \times Z_+ \times Z_+, \quad D_x^\alpha \equiv \prod_{j=0}^{3} (\partial_j)^{\alpha^j}.$$

Inequalities (10) and (11) are sufficient to prove the following useful criterion for conditions (i) and (ii) of Definition 1.

Theorem 2: Let $k_\pm$ be a multiplier in $\mathcal{S}(R^8)$, i.e., $k_\pm \in \mathcal{O}_M(R^8)$, with

$$\lim_{\mu \to +\infty} \max_{(x,y) \in G_\pm(\mu)} \mu^r |D_x^\alpha D_y^\beta k_\pm(x,y)| = 0 \tag{C}$$

for all $r \in Z_+$ and $\alpha, \beta \in Z_+^4$, where

$$G_\pm(\mu) \equiv \{(x,y) \in R^8 : (x-y)^2 > \mu, \pm(x^0 - y^0) > 0\}.$$

Then

$$\lim_{t \to \pm\infty} k_\pm(x,y) h(\hat{x},y) \varphi^\#(x;t) = 0$$

holds in $\mathcal{C}(R^{4(n+2)})$ for arbitrary $h \in \mathcal{C}(R^{4(n+1)})$ and $\tilde{\varphi} \in \tilde{\mathcal{C}}(R^4)$.

Proof: We prove Theorem 2 for the "–" sign only, the proof for the "+" sign being analogous.

Since no restriction was imposed on the tempered test function h and since partial differentiation as well as complex conjugation does not change the required properties of k_-, it is clearly sufficient to show, e.g.,

$$\lim_{t \to -\infty} \max_{\hat{x} \in R^{4n}} \max_{(x,y) \in R^8} \|x\|^r \times |k_-(x,y) h(\hat{x},y) D_x^\alpha \varphi(x;t)| = 0$$

for arbitrary $r \in Z_+$, $\alpha \in Z_+^4$. As $k_- \in \mathcal{O}_M(R^8)$ and $h \in \mathcal{S}(R^{4(n+1)})$, an even stronger statement is obtained if one or both of the factors k_-, h are dropped. Thus, introducing the notation

$$x + t \equiv (x^0 + t, \mathbf{x}) \quad \text{for } x \in R^4, \quad t \in R^1,$$

and defining the four regions [$s \equiv (2r+2)^{-1}$]

$$G_1(t) \equiv \{(x,y) \in R^8 : y^0 - x^0 \leq |t|^s\},$$

$$G_2(t) \equiv \{(x,y) \in R^8 : |x^0 - t| \geq |t|^s\},$$

$$G_3(t) \equiv \{(x,y) \in R^8 : y^0 - x^0 > |t|^s, \ (x - y + |t|^s)^2 \geq 0\},$$

$$G_4(t) \equiv \{(x,y) \in R^8 : y^0 - x^0 > |t|^s, \ (x - y + |t|^s)^2 < 0, \ |x^0 - t| < |t|^s\}$$

fulfilling

$$G_1(t) \cup \cdots \cup G_4(t) = R^8,$$

we only have to prove the four relations:

$$\lim_{t \to -\infty} \max_{\hat{x} \in R^{4n}} \max_{(x,y) \in G_1(t)} \|x\|^r |h(\hat{x},y) D_x^\alpha \varphi(x;t)| = 0, \tag{12}$$

$$\lim_{t \to -\infty} \max_{(x,y) \in G_2(t)} \|x\|^r |D_x^\alpha \varphi(x;t)| = 0, \tag{13}$$

$$\lim_{t \to -\infty} \max_{(x,y) \in G_3(t)} \|x\|^r |k_-(x,y) D_x^\alpha \varphi(x;t)| = 0, \tag{14}$$

$$\lim_{t \to -\infty} \max_{\hat{x} \in R^{4n}} \max_{(x,y) \in G_4(t)} \|x\|^r |h(\hat{x},y) D_x^\alpha \varphi(x;t)| = 0. \tag{15}$$

Proof of (12): (12) is just a consequence of the fact that by virtue of (11)

$$\max_{\hat{x} \in R^{4n}} \max_{(x,y) \in G_1(t)} (1 + |y^0|^{r+1})(1 + |y^0 - t|^{r+1}) \times \|x\|^r |h(\hat{x},y) D_x^\alpha \varphi(x;t)|$$

cannot increase faster than $|t|^r$ when $t \to -\infty$, whereas

$$\min_{(x,y) \in G_1(t)} (1 + |y^0|^{r+1})(1 + |y^0 - t|^{r+1})$$

increases like $|t|^{r+1}$ because of

$$\max\{|y^0|, |x^0 - t|\} \geq (|t| - |t|^s)/s \text{ in } G_1(t) \quad \text{for } t < 0.$$

Proof of (13): This relation holds because

$$\max_{(x,y) \in G_2(t)} |x^0 - t|^{(r+1)/s} \|x\|^r |D_x^\alpha \varphi(x;t)|$$

cannot, again due to (11), grow faster than $|t|^r$ when $t \to -\infty$, whereas

$$\min_{(x,y) \in G_2(t)} |x^0 - t|^{(r+1)/s}$$

will increase like $|t|^{r+1}$.

Proof of (14): Inequality (11) shows that

$$\max_{(x,y) \in G_3(t)} \|x\|^r |D_x^\alpha \varphi(x;t)|$$

can be majorized by $A + B|t|^r$ with suitable constants

A, B independent of t. Therefore (14) is an immediate consequence of our assumptions on k_- because

$$G_3(t) \subset G_-(|t|^{2s}).$$

Proof of (15): Since $\|x\|^r$ can be majorized by a finite linear combination of expressions of the type

$$|t|^{r_1 s}\|y\|^{r_2}\|x - y + |t|^s\|^{r_3}$$

(r_1, r_2, r_3 being nonnegative integers not greater than r), it is sufficient to prove

$$\lim_{t\to-\infty} \max_{\hat{x}\in R^{4n}} \max_{(x,y)\in G_4(t)} |t|^{rs}(1 + \|x - y + |t|^s\|)^r \times |h(\hat{x}, y) D_x^\alpha \varphi(x; t)| = 0 \tag{16}$$

instead of (15). Using inequality (10) we get the inequality

$$(1 + |y^0|)(1 + \|x - y + |t|^s\|)|t|^{rs} \times (1 + \|x - y + |t|^s\|)^r |h(\hat{x}, y) D_x^\alpha \varphi(x; t)| \leq A|(1 + |y^0|)h(\hat{x}, y)||t|^{rs}(1 + \|(t - (x^0 - y^0 + |t|^s), \mathbf{y})\|^{r+1})$$

for all $(x, y) \in G_4(t)$, $t \in R^1$, where A is a suitable constant independent of (x, y) and t. Because $h(\hat{x}, y)$ is an element of $\mathcal{S}(R^{4(n+1)})$, the maximum value of the right-hand side can not increase faster than $|t|^{s(2r+1)}$ for $(x, y) \in G_4(t)$ when $t \to -\infty$. The inequality

$$\min_{(x,y)\in G_4(t)} \max\{|y^0|, |x^0 - y^0 + |t|^s|\} > |t|/2 - |t|^s \quad \text{for } t < 0$$

shows that the increase of

$$\min_{(x,y)\in G_4(t)} (1 + |y^0|)(1 + \|x - y + |t|^s\|)$$

is at least as $|t|$ when $t \to -\infty$. Thus (16) must be valid and the proof of Theorem 2 is complete. ∎

B. Construction of Poincaré-invariant causal multipliers in $\mathcal{S}(R^8)$

By Theorem 3 we may easily construct a causal multiplier in $\mathcal{S}(R^8)$ which is also Poincaré invariant:

Choose $a, b > 0$ and $\varphi \in \mathcal{D}_M(R^1)$ with $\varphi(q) = 1$ for $q > a$ and $\varphi(q) = 0$ for $q < b$. Then $z(x, y) \equiv \theta(x^0 - y^0)\varphi((x - y)^2)$ is a Poincaré invariant multiplier in $\mathcal{S}(R^8)$ with support contained in V_+. Therefore we only have to check conditions (i) and (ii) of Definition 1. Theorem 3 tells us that both conditions are fulfilled if (C) holds for $k_\pm(x, y) = z_\mp(x, y)$ as well as for $k_\pm(x, y) = 1 - z_\pm(x, y)$, ($z_-(y, x) \equiv z_+(x, y) \equiv z(x, y)$). But (C) is trivially fulfilled since in both cases $k_\pm = 0$ in $G_\pm(a)$. Thus $z(x, y)$ is a *causal* Poincaré invariant multiplier in $\mathcal{S}(R^8)$.

5. CONCLUSIONS

We have seen that a rigorous version of Toll's reduction formalism is already a consequence of Hepp's rigorous version of the LSZ asymptotic condition. Although this leads to primitive domains of analyticity even larger than those derived by locality, there are serious difficulties in trying to enlarge these domains for derivation of dispersion relations.

Complete reduction of the whole S matrix by the new reduction formalism showed that for a calculation of the S matrix we need only know the n-point Wightman distributions at arguments $x_1, \ldots, x_n$ with large *timelike* separation, e.g.,

$$(x_j - x_k)^2 > \mu \quad \text{for } j \neq k, \ 0 < \mu \text{ fixed.}$$

From this point of view the axiom of locality seems very far removed from any direct physical interpretation concerning the S matrix. Up to now, we have not been able to give a satisfactory analysis of what our results precisely imply. We have the feeling, however, that they indicate some macrocausality property—perhaps even of the type investigated by Chandler, Iagolnitzer, and Stapp[25]—as a consequence of microcausality of the asymptotic field to which the interpolating field is to converge in some well specified way by the LSZ asymptotic condition. It should be worthwhile to further analyze this problem.

We conclude with the remark that our results may appear surprising for the following reasons:

(i) If the cluster property holds, then by Haag's asymptotic condition[1] for nonoverlapping states (in connection with a lemma by Ruelle on smooth Klein—Gordon solutions[11]) one can show that for a calculation of the S matrix it would be alternatively sufficient to know the vector valued distributions $A(x_1)\cdots A(x_n)\Omega$ at arguments with large pairwise *spacelike* separation.

(ii) For the usual T products Hepp proved[20] that the off-shell extrapolations

$$\prod_{j=1}^{n} (p_j^2 - m^2)\langle\Omega|\tilde{T}(p_1, \ldots, p_k, -p_{k+1}, \ldots, -p_n)|\Omega\rangle$$

of the S matrix are C^∞ in all the variables $p_j^0 - \omega_{\mathbf{p}_j}$ simultaneously within some neighborhood of the origin when integrated over $\mathbf{p}_1, \ldots, \mathbf{p}_n$ with nonoverlapping test functions $f_j(\mathbf{p}_j) \in \mathcal{D}(R^3)$. If this were also the case for the T_z products, then, by Theorem 1, the S matrix would be trivial. Actually, there is no analogous proof for the T_z products.

[1] R. Haag, Phys. Rev. **112**, 669 (1958); J. Bümmerstede and W. Lücke, Commun. Math. Phys. **37**, 121 (1974).
[2] R. Jost, *The general theory of quantized fields* (Am. Math. Soc., Providence, R.I., 1965), p. 130; see also J. Bümmerstede and W. Lücke, J. Math. Phys. **16**, 1203 (1975).
[3] P.N. Dobson, J. Math. Phys. **11**, 1750 (1970).
[4] J.S. Toll, "A new reduction technique in quantum field theory," Rochester Conference (1960).
[5] H. Lehmann, K. Symanzik, and W. Zimmermann, Nuovo Cimento **1**, 205 (1955).
[6] I.e., tensor products, derivatives, Fourier transform, affine transformations, etc., can be defined as usual.
[7] See, for example, A. Pietsch, *Nukleare lokkonvexe Räume* (Akademie-Verlag, Berlin, 1969).
[8] See, for example, R.F. Streater and A.S. Wightman, *PCT spin and statistics and all that* (Benjamin, New York, 1964).

[9] The notation is:

$$px \equiv p^0x^0 - \mathbf{px}, \quad \omega_{\mathbf{p}} \equiv (\mathbf{p}^2 + m^2)^{1/2},$$
$$\hat{x} \equiv (x_1, \ldots, x_n), \quad d\hat{x} \equiv dx_1 \cdots dx_n,$$
$$\tilde{\varphi}(\hat{p}) \equiv (2\pi)^{-2n} \int d\hat{x}\varphi(\hat{x}) \exp(i\sum_{j=1}^{n} p_j x_j).$$

[10]For comparison with the original (nonrigorous) LSZ asymptotic condition note that, with $\tilde{\varphi}(p) = (2\pi)^2 \hat{f}(\mathbf{p})\tilde{\psi}(p)$,

$$\int dx\, A(x)\varphi(x;t) = i\int_{x^0=t} dx \left(\int dx'\, A(x')\psi(x'-x) \right) \overleftarrow{\partial_0} f(x) \text{ holds.}$$

[11]D. Ruelle, Helv. Phys. Acta **35**, 147 (1962); H. Araki, Zürich lectures, 1962 (unpublished); R. Jost, Helv. Phys. Acta **39**, 21 (1966); W. Lücke, Int. J. Theor. Phys. **10**, 91 (1974).

[12]The notation $\varphi^{\#}(x;t)$ means that the relation under consideration holds for $\varphi^{\#}(x;t) = \varphi(x;t)$ as well as for $\varphi^{\#}(x;t) = \varphi^*(x;t)$.

[13]Note that $\lim_{a,b\to+\infty}\varphi_j(x;a,b) = f_j(x)$.

[14]The second equality can be seen by also considering the derivatives with respect to y of conditions (i) and (ii) of Definition 1.

[15]Compare Sec. 4, Part B.

[16]It will prove convenient to define the T products as a set function and thus explicitly exhibit the corresponding symmetry property.

[17]We assume stability of the vacuum and of the 1-particle states, of course.

[18]D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969); for further applications of this product see also G.C. Hegerfeldt, Commun. Math. Phys. **45**, 137 (1975).

[19]Compare H. Araki, Varenna lectures (1968).

[20]K. Hepp, Commun. Math. Phys. **1**, 95 (1965); H. Araki and R. Haag, Commun. Math. Phys. **4**, 77 (1967); W. Lücke, Ref. 11; J. Bümmerstede and W. Lücke, Ref. 1.

[21]W. Lücke, Ref. 11.

[22]H. Araki, Ref. 11; W. Lücke, Ref. 11.

[23]For a simpler, though physically not as transparent, self-contained proof of inequality (10) see also H.D. Doebner and W. Lücke, internal report, International Centre For Theoretical Physics, Trieste, IC/72/20.

[24]Compare K. Hepp, Helv. Phys. Acta **37**, 639 (1964), Eq. (1.2.5).

[25]C. Chandler, Phys. Rev. **174**, 1749 (1968); D. Iagolnitzer and H.P. Stapp, Commun. Math. Phys. **14**, 15 (1969).

Graded Lie algebras: Generalization of Hermitian representations

M. Scheunert, W. Nahm, and V. Rittenberg

Physikalisches Institut der Universität Bonn, Bonn, Germany
(Received 30 August 1976)

Hermitian representations play a fundamental role in the study of the representations of simple Lie algebras. We show how this concept generalizes for classical simple graded Lie algebras. Star and grade star representations are defined through adjoint and grade adjoint operations. Each algebra admits at most two adjoint and two grade adjoint operations (we list the various possibilities for all classical simple graded Lie algebras). To each adjoint (grade adjoint) operation corresponds a class of star (grade star) representations. The tensor product of two star representations belonging to one class is completely reducible into irreducible representations belonging to the same class. This property is very useful since in general the finite-dimensional representations of classical simple graded Lie algebras are not completely reducible.

1. INTRODUCTION

Much work has been done lately on the classification of graded Lie algebras. All simple graded Lie algebras are now known.[1-4] An important part of the simple graded Lie algebras are the so-called *classical* simple graded Lie algebras: For them the odd part of the graded Lie algebra is completely reducible into one or two irreducible subspaces. The classical simple graded Lie algebras are the $\mathrm{spl}(n,m)$ $(n>m\geq 1)$, $\mathrm{spl}(n,n)/z_n$ $(n\geq 2)$, $\mathrm{osp}(2p,m)$ $(p,m\geq 1)$, $\mathrm{b}(n)$ $(n\geq 3)$, and $\mathrm{d}(n)/z_n$ $(n\geq 3)$ series and the exceptional algebras whose Lie algebras are $\mathrm{sl}(2)\times G_2$, $\mathrm{sl}(2)\times\mathrm{o}(7)$ and $\mathrm{sl}(2)\times\mathrm{sl}(2)\times\mathrm{sl}(2)$, respectively.[2] In this paper we shall deal mainly with the classical simple graded Lie algebras.

A natural question is: Do the representations of simple graded Lie algebras have properties similar to those of simple Lie algebras? Such a possibility could be expected at least for the classical simple graded Lie algebras because in this case the underlying Lie algebra is reductive (semisimple times Abelian). An indication that this is not the case comes from a theorem by Djoković and Hochschild[5] which states:

The only graded Lie algebras for which all finite-dimensional representations are completely reducible are the direct products of $\mathrm{osp}(2p,1)$ *algebras and semisimple Lie algebras.*

This theorem singles out the $\mathrm{osp}(2p,1)$ algebras as possible candidates for our analogy. As will be shown, however, in this paper and through an example in the next one,[6] the properties of the representations of these algebras are different in several aspects from those of simple Lie algebras.

We are still left with the problem: Do the representations of the classical simple graded Lie algebras have some common properties? As will be shown the answer to this question is rather negative.

In the present paper we generalize the concept of Hermitian representations. (Remember that the finite-dimensional representations of the compact simple Lie algebras are equivalent to Hermitian representations.) In this way we not only try to get a better insight into the properties of the representations of graded Lie algebras but also to deal with one consequence of the Djoković–Hochschild theorem: The tensor product of two irreducible representations is in general not completely reducible. We show that it is possible to find classes of representations for which complete reducibility holds.

We first define the generalization of the usual adjoint operations for graded Lie algebras (see Sec. 2). For the even generators this is done in the usual way. For the odd generators this can be done either through a trivial generalization [see (2.8)–(2.11)] leading to an *adjoint operation* or through a specific new procedure [see (2.14)–(2.17)] leading to a *grade adjoint operation*. To adjoint (grade adjoint) operations correspond *star* (*grade star*) *representations*. One then shows that every star (grade star) representation in a graded *Hilbert space* is completely reducible. The tensor product of two star (grade star) representations can be made a star (grade star) representation by choosing a suitable scalar product. There is, however, a crucial difference between star and grade star representations: Whereas the scalar product on the tensor space will be positive definite in the former case it will be indefinite in the latter (i.e., the tensor product of two grade star representations will no longer be a representation in a Hilbert space). Thus the tensor product of two star representations is always completely reducible but this is generally not the case for grade star representations.

In Sec. 3 we examine the possibility of defining other "generalized adjoint operations;" it turns out that for the classical simple graded Lie algebras we are essentially left with the possibilities considered in Sec. 2, namely the adjoint and grade adjoint operations. There is, however, a major difference between the adjoint operations for simple Lie algebras and the adjoint (grade adjoint) operations for classical simple graded Lie algebras, namely the latter are "not uniquely defined" (more precisely, if we choose any adjoint operation for the even generators one can define in general several adjoint (grade adjoint) operations for the odd generators). This will also be evident from Sec. 4 which deals with the explicit construction of adjoint (grade adjoint) operations. Since to each adjoint (grade adjoint) operation corresponds a class of star (grade star) representations we have also to discuss the problem of the equivalence of the various classes. This is done in Sec. 5.

Our results are summarized in Sec. 6 where we list for almost all the classical simple graded Lie algebras [the $\Gamma(\sigma_1, \sigma_2, \sigma_3)$ and the $\mathrm{spl}(2,2)/z_2$ algebras have not been studied] the possible classes of representations. It turns out that [perhaps with the exception of certain $\mathrm{spl}(n,n)/z_n$ algebras] we can have at most two classes of star and two classes of grade star representations. Some algebras, however, have no star and/or no grade star representations. The problem of constructing explicitly the various classes was not considered. In the subsequent paper[6] we shall consider in full detail the examples of the osp(2,1) and spl(2,1) algebras.

2. STAR AND GRADE STAR REPRESENTATIONS

To begin with let us describe the background which led us to define the star and grade star representations. It is a classical theorem by Weyl that all finite-dimensional representations of a semisimple Lie algebra are completely reducible. This important theorem is not true for simple graded Lie algebras. In fact[5] all finite-dimensional representations of a simple graded Lie algebra L are completely reducible if and only if L is isomorphic to one of the algebras $\mathrm{osp}(2p,1)$, $p \geq 1$.

Because of this result it is natural to look for other criteria which imply complete reducibility. In classical group theory there is a simple means to ensure complete reducibility: One demands that the representation space should be a Hilbert space and that the representation should be unitary. The corresponding requirement for a representation of a real Lie algebra to be completely reducible is that the Lie algebra should be represented by skew-Hermitian operators.

Finally, let L_0 be a complex Lie algebra and let $Q \to Q^*$ be an *adjoint operation* in L_0, i.e., a mapping of L_0 into itself which satisfies

$$(aP + bQ)^* = a^* P^* + b^* Q^*, \tag{2.1}$$

$$[P,Q]^* = [Q^*, P^*], \tag{2.2}$$

$$(Q^*)^* = Q, \tag{2.3}$$

for all $P, Q \in L_0$ and for all complex numbers a, b. (The asterisk denotes complex conjugation.) It is well known that

$$L_0^R = \{Q \in L_0 \mid Q^* = -Q\} \tag{2.4}$$

is a real form of L_0 and that, conversely, every real form L_0^R of L_0 determines in a natural way an adjoint operation in L_0 such that (2.4) is valid.

Now let ρ be any representation of L_0 in a finite-dimensional Hilbert space. Then the restriction ρ^R of ρ to L_0^R is skew-Hermitian if and only if

$$\rho(Q^*) = \rho(Q)^* \tag{2.5}$$

for all $Q \in L_0$, and the representation ρ is completely reducible if and only if ρ^R is completely reducible. A representation with the property (2.5) will be called a *star representation*.

We remind the reader that this is a well-known concept in the theory of (associative) star algebras. Furthermore, we note that in the mathematical literature it is customary to consider conjugations of L_0 instead of adjoint operations. A conjugation of L_0 is a mapping τ of L_0 into itself which satisfies

$$\tau(aP + bQ) = a^* \tau(P) + b^* \tau(Q), \tag{2.1'}$$

$$\tau([P,Q]) = [\tau(P), \tau(Q)], \tag{2.2'}$$

$$\tau^2(Q) = Q, \tag{2.3'}$$

for all $P, Q \in L_0$ and all complex numbers a, b. However, it is obvious that τ is a conjugation if and only if $-\tau$ is an adjoint operation.

In the present work we try to generalize the concepts of an adjoint operation and of a star representation to graded Lie algebras. It turns out that because of the grading the Eqs. (2.1)–(2.3) may be generalized in essentially two different ways. Similarly (and connected with this fact) for a linear operator acting in a graded Hilbert space there is a natural modified definition of the adjoint operator [see Eq. (2.12)].

Let us now give the details. Suppose that $V = V_0 \oplus V_1$ is a finite-dimensional graded vector space with even subspace V_0 and odd subspace V_1. We assume that on V there exists a nondegenerate Hermitian form, denoted by a bracket $\langle | \rangle$, such that V_0 and V_1 are orthogonal with respect to this form, i.e.,

$$\langle V_0 \mid V_1 \rangle = \{0\}. \tag{2.6}$$

At the present stage we would gain nothing by demanding that $\langle | \rangle$ should be positive definite. Nevertheless, if $\langle | \rangle$ is positive definite then we call V a *graded Hilbert space*.

As is well known for any linear operator A in V the adjoint operator A^* (with respect to $\langle | \rangle$) is defined by

$$\langle A^* x \mid y \rangle = \langle x \mid Ay \rangle \tag{2.7}$$

for all $x, y \in V$.

Now recall that there is a natural procedure to convert the vector space of all linear operators in V into a graded Lie algebra, denoted by $\mathrm{pl}(V)$. The well-known rules for the adjoint operator imply that

The adjoint of an even (resp. odd) element is even (resp. odd), (2.8)

$$(aA + bB)^* = a^* A^* + b^* B^*, \tag{2.9}$$

$$\langle A, B \rangle^* = \langle B^*, A^* \rangle, \tag{2.10}$$

$$(A^*)^* = A, \tag{2.11}$$

for all elements A, B of $\mathrm{pl}(V)$ and all complex numbers a, b.

Definition 1: *An adjoint operation in a graded Lie algebra L is a mapping $A \to A^*$ of L into itself which satisfies the conditions* (2.8)–(2.11).

On the other hand, a well-known rule of thumb says that in going from normal Lie algebras to graded Lie algebras it is natural to make a change of sign at every place where "two odd objects have been interchanged."

Suppose, then, that A is a homogeneous (i.e., even or odd) linear operator in V of degree α. We define the *grade adjoint* operator $A^\dagger$ of A by

$$\langle A^\dagger x | y\rangle = (-1)^{\alpha\xi}\langle x | Ay\rangle \tag{2.12}$$

for all homogeneous elements $x, y \in V$ of degrees ξ, η. One should note that this definition depends on which of the subspaces V_0, V_1 is even and which is odd.

In the special case where $\langle\ |\ \rangle$ is positive definite let us give the grade adjoint operator also in matrix notation. Choose an orthonormal basis $e_1, \ldots, e_n, e_{n+1}, \ldots, e_{n+m}$ of V such that $e_1, \ldots, e_n$ is a basis of V_0 and $e_{n+1}, \ldots, e_{n+m}$ is a basis of V_1. If A is any linear operator in V and if $\begin{pmatrix} P & Q \\ R & S \end{pmatrix}$ is its matrix (written in block form) with respect to our basis, then the matrix of $A^\dagger$ is equal to $\begin{pmatrix} P^* & -R^* \\ -Q^* & S^* \end{pmatrix}$, where the plus sign denotes the normal Hermitian conjugation of a matrix.

Obviously, we have $A^\dagger = A^*$ if A is even. It is easy to check that

$$(AB)^\dagger = (-1)^{\alpha\beta} B^\dagger A^\dagger \tag{2.13}$$

for all homogeneous linear operators A, B of degrees α, β. Furthermore, for the graded Lie algebra $\mathrm{pl}(V)$ we derive the following rules:

The grade adjoint of an even (resp. odd) element is even (resp. odd), (2.14)

$$(aA + bB)^\dagger = a^* A^\dagger + b^* B^\dagger, \tag{2.15}$$

$$\langle A, B\rangle^\dagger = (-1)^{\alpha\beta}\langle B^\dagger, A^\dagger\rangle, \tag{2.16}$$

$$(A^\dagger)^\dagger = (-1)^\alpha A, \tag{2.17}$$

for all homogeneous elements A, B of $\mathrm{pl}(V)$ with degrees α, β and for all complex numbers a, b. Note that because of condition (2.14), Eq. (2.16) is equivalent to

$$\langle A, B\rangle^\dagger = -\langle A^\dagger, B^\dagger\rangle. \tag{2.16'}$$

Definition 2: *A grade adjoint operation in a graded Lie algebra L is a mapping $A \to A^\dagger$ of L into itself which satisfies the conditions* (2.14)–(2.17).

Definition 3: *Let L be a graded Lie algebra equipped with an adjoint (resp. grade adjoint) operation, and let V be a graded vector space as described above. A star representation (resp. a grade star representation) of L in V is a graded representation ρ of L in V which satisfies*

$$\rho(A^*) = \rho(A)^*, \quad [\text{resp. } \rho(A^\dagger) = \rho(A)^\dagger] \tag{2.18}$$

for all elements $A \in L$.

As explained above we expect the following proposition to be valid.

Proposition 1: *Every star (resp. grade star) representation ρ of L in a graded Hilbert space V is completely reducible.*

Proof: The argument is standard. Let U be a graded subspace of V which is invariant under the representation ρ. We have to find a complementary invariant graded subspace U' of V, i.e., an invariant graded subspace U' of V such that V is the direct sum of U and U'. Evidently there is a natural candidate for U', namely the orthocomplement $U^\perp$ of U in V,

$$U^\perp = \{x \in V \mid \langle x | y\rangle = 0 \text{ for all } y \in U\}. \tag{2.19}$$

It is easy to check that $U^\perp$ is graded. Furthermore, V is the direct sum of U and $U^\perp$ (since $\langle\ |\ \rangle$ is positive definite) and the star (resp. grade star) property of ρ implies that $U^\perp$ is invariant.

Next we shall discuss tensor products of representations. Let ρ, ρ' be two graded representations of a graded Lie algebra L in some graded vector spaces V, resp. V'. Then the tensor product $\rho \otimes \rho'$ of ρ and ρ' is the graded representation in $V \otimes V'$ defined by

$$(\rho \otimes \rho')(A)(x \otimes x') = (\rho(A)x) \otimes x' + (-1)^{\alpha\xi} x \otimes (\rho'(A)x'), \tag{2.20}$$

for all homogeneous elements $A \in L$, $x \in V$, $x' \in V'$ of degrees, respectively α, ξ, ξ'.

Suppose now that ρ and ρ' are star representations. We denote both the Hermitian form on V and that on V' by the same symbol $\langle\ |\ \rangle$ and define as usual a Hermitian form on $V \otimes V'$ by

$$\langle x \otimes x' | y \otimes y'\rangle = \langle x | y\rangle\langle x' | y'\rangle, \tag{2.21}$$

for all $x, y \in V$ and $x', y' \in V'$.

Then it is easy to check that *the tensor product $\rho \otimes \rho'$ is a star representation*, too. Note that if V and V' are graded Hilbert spaces, then $V \otimes V'$ is also a graded Hilbert space. Hence in this case we can use Proposition 1 to prove the complete reducibility of $\rho \otimes \rho'$.

Let us now consider the tensor product of two grade star representations ρ and ρ'. In that case we define (in agreement with the rule of thumb mentioned above) the Hermitian form on $V \otimes V'$ by

$$\langle x \otimes x' | y \otimes y'\rangle = (-1)^{\xi'\eta}\langle x | y\rangle\langle x' | y'\rangle, \tag{2.22}$$

for all homogeneous elements $x, y \in V$ and $x', y' \in V'$ of degrees, respectively, ξ, η, ξ', η'.

Once again it is easy to see that *the tensor product $\rho \otimes \rho'$ is a grade star representation* with respect to the Hermitian form (2.22). However, even if V and V' are graded Hilbert spaces, the Hermitian form defined in (2.22) is not positive definite and hence Proposition 1 cannot be applied in this case. Nevertheless, the form (2.22) can be useful to reduce a given tensor product of grade star representations.

One might suspect that the definition (2.22) is not appropriate. However, as will be shown in the subsequent paper,[6] there exists an irreducible grade star representation ρ of $\mathrm{spl}(2,1)$ in a Hilbert space V such that $\rho \otimes \rho$ is not completely reducible. In this case there cannot exist any positive definite scalar product on $V \otimes V$ with respect to which $\rho \otimes \rho$ is a grade star representation.

3. GENERALIZED ADJOINT OPERATIONS IN THE CLASSICAL SIMPLE GRADED LIE ALGEBRAS

Let $L = L_0 \oplus L_1$ be a graded Lie algebra with Lie algebra L_0 and odd subspace L_1. We are going to discuss generalized adjoint operations in L. As shown in Sec. 2 there are at least two possible definitions of such an operation (see Definitions 1 and 2). In order to see whether there are other "reasonable" possibilities we shall list four properties which every generalized adjoint operation should have. Of course, the adjoint and

grade adjoint operations as defined in Sec. 2 meet these requirements. It will turn out that for the classical simple graded Lie algebras there are "essentially no other possibilities." Moreover, our discussion will provide the means for the construction of all generalized adjoint operations.

Our four conditions are the following:

(I) A generalized adjoint operation is additive and maps L_i into L_i, $i=0,1$.

According to (I) a generalized adjoint operation is completely determined by its restrictions to L_0 and to L_1.

(II) The restriction of a generalized adjoint operation to L_0 is an adjoint operation of the Lie algebra L_0, i.e., it is a mapping $Q \to Q^*$ of L_0 into itself which satisfies Eqs. (2.1)–(2.3).

(III) The restriction of a generalized adjoint operation to L_1 is a bijective semilinear mapping σ of L_1 into itself, i.e., it satisfies

$$\sigma(aU+bV)=a^*\sigma(U)+b^*\sigma(V), \tag{3.1}$$

for all $U,V\in L_1$ and all complex numbers a,b.

Finally, the adjoint representation of L_0 in L_1 should be compatible (in some sense) with the adjoint operation $Q\to Q^*$ in L_0 and the mapping σ. We shall demand that a generalized adjoint operation acts on the commutator of an even element Q with an odd element U like a normal adjoint operation, i.e.,

(IV) We have

$$\sigma(\langle Q,U\rangle)=-\langle Q^*,\sigma(U)\rangle, \tag{3.2}$$

for all elements $Q\in L_0$ and $U\in L_1$.

As we have mentioned in Sec. 2 the real subalgebra

$$L_0^R=\{Q\in L_0\,|\,Q^*=-Q\} \tag{3.3}$$

is a real form of L_0. The existence of a mapping σ: $L_1\to L_1$ which satisfies the conditions (III) and (IV) is then equivalent to the requirement that *the adjoint representation of L_0 in L_1 should be self-conjugate* (with respect to L_0^R), i.e., that it should be equivalent to its complex conjugate representation (with respect to L_0^R). In particular we conclude that σ^2 *commutes with the adjoint representation of L_0 in L_1.*

In the following we assume that we are given an adjoint operation in L_0 and a mapping σ of L_1 into itself which satisfy the conditions (I)–(IV). To exploit these conditions we construct a new graded Lie algebra $\tilde{L}$, whose Lie algebra and odd subspace are again L_0, resp. L_1, but whose multiplication $\langle\,,\rangle_\sim$ is defined by

$$\begin{aligned}
&\langle P,Q\rangle_\sim=\langle P,Q\rangle \quad \text{if } P,Q\in L_0,\\
&\langle Q,U\rangle_\sim=-\langle U,Q\rangle_\sim=\langle Q,U\rangle \quad \text{if } Q\in L_0,\ U\in L_1,\\
&\langle U,V\rangle_\sim=-\langle\sigma^{-1}(U),\sigma^{-1}(V)\rangle^* \quad \text{if } U,V\in L_1.
\end{aligned} \tag{3.4}$$

It is easy to check that $\tilde{L}$ is indeed a graded Lie algebra. Let us define a mapping

$$\omega: L\to\tilde{L} \tag{3.5a}$$

by

$$\omega(Q+U)=-Q^*+\sigma(U) \quad \text{if } Q\in L_0,\ U\in L_1. \tag{3.5b}$$

Then it is easy to see that ω is a bijective mapping which satisfies

$$\begin{aligned}
&\omega(aA+bB)=a^*\omega(A)+b^*\omega(B),\\
&\omega(\langle A,B\rangle)=\langle\omega(A),\omega(B)\rangle_\sim,
\end{aligned} \tag{3.6}$$

for all $A,B\in L$ and all complex numbers a,b. In particular $\tilde{L}$ *is simple if and only if L is simple.*

From now on *we shall assume that L is a classical simple graded Lie algebra.* (These algebras have been studied in Refs. 1 and 2; a simple graded Lie algebra is called classical if the underlying Lie algebra is reductive.) Disregarding the possibility $L=\Gamma(\sigma_1,\sigma_2,\sigma_3)$ for a moment, we know from Ref. 2 that the product mapping $L_1\times L_1\to L_0$ of a classical simple graded Lie algebra is determined up to a factor once the Lie algebra L_0 and the adjoint representation of L_0 in the odd subspace L_1 are given. By construction both items agree for L and $\tilde{L}$, hence we conclude that there exists a nonzero constant t such that

$$\langle U,V\rangle_\sim=-t\langle U,V\rangle, \tag{3.7}$$

i.e., such that

$$\langle U,V\rangle^*=t\,\langle\sigma(U),\sigma(V)\rangle, \tag{3.8}$$

for all $U,V\in L_1$.

The case $L=\Gamma(\sigma_1,\sigma_2,\sigma_3)$ is more complicated because of the parameters $\sigma_1,\sigma_2,\sigma_3$ and because of the automorphisms of $L_0=\mathrm{sl}(2)\times\mathrm{sl}(2)\times\mathrm{sl}(2)$ which interchange the three Lie algebras sl(2). We shall not go into the details here but state without proof that in this case Eq. (3.8) is valid if (for example) the adjoint operation in $L_0=\mathrm{sl}(2)\times\mathrm{sl}(2)\times\mathrm{sl}(2)$ maps every factor sl(2) into itself and if, furthermore, the parameters $\sigma_1,\sigma_2,\sigma_3$ are real.

The result (3.8) is very agreeable; it says that an adjoint operation in L_0 and a mapping $\sigma: L_1\to L_1$ which satisfy the conditions (I)–(IV) automatically act on the anticommutator of two odd elements as they should.

As a consequence of (3.8) we note that

$$\langle U,V\rangle=|t|^2\langle\sigma^2(U),\sigma^2(V)\rangle, \tag{3.9}$$

for all $U,V\in L_1$.

It is obvious that the mapping σ (if it exists at all) is not completely fixed by the conditions (III) and (IV), since its "normalization" is completely arbitrary up to now. We shall use this remaining freedom to redefine σ in such a way that an adjoint or grade adjoint operation in the sense of Sec. 2 emerges. To this end, recall that L_1 is either irreducible (under L_0) or it decomposes into the direct sum of two irreducible subspaces. We consider both cases separately.

Case (A) L_1 is irreducible

Then σ^2 is a scalar multiple of the identity

$$\sigma^2=s, \tag{3.10}$$

with some nonzero complex number s. In fact s must be *real* since

$$s^2 = \sigma^2\sigma^2 = \sigma\sigma^2\sigma = \sigma s\sigma = s^*\sigma^2 = s^*s. \tag{3.11}$$

The reality of s follows also from the relation

$$s^2|t|^2 = 1 \tag{3.12}$$

which is equivalent to Eq. (3.9).

If $s > 0$ (resp. $s < 0$) the adjoint representation of L_0 in L_1 is called *real* (resp. *quaternionic*).[7]

Evidently the mapping σ is only fixed up to a nonzero factor. Introducing

$$\hat{\sigma} = c\sigma \tag{3.13}$$

with a nonzero complex number c it is easy to see that in the real (resp. quaternionic) case the number c can be choosen in such a way that

$$\langle U, V\rangle^* = \langle\hat{\sigma}(U), \hat{\sigma}(V)\rangle \quad \text{and} \quad \hat{\sigma}^2 = 1$$
$$\text{(resp. } \langle U, V\rangle^* = -\langle\hat{\sigma}(U), \hat{\sigma}(V)\rangle \quad \text{and} \quad \hat{\sigma}^2 = -1), \tag{3.14}$$

for all $U, V \in L_1$. In this way we have obtained an adjoint (resp. grade adjoint) operation in L. Note that the number c and, therefore, the mapping $\hat{\sigma}$ are fixed up to the sign.

Case (B) L_1 decomposes into the direct sum of two irreducible subspaces L_1^1, L_1^2

$$L_1 = L_1^1 \oplus L_1^2; \tag{3.15a}$$

recall that in this case

$$\langle L_1^1, L_1^1\rangle = \langle L_1^2, L_1^2\rangle = \{0\}, \quad \langle L_1^1, L_1^2\rangle = L_0. \tag{3.15b}$$

From our list of classical simple graded Lie algebras[2] we know that the representations ρ_1 and ρ_2 of L_0 induced in L_1^1, resp. L_1^2, are inequivalent except in the case $L = \mathrm{spl}(2,2)/z_2$. This special case is more complicated than the others and will be disregarded. Hence we may suppose that ρ_1 and ρ_2 are inequivalent.

Since we assume the existence of the mapping σ, i.e., since we assume that the adjoint representation of L_0 in L_1 is self-conjugate, there exist the following two possibilities (a) and (b):

(a) *The representations ρ_1 and ρ_2 are self-conjugate.*

In this case it follows that

$$\sigma(L_1^1) = L_1^1, \quad \sigma(L_1^2) = L_1^2. \tag{3.16}$$

Let

$$\sigma_1 : L_1^1 \to L_1^1, \quad \sigma_2 : L_1^2 \to L_1^2 \tag{3.17}$$

be the mappings induced by σ. Because of the irreducibility of ρ_1 and ρ_2 it follows that

$$\sigma_1^2 = s_1, \quad \sigma_2^2 = s_2 \tag{3.18}$$

with some nonzero *real* numbers s_1, s_2, and condition (3.9) is equivalent to

$$s_1 s_2 |t|^2 = 1. \tag{3.19}$$

From this equation we conclude that s_1 and s_2 are either both positive or both negative, i.e., that the representations ρ_1 and ρ_2 are either both real or both quaternionic.

The mappings σ_1 and σ_2 are only fixed up to a nonzero factor. Let us introduce the new mappings

$$\hat{\sigma}_1 = c_1\sigma_1, \quad \hat{\sigma}_2 = c_2\sigma_2 \tag{3.20}$$

with nonzero complex numbers c_1, c_2 and let us combine $\hat{\sigma}_1$ and $\hat{\sigma}_2$ to a mapping $\hat{\sigma}$ of L_1 into itself. Then it is easy to see that we can choose c_1 and c_2 in such a way that in the real (resp. quaternionic) case we have

$$\langle U, V\rangle^* = \langle\hat{\sigma}(U), \hat{\sigma}(V)\rangle \quad \text{and} \quad \hat{\sigma}^2 = 1$$
$$\text{(resp. } \langle U, V\rangle^* = -\langle\hat{\sigma}(U), \hat{\sigma}(V)\rangle \quad \text{and} \quad \hat{\sigma}^2 = -1), \tag{3.21}$$

for all $U, V \in L_1$. In this way we have obtained an adjoint (resp. grade adjoint) operation in L. Note that in this case a free phase factor $e^{i\phi}$ (ϕ real) is left: Our conclusions remain unchanged if we choose $e^{i\phi}c_1$ and $e^{-i\phi}c_2$ instead of c_1, resp. c_2.

(b) *The representation ρ_2 is equivalent to the complex conjugate of ρ_1 and vice versa.*

In this case our assumptions imply

$$\sigma(L_1^1) = L_1^2, \quad \sigma(L_1^2) = L_1^1; \tag{3.22}$$

let

$$\sigma_1 : L_1^1 \to L_1^2, \quad \sigma_2 : L_1^2 \to L_1^1 \tag{3.23}$$

be the mappings induced by σ. Since ρ_1 and ρ_2 are irreducible we conclude that

$$\sigma_2\sigma_1 = s, \quad \sigma_1\sigma_2 = s^* \tag{3.24}$$

with some nonzero complex number s, and Eq. (3.9) is equivalent to

$$|st| = 1. \tag{3.25}$$

Here again the mappings σ_1, σ_2 are only fixed up to a nonzero factor. Introducing the new mappings

$$\hat{\sigma}_1 = c_1\sigma_1, \quad \hat{\sigma}_2 = c_2\sigma_2 \tag{3.26}$$

with some nonzero complex numbers c_1, c_2 and combining $\hat{\sigma}_1$ and $\hat{\sigma}_2$ to a mapping $\hat{\sigma} : L_1 \to L_1$ it is easy to check that the numbers c_1, c_2 can be chosen in such a way that

$$\langle U, V\rangle^* = \pm\langle\hat{\sigma}(U), \hat{\sigma}(V)\rangle \quad \text{and} \quad \hat{\sigma}^2 = \pm 1, \tag{3.27}$$

for all $U, V \in L_1$. Note that *both choices of the sign are possible*, i.e., that we can construct an adjoint operation as well as a grade adjoint operation in L. Furthermore, there is still a free nonzero real parameter d left: Our conclusions remain unchanged if we choose $c_1 \cdot d$ and c_2 / d instead of c_1, resp. c_2.

Our results may be summarized as follows:

If a mapping $\sigma : L_1 \to L_1$ which satisfies the conditions (III) and (IV) exists at all, then we may suppose that

$$\text{(V)} \quad \langle U, V\rangle^* = \pm\langle\sigma(U), \sigma(V)\rangle, \quad \text{for all } U, V \in L_1, \tag{3.28}$$

and, furthermore, that

$$\text{(VI)} \quad \sigma^2 = \pm 1. \tag{3.29}$$

Stated differently, we may then suppose that the adjoint operation in L_0 and the mapping σ define an adjoint/grade adjoint operation of the graded Lie algebra L.

As we have seen, the mapping σ is not uniquely determined even if we demand that the conditions (V) and (VI) should also be satisfied. This is in part obvious: If σ satisfies the conditions (III)–(VI), then $-\sigma$ does as well.

In the case where L_1 is irreducible there is no other

freedom. On the other hand, if L_1 decomposes into the direct sum of two irreducible subspaces

$$L_1 = L_1^1 \oplus L_1^2 \tag{3.30}$$

[see (3.15a) and (3.15b)], then there is a free parameter left.

This latter nonuniqueness is connected with some trivial automorphisms of our algebra. In fact, let g be any nonzero complex number. We define a linear mapping ψ_1 of L_1 into itself by

$$\begin{aligned} &\psi_1(U_1) = gU_1 \quad \text{if} \quad U_1 \in L_1^1, \\ &\psi_1(U_2) = \frac{1}{g} U_2 \quad \text{if} \quad U_2 \in L_1^2. \end{aligned} \tag{3.31}$$

Then it is easy to see that ψ_1 combined with the identity mapping of L_0 onto itself yields an automorphism ψ of the graded Lie algebra L.

Consequently, the mapping

$$\sigma' = \psi_1 \circ \sigma \circ \psi_1^{-1} \tag{3.32}$$

also satisfies our conditions (III)—(VI). Moreover, if we are in Case (B), (a) then the mappings σ_1, σ_2 are changed into

$$\sigma_1' = (g/g^*)\sigma_1, \quad \sigma_2' = (g^*/g)\sigma_2, \tag{3.33}$$

whereas in Case (B), (b) the mappings σ_1, σ_2 are replaced by

$$\sigma_1' = (1/|g|^2)\sigma_1, \quad \sigma_2' = |g|^2 \sigma_2. \tag{3.34}$$

The Eqs. (3.33) and (3.34) define what we call a *one-parameter family* of adjoint or grade adjoint operations. Note that the scalar factors appearing in (3.33) [resp. (3.34)] are *phase factors* (resp. are *positive*). Hence whereas the change of sign of σ can be achieved via (3.33) this is not possible using (3.34).

With the change of sign of the mapping σ and with the procedures described in (3.33) and (3.34) we have reobtained exactly the free parameters which were mentioned in the discussion above. "Normally" the adjoint (grade adjoint) operations belonging to a one-parameter family (i.e., σ and σ') should be equivalent. In Case (B), (a) there is no additional freedom, whereas in Case (B), (b) we are left with the change of sign of the mapping σ. With this understanding we have in Case (A) two adjoint *or* two grade adjoint operations, in Case (B), (a) there is essentially one adjoint *or* one grade adjoint operation, whereas in Case (B), (b) there are essentially two adjoint *and* two grade adjoint operations. We shall come back to this result in Sec. 5.

4. CONSTRUCTION OF ADJOINT AND GRADE ADJOINT OPERATIONS IN THE CLASSICAL SIMPLE GRADED LIE ALGEBRAS

In view of the results derived in Sec. 3 we are now ready to construct all possible adjoint and grade adjoint operations in the classical simple graded Lie algebras (with the exception of the two algebras $\mathrm{spl}(2,2)/z_2$ and, partly, $\Gamma(\sigma_1, \sigma_2, \sigma_3)$, which, however, can be treated separately).[8]

In fact, given any adjoint operation in the Lie algebra L_0 (or, equivalently, the corresponding real form L_0^R of L_0) we have to decide whether the adjoint representation of L_0 in L_1 is self-conjugate with respect to L_0^R or not. Hence all that we need is the following information:

(1) Which are the real forms of the (complex) semisimple Lie algebras?

(2) Given an irreducible representation of such a real form, determine the complex conjugate of the representation.

(3) Suppose that an irreducible representation of a real form is self-conjugate; is the representation real or quaternionic?

The answers to all these questions are well known and are contained in the tables of Ref. 7.

Furthermore, the following additional remark is needed if L_0 has a (necessarily one-dimensional) center, i.e., if L is one of the algebras $\mathrm{spl}(n,m)$ with $n \neq m$ or $\mathrm{osp}(2p,2)$. Recall that in this case L_1 decomposes into two irreducible subspaces

$$L_1 = L_1^1 \oplus L_1^2 \tag{4.1}$$

and that there exists a unique element E in the center of L_0 such that

$$\begin{aligned} &\langle E, U_1 \rangle = U_1 \quad \text{if} \quad U_1 \in L_1^1, \\ &\langle E, U_2 \rangle = -U_2 \quad \text{if} \quad U_2 \in L_1^2. \end{aligned} \tag{4.2}$$

Evidently the adjoint operation in L_0 maps the center of L_0 into itself, i.e., E^* must be proportional to E. As a consequence of Eq. (3.2) it is then easy to see that either $E^* = E$ and we are in Case (B), (b) or else $E^* = -E$ and we are in Case (B), (a).

Assuming that the real form L_0^R is compact we can say even more. In fact, for the algebras $\mathrm{spl}(n,m)$ with $n \neq m$ and $\max(n,m) > 3$ the representations of $\mathrm{sl}(n) \times \mathrm{sl}(m)$ in L_1^1 and L_1^2 are not self-conjugate (but one is the complex conjugate of the other); hence an adjoint or grade adjoint operation on L (extending that on L_0) exists only if we are in Case (B), (b) and if $E^* = E$. On the other hand, for the algebras $\mathrm{osp}(2p,2)$ and for $\mathrm{spl}(2,1)$ [which is isomorphic to $\mathrm{osp}(2,2)$] both definitions $E^* = E$ and $E^* = -E$ are possible.

We shall not pursue the problem of constructing all adjoint (grade adjoint) operations in full generality but restrict our attention to some interesting cases. In the following *we shall assume that the real form L_0^R is compact*. Note that in all other cases it is impossible to find star or grade star representations of L in a *finite-dimensional Hilbert space*.

Now it is easy to prove the following lemma.

Lemma 4.1: *Let L_0 be a complex semisimple Lie algebra and let L_0^R be a compact real form of L_0. A finite-dimensional irreducible representation of L_0 is self-conjugate with respect to L_0^R if and only if it is self-contragredient*. More precisely: *The irreducible representation is real (resp. quaternionic) if and only if it is orthogonal (resp. symplectic).*

Using this lemma as well as the results of Ref. 2, the case where the odd subspace L_1 is irreducible [Case

(A)] can be settled immediately. For the orthosymplectic algebras $\mathrm{osp}(2p,m)$, $p,m\geq 1$, $m\neq 2$, as well as for the two exceptional graded Lie algebras whose Lie algebras are respectively $\mathrm{sl}(2)\times G_2$ and $\mathrm{sl}(2)\times o(7)$ there are two grade adjoint operations. On the other hand, for the (f,d) algebras $d(n)/z_n$, $n\geq 3$, there exist two adjoint operations. Finally, it is easy to see that for the algebras $b(n)$, $n\geq 3$, there exist no adjoint and no grade adjoint operations.

Next we have to consider the case where the odd subspace L_1 decomposes into two irreducible subspaces [Case (B)]. Contrary to the abstract argument given above we shall now use a different approach and construct the adjoint or grade adjoint operations explicitly in matrix notation [this method applies equally well in Case (A)].

Let us first consider the general linear graded Lie algebra $\mathrm{pl}(n,m)$, $n,m\geq 1$. Recall[2] that the elements of $\mathrm{pl}(n,m)$ are the $(n+m)\times(n+m)$ matrices, written in block form

$$X=\begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{4.3}$$

with A an arbitrary $n\times n$ matrix, B an arbitrary $n\times m$ matrix, C an arbitrary $m\times n$ matrix, and D an arbitrary $m\times m$ matrix.

The Lie algebra L_0 of $\mathrm{pl}(n,m)$ consists of the "diagonal" block matrices $\left(\begin{smallmatrix} A & 0 \\ 0 & D \end{smallmatrix}\right)$, and the odd subspace L_1 consists of the "off-diagonal" block matrices $\left(\begin{smallmatrix} 0 & B \\ C & 0 \end{smallmatrix}\right)$. The adjoint operation in L_0 will be choosen to be the normal Hermitian conjugation of a matrix, i.e.,

$$\begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}^{*}=\begin{pmatrix} A^{*} & 0 \\ 0 & D^{*} \end{pmatrix}. \tag{4.4}$$

Let d be any nonzero *real* number. Then

$$\sigma\begin{pmatrix} 0 & B \\ C & 0 \end{pmatrix}=\begin{pmatrix} 0 & dC^{*} \\ d^{-1}B^{*} & 0 \end{pmatrix} \tag{4.5}$$

defines an adjoint operation in $\mathrm{pl}(n,m)$, whereas

$$\sigma\begin{pmatrix} 0 & B \\ C & 0 \end{pmatrix}=\begin{pmatrix} 0 & -dC^{*} \\ d^{-1}B^{*} & 0 \end{pmatrix} \tag{4.6}$$

defines a grade adjoint operation in $\mathrm{pl}(n,m)$. (Here again the plus sign on the right-hand side denotes the normal Hermitian conjugation of a matrix.)

Note that corresponding to the two possible signs of d the Eqs. (4.4) and (4.5) [resp. (4.4) and (4.6)] yield two one-parameter families of adjoint (resp. grade adjoint) operations.[9] In particular, choosing $d=1$ in Eq. (4.5) we obtain the normal Hermitian conjugation in $\mathrm{pl}(n,m)$. On the other hand, the choice $d=\pm 1$ in (4.6) leads to the grade adjoint as defined in Sec. 2; more precisely, if $d=1$ (resp. $d=-1$), then the first n rows and columns (resp. the last m rows and columns) belong to the even subspace.

Let us now discuss the special linear graded Lie algebra $\mathrm{spl}(n,m)$ which is the subalgebra of $\mathrm{pl}(n,m)$ defined by the equation

$$\mathrm{Tr}(A)=\mathrm{Tr}(D). \tag{4.7}$$

By restriction the mappings given in (4.4) and (4.5) [resp. (4.4) and (4.6)] define adjoint (resp. grade adjoint) operations in the algebra $\mathrm{spl}(n,m)$ [and, if $n=m$, by going to the quotient, also in the algebra $\mathrm{spl}(n,n)/z_n$; recall that z_n consists of the scalar multiples of the $2n\times 2n$ unit matrix].

Suppose now that $n=2p$ is even and recall that the orthosymplectic graded Lie algebra is the subalgebra of $\mathrm{pl}(2p,m)$ defined by the equations

$${}^{t}AG+GA=0,\quad {}^{t}D+D=0,\quad C={}^{t}BG, \tag{4.8}$$

where G is the $2p\times 2p$ matrix

$$G=\begin{pmatrix} 0 & I_p \\ -I_p & 0 \end{pmatrix} \tag{4.9}$$

and I_p is the $p\times p$ unit matrix.

Then the adjoint operations (4.4), (4.5) do not map the algebra $\mathrm{osp}(2p,m)$ into itself. On the other hand, the grade adjoint operation (4.4), (4.6) maps the algebra $\mathrm{osp}(2p,m)$ into itself if and only if $d=\pm 1$, thus defining a a grade adjoint operation in $\mathrm{osp}(2p,m)$.

The cases $m=2$ require special attention since in these cases the odd space L_1 decomposes into two irreducible subspaces. One can show that with the definition (4.4) we are in Case (B), (b) described in Sec. 3; there exist two one-parameter families of adjoint operations and two one-parameter families of grade adjoint operations[9] in $\mathrm{osp}(2p,2)$ which satisfy (4.4). On the other hand, for the algebra $\mathrm{osp}(2p,2)$ we could equally well define the adjoint operation on the Lie algebra by

$$\begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}\to\begin{pmatrix} A^{*} & 0 \\ 0 & -D^{*} \end{pmatrix} \tag{4.10}$$

(recall that $E^{**}=\pm E$). With this definition we are in Case (B), (a) and obtain a one-parameter family of grade adjoint operations.

Finally, let $n=m\geq 3$ and consider the subalgebra $\mathrm{d}(n)$ of $\mathrm{pl}(n,n)$ defined by the conditions

$$A=D,\quad B=C,\quad \mathrm{Tr}(B)=0. \tag{4.11}$$

Note that $\mathrm{d}(n)/z_n$ is the (f,d) algebra. Evidently the adjoint operation (4.4), (4.5) maps the algebra $\mathrm{d}(n)$ into itself if and only if $d=\pm 1$, thus defining an adjoint operation of $\mathrm{d}(n)$ and, by going to the quotient, of the (f,d) algebra. On the other hand, the grade adjoint operations (4.4), (4.6) do not map $\mathrm{d}(n)$ into itself.

5. STAR AND GRADE STAR REPRESENTATIONS WHEN L_1 IS DECOMPOSABLE

As we have seen in Secs. 3 and 4, when the odd subspace L_1 decomposes into two irreducible subspaces

$$L_1=L_1^1\oplus L_1^2 \tag{5.1}$$

there is a free parameter in the adjoint, resp. grade adjoint, operations (apart from the freedom in choosing the sign of the mapping σ). Let us assume that we fix the parameter and, therefore, a certain mapping σ. In this way we define a class of star, resp. grade star, representations. For another choice of the parameter (mapping σ') we get another class of representations. It turns out that there is a natural transformation [see Eq. (5.2)] which associates with every representation of the first class a certain representation of the second

class but it is not obvious that all the representations belonging to one class are equivalent to those belonging to the other. We have not studied the problem in full detail but we have shown that for the spl(n,m), $n \neq m$, and osp$(2p,2)$ algebras the original and the transformed representation are indeed equivalent.

In the following discussion we use the notation of Sec. 3. To begin with, we recall that two adjoint, resp. grade adjoint, operations which belong to different values of the free parameter are connected by an automorphism ψ of the algebra L [see Eqs. (3.32)–(3.34)]. Suppose now that ρ is a star, resp. grade star, representation of L with respect to the adjoint, resp. grade adjoint, operation given by the mapping σ and which acts in a vector space V equipped with a nondegenerate Hermitian form $\langle\,|\,\rangle$. Then it is obvious that

$$\rho' = \rho \circ \psi^{-1} \tag{5.2}$$

is a star, resp. grade star, representation of L with respect to the adjoint, resp. grade adjoint, operation given by σ'. In this sense we are allowed to restrict our attention to a special adjoint, resp. grade adjoint, operation out of the one-parameter family.

It will be useful to look at this result from a different viewpoint. We note that it is not obvious *a priori* whether the representations ρ and ρ' are equivalent. Nevertheless, there is a class of algebras for which this is easy to prove. In fact, assume that L is one of the classical simple graded Lie algebras for which L_0 has a (necessarily one-dimensional) center, i.e., one of the algebras spl(n,m) with $n \neq m$ or osp$(2p,2)$. Then L_1 is automatically decomposable and there exists a unique element E in the center of L_0 such that

$$\begin{aligned} &\langle E, U_1\rangle = U_1 \quad \text{if} \quad U_1 \in L_1^1, \\ &\langle E, U_2\rangle = -U_2 \quad \text{if} \quad U_2 \in L_1^2. \end{aligned} \tag{5.3}$$

Recalling that ad$E$ is the linear operator in L defined by

$$(\mathrm{ad}E)(A) = \langle E, A\rangle \quad \text{for all} \quad A \in L \tag{5.4}$$

we see that ad$E$ is the generator of the (complex) one-parameter group of automorphisms ψ. Stated differently, if (with the notation of Sec. 3)

$$g = \exp(h), \tag{5.5}$$

then

$$\psi = \exp(h\mathrm{ad}E). \tag{5.6}$$

Defining the linear operator

$$H = \exp(h\rho(E)) \tag{5.7}$$

in the representation space V, we conclude that

$$\rho'(A) = H^{-1}\rho(A)H, \tag{5.8}$$

for all $A \in L$. Therefore, ρ and ρ' are equivalent.

On the other hand, defining a new Hermitian form $\langle\,|\,\rangle'$ on V by

$$\langle x|y\rangle' = \langle H^{-1}x|H^{-1}y\rangle, \tag{5.9}$$

for all $x, y \in V$, it is easy to check that ρ is a star, resp. grade star representation of L with respect to the adjoint, resp. grade adjoint, operation given by the mapping σ' and with respect to the new Hermitian form $\langle\,|\,\rangle'$.

The results of this section indicate that for the study of star and grade star representations it should "normally" be sufficient to fix the free parameter in the adjoint, resp. grade adjoint, operations arbitrarily.

6. CLASSES OF STAR AND GRADE STAR REPRESENTATIONS

As was shown in Secs. 3 and 4, in a classical simple graded Lie algebra $L = L_0 \oplus L_1$ there exist in general various adjoint and/or grade adjoint operations which all belong to the same adjoint operation in the Lie algebra L_0. To every one of these adjoint (grade adjoint) operations belongs a class of star (grade star) representations of the algebra L.

At this place we have to include a trivial remark on grade star representations. Let $V = V_0 \oplus V_1$ be a graded vector space as in Sec. 2, i.e., equipped with a nondegenerate Hermitian form $\langle\,|\,\rangle$ such that $\langle V_0|V_1\rangle = \{0\}$. We convert V into a new graded vector space V' by demanding that $V_0' = V_1$ and $V_1' = V_0$ should be the even, resp. odd, subspace of V'.

Suppose now that we are given a grade adjoint operation in L, described by a mapping $\sigma: L_1 \to L_1$, and let ρ be a grade star representation of L in V with respect to σ. Then it is obvious that ρ is also a grade star representation of L in V' with respect to the grade adjoint operation in L described by the mapping $-\sigma$.

It depends on the circumstances whether the redefinition of the grading of V is adequate or not. If this redefinition is convenient, then our remark implies that the classes of grade star representations which belong to σ, resp. $-\sigma$, coincide. In the present work we distinguish between these two classes of representations.

In the following we shall always assume that the adjoint operation on L_0 defines a *compact* real form L_0^R of L_0 and, furthermore, that the star and grade star representations act in a graded *Hilbert space*.

For each class of star representations the tensor product of two representations belonging to the class is completely reducible into irreducible representations belonging to the same class. This statement is generally not valid for classes of grade star representations. As is shown in Ref. 6 for example, the osp$(2,1)$ algebra admits two classes of grade star representations for which the tensor product is always completely reducible but the spl$(2,1)$ algebra has grade star representations such that the tensor product is not completely reducible.

Using the results of Secs. 3–5 we can summarize the situation as follows:

(a) The spl(n,m), $n > m \geq 1$, $n \neq 2$, algebras have two classes of star representations and two classes of grade star representations.

(b) For the spl$(n,n)/z_n$, $n \geq 3$, algebras we have constructed two one-parameter families of adjoint operations and two one-parameter families of grade adjoint operations. To all these operations there may cor-

respond a class of star, resp. grade star, representations. Since Sec. 5 does not apply in this case we do not know whether the free parameters are really relevant or not.

(c) The $\mathrm{osp}(2p,m)$, $p,m \geq 1$, $m \neq 2$, algebras have two classes of grade star representations. In the case of the $\mathrm{osp}(2p,2)$, $p \geq 1$, algebras we have two distinct possibilities depending on the definition of the adjoint operation in the Lie algebra (see Sec. 4). For the choice (4.4) one has two classes of star representations and two classes of grade star representations, for the other choice (4.10) one has one class of grade star representations. [Recall that $\mathrm{spl}(2,1)$ is isomorphic to $\mathrm{osp}(2,2)$.]

(d) The $\mathrm{b}(n)$, $n \geq 3$, algebras have neither star nor grade star representations.

(e) The $\mathrm{d}(n)/z_n$, $n \geq 3$, algebras have two classes of star representations.

(f) The exceptional simple graded Lie algebras whose underlying Lie algebras are $\mathrm{sl}(2)\times G_2$ and $\mathrm{sl}(2)\times \mathrm{o}(7)$, respectively, have two classes of grade star representations.

We have left out from our study the $\mathrm{spl}(2,2)/z_2$ and the $\Gamma(\sigma_1,\sigma_2,\sigma_3)$ algebras (see Sec. 3 for the reasons). Let us stress that our conditions for the existence of the various classes of representations are only necessary. It may turn out that some of them are actually empty or that two different classes contain equivalent representations.

[1]V.G. Kac, Funct. Anal. Appl. 9, No. 3, 91 (1975).
[2]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. 17, 1626, 1640 (1976).
[3]W. Nahm, V. Rittenberg, and M. Scheunert, Phys. Lett. B 61, 383 (1976).
[4]P.G.O. Freund and I. Kaplansky, J. Math. Phys. 17, 228 (1976).
[5]G. Hochschild, III, J. Math. 20, 107 (1976); D. Ž. Djoković and G. Hochschild, III, J. Math. 20, 134 (1976); D. Ž. Djoković, J. Pure Appl. Alg. 7, 217 (1976).
[6]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. 18, 155 (1977).
[7]J. Tits, "Tabellen zu den einfachen Liegruppen und ihren Darstellungen," Lecture Notes in Mathematics 40, (Springer, Berlin, 1967).
[8]Note the relevance of our results for the construction of real forms of the classical simple graded Lie algebras.
[9]See the discussion at the end of Sec. 3.

Irreducible representations of the osp(2,1) and spl(2,1) graded Lie algebras

M. Scheunert, W. Nahm, and V. Rittenberg

Physikalisches Institut der Universität Bonn, Bonn, Germany
(Received 30 August 1976)

We illustrate through the examples of the osp(2,1) and spl(2,1) algebras the differences between the properties of the irreducible representations of simple graded Lie algebras and simple Lie algebras.

1. INTRODUCTION

In this paper we present a detailed study of the finite-dimensional irreducible representations of the osp(2,1) and spl(2,1) graded Lie algebras. These algebras are strictly simple: They are simple and have a nondegenerate Killing form.[1] We have choosen to study these algebras because they have a small number of generators and because the underlying Lie algebras are sl(2) (call it isospin), respectively sl(2)×gl(1) (call it isospin and baryon number), which have many physical applications and are well known. These examples will also illustrate the importance of star and grade star representations defined in the preceding paper.[2] As will be shown, the properties of the irreducible representations are quite different from those of ordinary simple Lie algebras.

The Casimir operators are defined as in the case of ordinary semisimple Lie algebras. Let X_μ be the generators of our graded Lie algebra, given in some matrix representation. If the commutation relations are

$$\langle X_\mu, X_\nu\rangle = C_{\mu\nu}^{\omega} X_\omega, \tag{1.1}$$

where

$$\langle X_\mu, X_\nu\rangle = X_\mu X_\nu - (-1)^{mn} X_\nu X_\mu, \tag{1.2}$$

and where $m, n \in \{0,1\}$ is the degree of X_μ, resp. X_ν, then one can define a trace metric form[3]

$$g_{\mu\nu} = \mathrm{Tr}(\gamma X_\mu X_\nu) \tag{1.3}$$

(γ is a diagonal matrix with eigenvalues $+1$, resp. -1, in the even, resp. odd, subspace).

The metric form is nondegenerate in our examples. It is easy to show[4-6] that

$$K_n = g_{\sigma_1\sigma_2\cdots\sigma_n} X^{\sigma_n}\cdots X^{\sigma_2} X^{\sigma_1}, \tag{1.4}$$

where

$$g_{\sigma_1\sigma_2\cdots\sigma_n} = \mathrm{Tr}(\gamma X_{\sigma_1} X_{\sigma_2}\cdots X_{\sigma_n}), \quad X^\sigma = g^{\sigma\tau} X_\tau \tag{1.5}$$

are Casimir operators, i.e.,

$$[K_n, X_\mu] = 0. \tag{1.6}$$

In the examples we study, the number of independent Casimir operators is the same as that of the underlying Lie algebra: One in the case of osp(2,1) and two in the case of spl(2,1).

The osp(2,1) algebra was studied in some detail in Ref. 5 where it was shown that the Casimir operator uniquely determines the irreducible representations and that we have complete reducibility. We shall show that the irreducible representations are equivalent to grade star representations. This observation has an interesting application when one computes the Kronecker product of two irreducible representations. If the state vectors for each representation separately form a Hilbert space, when one decomposes their product into irreducible representations, the state vectors of the various irreducible subrepresentations are not orthogonal. One can, however, define a bilinear form which is invariant with respect to the product of the given representations, and the state vectors are orthogonal with respect to this form. In this way we give explicit expressions for the Clebsch–Gordan coefficients which are fixed up to an overall sign for each irreducible subrepresentation.

We next study the irreducible representations of the spl(2,1) algebra. It turns out that the eigenvalues of the two Casimir operators do not always specify the irreducible representations. This happens when the eigenvalues of the Casimir operators are both zero. In this case an interesting relation between the "isospin" and the "baryon number" content of the irreducible representation appears.

In the case of the spl(2,1) algebra we do not have complete reducibility, as it shown by examples. We show, however, that one can define two classes of star representations[2] and inside each class one has indeed complete reducibility. The Clebsch–Gordan series for each class are also given.

We think that the examples presented give an insight into the basic differences between the properties of the irreducible representations of simple Lie algebras and simple graded Lie algebras.

2. IRREDUCIBLE REPRESENTATIONS OF THE osp(2, 1) ALGEBRA

Let us first recall that all finite-dimensional representations of osp(2,1) are completely reducible[5,7] and that the irreducible representations of osp(2,1) have been constructed in Ref. 5. In this section we shall give a more detailed discussion of these representations.

A. Definition of the irreducible representations

The even part of the osp(2,1) algebra is sl(2). We denote by Q_m, $m = 1,2,3$ the usual generators of sl(2) and call them isospin. The odd generators $V_\pm$ are sl(2) spinors. Defining as usual $Q_\pm = Q_1 \pm iQ_2$, the commutation relations of our algebra read as follows:

$$[Q_3, Q_\pm] = \pm Q_\pm, \quad [Q_+, Q_-] = 2Q_3, \tag{2.1}$$

$$[Q_3, V_\pm] = \pm\tfrac{1}{2} V_\pm, \quad [Q_\pm, V_\pm] = 0, \quad [Q_\pm, V_\mp] = V_\pm, \tag{2.2}$$

$$\{V_+, V_+\} = \pm\tfrac{1}{2}Q_+, \quad \{V_+, V_-\} = -\tfrac{1}{2}Q_3. \tag{2.3}$$

The irreducible representations of this algebra are characterized by a quantum number q which takes the values $q = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$. The eigenvalues of the Casimir operator

$$K_2 = Q_m Q_m + V_+ V_- - V_- V_+ \tag{2.4}$$

are $q(q+\frac{1}{2})$.

The representation corresponding to $q=0$ is the trivial one-dimensional representation. Suppose now that $q \geq \frac{1}{2}$. Then the q representation contains two isospin multiplets which belong to isospin q, resp. $q-\frac{1}{2}$, and which are denoted by $|q,q,q_3\rangle$, resp. $|q,q-\frac{1}{2},q_3\rangle$; the first quantum number characterizes the representation and the second and third quantum numbers give the isospin and its three component. With a convenient normalization of the state vectors the q representation is defined as follows

$$\begin{aligned}
&Q_3|q,q,q_3\rangle = q_3|q,q,q_3\rangle,\\
&Q_3|q,q-\tfrac{1}{2},q_3\rangle = q_3|q,q-\tfrac{1}{2},q_3\rangle,\\
&Q_\pm|q,q,q_3\rangle = \sqrt{(q\mp q_3)(q\pm q_3+1)}\,|q,q,q_3\pm 1\rangle,\\
&Q_\pm|q,q-\tfrac{1}{2},q_3\rangle = \sqrt{(q-1/2\mp q_3)(q+1/2\pm q_3)}\\
&\qquad\qquad \times|q,q-\tfrac{1}{2},q_3\pm 1\rangle,\\
&V_\pm|q,q,q_3\rangle = \mp\tfrac{1}{2}\sqrt{q\mp q_3}\,|q,q-\tfrac{1}{2},q_3\pm\tfrac{1}{2}\rangle,\\
&V_\pm|q,q-\tfrac{1}{2},q_3\rangle = -\tfrac{1}{2}\sqrt{q+1/2\pm q_3}\,|q,q,q_3\pm\tfrac{1}{2}\rangle.
\end{aligned} \tag{2.5}$$

To simplify the notation we use the same symbol for an element of our algebra and for the linear operator representing it.

In the following it will be necessary to know which subspace of the representation space is defined to be even or odd. In the present case one of the isospin multiplets generates the even subspace and the other generates the odd subspace; however, it is completely up to our choice which of these multiplets is considered to be even, resp. odd. If the $|q,q,q_3\rangle$ multiplet has degree $\lambda \in \{0,1\}$, then the $|q,q-\frac{1}{2},q_3\rangle$ multiplet has degree $\lambda+1$ and the states will be denoted by

$$|q,q,q_3,\lambda\rangle, \quad \text{resp.} \quad |q,q-\tfrac{1}{2},q_3,\lambda+1\rangle. \tag{2.6}$$

[Recall that an even (resp. odd) vector is said to have degree 0 (resp. 1) and that degrees are added and multiplied modulo 2.]

B. Grade star representations

We shall now discuss whether the q representation of osp(2,1) may be considered as a star or grade star representation as explained in Ref. 2. Since we shall require that the generators Q_m are represented by Hermitian operators the adjoint operation in sl(2) should satisfy

$$Q_m^+ = Q_m. \tag{2.7}$$

As shown in Ref. 2 there is no (normal) adjoint operation in osp(2,1) which is consistent with (2.7). However, (2.7) may be extended to a grade adjoint operation of osp(2,1) in just two different ways, namely[8]

$$V_+^\dagger = \mp V_-, \quad V_-^\dagger = \pm V_+. \tag{2.8}$$

Suppose now that $q \geq \frac{1}{2}$ and that the $|q,q,q_3\rangle$ multiplet has degree λ. Let us assume that the q representation is a grade star representation with respect to a non-degenerate Hermitian form (denoted by a bracket $\langle\,|\,\rangle$). Then (2.7) implies that the two isospin multiplets are orthogonal and that the restriction of $\langle\,|\,\rangle$ to an isospin multiplet is proportional to the well-known positive definite scalar product. The latter statement means that there exist real nonzero constants g,h such that

$$\begin{aligned}
&\langle q,q,q_3,\lambda|q,q,q_3',\lambda\rangle = g\delta_{q_3q_3'},\\
&\langle q,q-\tfrac{1}{2},q_3,\lambda+1|q,q-\tfrac{1}{2},q_3',\lambda+1\rangle = h\delta_{q_3q_3'}.
\end{aligned} \tag{2.9}$$

It follows that the conditions (2.8) are fulfilled if and only if

$$h = \pm(-1)^\lambda g. \tag{2.10}$$

We conclude that every q representation is a grade star representation with respect to a suitable *positive definite* scalar product on the representation space provided that the sign in (2.8) and the degree of the isospin multiplets have been choosen such that

$$\pm(-1)^\lambda = 1. \tag{2.11}$$

If this condition is fulfilled, then our state vectors are orthogonal and of equal length with respect to the appropriate scalar product.

C. Tensor products of irreducible representations

In this subsection we shall give the Clebsch–Gordan coefficients for the tensor product of two irreducible representations of osp(2,1). To begin with we recall that an odd generator V acts on the product of an even/odd state $|f\rangle$ of the first representation with an arbitrary state $|g\rangle$ of the second representation like

$$V(|f\rangle\otimes|g\rangle) = (V|f\rangle)\otimes|g\rangle \pm |f\rangle\otimes(V|g\rangle). \tag{2.12}$$

Hence we have to specify once again which subspace of our irreducible representation is even and which is odd.

We shall consider the tensor product of a q representation with a q' representation. From Ref. 5 we know that this tensor product decomposes into the direct sum of the p representations with

$$p \in \{q+q', q+q'-\tfrac{1}{2}, q+q'-1, \ldots, |q-q'|\}. \tag{2.13}$$

Using the well-known recoupling techniques for product representations of sl(2) it is straightforward to compute the Clebsch–Gordan coefficients.

To formulate our results we shall first introduce some notations. Let us consider the two isospin multiplets $|q,q-\frac{1}{2},q_3,\lambda+1\rangle$ and $|q',q'-\frac{1}{2},q_3',\lambda'+1\rangle$. The usual coupling of these multiplets to a state with isospin q'' and three component q_3'' will be denoted by

$$\begin{aligned}
&|q,q';q-\tfrac{1}{2},q'-\tfrac{1}{2};q'',q_3'',\lambda+\lambda'\rangle\\
&\quad = (q-\tfrac{1}{2},q'-\tfrac{1}{2},q_3,q_3'|q'',q_3'')\,|q,q-\tfrac{1}{2},q_3,\lambda+1\rangle\\
&\qquad\otimes|q',q'-\tfrac{1}{2},q_3',\lambda'+1\rangle,
\end{aligned} \tag{2.14}$$

where $(q-\frac{1}{2},q'-\frac{1}{2},q_3,q_3'|q'',q_3'')$ is a Clebsch–Gordan

coefficient of sl(2). The other possibilities are treated similarly: The first and second quantum numbers give the representations of osp(2,1), whose tensor product is considered, the third and fourth quantum numbers denote the isospin of the two isospin multiplets which are coupled to a state with isospin q'' and three component q_3'', finally in the last place we give the degree of the product state (in our example this degree is equal to $\lambda+1+\lambda'+1=\lambda+\lambda' \bmod 2$).

Furthermore, we denote the states of the p subrepresentation of our tensor product by $|q, q'; p, p', p_3', \mu\rangle$. The first and second quantum numbers indicate the representations whose tensor product is considered; the remaining four quantum numbers specify the state: It belongs to the p subrepresentation, its isospin and three component are respectively p' and p_3', and its degree is equal to μ.

The osp(2,1) Clebsch-Gordan coefficients are then given by the following formulas.

Suppose first that

$$p \in \{q+q',\ q+q'-1,\ \ldots,\ |q-q'|\}. \tag{2.15a}$$

Then

$$\begin{aligned}|q, q'; p, p, p_3, \lambda+\lambda'\rangle &= \frac{1}{\sqrt{2p+1}}\{\sqrt{p+q+q'+1}\,|q, q'; q, q'; p, p_3, \lambda+\lambda'\rangle \\ &\quad -(-1)^{\lambda}\sqrt{q+q'-p}\,|q, q'; q-\tfrac{1}{2}, q'-\tfrac{1}{2}; p, p_3, \lambda+\lambda'\rangle\},\end{aligned} \tag{2.15b}$$

$$\begin{aligned}|q, q'; p, p-\tfrac{1}{2}, p_3, \lambda+\lambda'+1\rangle &= (-1)^{\lambda}\frac{1}{\sqrt{2p}}\{\sqrt{p+q'-q}\,|q, q'; q, q'-\tfrac{1}{2}; p-\tfrac{1}{2}, \\ &\quad p_3, \lambda+\lambda'+1\rangle + (-1)^{\lambda}\sqrt{p+q-q'} \\ &\quad \times |q, q'; q-\tfrac{1}{2}, q'; p-\tfrac{1}{2}, p_3, \lambda+\lambda'+1\rangle\}.\end{aligned} \tag{2.15c}$$

Of course, if $q=q'$ and $p=0$, Eq. (2.15c) has to be omitted.

On the other hand, if

$$p \in \{q+q'-\tfrac{1}{2},\ q+q'-\tfrac{3}{2},\ \ldots,\ |q-q'|+\tfrac{1}{2}\}, \tag{2.16a}$$

then

$$\begin{aligned}|q, q'; p, p, p_3, \lambda+\lambda'+1\rangle &= \frac{1}{\sqrt{2p+1}}\{\sqrt{p+q-q'+1/2}\,|q, q'; q, q'-\tfrac{1}{2}; \\ &\quad p, p_3, \lambda+\lambda'+1\rangle - (-1)^{\lambda}\sqrt{p+q'-q+1/2} \\ &\quad \times |q, q'; q-\tfrac{1}{2}, q'; p, p_3, \lambda+\lambda'+1\rangle\},\end{aligned} \tag{2.16b}$$

$$\begin{aligned}|q, q'; p, p-\tfrac{1}{2}, p_3, \lambda+\lambda'\rangle &= -(-1)^{\lambda}\frac{1}{\sqrt{2p}}\{\sqrt{q+q'-p+1/2}\,|q, q'; q, q'; \\ &\quad p-\tfrac{1}{2}, p_3, \lambda+\lambda'\rangle - (-1)^{\lambda}\sqrt{p+q+q'+1/2} \\ &\quad \times |q, q'; q-\tfrac{1}{2}, q'-\tfrac{1}{2}; p-\tfrac{1}{2}, p_3, \lambda+\lambda'\rangle\}.\end{aligned} \tag{2.16c}$$

In the next subsection we shall comment on the normalization of the states $|q, q'; p, p', p_3', \mu\rangle$.

D. Invariant bilinear forms for the q representations

From Eqs. (2.15) we conclude that the tensor product of a q representation with itself contains the trivial one-dimensional representation. Hence it is evident that every q representation admits an even (with respect to the grading) invariant nondegenerate bilinear form.[3] It is easy to construct this form directly; let us denote it by $(\,|\,)$. From sl(2)-invariance we derive that the two isospin multiplets are orthogonal with respect to $(\,|\,)$ and that

$$\begin{aligned}&(q, q, q_3, \lambda\,|\,q, q, q_3', \lambda) = a(-1)^{q-q_3}\delta_{q_3,-q_3'}, \\ &(q, q-\tfrac{1}{2}, q_3, \lambda+1\,|\,q, q-\tfrac{1}{2}, q_3', \lambda+1) = b(-1)^{q-1/2-q_3}\delta_{q_3,-q_3'},\end{aligned} \tag{2.17}$$

with some nonzero constants a, b. It turns out that this form is osp(2,1)-invariant if and only if

$$b=(-1)^{\lambda}a. \tag{2.18}$$

In the following we shall choose

$$a=1, \quad b=(-1)^{\lambda}. \tag{2.19}$$

We shall now make use of the following general remark.

Suppose we are given any graded Lie algebra L and two graded representations of L in some graded vector spaces V_1 and V_2. Assume, furthermore, that there exists an L-invariant bilinear form ϕ_i on V_i, $i=1,2$. We define a bilinear form ϕ on $V_1 \otimes V_2$ by

$$\phi(x_1 \otimes x_2,\ y_1 \otimes y_2) = (-1)^{\xi_2\eta_1}\phi_1(x_1,\ y_1)\,\phi_2(x_2,\ y_2) \tag{2.20}$$

if $x_1, y_1 \in V_1$ and $x_2, y_2 \in V_2$ are homogeneous (i.e., even or odd) elements of degrees, respectively, ξ_1, η_1, ξ_2, η_2. Then it is easy to check that this "product form" ϕ is invariant with respect to the tensor product of the given representations.

Let us apply this remark to the tensor products studied in subsection C. As expected, the p subrepresentations are orthogonal with respect to the product form.

Furthermore, the product form induces on every p subrepresentation the bilinear form defined by Eqs. (2.17) and (2.19) apart, possibly, from a sign. By this latter property our Clebsch–Gordan coefficients are fixed up to an overall sign for each p subrepresentation (provided we demand that they should be real).

3. IRREDUCIBLE REPRESENTATIONS OF THE spl(2, 1) ALGEBRA

A. Construction of the irreducible representations

The even part of spl(2,1) is sl(2)×gl(1). We denote by Q_m, $m=1,2,3$ the usual generators of sl(2) (call it isospin) and by B the generator of gl(1) (call it baryon number).

The odd generators $V_\pm$, resp. $W_\pm$, carry baryon number $+\frac{1}{2}$, resp. $-\frac{1}{2}$, and are sl(2) spinors.

If we define

$$U_1=V_+, \quad U_2=V_-, \quad U_3=W_+, \quad U_4=W_-, \tag{3.1}$$

the commutation relations of our algebra read as follows:

$$[Q_m, Q_n] = i\epsilon_{mnp}Q_p, \quad [Q_m, B] = 0,$$
$$[Q_m, U_\alpha] = \tfrac{1}{2}\hat{\tau}^m_{\beta\alpha}U_\beta, \quad [B, U_\alpha] = \tfrac{1}{2}\hat{\epsilon}_{\beta\alpha}U_\beta, \tag{3.2}$$
$$\{U_\alpha, U_\beta\} = (\hat{C}\hat{\tau}^m)_{\alpha\beta}Q_m - (\hat{C}\hat{\epsilon})_{\alpha\beta}B,$$

where the 4×4 matrices $\hat{\tau}^m$, $\hat{C}$, $\hat{\epsilon}$ are defined by

$$\hat{\tau}^m = \begin{pmatrix} \tau^m & 0 \\ 0 & \tau^m \end{pmatrix}, \quad \hat{C} = \begin{pmatrix} 0 & C \\ C & 0 \end{pmatrix}, \quad \hat{\epsilon} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.3}$$

τ^m are the usual Pauli matrices and $C = i\tau^2$ is the charge conjugation matrix

$$\tau^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{3.4}$$

One can show that the algebra spl(2,1) has just two Casimir operators K_2 and K_3,

$$[K_i, B] = [K_i, Q_m] = [K_i, U_\alpha] = 0, \tag{3.5}$$

where

$$i = 2,3, \quad m = 1,2,3, \quad \alpha = 1,2,3,4,$$

and their expression is [see (1.4)]

$$K_2 = \mathbf{Q}^2 - B^2 + \tfrac{1}{2}U\hat{C}U,$$
$$K_3 = BK_2 + \tfrac{1}{2}BU\hat{C}U + \tfrac{1}{6}U\mathbf{Q}\hat{\epsilon}\hat{\tau}\hat{C}U + \tfrac{1}{12}U\hat{\epsilon}\hat{\tau}\hat{C}U\mathbf{Q}. \tag{3.6}$$

In the following we shall classify the states of a representation according to their isospin and baryon number. Therefore, we shall first rewrite the commutation relations (3.2) appropriately. Introducing as usual

$$Q_\pm = Q_1 \pm iQ_2 \tag{3.7}$$

we obtain

$$[Q_3, Q_\pm] = \pm Q_\pm, \quad [Q_+, Q_-] = 2Q_3,$$
$$[B, Q_\pm] = [B, Q_3] = 0, \tag{3.8}$$

$$[B, V_\pm] = \tfrac{1}{2}V_\pm, \quad [B, W_\pm] = -\tfrac{1}{2}W_\pm,$$
$$[Q_3, V_\pm] = \pm\tfrac{1}{2}V_\pm, \quad [Q_3, W_\pm] = \pm\tfrac{1}{2}W_\pm,$$
$$[Q_\pm, V_\mp] = V_\pm, \quad [Q_\pm, W_\mp] = W_\pm, \tag{3.9}$$
$$[Q_\pm, V_\pm] = [Q_\pm, W_\pm] = 0,$$

$$\{V_\pm, V_\pm\} = \{V_\pm, V_\mp\} = 0,$$
$$\{W_\pm, W_\pm\} = \{W_\pm, W_\mp\} = 0,$$
$$\{V_\pm, W_\pm\} = \pm Q_\pm, \tag{3.10}$$
$$\{V_\pm, W_\mp\} = -Q_3 \pm B.$$

From the commutation relations it is evident that the substitution

$$Q_i \to Q_i, \quad B \to -B, \quad V_\pm \to W_\pm, \quad W_\pm \to V_\pm \tag{3.11}$$

defines an automorphism of our algebra.

Let us now construct all finite-dimensional irreducible representations of spl(2,1).

Suppose we are given any irreducible representation of our algebra, acting in some vector space V. For simplicity we denote an element of our algebra and its representative (which is a linear operator in V) by the same symbol. As is well known we can decompose V according to isospin; let q be the maximal isospin contained in V and define

$$V' = \{\phi \in V \mid \mathbf{Q}^2\phi = q(q+1)\phi, \ Q_3\phi = q\phi\}. \tag{3.12}$$

Evidently B maps V' into itself, hence V' contains an eigenvector ϕ_0 of B; let us denote the corresponding eigenvalue by b. Then

$$\mathbf{Q}^2\phi_0 = q(q+1)\phi_0, \quad Q_3\phi_0 = q\phi_0, \quad B\phi_0 = b\phi_0. \tag{3.13}$$

Note that $q = 0, \frac{1}{2}, 1, \cdots$, but that b may be any complex number.

Since our representation is supposed to be irreducible, the vector ϕ_0 must be cyclic, i.e., applying arbitrary polynomials in the generators to ϕ_0 we must generate the whole space V.

According to the choice of q we have

$$V_+\phi_0 = W_+\phi_0 = Q_+\phi_0 = 0, \tag{3.14}$$

furthermore, the commutation relations (3.10) yield

$$V_-^2 = W_-^2 = 0, \quad \{V_-, W_-\} = -Q_-. \tag{3.15}$$

Using the commutation relations once again it is now easy to see that the vectors

$$Q_-^m\phi_0, \quad Q_-^m V_-\phi_0, \quad Q_-^m W_-\phi_0, \quad Q_-^m W_-V_-\phi_0, \tag{3.16}$$
$$\text{with integers } m \geq 0$$

generate the representation space V.

As a consequence of (3.14) one can prove that

$$Q_+V_-W_-\phi_0 = -(q-b)\phi_0, \quad Q_+W_-V_-\phi_0 = -(q+b)\phi_0, \tag{3.17}$$

whereas

$$Q_+Q_-\phi_0 = 2q\phi_0.$$

It is now obvious that our representation contains a multiplet with isospin q and baryon number b, at most one multiplet with isospin $q-\frac{1}{2}$ and baryon number $b+\frac{1}{2}$, at most one multiplet with isospin $q-\frac{1}{2}$ and baryon number $b-\frac{1}{2}$, and at most one multiplet with isospin $q-1$ and baryon number b. The states which belong to these quantum numbers as well as to the eigenvalue q_3 of Q_3 will be denoted by

$$|b, q, q_3\rangle, \quad |b+\tfrac{1}{2}, q-\tfrac{1}{2}, q_3\rangle,$$
$$|b-\tfrac{1}{2}, q-\tfrac{1}{2}, q_3\rangle, \quad |b, q-1, q_3\rangle, \tag{3.18}$$

respectively.

One can now use the commutation relations (3.8)–(3.10) to construct the representation explicitly. To facilitate the discussion we shall employ the Wigner–Eckart theorem for sl(2). To do this we assume that our states are normalized according to the usual conventions, e.g.,

$$Q_3|b, q, q_3\rangle = q_3|b, q, q_3\rangle,$$
$$Q_\pm|b, q, q_3\rangle = \sqrt{(q \mp q_3)(q \pm q_3 + 1)}\,|b, q, q_3 \pm 1\rangle. \tag{3.19}$$

Then the Wigner–Eckart theorem yields

$$
\begin{aligned}
&V_\pm|b,q,q_3\rangle = \pm\alpha\sqrt{q\mp q_3}\,|b+\tfrac{1}{2},\ q-\tfrac{1}{2},\ q_3\pm\tfrac{1}{2}\rangle,\\
&W_\pm|b,q,q_3\rangle = \pm\beta\sqrt{q\mp q_3}\,|b-\tfrac{1}{2},\ q-\tfrac{1}{2},\ q_3\pm\tfrac{1}{2}\rangle,\\
&V_\pm|b+\tfrac{1}{2},\ q-\tfrac{1}{2},\ q_3\rangle = 0,\\
&W_\pm|b+\tfrac{1}{2},\ q-\tfrac{1}{2},\ q_3\rangle = \gamma\sqrt{q\pm q_3+1/2}\,|b,q,q_3\pm\tfrac{1}{2}\rangle\\
&\qquad\qquad \pm\delta\sqrt{q\mp q_3-1/2}\,|b,\ q-1,\ q_3\pm\tfrac{1}{2}\rangle,\\
&V_\pm|b-\tfrac{1}{2},\ q-\tfrac{1}{2},\ q_3\rangle = \epsilon\sqrt{q\pm q_3+1/2}\,|b,q,q_3\pm\tfrac{1}{2}\rangle\\
&\qquad\qquad \pm\zeta\sqrt{q\mp q_3-1/2}\,|b,q-1,q_3\pm\tfrac{1}{2}\rangle,\\
&W_\pm|b-\tfrac{1}{2},\ q-\tfrac{1}{2},\ q_3\rangle = 0,\\
&V_\pm|b,\ q-1,\ q_3\rangle = \tau\sqrt{q\pm q_3}\,|b+\tfrac{1}{2},\ q-\tfrac{1}{2},\ q_3\pm\tfrac{1}{2}\rangle,\\
&W_\pm|b,\ q-1,\ q_3\rangle = \omega\sqrt{q\pm q_3}\,|b-\tfrac{1}{2},\ q-\tfrac{1}{2},\ q_3\pm\tfrac{1}{2}\rangle,
\end{aligned}
\tag{3.20}
$$

where $\alpha,\beta,\ldots,\tau,\omega$ are some numbers which are independent of q_3 but which may depend on b and q.

The commutation relations (3.8) and (3.9) are then satisfied. Let us now first consider some degenerate cases.

(a) Suppose that

$$V_-\phi_0 = W_-\phi_0 = 0. \tag{3.21}$$

Then $Q_-\phi_0 = 0$ and, therefore, $q=0$. Conversely, if $q=0$, then our representation is the trivial one-dimensional representation of spl(2,1).

In the following we may assume, therefore, that $q\neq 0$.

(b) Suppose that

$$V_-\phi_0 \neq 0, \quad W_-\phi_0 = 0. \tag{3.22}$$

Then our representation contains the multiplets $|b,q,q_3\rangle$ and $|b+\frac{1}{2},q-\frac{1}{2},q_3\rangle$ but not the multiplets $|b-\frac{1}{2},q-\frac{1}{2},q_3\rangle$ and $|b,q-1,q_3\rangle$. With the appropriate definitions

$$
\begin{aligned}
&|b-\tfrac{1}{2},q-\tfrac{1}{2},q_3\rangle = |b,q-1,q_3\rangle = 0,\\
&\beta=\delta=\epsilon=\zeta=\tau=\omega=0
\end{aligned}
\tag{3.23}
$$

it is easy to see that (3.20) defines a representation of spl(2,1) if and only if

$$\alpha\gamma = 1, \quad b = q. \tag{3.24}$$

The equation $b=q$ may also be read off from (3.17). Note that the remaining free parameter (choose for example α) reflects the fact that the relative normalization of the two isospin multiplets is not fixed. Representations with different choices of α are equivalent.

This representation will be called the (q,q) representation; it is readily shown to be irreducible of dimension $4q+1$. However, we would like to stress that both Casimir operators K_2 and K_3 are zero in this representation. *Thus the eigenvalues of the Casimir operators do not specify the irreducible representations.*

(c) Suppose similarly that

$$V_-\phi_0 = 0, \quad W_-\phi_0 \neq 0. \tag{3.25}$$

Then our representation contains the multiplets $|b,q,q_3\rangle$ and $|b-\frac{1}{2},q-\frac{1}{2},q_3\rangle$ but not the multiplets $|b+\frac{1}{2},q-\frac{1}{2},q_3\rangle$ and $|b,q-1,q_3\rangle$. With the appropriate definitions

$$
\begin{aligned}
&|b+\tfrac{1}{2},q-\tfrac{1}{2},q_3\rangle = |b,q-1,q_3\rangle = 0,\\
&\alpha=\gamma=\delta=\zeta=\tau=\omega=0,
\end{aligned}
\tag{3.26}
$$

it is easy to see that (3.20) defines a representation of spl(2,1) if and only if

$$\beta\epsilon = 1, \quad b = -q. \tag{3.27}$$

The equation $b=-q$ may also be obtained from (3.17). Note that the remaining free parameter (choose for example β) reflects the fact that the relative normalization of the two isospin multiplets is not fixed. Representations with different choices of β are equivalent.

This representation will be called the $(-q,q)$ representation; it is irreducible of dimension $4q+1$ and (once again) the Casimir operators K_2 and K_3 are zero.

(d) Suppose now that

$$V_-\phi_0 \neq 0, \quad W_-\phi_0 \neq 0, \tag{3.28}$$

but that $W_-V_-\phi_0$ or $V_-W_-\phi_0$ (and, therefore, $W_-V_-\phi_0$ and $V_-W_-\phi_0$) are scalar multiples of $Q_-\phi_0$. This latter condition is fulfilled if and only if the multiplet $|b,q-1,q_3\rangle$ is not contained in our representation. Using (3.15) it is easy to see that this is the case if and only if $q=\frac{1}{2}$. With the appropriate definitions,

$$|b,q-1,q_3\rangle = 0, \quad \delta=\zeta=\tau=\omega=0, \tag{3.29}$$

we derive that (3.20) defines a representation of spl(2,1) if and only if

$$\alpha\gamma+\beta\epsilon = 1, \quad \alpha\gamma-\beta\epsilon = 2b, \tag{3.30}$$

i.e., if and only if

$$\alpha\gamma = \tfrac{1}{2}+b, \quad \beta\epsilon = \tfrac{1}{2}-b. \tag{3.31}$$

Note that in this case we have two free (nonzero) parameters α and β which are due to the freedom in normalizing the three isospin multiplets independently. Representations with different α,β are equivalent. Since these representations fit quite well into the general case we postpone their discussion to that place.

(e) We are now ready to discuss the general case in which our representation contains all four isospin multiplets. As we have seen we may suppose $q\geq 1$. Then Eqs. (3.20) define a representation of spl(2,1) if and only if

$$\alpha\epsilon+\zeta\tau = 0, \tag{3.32}$$

$$\beta\gamma+\delta\omega = 0, \tag{3.33}$$

$$
\begin{aligned}
&\alpha\gamma+\beta\epsilon = 1, \quad \alpha\delta+\beta\zeta = 0, \quad \alpha\gamma+\delta\tau = 1,\\
&\beta\epsilon+\zeta\omega = 1, \quad \delta\tau+\zeta\omega = 1, \quad \gamma\tau+\epsilon\omega = 0,
\end{aligned}
\tag{3.34}
$$

$$
\begin{aligned}
&\alpha\gamma+\beta\epsilon = 1, \quad \alpha\gamma-\beta\epsilon = b/q, \quad \alpha\delta+\beta\zeta = 0,\\
&\alpha\gamma+\delta\tau = 1, \quad \alpha\gamma(q+\tfrac{1}{2})-\delta\tau(q-\tfrac{1}{2}) = b+\tfrac{1}{2},\\
&\beta\epsilon+\zeta\omega = 1, \quad -\beta\epsilon(q+\tfrac{1}{2})+\zeta\omega(q-\tfrac{1}{2}) = b-\tfrac{1}{2},\\
&\delta\tau+\zeta\omega = 1, \quad -\delta\tau+\zeta\omega = b/q, \quad \gamma\tau+\epsilon\omega = 0.
\end{aligned}
\tag{3.35}
$$

It is easy to compute the general solution of these equations. In fact, choosing three arbitrary nonzero constants α,β,δ, the Eqs. (3.32)–(3.35) are fulfilled if and only if

$$
\begin{aligned}
&\zeta = -\frac{\alpha\delta}{\beta}, \quad \gamma = \frac{1}{\alpha}\frac{q+b}{2q}, \quad \epsilon = \frac{1}{\beta}\frac{q-b}{2q},\\
&\tau = \frac{1}{\delta}\frac{q-b}{2q}, \quad \omega = -\frac{\beta}{\alpha\delta}\frac{q+b}{2q}.
\end{aligned}
\tag{3.36}
$$

The three free parameters α, β, δ reflect the fact that the relative normalization of our four isospin multiplets is not fixed. Representations with different α, β, δ are equivalent.

The cases $q=\frac{1}{2}$ are included if we choose $\delta=0$ and define $\tau=\omega=0$. The equations for ζ, γ, ϵ in (3.36) are then valid [see (3.31)].

If $b \neq \pm q$ then the representation which we have obtained will be called the (b,q) representation; it is irreducible of dimension $8q$ and the Casimir operators K_2 and K_3 take the values

$$K_2=q^2-b^2, \quad K_3=b(q^2-b^2). \tag{3.37}$$

Recall that $q=\frac{1}{2}, 1, \frac{3}{2}, \cdots$ and that b is an arbitrary complex number.

Next we remark that our construction works in the cases $b=\pm q$, too, but the nominators $q \mp b$ appearing in (3.36) lead to several complications. In fact, in this case the representation is not irreducible, since the states $|b \mp \frac{1}{2}, q-\frac{1}{2}, q_3\rangle$ and $|b, q-1, q_3\rangle$ span an invariant subspace which carries the $(\pm(q-\frac{1}{2}), q-\frac{1}{2})$ representation. On the other hand, the vector $\phi_0=|\pm q, q, q\rangle$ is still cyclic. It is now easy to show that there is no subrepresentation which is complementary to the $(\pm(q-\frac{1}{2}), q-\frac{1}{2})$ subrepresentation. Hence we have obtained a sequence of representations of spl(2,1) which are not completely reducible.

In the next subsection we shall show that even the tensor product of two irreducible representations is not necessarily completely reducible.

We conclude this subsection with a remark on the action of the automorphism (3.11) on a representation given by (3.20). It is obvious that this automorphism maps the representation (3.20) into a similar one whose quantum numbers are

$$b'=-b, \quad q'=q, \tag{3.38}$$

whose isospin multiplets are

$$\begin{aligned} &|b',q',q_3\rangle' = |b,q,q_3\rangle, \\ &|b' \pm \tfrac{1}{2}, q'-\tfrac{1}{2}, q_3\rangle' = |b \mp \tfrac{1}{2}, q-\tfrac{1}{2}, q_3\rangle, \\ &|b',q'-1,q_3\rangle' = |b,q-1,q_3\rangle, \end{aligned} \tag{3.39}$$

and which belongs to the parameters

$$\begin{aligned} &\alpha'=\beta, \quad \beta'=\alpha, \quad \gamma'=\epsilon, \quad \epsilon'=\gamma, \\ &\delta'=\zeta, \quad \zeta'=\delta, \quad \tau'=\omega, \quad \omega'=\tau. \end{aligned} \tag{3.40}$$

B. The product of two irreducible representations is not necessarily completely reducible

To prove the statement in the title we consider the product $(0,\frac{1}{2})\otimes(0,\frac{1}{2})$. The states are denoted according to the following table:

state			
1. rep.	2. rep.	B	Q_3
a	b	$\frac{1}{2}$	0
ξ_+	η_+	0	$\frac{1}{2}$
ξ_-	η_-	0	$-\frac{1}{2}$
$\bar{a}$	$\bar{b}$	$-\frac{1}{2}$	0

We choose $\alpha=\beta=\gamma=\epsilon=1/\sqrt{2}$ in (3.20) and get the matrices representing the odd generators

$$V_+=\frac{1}{\sqrt{2}}\begin{pmatrix} 0&0&1&0\\ 0&0&0&1\\ 0&0&0&0\\ 0&0&0&0 \end{pmatrix}, \quad V_-=\frac{1}{\sqrt{2}}\begin{pmatrix} 0&-1&0&0\\ 0&0&0&0\\ 0&0&0&1\\ 0&0&0&0 \end{pmatrix}, \tag{3.41}$$

$$W_+=\frac{1}{\sqrt{2}}\begin{pmatrix} 0&0&0&0\\ 1&0&0&0\\ 0&0&0&0\\ 0&0&1&0 \end{pmatrix}, \quad W_-=\frac{1}{\sqrt{2}}\begin{pmatrix} 0&0&0&0\\ 0&0&0&0\\ 1&0&0&0\\ 0&-1&0&0 \end{pmatrix}.$$

Next we recall that an odd generator U acts on the product of an even/odd state $|f\rangle$ of the first representation with an arbitrary state $|g\rangle$ of the second representation like

$$U(|f\rangle \otimes |g\rangle) = (U|f\rangle)\otimes|g\rangle \pm |f\rangle \otimes (U|g\rangle). \tag{3.42}$$

To define the product of our representations we have, therefore, to make a choice which states in the first representation should be considered as even (resp. odd). There are two possibilities which will be treated simultaneously using the convention that the upper (resp. lower) sign is valid if we choose ξ_+, ξ_- (resp. $a, \bar{a}$) to be even and hence $a, \bar{a}$ (resp. ξ_+, ξ_-) to be odd.

Then it is easy to check that the states

$$\begin{aligned} &\xi_+\otimes\eta_+, \quad \xi_-\otimes\eta_-, \quad \frac{1}{\sqrt{2}}(\xi_+\otimes\eta_- + \xi_-\otimes\eta_+), \\ &\frac{1}{\sqrt{2}}(a\otimes\eta_+ \pm \xi_+\otimes b), \quad \frac{1}{\sqrt{2}}(a\otimes\eta_- \pm \xi_-\otimes b), \\ &\frac{1}{\sqrt{2}}(\bar{a}\otimes\eta_+ \pm \xi_+\otimes\bar{b}), \quad \frac{1}{\sqrt{2}}(\bar{a}\otimes\eta_- \pm \xi_-\otimes\bar{b}), \\ &\frac{1}{\sqrt{2}}(a\otimes\bar{b} - \bar{a}\otimes b) \end{aligned} \tag{3.43}$$

span an invariant subspace corresponding to the representation $(0,1)$; the normalization parameters are $\alpha=\beta=\mp\delta=1/\sqrt{2}$.

The remaining states are combined as follows:

$$\begin{aligned} &\lambda_+=\frac{1}{\sqrt{2}}(a\otimes\eta_+ \mp \xi_+\otimes b), \quad \lambda_-=\frac{1}{\sqrt{2}}(a\otimes\eta_- \mp \xi_-\otimes b), \\ &p=a\otimes b, \\ &\bar{\lambda}_+=\frac{1}{\sqrt{2}}(\bar{a}\otimes\eta_+ \mp \xi_+\otimes\bar{b}), \quad \bar{\lambda}_-=\frac{1}{\sqrt{2}}(\bar{a}\otimes\eta_- \mp \xi_-\otimes\bar{b}), \\ &\bar{p}=\bar{a}\otimes\bar{b}, \\ &s=\tfrac{1}{2}(a\otimes\bar{b} \mp \xi_+\otimes\eta_- \pm \xi_-\otimes\eta_+ + \bar{a}\otimes b), \\ &t=\tfrac{1}{2}(a\otimes\bar{b} \pm \xi_+\otimes\eta_- \mp \xi_-\otimes\eta_+ + \bar{a}\otimes b). \end{aligned} \tag{3.44}$$

It is easy to prove the following statements:

(I) The eight vectors (3.44) span an invariant subspace and t is a cyclic vector for the corresponding representation.

(II) The states $\lambda_+, \lambda_-, p, s$ (resp. $\bar{\lambda}_+, \bar{\lambda}_-, \bar{p}, s$) span an invariant subspace which corresponds to the representation defined in Sec. 3 A(d) and whose quantum numbers are $b=\frac{1}{2}$, $q=\frac{1}{2}$ (resp. $b=-\frac{1}{2}$, $q=\frac{1}{2}$). The normalization

parameters are $\alpha=\beta=\mp 1$. Recall that this representation is not completely reducible; in particular s is invariant, i.e., it is mapped into zero by all generators of our algebra.

C. Star and grade star representations

We have seen that in general the product of two irreducible representations is not completely reducible. In order to find out if there are classes of representations for which we have complete reducibility we look for star and grade star representations as explained in Ref. 2.

In the following we shall require that the generators Q_m and B are represented by Hermitian operators. Hence the adjoint operation in sl(2)×gl(1) should satisfy

$$Q_m^*=Q_m, \quad B^*=B. \tag{3.45}$$

In the normal adjoint case the appropriate extensions of this operation to spl(2,1) are then given by[8]

$$U_\alpha^*=D_{\beta\alpha}U_\beta \tag{3.46a}$$

with

$$D=\pm\begin{pmatrix}0 & C\\ -C & 0\end{pmatrix}, \tag{3.46b}$$

whereas in the grade adjoint case we have

$$U_\alpha^\dagger=D_{\beta\alpha}U_\beta \tag{3.47a}$$

with

$$D=\pm\begin{pmatrix}0 & C\\ C & 0\end{pmatrix}. \tag{3.47b}$$

These equations are equivalent to

$$\begin{aligned} &V_+^*=\pm W_-, \quad V_-^*=\mp W_+,\\ &W_-^*=\pm V_+, \quad W_+^*=\mp V_-, \end{aligned} \tag{3.46}$$

(normal adjoint)

$$\begin{aligned} &V_+^\dagger=\mp W_-, \quad V_-^\dagger=\pm W_+,\\ &W_-^\dagger=\pm V_+, \quad W_+^\dagger=\mp V_-. \end{aligned} \tag{3.47}$$

(grade adjoint)

We shall now answer the following question.

Suppose we are given one of the (b,q) representations which have been constructed in Sec. 3A. Is it possible to find a nondegenerate (but for the present not necessarily positive definite) Hermitian scalar product (denoted by a bracket $\langle\ |\ \rangle$) on the representation space such that the representation is a star representation (resp. a grade star representation) with respect to one of the adjoint (resp. grade adjoint) operations defined in (3.46) [resp. (3.47)]?

As an example we consider a (b,q) representation with $b\neq q$, $-q$; $q\geq\frac{1}{2}$, and try to convert it into a star representation; the other cases are treated similarly.

To begin with we exploit (3.45). It is easy to see that these equations are fulfilled if and only if

(I) b is real,

(II) any two different of our isospin multiplets are orthogonal,

(III) the restriction of $\langle\ |\ \rangle$ to an isospin multiplet is proportional to the well-known positive definite scalar product.

The last statement means that there exist real nonzero constants g, g_+, g_-, g_1 such that

$$\begin{aligned} &\langle b,q,q_3|b,q,q_3'\rangle=g\delta_{q_3q_3'},\\ &\langle b+\tfrac{1}{2},q-\tfrac{1}{2},q_3|b+\tfrac{1}{2},q-\tfrac{1}{2},q_3'\rangle=g_+\delta_{q_3q_3'},\\ &\langle b-\tfrac{1}{2},q-\tfrac{1}{2},q_3|b-\tfrac{1}{2},q-\tfrac{1}{2},q_3'\rangle=g_-\delta_{q_3q_3'},\\ &\langle b,q-1,q_3|b,q-1,q_3'\rangle=g_1\delta_{q_3q_3'}. \end{aligned} \tag{3.48}$$

The conditions (3.46) are then fulfilled if and only if

$$\begin{aligned} &\alpha g_+=\pm\gamma^* g, \quad \beta g_-=\mp\epsilon^* g,\\ &\delta g_1=\mp\tau^* g_+, \quad \zeta g_1=\pm\omega^* g_-. \end{aligned} \tag{3.49}$$

Because of (3.36) the solution of these equations is

$$\begin{aligned} g_+&=\pm\frac{1}{|\alpha|^2}\frac{b+q}{2q}g,\\ g_-&=\pm\frac{1}{|\beta|^2}\frac{b-q}{2q}g,\\ g_1&=\frac{1}{|\alpha\delta|^2}\frac{b^2-q^2}{4q^2}g, \end{aligned} \tag{3.50}$$

with an arbitrary nonzero real number g.

[In the cases $q=\frac{1}{2}$ one has to drop the last equation in (3.48), the third and fourth equation in (3.49) and the last equation in (3.50).]

Hence our problem always has a solution (which is unique up to a real factor). However, we would like to get a *positive definite* scalar product, i.e., we would like to require that g, g_+, g_-, g_1 should be (strictly) positive. This is possible if and only if $\pm b>q$.

The necessity of this latter condition is easily understood. From (3.46) and the commutation relations we infer that

$$V_+V_+^*+V_+^*V_+=\pm(B-Q_3). \tag{3.51}$$

If our scalar product is positive definite the left-hand side is a positive (semidefinite) operator which implies $\pm b\geq q$.

Assuming that the condition $\pm b>q$ is fulfilled it is natural to arrange that our state vectors are normalized, i.e., that

$$g=g_+=g_-=g_1=1. \tag{3.52}$$

This will be the case if we choose $g=1$ and

$$\alpha=\left(\pm\frac{b+q}{2q}\right)^{1/2}, \quad \beta=\delta=\left(\pm\frac{b-q}{2q}\right)^{1/2}. \tag{3.53}$$

(In the cases $q=\frac{1}{2}$ we choose of course $\delta=0$.)

Without further comments we shall now describe our results concerning star and grade star representations of spl(2,1). The scalar product in the representation space will always be supposed to be *positive definite*.

We first look for the two classes of star representations S^* corresponding to the two signs in (3.46). The S^* class is composed of the (b,q) representations for which b is real and $\pm b\geq q$. Choosing

$$\alpha=1 \text{ or } \beta=1, \quad \text{if } \pm b=q, \text{ case (b) or (c), Sec. 3 A,}$$
$$\alpha=\left(\pm\frac{b+q}{2q}\right)^{1/2}, \quad \beta=\left(\pm\frac{b-q}{2q}\right)^{1/2}, \quad \text{if } \pm b>q=\tfrac{1}{2}, \text{ case (d), Sec. 3 A,}$$
$$\alpha=\left(\pm\frac{b+q}{2q}\right)^{1/2}, \quad \beta=\delta=\left(\pm\frac{b-q}{2q}\right)^{1/2} \quad \text{if } \pm b>q\geq 1, \text{ case (e), Sec. 3 A,} \tag{3.54}$$

our state vectors will be orthogonal and of the same length with respect to the appropriate scalar product.

As was stressed in Ref. 2 the product of two representations belonging to the S^* class is completely reducible into (irreducible) representations belonging to the $S^\pm$ class.

The Clebsch–Gordan series are

$$\begin{aligned}(b,q)\otimes(b',q') &= (b+b', |q-q'|)\oplus(b+b', |q-q'|+1)\oplus\cdots \\ &\oplus(b+b', q+q') \\ &\oplus(b+b', |q-q'|+1)\oplus\cdots\oplus(b+b', q+q'-1) \\ &\oplus(b+b'+\tfrac{1}{2}, |q-q'|+\tfrac{1}{2})\oplus\cdots\oplus(b+b'+\tfrac{1}{2}, q+q'-\tfrac{1}{2}) \\ &\oplus(b+b'-\tfrac{1}{2}, |q-q'|+\tfrac{1}{2})\oplus\cdots\oplus(b+b'-\tfrac{1}{2}, q+q'-\tfrac{1}{2}),\end{aligned} \tag{3.55}$$

if

$$\pm b>q\geq\tfrac{1}{2}, \quad \pm b'>q'\geq\tfrac{1}{2}.$$

In the case $q=q'$ the first term on the right-hand side [i.e., $(b+b', |q-q'|)$] has to be dropped. If $q=\frac{1}{2}$ or $q'=\frac{1}{2}$ the third line on the right-hand side is empty,

$$\begin{aligned}(\pm q,q)\otimes(b',q') &= (\pm q+b', |q-q'|)\oplus(\pm q+b', |q-q'|+1)\oplus\cdots \\ &\oplus(\pm q+b', q+q') \\ &\oplus(\pm q+b'\pm\tfrac{1}{2}, |q-q'|+\tfrac{1}{2})\oplus\cdots\oplus(\pm q+b'\pm\tfrac{1}{2}, q+q'-\tfrac{1}{2}),\end{aligned} \tag{3.56}$$

if

$$q\geq 0, \quad \pm b'>q'\geq\tfrac{1}{2}.$$

If $q\geq q'$ the first term on the right-hand side [i.e., $(\pm q+b', |q-q'|)$] has to be dropped,

$$\begin{aligned}(\pm q,q)\otimes(\pm q',q') &= (\pm q\pm q', q+q')\oplus(\pm q\pm q'\pm\tfrac{1}{2}, |q-q'|+\tfrac{1}{2})\oplus\cdots \\ &\oplus(\pm q\pm q'\pm\tfrac{1}{2}, q+q'-\tfrac{1}{2}), \quad \text{for all } q,q'\geq 0.\end{aligned} \tag{3.57}$$

The Clebsch–Gordan series (3.55)–(3.57) may be obtained just from the isospin and baryon number content of the product representations. In particular they are independent of which subspace in the (b,q), resp. $(\pm q,q)$, representations has been defined to be even or odd.

Note, furthermore, that the properties of the S^- class can be read off directly from those of the S^+ class and vice versa by applying the automorphism (3.11). In fact, this automorphism interchanges the two possible adjoint operations given in (3.46).

Let us next comment on the grade star representations and try to convert a (b,q) representation with $b\neq q$, $-q$; $q\geq\frac{1}{2}$, into a grade star representation. An argument which is completely analogous to that given above yields the equations [corresponding to (3.50)]

$$\begin{aligned}g_+ &= \pm(-1)^\lambda\frac{1}{|\alpha|^2}\frac{q+b}{2q}g, \\ g_- &= \pm(-1)^\lambda\frac{1}{|\beta|^2}\frac{q-b}{2q}g, \\ g_1 &= \frac{1}{|\alpha\delta|^2}\frac{b^2-q^2}{4q^2}g,\end{aligned} \tag{3.58}$$

with an arbitrary nonzero real number g. Here $\lambda\in\{0,1\}$ is the degree of the isospin q multiplet of the first representation. The last equation in (3.58) has to be dropped if $q=\frac{1}{2}$.

It is now evident that our Hermitian form can be choosen to be positive definite if and only if $q=\frac{1}{2}$ and $-\frac{1}{2}<b<\frac{1}{2}$. If this is the case we have to arrange, furthermore, that the sign in the grade adjoint operation (3.47) and the degree λ satisfy

$$\pm(-1)^\lambda=1. \tag{3.59}$$

This result yields a counterexample which is interesting for the general theory of grade star representations. We have shown that the $(0,\frac{1}{2})$ representation is a grade star representation with respect to a positive definite scalar product. On the other hand, the tensor product of the $(0,\frac{1}{2})$ representation with itself is not completely reducible (see Sec. 3 B). Hence[2] there cannot exist any positive definite scalar product which converts $(0,\frac{1}{2})\otimes(0,\frac{1}{2})$ into a grade star representation.

The (q,q) and $(-q,q)$ representations can be converted into a grade star representation provided that condition (3.59) is satisfied. In fact, these representations are then star representations *and* grade star representations with respect to the same positive definite scalar product. It is evident from (3.57) that in this case the tensor product of two irreducible grade star representations is not decomposable into irreducible grade star representations. Let us recall, furthermore, that the (q,q) and $(-q,q)$ representations are those which are not characterized by the Casimir operators. Presumably all these pathologies are connected somehow.

From the results derived above we conclude that considering grade star representations of spl(2,1) is almost useless.

[1] V.G. Kac, Funct. Anal. Appl. 9, 91 (1975); W. Nahm, V. Rittenberg, and M. Scheunert, Phys. Lett. B 61, 383 (1976); P.G.O. Freund and I. Kaplansky, J. Math. Phys. 17, 228 (1976).

[2] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. 18, 146 (1977).

[3] W. Nahm and M. Scheunert, J. Math. Phys. 17, 868 (1976).

[4] The proof is trivial and is based on the observation that $\mathrm{Tr}(\gamma\{A,B\})=0$ (which is the generalization of $\mathrm{Tr}([A,B])=0$). In order to prove Eq. (1.6) all we have to do is to bring X_μ under the trace sign.

[5] A. Pais and V. Rittenberg, J. Math. Phys. 16, 2062 (1975).

[6] N. Backhouse, J. Math. Phys. 18, (Feb. 1977).

[7] G. Hochschild, Ill. J. Math. 20, 107 (1976); D. Ž. Djoković and G. Hochschild, Ill. J. Math. 20, 134 (1976); D. Ž. Djoković, J. Pure Appl. Alg. 7, 217 (1976).

[8] See Eqs. (2.8)–(2.11) and (2.14)–(2.17) of Ref. 2 for the definitions of the adjoint and grade adjoint operations. The star and grade star representations are defined in (2.18).[2]

Solitons, pseudopotentials, and certain Lie algebras

James Corones

Department of Mathematics, Iowa State University, Ames, Iowa 50011
(Received 24 May 1976)

It is shown that there is a common algebraic structure in the pseudopotentials of equations solvable by the generalized Zakharov–Shabat eigenvalue problem. It follows that an arbitrarily large number of prolongation variables can be associated with these equations and that a recently developed geometric interpretation of solitons can be given for each of these equations.

1. INTRODUCTION

The idea of a pseudopotential was introduced into the study of nonlinear partial differential equations by Wahlquist and Estabrook.[1] It has previously been shown that the simple pseudopotential[2] associated with the Korteweg–de Vries (KdV) equation and several other equations that have soliton solutions define a Lie algebra.[2,3] For the KdV equation this observation was also made in Ref. 4 where the algebra was identified and used in a geometric interpretation of solitons.

In this note it is shown that there is a common algebraic structure in the linear pseudopotentials (see Ref. 3 and below) associated with equations soluable by the inverse scattering method via the generalized Zakharov–Shabat eigenvalue problem.[5] It is shown that the Zakharov–Shabat eigenvalue problem corresponds to a 2×2 matrix representation of this structure. Higher dimensional representations correspond to new eigenvalue problems and associated isospectral flows.

The results of this note show that an arbitrarily large number of prolongation variables, in the sense of Ref. 1, can be associated with the equations soluable via the Zakharov–Shabat problem. That this might be possible is suggested, for KdV, in Ref. 1 in connection with a new infinity of conservation laws for KdV. Here the new prolongation variables are defined by seeking higher dimensional representation of the algebraic structure. In addition the results show that the geometric construction of Ref. 4 can be applied, using the same algebraic structure, to all equations, treated via the generalized Zakharov–Shabat problem.

2. DEFINITIONS

In a previous paper, Ref. 2, the classical definition of the set of all pseudopotentials associated with a given scalar evolution equation was given. Here an (obvious) extension to pairs of coupled equations is needed. Hence consider

$$\phi_t = K_1(\phi, \psi, \phi_x, \psi_x, \cdots), \tag{1a}$$

$$\psi_t = K_2(\phi, \psi, \phi_x, \psi_x, \cdots), \tag{1b}$$

where K_1 and K_2 are functions of ϕ and its first $m+1$ spatial derivatives as well as ψ and its first $l+1$ spatial derivatives. Let S be the set of ϕ and its first spatial derivatives and ψ and its first l spatial derivatives. The set of all *pseudopotentials* associated with system (1) is the set of all functions $q^i(x,t)$, $l=1,\ldots,n$, n arbitrary such that

$$q^i_x = F^i(S, q^1, \ldots, q^N; x, t), \tag{2a}$$

$$q^i_t = G^i(S, q^1, \ldots, q^N, x, t) \tag{2b}$$

are integrable subject to constraint (1).

A subset of (2), of special interest, is defined as follows. Let $\hat{q}$ denote an n component column vector. The set of all *linear pseudopotentials* associated with (1) is the set of all n vectors $\hat{q}$ (n arbitrary) such that

$$\hat{q}_x = \hat{F}(S)\hat{q}, \tag{3a}$$

$$\hat{q}_t = \hat{G}(S)\hat{q} \tag{3b}$$

are integrable subject to constraint (1). Here $\hat{F}(S)$, $\hat{G}(S)$ are $n\times n$ matrix functions on S.

Now suppose $\hat{F}(S)$ and $\hat{G}(S)$ have a definite structure. In particular suppose

$$\hat{F}(S) = \sum_{i=1}^{j} a^i(S) X^i, \tag{4a}$$

$$G(S) = \sum_{i=1}^{j} b^i(S) X^i, \tag{4b}$$

where the a^i and b^i are scalar functions and the X^i are constant matrices such that

$$[X^r, X^s] = \sum_{p=1}^{j} c^{rs}{}_p X^p, \tag{5}$$

where the bracket denotes the matrix commutator. Since the X^i are used to specify $\hat{q}$ via the first order system (3), Eq. (5) is interpreted as defining a Lie algebra.

3. CONNECTION WITH INVERSE METHOD

It happens that the case when three generators are present and they satisfy

$$[X^1, X^2] = X^3, \tag{6a}$$

$$[X^1, X^3] = -X^2, \tag{6b}$$

$$[X^2, X^3] = X^1 \tag{6c}$$

is of particular interest. In this instance it is readily seen that the scalar functions a^i and b^i, $i=1,2,3$ must satisfy

$$a^1_t + a^2 b^3 - a^3 b^2 = b^1_x, \tag{7a}$$

$$a^2_t + a^3 b^1 - a^1 b^3 = b^2_x, \tag{7b}$$

$$a^3_t + a^1 b^2 - a^2 b^1 = b^3_x, \tag{7c}$$

if the system (3) is to be integrable.

A specialization of system (7) has been fully investigated in Ref. 5 in connection with the generalized

Zakharov–Shabat eigenvalue problem. To see this it is only necessary to let

$$a^1 = -i(\psi + \phi), \tag{8a}$$

$$a^2 = \psi - \phi, \tag{8b}$$

$$a^3 = -2\lambda, \tag{8c}$$

$$(ib^1 - b^2)/2 = B, \tag{8d}$$

$$(ib^1 + b^2)/2 = C, \tag{8e}$$

$$ib^3/2 = A. \tag{8f}$$

It then easily follows that

$$A_x = \phi C - \psi B, \tag{9a}$$

$$B_x + 2i\lambda B = \phi_t - 2\phi A, \tag{9b}$$

$$C_x - 2i\lambda C = \psi_t + 2\psi A. \tag{9c}$$

This set of scalar equations has been used in Ref. 4 to find equations that are soluable by the inverse scattering method via the Zakharov–Shabat eigenvalue problem. The infinity of nonlinear partial differential equations that can thereby be solved include, for example, the KdV, modified KdV, and nonlinear Schrödinger equations.

To this point no representation of (6) has been used. The derivation leading to (9) used only the algebraic properties of the X^i. The 2×2 matrix representation of su(2),

$$X^1 = \frac{i}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad X^2 = \frac{1}{2}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad X^3 = \frac{i}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{10}$$

satisfies (6). It is easily checked that the standard generalized Zakharov–Shabat eigenvalue problem follows by using (8) and (10) in (4).

The algebra of the linear pseudopotentials considered above can be identified as su(2). However, the various connections[6] between the (real and complex) Lie algebras su(2), so(3), and sl(2, R) make other identifications possible. The application to the inverse scattering method suggests the su(2) identification. It is not difficult to show that simple pseudopotentials have an algebraic structure specified by (6) (with the appropriate definition of the bracket, see Refs. 2 and 3). In this case the sl(2, R) identification seems appropriate.

Finally, it should be noted that the key results above do not depend on (1) being evolution equations. All the equations considered in Ref. 4 yield the same algebraic structure. This is clear from (7), (8), and (9).

[1]H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. **16**, 1 (1975).
[2]J. Corones, J. Math. Phys. **17**, 756 (1976).
[3]J. Corones and F. Testa, to appear in the Proceedings of the NSF Workshop on Contact Transformations, Vanderbilt University, September 1974.
[4]R. Hermann, Phys. Rev. Lett. **36**, 835 (1976).
[5]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[6]W. Miller, *Symmetry Groups and Their Applications* (Academic, New York, 1972).

Geometric formulation of electrodynamics and general relativity in discrete space–time*

Don Weingarten†

Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627
(Received 1 March 1976)

Classical electrodynamics is reconstructed in discrete space–time, then combined with Regge's discrete version of general relativity. Quantization using path integrals is considered briefly.

1. INTRODUCTION

By using what might be described as a "geometric interpretation" of Maxwell's equations we will show that electrodynamics can be transferred, in a natural way, from continuous space—time parametrized by coordinates to a discrete version of space—time which requires no coordinates but retains some of the geometric structure of continuous space—time. Then we will combine this result with a formulation of general relativity proposed by Regge[1] and obtain a discrete, coordinate-free model incorporating both electrodynamics and gravitation.

One of the motivations of this work is simply that it provides an interesting way of looking at interacting electrodynamics and gravitation. Another is the familiar suspicion that a fully satisfactory cure for the ultraviolet divergences of relativistic quantum field theory will eventually require some form of discrete space—time. Thus it seems useful to see what discrete theories suggest themselves as natural extensions of existing continuous theories. Finally, the theory we will construct may also have some practical value. The dynamical equations we will obtain are algebraic rather than differential and are therefore convenient for solution by computer.[2] The solutions gotten in this way should approximate solutions to the differential equations of the continuous theory, but the theory we will derive is not simply the result of replacing space—time by a lattice of points and derivatives by finite differences. Instead, the discrete theory retains geometric features of the continuous theory, and we expect that as a consequence it will approximate the qualitative behavior of the continuous theory more efficiently.

The theory we will describe is classical; the extent to which it might be quantized using Feynman path integration will be discussed briefly toward the end of this paper.

2. CHAINS, BOUNDARIES, AND COBOUNDARIES

We will begin by describing an interpretation of some operations which can be performed with arbitrary antisymmetric tensor fields on four-dimensional curved space—time. Then we will apply these ideas to electrodynamics. Most of the definitions we will use are familiar notions of differential geometry.[3] For convenience we will assume space—time is topologically trivial. The extension of our results to nontrivial topologies will be fairly obvious.

A four-dimensional manifold can support five different sorts of covariant tensor fields antisymmetric in all indices: scalar fields, vector fields, and antisymmetric tensor fields with two, three, or four indices. Antisymmetric tensor fields with five or more indices vanish identically. We will call each such field an n-chain, where n is the number of its indices. By exterior differentiation we can map the set of n-chains into the set of $(n+1)$-chains. That is if we represent an arbitrary n-chain by $f^{(n)}$, exterior differentiation by ∇, and the components of $f^{(n)}$ in a coordinate basis by $f^{(n)}_{\mu_1\cdots\mu_n}$ we have

$$
\begin{aligned}
&[\nabla f^{(0)}]_\mu = \frac{\partial}{\partial x^\mu} f^{(0)},\\
&[\nabla f^{(1)}]_{\mu\nu} = \frac{\partial}{\partial x^\mu} f^{(1)}_\nu - \frac{\partial}{\partial x^\nu} f^{(1)}_\mu,\\
&[\nabla f^{(2)}]_{\mu\nu\lambda} = \frac{\partial}{\partial x^\mu} f^{(2)}_{\nu\lambda} - \frac{\partial}{\partial x^\nu} f^{(2)}_{\mu\lambda} - \frac{\partial}{\partial x^\lambda} f^{(2)}_{\nu\mu},\\
&[\nabla f^{(3)}]_{\mu\nu\lambda\phi} = \frac{\partial}{\partial x^\mu} f^{(3)}_{\nu\lambda\phi} - \frac{\partial}{\partial x^\nu} f^{(3)}_{\mu\lambda\phi} - \frac{\partial}{\partial x^\lambda} f^{(3)}_{\nu\mu\phi} - \frac{\partial}{\partial x^\phi} f^{(3)}_{\lambda\mu}.
\end{aligned}
\tag{2.1}
$$

Each $f^{(4)}$ is mapped to 0 by this operation. On the other hand, by taking divergences, each n-chain can be mapped to an $(n-1)$-chain. Representing the divergence operator by $-\Delta$ and converting it to a form convenient for use on antisymmetric tensors we have

$$
[\Delta f^{(n)}]^{\mu_1\cdots\mu_{n-1}} = -(-g)^{-1/2} \frac{\partial}{\partial x^\nu} (-g)^{1/2} f^{(n)\nu\mu_1\cdots\mu_{n-1}},
\tag{2.2}
$$

where the quantity g is the determinant of the covariant metric tensor $g_{\mu\nu}$ which we choose to have signature $(+---)$. The result of Δ acting on any 0-chain is defined to be 0. Definitions (2.1) and (2.2) yield the identities

$$
\nabla\nabla f^{(n)} = 0,
\tag{2.3}
$$

$$
\Delta\Delta f^{(n)} = 0,
\tag{2.4}
$$

for all $f^{(n)}$.

Next let us define an invariant inner product between pairs of n-chains

$$
\langle f^{(n)} | h^{(n)} \rangle = \frac{1}{n!} \int d^4x\, (-g)^{1/2} f^{(n)\mu_1\cdots\mu_n} h^{(n)}_{\mu_1\cdots\mu_n}.
\tag{2.5}
$$

It follows that if all the chains we consider are required to vanish sufficiently rapidly at ∞, the operation ∇ is the adjoint of Δ:

$$
\langle f^{(n-1)} | \Delta h^{(n)} \rangle = \langle \nabla f^{(n-1)} | h^{(n)} \rangle.
\tag{2.6}
$$

This relation is a version of Stokes theorem.

But at each point in space—time an n-chain with $n \geq 1$ may be considered the tangent to an oriented n-dimensional surface. Thus the full n-chain may be thought of as a formal superposition of n-surface tangents each weighted with an appropriate coefficient. This picture can be extended to include $n=0$ by considering a point to be a 0-dimensional surface and defining the tangent to a point to be a real number. Consider the n-chain, $n \geq 1$, giving the tangent to a single oriented n-dimensional surface Σ parametrized by the map $y_i \to x^\mu(y_i)$, where i runs from 1 to n,

$$f^{(n)\mu_1\cdots\mu_n}(z) = (-g)^{-1/2}\int_\Sigma d^n y\, \delta^{(4)}[z - x(y)] \times \frac{\partial(x^{\mu_1}\cdots x^{\mu_n})}{\partial(y_1\cdots y_n)}.$$

The multiplicative factor $(-g)^{1/2}$ is necessary to make $f^{(n)}$ a tensor and

$$\frac{\partial(x^{\mu_1}\cdots x^{\mu_n})}{\partial(y_1\cdots y_n)}$$

is the Jacobian of the map $(y_i,\ldots,y_n) \to (x^{\mu_1},\ldots,x^{\mu_n})$. For a point x we will choose the corresponding $f^{(0)}$ to be

$$f^{(0)}(z) = (-g)^{-1/2}\delta^{(4)}(z - x).$$

If we now evaluate $[\Delta f^{(n)}]^{\mu_1\cdots\mu_{n-1}}$ for the tensor $f^{(n)}$ of the surface Σ, we find we have obtained the $(n-1)$-surface tensor of the boundary of Σ. For example, for a tensor $f^{(2)}$ representing a 2-surface parameterized by y_1 and y_2 both running from 0 to 1, we have

$$\begin{aligned}
&[\Delta f^{(2)}]^\nu(z) \\
&= -(-g)^{-1/2}\frac{\partial}{\partial z^\mu}\int_0^1\int_0^1 dy_1\,dy_2\,\delta^{(4)}[z - x(y)] \times\left(\frac{\partial x^\mu}{\partial y_1}\frac{\partial x^\nu}{\partial y_2} - \frac{\partial x^\nu}{\partial y_1}\frac{\partial x^\mu}{\partial y_2}\right) \\
&= (-g)^{-1/2}\int_0^1\int_0^1 dy_1\,dy_2\left(\frac{\partial\delta^{(4)}[z - x(y)]}{\partial y_1}\frac{\partial x^\nu}{\partial y_2} - \frac{\partial\delta^{(4)}[z - x(y)]}{\partial y_2}\frac{\partial x^\nu}{\partial y_1}\right) \\
&= (-g)^{-1/2}\int_0^1 dy_1\,\delta^{(4)}[z - x(y)]\frac{\partial x^\nu}{\partial y_1}\Big]_{y_2=0} \\
&\quad + (-g)^{-1/2}\int_0^1 dy_2\,\delta^{(4)}[z - x(y)]\frac{\partial x^\nu}{\partial y_2}\Big]_{y_1=1} \\
&\quad + (-g)^{-1/2}\int_1^0 dy_1\,\delta^{(4)}[z - x(y)]\frac{\partial x^\nu}{\partial y_1}\Big]_{y_2=1} \\
&\quad + (-g)^{-1/2}\int_1^0 dy_2\,\delta^{(4)}[z - x(y)]\frac{\partial x^\nu}{\partial y_2}\Big]_{y_1=0}.
\end{aligned}$$

Thus we will refer to Δ as the boundary operator and its adjoint ∇ as the coboundary operator.

Finally let us examine the action of the inner produce on the tensor $f^{(n)}$ of a single surface Σ for which the number, s, of linearly independent spacelike tangent vectors is the same at all points. Now the inner product we have defined actually gives infinite norm for a surface of 0 thickness and dimension less than 4. However, by giving the surface a small thickness greater than 0 and then compensating for this thickness with an appropriate normalization, a useful finite result can be obtained. Let $y_1,\ldots,y_n$ parametrize Σ and let $y_{n+1},\ldots,y_4$ parametrize a family of nearly identical copies of Σ displaced from each other slightly in the direction orthogonal to Σ. Define the proper $(4-n)$-dimensional measure of the region Σ' swept out by $y_{n+1},\ldots,y_4$ at fixed $y_1,\ldots,y_n$ to be $m^{(4-n)}(\Sigma')$,

$$m^{(4-n)}(\Sigma') = \int_{\Sigma'} dy_{n+1}\cdots dy + \frac{1}{(4-n)!}\left|\frac{\partial(x^{\mu_{n+1}}\cdots x^{\mu_4})}{\partial(y_{n+1}\cdots y_4)}\frac{\partial(x^{\nu_{n+1}}\cdots x^{\nu_4})}{\partial(y_{n+1}\cdots y_4)} \times g_{\mu_{n+1}\nu_{n+1}}\cdots g_{\mu_4\nu_4}\right|^{1/2}, \tag{2.7}$$

and require $m^{(4-n)}(\Sigma')$ to be the same for all $y_1,\ldots,y_n$. For $n=4$, $m^{(4-n)}(\Sigma')$ is defined to be 1. Then the normalized n-surface tensor

$$f^{(n)\mu_1\cdots\mu_n}(z) = \lambda(-g)^{1/2}\int d^4y\,\delta^{(4)}[z - x(y)]\frac{\partial(x^{\mu_1}\cdots x^{\mu_n})}{\partial(y_1\cdots y_n)},$$

$$\lambda = [(4-n)!\,m^{(4-n)}(\Sigma')]^{-1/2} \times\left|\frac{\partial(x^{\mu_{n+1}}\cdots x^{\mu_4})}{\partial(y_{n+1}\cdots y_4)}\frac{\partial(x^{\nu_{n+1}}\cdots x^{\nu_4})}{\partial(y_{n+1}\cdots y_4)}g_{\mu_{n+1}\nu_{n+1}}\cdots g_{\mu_4\nu_4}\right|^{1/2},$$

yields the value

$$\langle f^{(n)} \mid f^{(n)}\rangle = (-1)^s m^{(n)}(\Sigma),$$

where, for $n \geq 1$, $m^{(n)}(\Sigma)$ is the proper n-measure of Σ given by an expression of the same form as (2.7), and the proper 0-measure of a point is defined to be 1.

The preceding ideas can be applied to electrodynamics immediately. The electromagnetic field F is a 2-chain and therefore corresponds to a weighted superposition of two-dimensional surfaces; the vector potential A is a 1-chain and therefore a superposition of lines. Maxwell's equations are

$$\nabla F = 0, \quad \Delta F = 0 \tag{2.8}$$

and state that both the boundary and coboundary of the electromagnetic field vanish, and the relation between F and A,

$$F = \nabla A, \tag{2.9}$$

states that F is the coboundary of A. An action integral which yields (2.8) is

$$\mathcal{A} = \langle \nabla A \mid \nabla A\rangle \tag{2.10}$$

if (2.9) is taken as the definition of F.

3. ELECTRODYNAMICS IN A SIMPLICIAL COMPLEX

We will now replace continuous space—time with a form of discrete space—time which can be equipped with a natural set of n-chains, a boundary operator, an inner product, and a coboundary operator, and these in turn will allow us to reconstruct electrodynamics.[4] The mathematical designation for the discrete space—time we will describe is a "simplicial complex.[5]"

Let S_0 be a set containing at most a countably infinite number of elements. For convenience we will take S_0 to be either the integers or a proper subset of the integers. For each $n=1,\ldots,4$ let S_n be a set of $(n+1)$-tuples of elements of S_0. We require that if $\{i_1,\ldots i_{n+1}\}$, $n\geq 1$, occurs in S_n, then each subset of $\{i_1,\ldots,i_{n+1}\}$ is an element of some S_m, $m\leq n$. From every element of S_n form all possible distinct ordered $(n+1)$-tuples. These will be written $[i_1,\ldots,i_{n+1}]$ and may be thought of as oriented n-surfaces. Each will be called an n-simplex. By allowing n-simplexes to be formally multiplied by arbitrary real numbers and then forming finite sums of such quantities, the collection of n-simplexes can be turned into a vector space. A vector in this space will be called an n-chain and written $c^{(n)}$. If $[j_1,\ldots,j_{n+1}]$ is a permutation of $[i_1,\ldots,i_{n+1}]$, we assume

$$[j_1,\ldots,j_{n+1}]=(-1)^P[i_1,\ldots,i_{n+1}],$$

where P is 1 for an odd permutation and 0 otherwise. Each $c^{(n)}$ can be written in the form

$$c^{(n)}=\sum_{[i_1,\ldots,i_{n+1}]\in U_n} c^{(n)}(i_1,\ldots,i_{n+1})[i_1,\ldots,i_{n+1}], \quad (3.1)$$

where $c^{(n)}(i_1,\ldots,i_{n+1})$ is a real number and the set of oriented simplexes U_n contains exactly one (arbitrarily chosen) ordering $[j_1,\ldots,j_{n+1}]$ for each $\{i_1,\ldots,i_{n+1}\}\in S_n$.

A boundary operator Δ taking any $c^{(n)}$ into some $c^{(n-1)}$ can be defined as follows. For an individual n-simplex assume

$$\Delta[i_1,\ldots,i_{n+1}]=\sum_{j=1}^{n+1}(-1)^{j+1}[i_1,\ldots,\hat{i}_j,\ldots,i_{n+1}], \quad (3.2)$$

where the symbol $\hat{i}_j$ indicates i_j has been omitted and the result of omitting i from the 0-simplex $[i]$ is 0. The action of Δ on an arbitrary n-chain is gotten from (3.2) by requiring Δ to be linear. It follows that

$$\Delta\Delta c^{(n)}=0 \quad (3.3)$$

for all $c^{(n)}$.

Next we will construct an inner product $\langle c^{(n)}|d^{(n)}\rangle$ on n-chains. For any individual n-simplex assume

$$\langle[i_1,\ldots,i_{n+1}]|[i_1,\ldots,i_{n+1}]\rangle=\mu(i_1,\ldots,i_{n+1}), \quad (3.4)$$

where $\mu(i_1,\ldots,i_{n+1})$ is a real number; for each pair of simplexes differing in the set $\{i_1,\ldots,i_{n+1}\}$ from which they are formed assume

$$\langle[i_1,\ldots,i_{n+1}]|[j_1,\ldots,j_{n+1}]\rangle=0. \quad (3.5)$$

For any 0-simplex $[i]$, $\mu(i)$ is chosen to be 1. For $n\geq 1$, $\mu(i_1,\ldots,i_{n+1})$ can be either positive or negative, but for convenience we exclude $\mu(i_1,\ldots,i_{n+1})=0$. Those n-simplexes with negative measure may be thought of as containing an odd number of linearly independent spacelike vectors. The inner product between arbitrary n-chains is gotten from (3.4) and (3.5) by requiring $\langle c^{(n)}|d^{(n)}\rangle$ to be linear in each of its arguments. The coboundary operator ∇ can now be defined as the adjoint of Δ with respect to the inner product. Equation (3.3) combined with the restriction $\mu(i_1,\ldots,i_{n+1})\neq 0$ for all $[i_1,\ldots,i_{n+1}]\in U_n$ implies

$$\nabla\nabla c^{(n)}=0 \quad (3.6)$$

for all $c^{(n)}$.

Finally we must impose a number of constraints on the structure of the simplicial complex to insure that it adequately approximates a curved four-dimensional space–time governed by a metric with signature $(+---)$. First we need a pair of "topological" requirements: each $(n-1)$-simplex must occur in the boundary of at least one n-simplex, for $n=1,\ldots,4$, and each 3-simplex must occur in the boundary of exactly two distinct 4-simplexes. These conditions imply, roughly speaking, that the complex will not have any isolated regions of dimension lower than four, any boundaries, or any "joints" which would require a space of dimension higher than four for a local embedding. The significance of the two conditions we have just given can be pictured more clearly by considering the effect of corresponding requirements on a complex composed only of 0-, 1-, 2- and 3-simplexes. It can be proved[5] that if these conditions are fulfilled then there is necessarily some curved, continuous four-dimensional space–time which can be covered with the simplicial complex in such a way that each 4-simplex is assigned to a unique singly connected 4-volume, each 3-simplex assigned to a unique singly connected 3-volume and so on.

We will also need a second set of constraints which are "metric." Consider a single 4-simplex $[i_1,\ldots,i_5]$ and all the 3-, 2-, 1-, and 0-simplexes which can be formed from subsets of $\{i_1,\ldots,i_5\}$. Let us embed this collection of simplexes in flat Minkowski space by assigning each 0-simplex $[j]$ to a point $p(j)$, each 1-simplex $[j_1,j_2]$ to a straight line $p(j_1,j_2)$ running from $p(j_1)$ to $p(j_2)$, and so on. If $p^\mu(j)$ are the Minkowski space components of $p(j)$, each n-simplex can be assigned to an n-indexed antisymmetric surface tensor by the relations

$$M^\mu(j_1,j_2)=p^\mu(j_2)-p^\mu(j_1),$$

$$\begin{aligned}M^{\mu_1\cdots\mu_n}(j_1,\ldots,j_{n+1})&=M^{\mu_1\cdots\mu_{n-1}}(j_1,j_2,\ldots,j_{n-1},j_n)M^{\mu_n}(j_1,j_{n+1})\\&\quad-M^{\mu_1\cdots\mu_{n-1}}(j_1,j_2,\ldots,j_{n-1},j_{n+1})M^{\mu_n}(j_1,j_n)\\&\quad\cdots-M^{\mu_1\cdots\mu_{n-1}}(j_1,j_{n+1},\ldots,j_{n-1},j_n)M^{\mu_n}(j_1,j_2).\end{aligned}$$

Then let the function $N(j_1,\ldots,j_{n+1})$ be given by

$$N(j_1,\ldots,j_{n+1})=\frac{1}{n!}M^{\mu_1\cdots\mu_n}(j_1,\ldots,j_{n+1})\times M^{\nu_1\cdots\nu_n}(j_1,\ldots,j_{n+1})\eta_{\mu_1\nu_1}\cdots\eta_{\mu_n\nu_n},$$

where $\eta_{\mu\nu}$ is the Minkowski metric. From $N(j_1,\ldots,j_{n+1})$ define a signed proper n-measure for each n-simplex by

$$\mu(j_1,\ldots,j_{n+1})=\frac{N(j_1,\ldots,j_{n+1})}{\sqrt{|N(j_1,\ldots,j_{n+1})|}}. \quad (3.7)$$

It follows that for each n-simplex, $[\mu(j_1,\ldots,j_{n+1})]^2$ can be expressed as the absolute value of a homogeneous polynomial of degree $2n$ in the set of 1-measures $\mu(j_1,j_2)$. Let us abstract this set of equations and use it to determine each $[\mu(j_1,\ldots,j_{n+1})]^2$, $n\geq 2$, in (3.4) from the set of $\mu(j_1,j_2)$ in (3.4). In addition, since the embedding we have made is in flat Minkowski space with signature $(+---)$, the quantities $\mu(j_1,\ldots,j_{n+1})$ in (3.7) fulfill a set of inequalities. For example, the 4-measure $\mu(i_1,\ldots,i_5)$ must be negative, and if three 1-simplexes

with positive measure form the boundary of 2-simplex, then the measure of the 2-simplex must be negative. Let us also abstract this set of inequalities and impose them on the measures in (3.4). Once the various metric constrains we have just described are imposed, the simplicial complex's inner product $\langle \cdots | \cdots \rangle$ acting on 1-chains becomes the discrete version of the metric tensor $g_{\mu\nu}$ of curved continuous space–time.

We can now construct electrodynamics on the simplicial complex simply by reinterpreting each of the quantities in Eqs. (2.8), (2.9), and (2.10). If F is a 2-chain on the simplicial complex, Δ the complex's boundary operator, and ∇ the coboundary operator, then Eqs. (2.8) become the simplicial complex's version of Maxwell's equations. If A is now a 1-chain on the complex, (2.9) is again the relation between the electromagnetic field and potential. Using the complex's inner product (2.10) becomes a satisfactory action integral. We have

$$\begin{aligned}\delta\mathcal{A} &= \langle \nabla\delta A \,|\, \nabla A\rangle + \langle \nabla A \,|\, \nabla\delta A\rangle \\ &= 2\langle \nabla\delta A \,|\, \nabla A\rangle = 2\langle \delta A \,|\, \Delta\nabla A\rangle = 0, \end{aligned} \tag{3.8}$$

but since we assumed none of the $\mu(i_1, \ldots, i_n)$ in (3.4) are 0, (3.8) implies

$$\Delta\nabla A = 0. \tag{3.9}$$

Using (3.9), (3.6), and (2.9) we obtain (2.8).

4. ELECTRODYNAMICS AND GENERAL RELATIVITY

A formulation of general relativity on a simplicial complex developed in Ref. 1 can be described as follows. The metric tensor of continuous space–time $g_{\mu\nu}$ is replaced by the inner product of the simplicial complex acting on the set of 1-chains, while the curvature tensor $R_{\mu\nu\alpha\beta}$ is replaced by a symmetric bilinear form $\mathcal{R}[c^{(2)}, c^{(2)\prime}]$ on the set of 2-chains of the simplicial complex. Using the inner product on the set of 2-chains, the bilinear form $\mathcal{R}[c^{(2)}, c^{(2)\prime}]$ can be related to a linear operator R on 2-chains:

$$\mathcal{R}[c^{(2)}, c^{(2)\prime}] = \langle c^{(2)} | Rc^{(2)\prime}\rangle.$$

If $[i,j,k]^*$ is the dual to the 2-simplex $[i,j,k]$ given by the linear map

$$[i,j,k]^* : c^{(2)} \to \langle [i,j,k] \,|\, c^{(2)}\rangle,$$

then, in obvious notation, R has the diagonal form

$$R = \sum_{[i,j,k]\in U_2} R(i,j,k)\frac{[i,j,k][i,j,k]^*}{\mu(i,j,k)},$$

where $R(i,j,k)$ is a real number which can be expressed as a function of the values of the inner product on the set of 1-simplexes, $\mu(i,j)$. Using R or $R(i,j,k)$ the action integral of general relativity becomes

$$\begin{aligned}\mathrm{Tr}(R) &= \sum_{[i,j,k]\in U_2} \langle [i,j,k] \,|\, R[i,j,k]\rangle \\ &= \sum_{[i,j,k]\in U_2} R(i,j,k)\mu(i,j,k). \end{aligned} \tag{4.1}$$

Einstein's equations are gotten by requiring

$$\frac{\delta \mathrm{Tr}(R)}{\delta\mu(i,j)} = 0$$

for all $[i,j]\in U_1$.

An action for interacting gravity and electrodynamics can now be obtained simply by adding (4.1) to the action for electrodynamics in Sec. 3,

$$\mathcal{A} = \mathrm{Tr}(R) + \langle \nabla A \,|\, \nabla A\rangle. \tag{4.2}$$

If we vary this with respect to A, Maxwell's equations (2.8) are reproduced again. Varying with respect to $\mu(i,j)$ for all $[i,j]\in U_1$ we obtain the interaction version of Einstein's equations. The first term on the right side of (4.2) yields

$$\begin{aligned}\frac{\delta \mathrm{Tr}(R)}{\delta\mu(i,j)} &= \sum_{[i',j',k']\in U_2} \frac{\delta R(i',j',k')}{\delta\mu(i,j)}\,\mu(i',j',k') \\ &\quad + \sum_{[i',j',k']\in U_2} R(i',j',k')\frac{\delta\mu(i',j',k')}{\delta\mu(i,j)}, \end{aligned} \tag{4.3}$$

but it is shown in Ref. 1 that the first term on the right side of (4.3) vanishes. In addition, the "metric" constraints imposed on the complex in Sec. 3 give

$$\begin{aligned}16\mu(i,j,k)^2 = \big|\, &2\mu(i,j)^2\mu(j,k)^2 + 2\mu(i,k)^2\mu(j,k)^2 \\ &+ 2\mu(i,k)^2\mu(i,j)^2 - \mu(i,j)^4 - \mu(j,k)^4 - \mu(i,k)^4 \big| \end{aligned} \tag{4.4}$$

and $\mu(i,j,k)$ has the same sign as the argument of the absolute value. The second term on the right of (4.3) becomes

$$\begin{aligned}\frac{\delta \mathrm{Tr}(R)}{\delta\mu(i,j)} &= \sum_{[i,j,k']\in U_2} R(i,j,k')\mu(i,j) \\ &\quad \times \frac{[\mu(j,k')^2 + \mu(i,k')^2 - \mu(i,j)^2]}{|\mu(i,j,k')|}. \end{aligned} \tag{4.5}$$

Varying $\langle \nabla A \,|\, \nabla A\rangle$ with respect to $\mu(i,j)$ yields

$$\begin{aligned}&\frac{\delta\langle \nabla A \,|\, \nabla A\rangle}{\delta\mu(i,j)} \\ &= \frac{\delta}{\delta\mu(i,j)} \sum_{[i',j',k']\in U_2} \frac{\langle A|\Delta[i',j',k']\rangle\langle\Delta[i',j',k']|A\rangle}{\mu(i',j',k')} \\ &= -\sum_{[i,j,k']\in U_2} \left\{ \frac{\langle \nabla A \,|\, [i,j,k']\rangle\langle [i,j,k'] \,|\, \nabla A\rangle}{|8\mu(i,j,k')^3|}\,\mu(i,j) \right. \\ &\quad \left. \times [\mu(j,k')^2 + \mu(i,k')^2 - \mu(i,j)^2] \right\} \\ &\quad + \frac{2\langle A \,|\, [i,j]\rangle\langle [i,j] \,|\, \Delta\nabla A\rangle}{\mu(i,j)^2}. \end{aligned} \tag{4.6}$$

Using Maxwell's equations and combining (4.5) and (4.6), the interacting version of Einstein's equations for the simplicial complex can be written

$$\begin{aligned}&\sum_{[i,j,k']\in U_2} \left\{ \frac{\langle [i,j,k'] \,|\, R[i,j,k']\rangle}{|\mu(i,j,k')|} \right. \\ &\quad \left. - \frac{\langle [i,j,k'] \,|\, \nabla A\rangle\langle \nabla A \,|\, [i,j,k']\rangle}{|\mu(i,j,k')^3|} \right\} \\ &\quad \times [\mu(j,k')^2 + \mu(i,k')^2 - \mu(i,j)^2] = 0 \end{aligned} \tag{4.7}$$

for each $[i,j]\in U_1$ where the sum is carried out over k'.

Solutions to (4.7) can be interpreted as approximate solutions to the differential equations of continuous interacting gravity and electrodynamics by a straight-

forward extension of methods described in Ref. 1. Roughly speaking, each 4-simplex is mapped into flat Minkowski space, then continuous fields are constructed from corresponding discrete quantities by smoothly interpolating each from the simplexes on which it is defined onto the remainder of Minkowski space. The details will not be given here.

5. QUANTIZATION

The theory described so far is classical. We will now consider quantization using a Feynman path integral.

If the simplicial complex and its inner product are assumed given, the electromagnetic field by itself can be quantized without great difficulty. For convenience let us restrict the set of points S_0 on which the complex is based to be finite. The various other requirements on the complex, given previously, remain unchanged.

The first candidate for an n-point function is the integral

$$T[A(i_1,j_1)\cdots A(i_n,j_n)] = \prod_{[i,j]\in U_1}\left[\int dA(i,j)\right]\exp(i\langle\nabla A\,|\,\nabla A\rangle)\prod_{l=1}^{n}A(i_l,j_l), \quad (5.1)$$

where $A(i,j)$ is defined by (3.1). But if λ is an arbitrary 0-chain, Eq. (3.6) implies the action $\langle\nabla A\,|\,\nabla A\rangle$ is invariant under the gauge transform

$$A' = A + \nabla\lambda. \quad (5.2)$$

The orbit obtained by applying all such transformations to a fixed A has infinite volume. Therefore, (5.1) diverges. This difficulty can be corrected by an adaptation of the work of Fadeev and Popov.[6]

Since the complex is finite the subspace of 1-chains, Q, which can be expressed in the form $\nabla\lambda$ is finite dimensional. Let the set of 1-chains $a_1, a_2, \cdots$ form a linearly independent basis for Q and let $a_1, a_2, \ldots, A_1, A_2, \cdots$ form a linearly independent basis for the entire space of 1-chains. There are many different ways $A_1, A_2, \cdots$ can be chosen. Each possible choice is equally good for our purposes. An arbitrary 1-chain $c^{(1)}$ can be written in two distinct forms

$$c^{(1)} = \sum_{[i,j]\in U_1} c^{(1)}(i,j)[i,j], \quad (5.3)$$

$$c^{(1)} = \sum_k c^{(1)}(k)A_k + \sum_k c^{(1)\prime}(k)\,a_k. \quad (5.4)$$

The relation between the two sets of coefficients is a linear transformation

$$c^{(1)}(i,j) = \sum_k L(i,j\,|\,k)\,c^{(1)}(k) + \sum_k L'(i,j\,|\,k)\,c^{(1)\prime}(k), \quad (5.5)$$

where

$$L(i,j\,|\,k) = \mu(i,j)^{-1}\langle[i,j]\,|\,A_k\rangle,$$
$$L'(i,j\,|\,k) = \mu(i,j)^{-1}\langle[i,j]\,|\,a_k\rangle.$$

If D is the determinant of the linear transformation (5.5), the integral in (5.1) can be rewritten

$$D\prod_{k_1}\left[\int dA(k_1)\right]\prod_{k_2}\left[\int dA'(k_2)\right]\exp\left[i\sum_{l_1,l_2}A(l_1)A(l_2)\right.$$
$$\left.\times\langle\nabla A_{l_1}\,|\,\nabla A_{l_2}\rangle\right]\prod_{m=1}^{n}\left[\sum_p L(i_m,j_m\,|\,p)A(p)\right.$$
$$\left.+\sum_p L'(i_m,j_m\,|\,p)A'(p)\right]. \quad (5.6)$$

The coefficients $A'(k)$ do not appear in the argument of the exponential in (5.6) since each a_k fulfills $\nabla a_k = 0$. Therefore, the integrals over $dA'(k)$ in (5.6) diverge. We now simply omit these integrals and set each $A'(k)$ in the remainder of (5.6) to 0. This procedure is equivalent to a choice of gauge in canonical quantization of continuous electrodynamics. The n-point functions we obtain will then depend on precisely how the set $A_1, A_2, \cdots$ was chosen, but if the n-point functions for A are used to construct n-point functions for quantities such as ∇A, which are invariant with respect to the gauge transform (5.2), the result can easily be shown independent of the choice of $A_1, A_2, \cdots$. Our replacement for (5.6) is now

$$\prod_k\left[\int dA(k)\right]\exp\left[i\sum_{l_1,l_2}A(l_1)A(l_2)\langle\nabla A_{l_1}\,|\,\nabla A_{l_2}\rangle\right]$$
$$\times\prod_{m=1}^{n}\left[\sum_p L(i_m,j_m\,|\,p)A(p)\right].$$

This expression is still not finite, however, since the equation

$$\sum_l A(l)\Delta\nabla A_l = 0$$

has nonzero solutions and therefore the matrix $\langle\nabla A_{l_1}/\nabla A_{l_2}\rangle$ has 0 as an eigenvalue. This further difficulty can be circumvented by introducing a small "photon mass" ϵ extremely close to 0.

Our final result becomes

$$T[A(i_1,i_1)\cdots A(i_n,j_n)]$$
$$=\prod_k\left[\int dA(k)\right]\exp\left\{i\sum_{l_1,l_2}A(l_1)A(l_2)\right.$$
$$\left.\times[\langle\nabla A_{l_1}\,|\,\nabla A_{l_2}\rangle - \epsilon^2\langle A_{l_1}\,|\,A_{l_2}\rangle]\right\}$$
$$\times\sum_{m=1}^{n}\left[\sum_p L(i_m,j_m\,|\,p)A(p)\right].$$

This is a Gaussian integral and can be expressed using the inverse matrix $M_{k_1k_2}$ defined by

$$\sum_{k_2} M_{k_1k_2}[\nabla A_{k_2}\,|\,\nabla A_{k_3}\rangle - \epsilon^2\langle A_{k_2}\,|\,A_{k_3}\rangle] = \delta_{k_1k_3}.$$

For example, if the vacuum to vacuum amplitude is

$$T = \prod_k\int dA(k)\exp\left\{i\sum_{l_1,l_2}A(l_1)A(l_2)\right.$$
$$\left.\times[\langle\nabla A_{l_1}\,|\,\nabla A_{l_2}\rangle - \epsilon^2\langle A_{l_1}\,|\,A_{l_2}\rangle]\right\},$$

then the normalized 2-point function becomes

$$T[A(i_1,j_1)A(i_2,j_2)]/T$$
$$=\sum_{k_1,k_2}L(i_1,j_1\,|\,k_1)\,L(i_2,j_2\,|\,k_2)\,M_{k_1k_2}.$$

The question which remains is whether path integration can also be used to quantize the simplicial complex's metric, given by the collection of 1-measures $\mu(i,j)$, or the complex's topology, described by the sets

$S_0, \ldots, S_4$. If the complex's topology is fixed, we would expect that the set of $\mu(i,j)$ could be quantized since path integral quantization has been carried through, at least formally, for the metric of the continuous theory.[6,7] Quantization of the discrete theory may actually be simplified to some extent by the absence of frame dependent quantities. Once a path integral has been put together its behavior could perhaps be determined using methods similar to those Wilson[8] has applied to conventional field theories on a lattice. Quantization of the complex's topology, however, is a more difficult question. I hope to return to this subject elsewhere.

After this work was completed, I was informed that a similar formulation of Maxwell's equations on a simplicial complex has been discussed by Sorkin.[9] Sorkin's version is dual to the construction given here in the sense that fields in the present paper are represented by linear combinations of simplexes while Sorkin's fields are functions over the set of simplexes.

*Supported in part by the U.S. Energy Research and Development Administration.

†Present Address: Physics Department, Indiana University, Bloomington, Indiana 47401.

[1]T. Regge, Nuovo Cimento **19**, 558 (1961).

[2]Computer solutions of Regge's theory by itself have been done by P.A. Collins and R.M. Williams, Phys. Rev. D **7**, 965 (1973). See also R. Sorkin, Phys. Rev. D **12**, 385 (1975).

[3]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963).

[4]This was suggested in part by D. Bohm, B.J. Hiley, and A.E.G. Stuart, Int. J. Theor. Phys. **3**, 171 (1970).

[5]A discussion of simplicial complexes is given by P.S. Alexandroff, *Combinatorial Topology* (Graylock, Rochester, 1957).

[6]V.N. Popov and L.D. Faddeev, "Perturbation Theory for Gauge Invariant Fields," Kiev ITP report (unpublished), L.D. Faddeev and V.N. Popov, Phys. Lett. B **25**, 29 (1967).

[7]L.D. Faddeev and V.N. Popov, Usp. Fiz. Nauk **111**, 427–50 (1973) [Sov. Phys.-Usp. **16**, 777 (1974)].

[8]K.G. Wilson and J. Kogut, Phys. Rep. **12**, 75 (1974).

[9]R. Sorkin, J. Math. Phys. **16**, 2452 (1975).

Is the Wick square infinitely divisible?

Klaus Baumann and Gerhard C. Hegerfeldt*

ZiF, Universität Bielefeld, 4800 Bielefeld-Wellenberg 1, Germany
(Received 12 July 1976)

We derive a sort of Lévy–Khintchine formula for the Wick square of the Euclidean free field in two and three dimensions and use it to show that the *Euclidean* Wick square is infinitely divisible. On the other hand, by analyzing the truncated four-point function we prove that the *relativistic* Wick square is not infinitely divisible in any space–time dimension.

1. INTRODUCTION

In a recent paper[1] it was shown that any field can be decomposed into (indecomposable) prime fields and an infinitely divisible field. A field was shown to be infinitely divisible if and only if its truncated n-point functions are conditionally positive definite. If the generating (or characteristic) functional $E(f)$ exists, this is equivalent to $E(f)^{1/n}$ again being a characteristic functional for each natural n.

Clearly, the free and generalized free fields are infinitely divisible since all the truncated n-point functions vanish except the two-point function. In the following we investigate the Wick square as another possible candidate. It turns out that in the Euclidean case (in two and three dimensions) it is infinitely divisible, whereas in the relativistic case (in any dimension) it is not.

This lends support to the conjecture that the relativistic free and generalized free fields are the only infinitely divisible fields satisfying all Wightman axioms. Without the cluster property, however, there are a host of such fields. For example, if $\mathbf{W}=(1, W_1, W_2, \cdots)$ is any Wightman functional then $\hat{W}=\exp(W-1)$ is infinitely divisible and satisfies all Wightman axioms except clustering.[1]

2. INFINITE DIVISIBILITY OF THE EUCLIDEAN WICK SQUARE

Let $H_0=-\Delta+m^2$, where Δ is the two- or three-dimensional Laplacian. The characteristic (generating, expectation) functional $E(f)$, $f\in\mathcal{D}$, of the Euclidean Wick square of a free Euclidean field with mass m is given by

$$E(f)=\langle\exp[i:\phi^2:(f)]\rangle$$
$$=\exp\{-\tfrac{1}{2}\operatorname{Tr}[\ln(1-2ifH_0^{-1})+2ifH_0^{-1}]\}. \tag{2.1}$$

The matrix identity $\det A=\exp(\operatorname{Tr}\ln A)$ implies

$$\exp(\operatorname{Tr}\ln A^{-1}B)=\exp(\operatorname{Tr}\ln B-\operatorname{Tr}\ln A)$$

and thus we obtain

$$E(f)=\exp\{-\tfrac{1}{2}\operatorname{Tr}[(H_0-2if)-\ln H_0+2ifH_0^{-1}]\}. \tag{2.2}$$

It suffices to show that the exponent is conditionally positive definite,[2] i.e., $\sum\lambda_i\bar{\lambda}_j\ln E(f_i-f_j)\geq 0$ if $\sum\lambda_i=0$. To do this we use the identity[3]

$$-\ln z=\int_0^\infty\frac{d\xi}{\xi}\{\exp(-z\xi)-\exp(-\xi)\}$$

valid for $\operatorname{Re}z>0$. Thus, with $H_0^{-1}=\int_0^\infty d\xi\exp(-H_0\xi)$, one gets

$$E(f)=\exp\{\tfrac{1}{2}\int_0^\infty\xi^{-1}d\xi\operatorname{Tr}[\exp(-H_0\xi+2i\xi f)$$
$$-(1+2i\xi f)\exp(-H_0\xi)]\}. \tag{2.3}$$

Let $K_\xi(x,y)$ be the kernel of $\exp(-H_0\xi)$. Note that $K_\xi(x,y)>0$. By the Trotter product formula[4] one has

$$\exp[-H_0\xi+2i(f_i-f_j)\xi](x,y)$$
$$=\lim_{n\to\infty}\int dx_1\cdots dx_{n-1}K_{\xi/n}(x,x_1)$$
$$\times\exp\{2i[f_i(x_1)-f_j(x_1)]\xi/n\}\cdots$$
$$\times K_{\xi/n}(x_{n-1},y)\exp\{2i[f_i(y)-f_j(y)]\xi/n\}.$$

Now, the integrand is positive definite in (i,j) since the $K_{\xi/n}$'s are positive. Equation (2.3) then shows[5] that $\ln E$ is conditionally positive, and thus E is infinitely divisible.

An alternative proof can be based on finite-dimensional approximations of $E(f)$.

Remark: Equation (2.3) resembles a Lévy–Khintchine decomposition of the Euclidean Wick square. The individual components also determine Euclidean fields. In general, however, they cannot satisfy T positivity[6,7] or Osterwalder–Schrader[8] positivity by the result of the next section. Their truncated n-point functions, i.e., the n-point functions of the integrand in Eq. (2.3) can be calculated from the Dyson series

$$\exp[(-H_0+2if)\xi]$$
$$=\sum_{n=0}^\infty\int_0^\xi dt_1\int_0^{t_1}dt_2\cdots\int_0^{t_{n-1}}dt_n\exp(-t_nH_0)(2if)$$
$$\times\exp[-(t_{n-1}-t_n)H_0](2if)\cdots(2if)\exp[-(\xi-t_1)H_0].$$

After removing the first two terms the series converges absolutely in trace norm since $\exp(-tH_0)f$ is the Hilbert–Schmidt. Integrating over $d\xi/\xi$ leads to the truncated n-point functions of the Euclidean Wick square with convergence in trace norm if $\max|f|<m^2$. This justifies, *a posteriori*, the previous steps.

3. THE RELATIVISTIC WICK SQUARE IS NOT INFINITELY DIVISIBLE

We consider the Wick square of a relativistic generalized free field $\phi(x)$ in arbitrary space–time dimension d. We denote $(\Omega,\phi(x_1)\phi(x_2)\Omega)$ by $W(x_2-x_1)$. We assume that the support of the Lehmann spectral function $\rho(m^2)$ has the lowest upper bound $M^2<\infty$. We will show that the truncated four-point function of $:\phi^2:(x)$ is not positive definite.

By elementary calculation we get

$$\tfrac{1}{16}W_4^T(\mathbf{x})$$
$$=W(x_2-x_1)W(x_4-x_3)\{W(x_3-x_2)W(x_4-x_1)$$

$$+W(x_3-x_1)W(x_4-x_2)]$$
$$+W(x_3-x_1)W(x_4-x_1)W(x_3-x_2)W(x_4-x_2)$$
$$=:U(\mathbf{x})+V(\mathbf{x}). \tag{3.1}$$

The Fourier transform of U

$$\tilde{U}(p)=\int\{\delta(p_1+q_1+q_4)\delta(p_2-q_1+q_3)\delta(p_3+q_2-q_3)$$
$$\times\delta(p_4-q_2-q_4)+\delta(p_1+q_1+q_3)\delta(p_2-q_1+q_4)$$
$$\times\delta(p_3+q_2-q_3)\delta(p_4-q_2-q_4)\}\prod_1^4\tilde{W}(q_i)\,dq_i \tag{3.2}$$

vanishes whenever $p_2^2>M^2$ or $p_3^2>M^2$. This is important because U by itself is positive definite. Indeed, for $h\in\mathcal{S}(\mathbb{R}^{2d})$,

$$U(h^*\otimes h)=\|\int dx_1\,dx_2 h(x_1,x_2)W(x_2-x_1)$$
$$\times:\phi(x_1)\phi(x_2):\Omega\|^2\geq 0.$$

Now take $\tilde{f}\in\mathcal{D}(\mathbb{R}^d)$, $\tilde{f}$ positive on $K_\epsilon:=\{p\mid|p|<\epsilon\}$, $\mathrm{supp}f=\bar{K}_\epsilon$, $\tilde{f}(-p)=\tilde{f}(p)$ and define

$$\tilde{g}_i(p):=\tilde{f}(p-a_i),\quad i=1,\ldots,4.$$

Choose[9] $\epsilon=\frac{1}{4}M$, and

$$a_1=(\sqrt{2}+\sqrt{5},\ -3,0,\cdots)M,$$
$$a_2=(\sqrt{2}+\sqrt{5},\ 3,0,\cdots)M,$$
$$a_3=(2\sqrt{5},\ 0,\cdots)M,\quad a_4=(2\sqrt{2},\ 0,\cdots)M.$$

This implies

$$\mathrm{supp}\tilde{g}_3\subset V_{2M}^+ \tag{3.3a}$$

and

$$a_3-\mathrm{supp}\tilde{W}+\bar{K}_\epsilon\cap a_4-\mathrm{supp}\tilde{W}+\bar{K}_\epsilon=\emptyset. \tag{3.3b}$$

With $g=g_1\otimes g_2+\lambda g_3\otimes g_4$, $\lambda\in\mathbb{R}$, we get by (3.3a)

$$\tfrac{1}{16}W_4^T(g^*\otimes g)=\tfrac{1}{16}W_4^T(g_2^*\otimes g_1^*\otimes g_1\otimes g_2)$$
$$+2\lambda\,\mathrm{Re}V(g_2^*\otimes g_1^*\otimes g_3\otimes g_4)$$
$$+\lambda^2V(g_4^*\otimes g_3^*\otimes g_3\otimes g_4). \tag{3.4}$$

We remark that

$$V(g_2^*\otimes g_1^*\otimes g_3\otimes g_4)$$
$$=\int\int dv\,dw\tilde{W}(v)\tilde{W}(w)\int\tilde{f}(u+v-a_2)\tilde{f}(u+w-a_3)\tilde{W}(u)\,du$$
$$\times\int\tilde{f}(u'+w-a_1)\tilde{f}(u'+v-a_4)\tilde{W}(u')\,du' \tag{3.5}$$

is positive, because the integrand is nonnegative, and a neighborhood of the points[10]

$$u=(\sqrt{5},-2,0,\cdots)M,\quad u'=(\sqrt{2},1,0,\circ\circ\circ)M,$$
$$v=(\sqrt{2},-1,0,\cdots)M,\quad w=(\sqrt{5},2,0,\circ\circ\circ)M$$

contributes to the integral.

On the other hand,

$$V(g_4^*\otimes g_3^*\otimes g_3\otimes g_4)$$
$$=\int\int dv\,dw\tilde{W}(v)\tilde{W}(w)\left|\int\tilde{f}(u+v-a_3)\tilde{f}(u+w-a_4)\right.$$
$$\left.\times\tilde{W}(u)\,du\right|^2 \tag{3.6}$$

vanishes by (3.3b). Hence for sufficiently negative λ the right-hand side of (3.4) becomes negative.

*Permanent address: Institut für Theoretische Physik, Universität Göttingen.

[1]G.C. Hegerfeldt, Commun. Math. Phys. **45**, 137—51 (1975).
[2]I.M. Gel'fand and N.Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1964), Vol. 4, Chap. III, §4.
[3]I.S. Gradshteyn and I.M. Ryzkik, *Table of Integrals, Series, and Products* (Academic, New York, 1965), formula 3.551.6.
[4]H.F. Trotter, Proc. Am. Math. Soc. **10**, 545—51 (1959).
[5]The details of this argument require some care since from positivity for almost all x and y one cannot conclude positivity for $x=y$. However, restriction to finite-dimensional spaces generated by characteristic functions of small intervals give the required result by taking suitable limits.
[6]G.C. Hegerfeldt, Commun. Math. Phys. **35**, 155—71 (1974).
[7]J. Fröhlich, Ann. Inst. Poincaré **21**, 271—317 (1974).
[8]J. Osterwalder and R. Schrader, Commun. Math. Phys. **31**, 83—112 (1973).
[9]If $M=0$ (and therefore $d\geq 3$), we choose $\epsilon=\frac{1}{4}$ and $a_1=(3,-3,0,\cdots)$, $a_2=(3,3,0,\cdots)$, $a_3=(4,0,\cdots)$, $a_4=(2,0,\cdots)$.
[10]If $M=0$, the corresponding points are $u=(2,-2,0,\cdots)$, $u'=(1,1,0,\cdots)$, $v=(1,-1,0,\cdots)$, $w=(2,2,0,\cdots)$.

General formalism solving linear recursion relations

Adel F. Antippa*

Département de Physique, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, G9A 5H7, Canada

Alain J. Phares†

Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138
(Received 12 April 1976)

We give the general solution of multiterm linear homogeneous recursion relations with nonconstant coefficients, by introducing a new class of special combinatorics functions based on the partition of an interval into classes. This provides power series solutions, whenever they exist, of ordinary linear differential equations, of any order, with polynomial coefficients.

I. INTRODUCTION

The purpose of this paper is to solve multiterm linear recursion relations with nonconstant coefficients. This provides power series solutions, whenever they exist, of linear differential equations, of any order, with polynomial coefficients. Applications to the nonrelativistic Schrödinger equation with polynomial type central potentials will be published elsewhere.[1]

Whenever the solution of an Nth order ordinary linear differential equation with polynomial coefficients can be expanded in a power series, the differential equation can be transformed into a multiterm linear recursion relation. Since, in general, the resulting recursion relation has nonconstant coefficients, the usefullness of this type of transformation has up till now been restricted to cases where the resulting recursion relation is a two-term one.

For linear recursion relations with *constant* coefficients, general solutions exist already, and are obtained via the roots of the associated characteristic equation.[2] On the other hand, for the case of nonconstant coefficients, the usual approach is to guess the solution and, then, prove it by mathematical induction. This is reasonably easy for two-term recursion relations, but becomes unfeasible for multiterm relations.[3]

The general solution presented here for multiterm linear recursion relations with nonconstant coefficients, is in terms of special combinatorics functions. These functions, rather functionals, are constructed from products of the coefficients with their arguments evaluated at a number of discrete points, determined by the partition of a certain interval into classes. The relation of a combinatorics function to the partitions of an interval, is analogous to the relation of the nth order term in a scattering cross section to the corresponding Feynman graphs.

The partitions of an interval into classes are related to the partitions of the subintervals, differing from it by one summond, by an abstract linear recursion relation. This can be most easily and concretely seen by considering a variation of the "postage problem" of combinatorics theory[2]:

> Given an unlimited number of 3¢, 5¢ and 8¢ stamps, what are the different ordered arrangements leading to a total of m¢?

The totality of ordered arrangements each of which adds up to m¢ can be separated into three sets, according to whether they end by a 3¢, a 5¢, or an 8¢ stamp. Thus,

$$\begin{aligned}\{\textstyle\sum(\text{stamps}) = m\} = &\{\{\textstyle\sum(\text{stamps}) = m-3\}+3\}\\ &\cup\{\{\textstyle\sum(\text{stamps}) = m-5\}+5\}\\ &\cup\{\{\textstyle\sum(\text{stamps}) = m-8\}+8\},\end{aligned}\tag{1.1}$$

where $\{\sum(\text{stamps}) = m\}$ is the set of all stamp arrangements, each of which adds up to m¢, and $\{\sum(\text{stamps}) = m-k\}+k$ means that a k¢ stamp is added to the right of each arrangement in the set $\{\sum(\text{stamps}) = m-k\}$ thus bringing up the total to m¢.

The above Eq. (1.1) is essentially an abstract linear recursion relation, and with a proper choice of mapping can be represented by a four-term linear recursion relation. The proper mapping in general, is shown in Fig. 1, and in the special case discussed above is obtained by the following rules:

(1) Corresponding to each k¢ stamp an arbitrary function $f_k(x)$ is introduced.

(2) Each arrangement of stamps is represented by a functional $F(m)$ formed as a product of the functions $f_k(x)$ in the following manner:

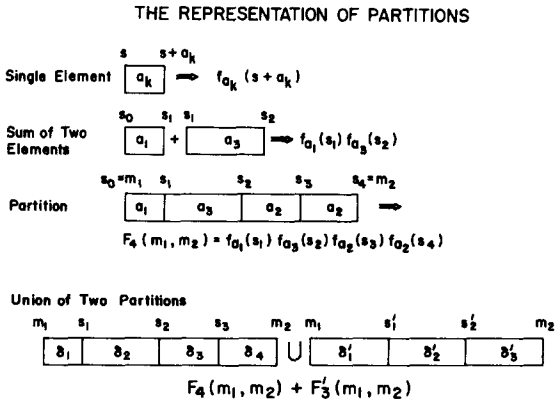


FIG. 1. The partitions of an interval (m_1, m_2) into n parts are represented by the functionals $F_n^q(m_1, m_2)$. The functions $f_{a_k}(x)$, corresponding to the partitioning subintervals a_k, are arbitrary.

(a) For each k¢ stamp in the arrangement, there is one corresponding $f_k(x)$ function in the product.

(b) The argument x, of the function f_k corresponding to the nth stamp in the arrangement is equal to the sum of the first n stamps.

(3) The set $\{\sum(\text{stamp}) = m\}$ is represented by a functional $C(m)$ formed as a sum of the functionals $F(m)$ corresponding to the arrangements in the set.

The above mapping transforms the addition and union operations of Eq. (1.1) into multiplication and addition respectively as shown in Fig. 1. Thus Eq. (1.1) is transformed into

$$C(m) = C(m-3)f_3(m) + C(m-5)f_5(m) + C(m-8)f_8(m), \tag{1.2}$$

which is a four-term linear recursion relation with arbitrary nonconstant coefficients. In Secs. II and III, we will formulate these intuitive ideas rigorously.

II. PARTITIONS AND COMBINATORICS FUNCTIONS

If one is ready to proceed intuitively and sacrifice rigor and precision, the solution of linear recursion relations can be arrived at rather fast and in a simple manner. Nonetheless, our choice is to develop the subject with a certain amount of mathematical rigor and precision of language. It is our belief that such an approach will facilitate further development of the theory.

A. Partitions

In this section we will establish the language relating to the partition of an interval into parts.[4] The parts available for the partition form a basic set, $\mathcal{A}$, of subintervals:

$$\mathcal{A} = \{a_1, \ldots, a_N\}, \quad 0 < a_1 < a_2 < \cdots < a_N. \tag{2.1}$$

In all subsequent work, whenever we mention the partition of an interval into parts, it is to be understood that these parts belong to $\mathcal{A}$.

An interval (m_1, m_2) can be partitioned into n parts belonging to the set $\mathcal{A}$ provided that there exists a set of nonnegative integers,

$$P = \{p_1, \ldots, p_N\}, \tag{2.2}$$

satisfying

$$\sum_{i=1}^{N} a_i p_i = m_2 - m_1. \tag{2.3a}$$

The number of parts, n, is given by

$$n = \sum_{i=1}^{N} p_i. \tag{2.3b}$$

When $m_2 = m_1$, Eq. (2.3a) can still be satisfied provided the p_i's are all zero. This leads through Eq. (2.3b) to $n = 0$. Thus, *an interval of length zero can be partitioned into zero parts.*

The infinite set of real numbers M that can be partitioned into parts belonging to $\mathcal{A}$ will be denoted by $\mathcal{M}$:

$$\mathcal{M} = \{M_1, M_2, \cdots\} \tag{2.4}$$

Furthermore, corresponding to each real number M we form the set $\mathcal{N}(M)$ given by

$$\mathcal{N}(M) = \begin{cases} \{n_1, n_2, \cdots\}, & M \in \mathcal{M}, \\ \emptyset, & M \notin \mathcal{M}, \end{cases} \tag{2.5}$$

where $n_i \in \mathcal{N}(M)$ if and only if M can be partitioned into n_i parts. It is evident that when $M \notin \mathcal{M}$ no partition is possible, by construction of $\mathcal{M}$, and consequently the set $\mathcal{N}(M)$ is empty.

The distinct partitions of $M \in \mathcal{M}$ into a given number of parts $n \in \mathcal{N}(M)$ will be labelled by an index q. We denote by $\mathbf{\Delta}_{nq}(M)$ the n-dimensional vector whose components are the positive subintervals δ_i of the (nq)th partition:

$$\mathbf{\Delta}_{nq}(M) = (\delta_1, \delta_2, \ldots, \delta_n), \quad q = 1, \ldots, q_{\max}(n) \tag{2.6a}$$

where

$$M \in \mathcal{M}, \ n \in \mathcal{N}(M), \ \delta_i \in \mathcal{A}, \tag{2.6b}$$

and

$$\sum_{i=1}^{n} \delta_i = M \ \text{ for } q = 1, \ldots, q_{\max}(n). \tag{2.6c}$$

Here $q_{\max}(n)$ is the number of distinct partitions of M into n parts. According to the above definition, there is a one to one correspondence between distinct partitions (nq) of M and distinct vectors $\mathbf{\Delta}_{nq}(M)$. Whenever the interval (m_1, m_2) cannot be partitioned into n subintervals belonging to $\mathcal{A}$, then, $q_{\max}(n) = 0$, and the vector $\mathbf{\Delta}_{nq}(m_2 - m_1)$ does not exist.

Let $\mathcal{D}(M, n)$ be the set of vectors corresponding to all partitions of M into n parts,

$$\mathcal{D}(M, n) = \{\mathbf{\Delta}_{n_1}(M), \ldots, \mathbf{\Delta}_{nq_{\max}}(M)\} \tag{2.7}$$

and $\bar{\mathcal{D}}(M)$ the set of vectors corresponding to all possible partitions of M:

$$\bar{\mathcal{D}}(M) = \{\mathbf{\Delta}_{nq}(M); \ n \in \mathcal{N}(M); \ q = 1, \ldots, q_{\max}(n)\}, \tag{2.8a}$$

$$\bar{\mathcal{D}}(M) = \bigcup_{n \in \mathcal{N}(M)} \mathcal{D}(M, n). \tag{2.8b}$$

Since there is a one-to-one correspondence between distinct partitions and distinct vectors $\mathbf{\Delta}$, the set of distinct partitions of M into n parts is isomorphic to $\mathcal{D}(M, n)$, and the set of all distinct partitions of M is isomorphic to $\bar{\mathcal{D}}(M)$. Furthermore, we denote by $\mathcal{D}^{\alpha}_{a_k}(M, n)$ the set of all n-dimensional vectors $\mathbf{\Delta}_{nq}(M)$ whose αth component is a_k. We also denote by $\bar{\mathcal{D}}^{in}_{a_k}(M)$ the set of all vectors $\mathbf{\Delta}$, irrespective of their dimension, which have a_k as their first component. Finally, the set of all vectors $\mathbf{\Delta}$, which have a_k as their last component, will be denoted by $\bar{\mathcal{D}}^{f}_{a_k}(M)$.

Let $\mathcal{A}''(M, n)$ be the subset of $\mathcal{A}$ containing all elements a_i not entering any partition of M into n parts, and $\mathcal{A}'(M, n)$ its complement. Then,

$$\mathcal{A} = \mathcal{A}'(M, n) \cup \mathcal{A}''(M, n) \ \text{ and } \ \mathcal{A}'(M, n) \cap \mathcal{A}''(M, n) = \emptyset. \tag{2.9}$$

Similarly, let $\bar{\mathcal{A}}''(M)$ be the subset of $\mathcal{A}$ containing all elements a_i not entering any partition of M whatsoever, and $\bar{\mathcal{A}}'(M)$ its complement. Then again

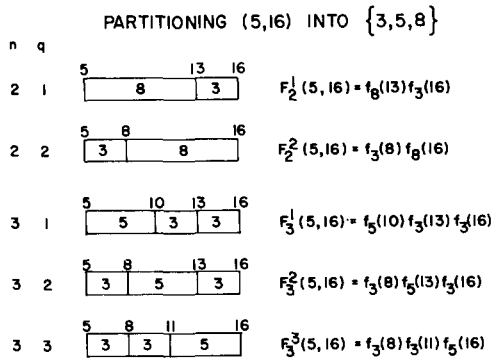


FIG. 2. The partitions of the interval (5,16), into the parts $\{3,5,8\}$, and their corresponding functionals. The parameters, vectors, and sets related to these partitions are given in Table I, and the corresponding combinatorics functions in Table II.

$$\mathcal{A} = \bar{\mathcal{A}}'(M) \cup \bar{\mathcal{A}}''(M) \quad \text{and} \quad \bar{\mathcal{A}}'(M) \cap \bar{\mathcal{A}}''(M) = \emptyset. \tag{2.10}$$

With each possible partition, (nq), corresponding to the vector $\Delta_{nq}(m_2 - m_1)$ we associate the set of $(n+1)$ numbers, $\{s_0, s_1, \ldots, s_n\}$, given by

$$s_i - s_{i-1} = \delta_i, \quad s_0 = m_1, \quad s_n = m_2 \tag{2.11a}$$

or, explicitly,

$$s_0 = m_1, \quad s_i = m_1 + \sum_{k=1}^{i} \delta_k$$

$$\text{for } n \geq 1 \text{ and } 1 \leq i \leq n. \tag{2.11b}$$

In the special case where $m_1 = m_2$, leading to $n = 0$, the set of s-numbers contains only one element, s_0. From the set of values $\{s_0, \ldots, s_n\}$ corresponding to a partition, (nq), of (m_1, m_2), we construct the vector

$$\mathbf{S}_{nq}(m_1, m_2) = (s_0, \ldots, s_n),$$

$$n \in \mathcal{N}(m_2 - m_1), \quad q = 1, \ldots, q_{\max}(n). \tag{2.12}$$

As an illustration, we show in Fig. 2 the partitions of the interval (5, 16) into the parts $\{3, 5, 8\}$. The corresponding parameters, vectors and sets are given in Table I.

B. Combinatorics functions

1. Special combinatorics functions

With the partitioning subintervals[5] $(a_1, \ldots, a_N)$, elements of the set $\mathcal{A}$, we associate a set of N functions conveniently denoted by $f_{a_1}, \ldots, f_{a_N}$. Furthermore, with a possible partition (nq) of (m_1, m_2), we associate the function[6]

$$F_n^q(m_1, m_2) = \prod_{i=1}^{n} f_{\delta_i}(s_i),$$

$$n \in \mathcal{N}(m_2 - m_1), \quad q = 1, \ldots, q_{\max}(n), \tag{2.13a}$$

where the δ_i, $i = 1, \ldots, n$, are the components of $\Delta_{nq}(m_2 - m_1)$, and the s_i, which are related to the δ_i by Eq. (2.11a), are the components of $\mathbf{S}_{nq}(m_1, m_2)$. When no partition of (m_1, m_2) into n parts belonging to $\mathcal{A}$ is possible, then, $n \notin \mathcal{N}(m_2 - m_1)$, $q_{\max}(n) = 0$, and, in this case, we define

$$F_n^q(m_1, m_2) = 0, \quad n \notin \mathcal{N}(m_2 - m_1). \tag{2.13b}$$

Since $q_{\max}(n) = 0$ implies that $n \notin \mathcal{N}(m_2 - m_1)$ and consequently, according to (2.13b), that $F_n^q(m_1, m_2)$ is zero, it is convenient to define:

$$F_n^0(m_1, m_2) \equiv 0. \tag{2.13c}$$

It is interesting to note that, when $m_1 = m_2$, the interval can be partitioned into zero parts, i.e., $n = 0$, in only one way, by setting $p_i = 0$ for $i = 1, \ldots, N$ in Eq. (2.3a). Consequently, $q_{\max}(0) = 1$ for the interval (m_1, m_1). This fact plays a dominant role in the theory and permits choosing $F_0^1(m_1, m_1) \neq 0$ since it is associated with a possible partition. It is convenient to normalize this function according to

$$F_0^1(m_1, m_1) = 1 \tag{2.13d}$$

Equations (2.13a), (2.13d) can be combined into a single equation by introducing an interval $\delta_0 \equiv 0$ and a function $f_0 \equiv 1$. Equations (2.13a), (2.13c), (2.13d) can then be written as

$$F_n^q(m_1, m_2) = \begin{cases} \prod_{i=0}^{n} f_{\delta_i}(s_i), & n \in \mathcal{N}(m_2 - m_1), \ q = 1, \ldots, q_{\max}(n), \\ 0, & n \notin \mathcal{N}(m_2 - m_1), \ q = 0. \end{cases} \tag{2.14}$$

By summing the functions $F_n^q(m_1, m_2)$ over all distinct partitions of (m_1, m_2) into n parts, we obtain the special combinatorics function of the first kind:

$$C_1(m_1, m_2, n) = \sum_{q=0}^{q_{\max}(n)} F_n^q(m_1, m_2). \tag{2.15}$$

On the other hand, by summing the functions, F_n^q, over

TABLE I. The partition parameters, and related vectors and sets, corresponding to the partitioning of the interval (5,16) into the set $\{3,5,8\}$. The partitions are shown in Fig. 2.

Partition parameters	
$a_1 = 3$ $a_2 = 5$ $a_3 = 8$ $N = 3$ $m_1 = 5$ $m_2 = 16$ $m_2 - m_1 = 11$	
Partition vectors	
$\Delta_{21}(11) = (8,3)$	$\mathbf{S}_{21}(5,16) = (5,13,16)$
$\Delta_{22}(11) = (3,8)$	$\mathbf{S}_{22}(5,16) = (5,8,16)$
$\Delta_{31}(11) = (5,3,3)$	$\mathbf{S}_{31}(5,16) = (5,10,13,16)$
$\Delta_{32}(11) = (3,5,3)$	$\mathbf{S}_{32}(5,16) = (5,8,13,16)$
$\Delta_{33}(11) = (3,3,5)$	$\mathbf{S}_{33}(5,16) = (5,8,11,16)$
Partition sets	
$\mathcal{D}(11,2) = \{\Delta_{21}(11), \Delta_{22}(11)\}$	$\mathcal{D}(11,3) = \{\Delta_{31}(11), \Delta_{32}(11), \Delta_{33}(11)$
$\mathcal{D}(11) = \mathcal{D}(11,2) \cup \mathcal{D}(11,3)$	
$\mathcal{D}_3^2(11,2) = \{\Delta_{21}(11)\}$ $\mathcal{D}_5^2(11,2) = \emptyset$	$\mathcal{D}_8^2(11,2) = \{\Delta_{22}(11)\}$
$\mathcal{D}_3^3(11,3) = \{\Delta_{31}(11), \Delta_{32}(11)\}$ $\mathcal{D}_5^3(11,3) = \{\Delta_{33}(11)\}$	$\mathcal{D}_8^3(11,3) = \emptyset$
$Q_1(11,2,k) = \begin{cases} \{1,2\} & \text{for } k = 0,1,2 \\ \{1\} & \text{for } k = 3,4,5,6,7 \\ \emptyset & \text{for } k = 8,9,10,11 \end{cases}$	
$Q_1(11,3,k) = \begin{cases} \{1,2,3\} & \text{for } k = 0,1,2 \\ \{1\} & \text{for } k = 3,4 \\ \emptyset & \text{for } k = 5,6,7,8,9,10,11 \end{cases}$	

all distinct partitions of the interval (m_1, m_2), we obtain the special combinatorics function of the second kind:

$$C_2(m_1, m_2) = \sum_n \sum_{q=0}^{q_{\max}(n)} F_n^q(m_1, m_2). \tag{2.16}$$

From Eq. (2.14) defining $F_n^q(m_1, m_2)$ it is seen that only values of $n \in \mathcal{N}(m_2 - m_1)$ will give a nonzero contribution to the summation in Eq. (2.16). Consequently, Eq. (2.16) can also be written as

$$C_2(m_1, m_2) = \sum_{n \in \mathcal{N}(m_2 - m_1)} C_1(m_1, m_2, n). \tag{2.17}$$

Using Eqs. (2.14), (2.15), and (2.17) and the fact that $f_{\delta_0} = f_0 \equiv 1$, we obtain

$$C_2(m_1, m_1) = C_1(m_1, m_1, 0) = f_0(m_1) \equiv 1. \tag{2.18}$$

2. Constrained combinatorics functions

The special combinatorics functions of the second kind were obtained by summing the functions $F_n^q(m_1, m_2)$ over all distinct partitions of (m_1, m_2). In general we can perform summations of the functions $F_n^q(m_1, m_2)$ over a set of partitions, of (m_1, m_2), satisfying certain constraints. When the constraint is that the partitions have n parts, the summation leads to the special combinatorics functions of the first kind. The class of constraints we will be concerned with in this section is that which requires the αth part in a partition to be greater than a certain value d.

By summing the functions $F_n^q(m_1, m_2)$ over all distinct partitions of (m_1, m_2) into n parts subject to the constraint that the αth part is greater than d, we obtain the constrained combinatorics functions of the first kind,

$$C_1^\alpha(m_1, m_2, n, d) = \sum_{q \in \mathcal{Q}_\alpha(m_2 - m_1, n, d)} F_n^q(m_1, m_2), \tag{2.19}$$

where $\mathcal{Q}_\alpha(m_2 - m_1, n, d)$ is the subset of q's labelling those partitions of (m_1, m_2) into n parts, whose αth part is greater than d.

On the other hand, by summing the functions $F_n^q(m_1, m_2)$ over all distinct partitions, of the interval (m_1, m_2), subject to the constraint that the first part is greater than d, we obtain the constrained combinatorics functions of the second kind:

$$\bar{C}_2(m_1, m_2, d) = \sum_n \sum_{q \in \mathcal{Q}_1(m_2 - m_1, n, d)} F_n^q(m_1, m_2). \tag{2.20}$$

From Eq. (2.14) it is seen that values of $n \notin \mathcal{N}(m_2 - m_1)$ give a zero contribution to the summation in Eq. (2.20). Consequently, Eq. (2.20) can also be written as

$$\bar{C}_2(m_1, m_2, d) = \sum_{n \in \mathcal{N}(m_2 - m_1)} C_1^1(m_1, m_2, n, d). \tag{2.21}$$

Using Eqs. (2.14), (2.19), and (2.21), we obtain the simple but, as will be seen, crucial result

$$\bar{C}_2(m_1, m_1, d) = C_1^1(m_1, m_1, 0, d) = f_0(m_1) \equiv 1. \tag{2.22}$$

Since $a_1 > 0$ is the smallest partitioning subinterval, the constrained combinatorics functions of the first and second kind reduce to the special ones whenever $0 \leq d < a_1$. Specifically, for $d = 0$, we have

TABLE II. Combinatorics functions corresponding to the partitioning of the interval (5,16) into the set $\{3,5,8\}$. The partitions are shown in Fig. 2.

Special—First kind
$C_1(5,16,1) = 0$
$C_1(5,16,2) = F_2^1(5,16) + F_2^2(5,16)$
$C_1(5,16,3) = F_3^1(5,16) + F_3^2(5,16) + F_3^3(5,16)$
$C_1(5,16,n) = 0$ for $n \geq 4$
Special—Second kind
$C_2(5,16) = C_1(5,16,2) + C_1(5,16,3)$
Constrained (with $j_0 = 8 \Rightarrow \delta_1 > 3$)—First kind
$C_1^1(5,16,2,3) = F_2^1(5,16)$
$C_1^1(5,16,3,3) = F_3^1(5,16)$
Constrained (with $j_0 = 8 \Rightarrow \delta_1 > 3$)—Second kind
$\bar{C}_2(5,16,3) = C_1^1(5,16,2,3) + C_1^1(5,16,3,3)$
$\bar{C}_2(5,16,3) = f_8(13) f_3(16) + f_5(10) f_3(13) f_3(16)$

$$C_1^\alpha(m_1, m_2, n, 0) = C_1(m_1, m_2, n), \tag{2.23a}$$

$$\bar{C}_2(m_1, m_2, 0) = C_2(m_1, m_2). \tag{2.23b}$$

According to Eq. (2.19) if $\mathcal{Q}_\alpha(m_2 - m_1, n, d)$ is empty, the constrained combinatorics function of the first kind is zero:

$$C_1^\alpha(m_1, m_2, n, d) = 0 \quad \text{for } \mathcal{Q}_\alpha(m_2 - m_1, n, d) = \emptyset \tag{2.24a}$$

when $n \notin \mathcal{N}(m_2 - m_1)$, then $\mathcal{Q}_\alpha(m_2 - m_1, n, d)$ is empty, and thus

$$C_1^\alpha(m_1, m_2, n, d) = 0 \quad \text{for } n \notin \mathcal{N}(m_2 - m_1). \tag{2.24b}$$

Furthermore, if (m_1, m_2) cannot be partitioned, then the set $\mathcal{N}(m_2 - m_1)$ is empty and consequently $n \notin \mathcal{N}(m_2 - m_1)$ for any n. Hence according to Eq. (2.21)

$$\bar{C}_2(m_1, m_2, d) = 0 \quad \text{for } (m_2 - m_1) \notin \mathcal{M}. \tag{2.25}$$

It is also worthwhile noting that when $m_2 < m_1$, there does not exist a set of nonnegative integers p_i, $i = 1, \ldots, N$ satisfying Eq. (2.3a). Consequently, the interval (m_1, m_2) is not partitionable. Thus according to Eqs. (2.24b) and (2.25),

$$C_1^\alpha(m_1, m_2, n, d) = 0 \quad \text{for } m_2 < m_1 \tag{2.26a}$$

and

$$\bar{C}_2(m_1, m_2, d) = 0 \quad \text{for } m_2 < m_1. \tag{2.26b}$$

As an illustration, the combinatorics functions corresponding to the interval (5, 16) are given in Table II.

C. Fundamental theorems

We will now prove two fundamental theorems concerning the recursion relations obeyed by the constrained combinatorics functions of the first and second kind respectively. To simplify the proofs, we first introduce four lemmas.

Lemma 1: For every $a_k \in \mathcal{A}'(M, n)$, there exists at least one partition, of the interval M into n parts, having its αth part equal to a_k.

Proof: Given a positive integer $\alpha \leq n$ and a partitioning subinterval $a_k \in \mathcal{A}'(M,n)$, we need to prove the existence of a partition (nq) of M into n parts $(\delta_1, \delta_2, \ldots, \delta_\alpha, \ldots, \delta_n)$ with $\delta_\alpha = a_k$. Since $a_k \in \mathcal{A}'(M,n)$, then by construction of $\mathcal{A}'(M,n)$, there exists at least one partition (nq'), of M into n parts, which includes a_k as one of its parts. Let this part be $\delta'_i = a_k$. If $i = \alpha$, then $q = q'$ and the lemma is proved. If $i \neq \alpha$, then the exchange of the intervals δ'_i and δ'_α leads to another possible partition of M into n parts, since the total number of partitions is still n and their sum is still M. The rearranged partition (nq) is given by $\delta_j = \delta'_j$ for $j \neq i$ and $j \neq \alpha$, $\delta_i = \delta'_\alpha$, and $\delta_\alpha = \delta'_i$. Thus $\delta_\alpha = a_k$, and the lemma is proved.

Lemma 2: The set of partitions of M into n parts can be divided into nonempty disjoint subsets which are in one-to-one correspondence with the elements of $\mathcal{A}'(m,n)$.

Proof: Corresponding to each $a_k \in \mathcal{A}'(M,n)$ we form the subset of partitions, of M into n parts, whose αth part is a_k. The elements of the subset corresponding to a_k are represented by the vectors $\Delta_{nq} \in \mathcal{D}^\alpha_{a_k}(M,n)$. On the other hand the partitions of M into n parts are represented by the vectors $\Delta_{nq} \in \mathcal{D}(M,n)$. What we need to prove is the following:

(i) The subsets $\mathcal{D}^\alpha_{a_k}(M,n)$ are nonempty and disjoint,

$$\mathcal{D}^\alpha_{a_k}(M,n) \neq \emptyset \quad \text{for } a_k \in \mathcal{A}'(M,n) \tag{2.27}$$

and

$$\mathcal{D}^\alpha_{a_i}(M,n) \cap \mathcal{D}^\alpha_{a_j}(M,n) = \emptyset \quad \text{for } i \neq j \tag{2.28}$$

(ii) $\mathcal{D}(M,n)$ can be divided into the subsets $\mathcal{D}^\alpha_{a_k}(M,n)$,

$$\mathcal{D}(M,n) = \bigcup_{a_k \in \mathcal{A}'(M,n)} \mathcal{D}^\alpha_{a_k}(M,n) \tag{2.29}$$

(iii) The correspondence between $\mathcal{D}^\alpha_{a_k}(M,n)$ and $a_k \in \mathcal{A}'(M,n)$ is one to one.

According to Lemma 1 the subsets $\mathcal{D}^\alpha_{a_k}(M,n)$ are nonempty for $a_k \in \mathcal{A}'(M,n)$. Thus Eq. (2.27) is established. Furthermore, if $\Delta_{nq} \in \mathcal{D}^\alpha_{a_i}(M,n)$ then its αth component is equal to a_i, $\delta_\alpha = a_i$. Since the a_k's are all different, then, for $j \neq i$, $a_j \neq a_i$ and hence $\delta_\alpha \neq a_j$. Consequently, $\Delta_{nq} \notin \mathcal{D}^\alpha_{a_j}(M,n)$ for any $j \neq i$. Equation (2.28) is thus established.

Every vector belonging to $\mathcal{D}(M,n)$ is made up of parts belonging to $\mathcal{A}'(M,n)$, by construction of the later. Specifically the αth part of every vector $\Delta_{nq} \in \mathcal{D}(M,n)$ is equal to some element $a_k \in \mathcal{A}'(M,n)$ and consequently $\Delta_{nq} \in \mathcal{D}^\alpha_{a_k}(M,n)$. Conversely, for every $a_k \in \mathcal{A}'(M,n)$, a vector $\Delta_{nq} \in \mathcal{D}^\alpha_{a_k}(M,n)$ corresponds to a partition of M into n parts and hence belongs to $\mathcal{D}(M,n)$. Equation (2.29) is thus established.

Finally, since for every $a_k \in \mathcal{A}'(M,n)$ there exists a nonempty subset $\mathcal{D}^\alpha_{a_k}(M,n)$, and since two subsets $\mathcal{D}^\alpha_{a_i}(M,n)$ and $\mathcal{D}^\alpha_{a_j}(M,n)$, corresponding to two different elements of $\mathcal{A}'(M,n)$, are distinct (actually disjoint), then the correspondence between the subsets $\mathcal{D}^\alpha_{a_k}(M,n)$ and the elements a_k of $\mathcal{A}'(M,n)$ is one to one.

Lemma 3: There is a one-to-one correspondence between the partitions of M into n parts and having their αth part equal to a_k, and the partitions of $(M - a_k)$ into $(n-1)$ parts.

Proof: (*i*) *Case* $n = 1$: A partition of M into one part is possible, when $M = a_k$ for some $a_k \in \mathcal{A}$. Since the a_k's are all different, then there exists only one possible partition of M into one part. Furthermore, the interval $M - a_k$ has length zero, and thus can be partitioned into $(n-1) = 0$ parts in one and only one way. Consequently, the one-to-one correspondence is trivially established in this case.

(*ii*) *Case* $n \geq 2$: Consider a partition $(\delta_1, \ldots, \delta_{\alpha-1}, \hat{a}_k, \delta_{\alpha+1}, \ldots, \delta_n)$ of M into n parts with $\delta_\alpha = a_k$. Then the partition

$$(\delta'_1, \ldots, \delta'_{n-1}) = (\delta_1, \ldots, \delta_{\alpha-1}, \delta_{\alpha+1}, \ldots, \delta_n) \tag{2.30}$$

is a partition of $M - a_k$ into $n-1$ parts, since

$$\sum_{i=1}^{n-1} \delta'_i = \left(\sum_{i=1}^{n} \delta_i\right) - \delta_\alpha = M - a_k. \tag{2.31}$$

Furthermore, two distinct partitions, (nq_1) and (nq_2), of M into n parts and having $\delta_\alpha = a_k$ must differ by at least one subinterval, say $\delta^1_l \neq \delta^2_l$, for some $l \neq \alpha$. Thus, they will lead to two distinct partitions, $(n-1, q'_1)$ and $(n-1, q'_2)$, of $(M - a_k)$ into $(n-1)$ parts, according to the correspondence established by Eq. (2.30).

Conversely consider a partition $(\delta'_1, \delta'_2, \ldots, \delta'_{n-1})$ of $(M - a_k)$ into $(n-1)$ parts. Then the partition

$$(\delta_1, \ldots, \delta_n) = (\delta'_1, \ldots, \delta'_{\alpha-1}, a_k, \delta'_\alpha, \ldots, \delta'_{n-1}) \tag{2.32}$$

is a partition of M into n parts since

$$\sum_{i=1}^{n} \delta_i = \left(\sum_{i=1}^{n-1} \delta'_i\right) + a_k = (M - a_k) + a_k. \tag{2.33}$$

Furthermore, it is obvious, according to Eq. (2.32), that two distinct partitions of $(M - a_k)$ into $(n-1)$ parts lead to two distinct partitions of M into n parts.

Thus the correspondence established in Eqs. (2.30) and (2.32) is one to one.

Lemma 4: An interval $(M - a_k)$, with $a_k \in \mathcal{A}$, can be partitioned into $(n-1)$ parts $(n \geq 1)$ if and only if the interval M can be partitioned into n parts and $a_k \in \mathcal{A}'(M,n)$.

Proof: The above lemma can be restated as follows:

$$(n-1) \in \mathcal{N}(M - a_k) \Leftrightarrow n \in \mathcal{N}(M) \quad \text{and} \quad a_k \in \mathcal{A}'(M,n). \tag{2.34}$$

(i) $n \in \mathcal{N}(M)$ *and* $a_k \in \mathcal{A}'(M,n)$: Then, by construction of $\mathcal{A}'(M,n)$, there exists at least one partition, of M into n parts, which includes a_k as one of its parts. The existence of a corresponding partition of $(M - a_k)$ into $(n-1)$ parts is then guaranteed by Lemma 3. Hence $(n-1) \in \mathcal{N}(M - a_k)$.

(ii) $(n-1) \in \mathcal{N}(M - a_k)$: In this case $(M - a_k)$ can be parti-

tioned into $(n-1)$ parts. Then, by Lemma 3, there exist into $(n-1)$ parts. Then, by Lemma 3, there exist corresponding partitions, of M into n parts, having one of their parts equal to a_k. This means that $n \in \mathcal{N}(M)$ and $a_k \in \mathcal{A}'(M,n)$.

Theorem 1: Given an interval (j,m), $j \leq j_0$, and a set $\mathcal{A}$ of partitioning subintervals, then

$$C_1^1(j,m,n,j_0-j) = \sum_{a_k \in \mathcal{A}} f_{a_k}(m) C_1^1(j, m-a_k, n-1, j_0-j) \tag{2.35}$$

for $m > j_0$ and $n \geq 1$.

Proof: According to Eq. (2.9), the summation over a_k in Eq. (2.35) can be separated into two parts;

$$\sum_{a_k \in \mathcal{A}} = \sum_{a_k \in \mathcal{A}'(m-j,n)} + \sum_{a_k \in \mathcal{A}''(m-j,n)}. \tag{2.36}$$

By Lemma 4, Eq. (2.34), if $a_k \notin \mathcal{A}'(m-j,n)$, then $(n-1) \notin \mathcal{N}(m-j-a_k)$. Thus by Eq. (2.24b), $C_1^1(j, m-a_k, n-1, j_0-j) = 0$ for $a_k \in \mathcal{A}''(m-j,n)$. Consequently

$$\sum_{a_k \in \mathcal{A}''(m-j,n)} f_{a_k}(m) C_1^1(j, m-a_k, n-1, j_0-j) = 0. \tag{2.37}$$

In evaluating the sum over the elements of $\mathcal{A}'(m-j,n)$, it is convenient to treat separately the cases $n \in \mathcal{N}(M-j)$ and $n \notin \mathcal{N}(m-j)$.

(i) *The case* $n \notin \mathcal{N}(m-j)$: In this case no partition, of (j,m) into n parts, is possible. Thus, by Lemma 4, Eq. (2.34), no partition of $(j, m-a_k)$ into $(n-1)$ parts is possible either, $(n-1) \notin \mathcal{N}(m-a_k-j)$. Consequently, according to Eq. (2.24b), $C_1^1(j,m,n,j_0-j)$ as well as $C_1^1(j, m-a_k, n-1, j_0-j)$, for all $a_k \in \mathcal{A}$, is zero. Equation (2.35) is thus trivially satisfied.

(ii) *The case* $n \in \mathcal{N}(m-j)$: In this case $\mathcal{N}(m-j) \neq \phi$ and, consequently, $(m-j) \in \mathcal{N}$. We will consider the cases $n=1$ and $n>1$ separately.

(a) $n=1$: Then (j,m) can be partitioned into one part; there exists an $a_t \in \mathcal{A}$ satisfying

$$m-j = a_t \tag{2.38}$$

Since the a_k's are all different then $m-j \neq a_k$ for $k \neq t$ and, hence, $\mathcal{A}'(m-j,1) = \{a_t\}$. Consequently, $\mathcal{Q}_1(m-j, 1, j_0-j)$ is empty unless $a_t > j_0-j$. Thus making use of Eq. (2.38) we have

$$\mathcal{Q}_1(m-j,1,j_0-j) = \begin{cases} \phi & \text{for } m \leq j_0, \\ \{1\} & \text{for } m > j_0. \end{cases} \tag{2.39}$$

Combining Eq. (2.39) with the defining Eqs. (2.14) and (2.19), we obtain

$$C_1^1(j,m,1,j_0-j) = \begin{cases} 0 & \text{for } m \leq j_0, \\ F_1^1(j,m) = f_{a_t}(m) & \text{for } m > j_0. \end{cases} \tag{2.40}$$

On the other hand

$$\sum_{a_k \in \mathcal{A}'(m-j,1)} f_{a_k}(m) C_1^1(j, m-a_k, n-1, j_0-j)$$

$$= f_{a_t}(m) C_1^1(j, m-a_t, 0, j_0-j). \tag{2.41}$$

From Eq. (2.38) we have $m-a_t = j$. Thus, according to Eq. (2.22)

$$C_1^1(j, m-a_t, 0, j_0-j) = 1. \tag{2.42}$$

Combining Eqs. (2.36), (2.37), (2.41), and (2.42), we obtain

$$\sum_{a_k \in \mathcal{A}} f_{a_k}(m) C_1^1(j, m-a_k, 0, j_0-j) = f_{a_t}(m). \tag{2.43}$$

Comparing Eqs. (2.40) and (2.43) we find that Eq. (2.35) holds for $n=1$ provided that $m > j_0$.

(b) $n > 2$: For this case we will present the main steps in the derivation in succession, and justify the passages from one step to the other subsequently. Our starting point is the definition of the constrained combinatorics function of the first kind:

$$C_1^1(j,m,n,j_0-j) = \sum_{q \in \mathcal{Q}_1(m-j,n,j_0-j)} F_n^q(j,m) \tag{2.44a}$$

$$= \sum_{\substack{\Delta_{nq} \in \mathcal{D}(m-j,n) \\ \delta_1 > j_0-j}} F_n^q(j,m) \tag{2.44b}$$

$$= \sum_{a_k \in \mathcal{A}'(m-j,n)} \sum_{\substack{\Delta_{nq} \in \mathcal{D}_{a_k}^n(m-j,n) \\ \delta_1 > j_0-j}} F_n^q(j,m) \tag{2.44c}$$

$$= \sum_{a_k \in \mathcal{A}'(m-j,n)} \left(f_{a_k}(m) \sum_{\substack{\Delta_{nq} \in \mathcal{D}_{a_k}^n(m-j,n) \\ \delta_1 > j_0-j}} \prod_{i=0}^{n-1} f_{\delta_i}(s_i) \right) \tag{2.44d}$$

$$= \sum_{a_k \in \mathcal{A}'(m-j,n)} \left(f_{a_k}(m) \sum_{\substack{\Delta_{n-1,q} \in \mathcal{D}(m-j-a_k,n-1) \\ \delta_1 > j_0-j}} \prod_{i=0}^{n-1} f_{\delta_i}(s_i) \right) \tag{2.44e}$$

$$= \sum_{a_k \in \mathcal{A}'(m-j,n)} \left(f_{a_k}(m) \sum_{\substack{\Delta_{n-1,q} \in \mathcal{D}(m-j-a_k,n-1) \\ \delta_1 > j_0-j}} F_{n-1}^q(j, m-a_k) \right) \tag{2.44f}$$

$$= \sum_{a_k \in \mathcal{A}'(m-j,n)} \left(f_{a_k}(m) \sum_{q \in \mathcal{Q}_1(m-j-a_k,n-1,j_0-j)} F_{n-1}^q(j, m-a_k) \right) \tag{2.44g}$$

or

$$C_1^1(j,m,n,j_0-j) = \sum_{a_k \in \mathcal{A}'(m-j,n)} f_{a_k}(m) C_1^1(j, m-a_k, n-1, j_0-j). \tag{2.45}$$

Summing over the elements of $\mathcal{Q}_1(m-j,n,j_0-j)$ in Eq. (2.44a) guarantees that the summation is performed only over those partitions, of (j,m) into n parts, whose first part δ_1 is greater than j_0-j. A partition (nq) of the interval (j,m) can be represented by the vector $\Delta_{nq}(m-j)$ whose components are the parts of (nq). Hence the summation can be equivalently performed over the subset of vectors Δ_{nq}, belonging to $\mathcal{D}(m-j,n)$, whose first component is greater than j_0-j, as in Eq. (2.44b).

The passage from Eq. (2.44b) to Eq. (2.44c) is allowed by Lemma 2 as expressed by Eqs. (2.27), (2.28), and (2.29), with $\alpha = n$. $\mathcal{D}_{a_k}^n(m-j,n)$ is the subset of vectors belonging to $\mathcal{D}(m-j,n)$ which have their nth com-

ponent equal to a_k. Since the constraint $\delta_1 > j_0 - j$ is on the first component of the vector $\Delta_{nq}(m-j)$, and since we are considering the case $n \geq 2$, then the nth component $\delta_n = a_k$ is unconstrained. Thus the first summation, over the elements of $\mathcal{A}'(m-j, n)$, in Eq. (2.44c) is unconstrained. It is rather the summation over $\Delta_{nq} \in \mathcal{D}^n_{a_k}(m-j, n)$ that is constrained by $\delta_1 > j_0 - j$.

Equation (2.44d) is obtained by using the definition of the function F^q_n as given by Eq. (2.14) and remembering that we are dealing with the case $n \in \mathcal{N}(m-j)$. Furthermore, $f_{\delta_n}(s_n) = f_{a_k}(m)$ has been taken outside the second summation, since the nth component δ_n of all vectors Δ_{nq} entering the summation is a_k, and since $s_n = m$ is the same for all partitions.

The passage from Eq. (2.44d) to Eq. (2.44e) is made using Lemma 3, according to which there is a one-to-one correspondence between the partitions of (j, m) into n parts ending by a_k, and the partitions of $(j, m-a_k)$ into $n-1$ parts. It is worth noting that inside the second summation in Eqs. (2.44d) and (2.44e), the last component of the vectors Δ_{nq} and $\mathbf{S}_{nq}$ is not needed to evaluate the summonds. Thus, in going from a summation over n-dimensional vectors to one over $(n-1)$-dimensional vectors, by dropping the nth component, we do not lose any of the information necessary to evaluate the sum.

Since we are considering the case $n \in \mathcal{N}(m-j)$, and since in Eq. (2.44e), $a_k \in \mathcal{A}'(m-j, n)$, then according to Lemma 4, $(n-1) \in \mathcal{N}(m-j-a_k)$. Thus, using the definition of the function F^q_n as given by Eq. (2.14), and noting that $s_{n-1} = m - a_k$ for all terms in the second summation of Eq. (2.44e), we obtain Eq. (2.45f).

The summation over the subset of vectors $\Delta_{n-1,q}$ whose first component is greater than $j_0 - j$, and which belong to $\mathcal{D}(m-j-a_k, n-1)$, is equivalent to a summation over the partitions $(n-1, q)$ belonging to $\mathcal{Q}_1(m-j-a_k, n-1, j_0-j)$. Thus Eq. (2.44f) can be rewritten as in Eq. (2.44g). Finally Eq. (2.45) is obtained by using the defining equation of the constrained combinatorics functions of the first kind, Eq. (2.19).

Combining Eqs. (2.37) and (2.45), we find that Eq. (2.35) is satisfied for $n \geq 2$. This completes the proof of Theorem 1.

Theorem 2: Given an interval (j, m), $j \leq j_0$, and a set $\mathcal{A}$ of partitioning subintervals, then

$$\bar{C}_2(j, m, j_0 - j) = \sum_{a_k \in \mathcal{A}} f_{a_k}(m) \bar{C}_2(j, m - a_k, j_0 - j) \qquad (2.46)$$

for $m > j_0$.

Proof: In order to achieve a better understanding of the limitations of the theorem, we will treat separately the four cases $m < j$, $m = j$, $j < m \leq j_0$, and $m > j_0$.

(i) *The case* $m < j$: In this case $\bar{C}_2(j, m, j_0 - j) = 0$ according to Eq. (2.26b). Furthermore, since all a_k's are greater than zero, then $m - a_k < j$ for all $a_k \in \mathcal{A}$. Consequently for all $a_k \in \mathcal{A}$, $\bar{C}_2(j, m - a_k, j_0 - j) = 0$ also. Hence Eq. (2.46) is trivially satisfied. Thus, the theorem is also true for $m < j$. But, in this range of m, the theorem has no content and is of no interest.

(ii) *The case* $m = j$: Then $m - a_k < j$, and hence $\bar{C}_2(j, m - a_k, j_0 - j) = 0$ for all $a_k \in \mathcal{A}$ according to Eq. (2.26b). Thus, the right-hand side of Eq. (2.46) is zero. On the other hand, the left-hand side is equal to one according to Eq. (2.22). Hence, Eq. (2.46) does not hold for $m = j$.

(iii) *The case* $j < m \leq j_0$: In this case $m - j \leq j_0 - j$, and hence there are no partitions of (j, m) with $\delta_1 > j_0 - j$. Thus $\mathcal{Q}_1(m-j, n, j_0-j) = \emptyset$ for all n, and $\bar{C}_2(j, m, j_0 - j) = 0$ according to the defining Eq. (2.20). On the other hand, there is no guarantee that $m - j - a_k$ will not turn out to be zero, for some $a_k \in \mathcal{A}$, thus leading to an interval which is partitionable into zero parts. This will permit bypassing the constraint $\delta_1 > j - j_0$ and leads to a $\bar{C}_2(j, m - a_k, j_0 - j) \neq 0$. Hence Eq. (2.46) is not guaranteed to hold for $j < m \leq j_0$.

(iv) *The case* $m > j_0$: Since $j \leq j_0$, and $m > j_0$, then $m - j > 0$, and no partitions of (j, m) into zero parts are possible. Hence in this case $n \geq 1$.

(*a*) $(m-j) \notin \mathcal{N}$: Then (j, m) cannot be partitioned. That is $\mathcal{A}'(m-j, n) = \emptyset$ for any n; none of the elements a_k of $\mathcal{A}$ belongs to $\mathcal{A}'(m-j, n)$. Thus, according to Lemma 4, $(j, m-a_k)$ cannot be partitioned into $n-1$ parts for any n. That is $(m-j-a_k) \notin \mathcal{N}$. Thus according to Eq. (2.25) we have $\bar{C}_2(j, m, j_0-j) = 0$ and $\bar{C}_2(j, m-a_k, j_0-j) = 0$ for $a_k \in \mathcal{A}$, and Eq. (2.46) is again trivially satisfied.

(*b*) $(m-j) \in \mathcal{N}$: In this case Eq. (2.46) is not trivial. We will present the main steps in the proof in succession and give their justification afterwards:

$$\bar{C}_2(j, m, j_0 - j)$$

$$= \sum_{n \in \mathcal{N}(m-j)} C^1_1(j, m, n, j_0 - j) \qquad (2.47a)$$

$$= \sum_{n \in \mathcal{A}(m-j)} \sum_{a_k \in \mathcal{A}} f_{a_k}(m) C^1_1(j, m - a_k, n-1, j_0 - j) \qquad (2.47b)$$

$$= \sum_{a_k \in \mathcal{N}} \left(f_{a_k}(m) \sum_{n \in \mathcal{N}(m-j)} C^1_1(j, m - a_k, n-1, j_0 - j) \right) \qquad (2.47c)$$

$$= \sum_{a_k \in \mathcal{N}} \left(f_{a_k}(m) \sum_{(n-1) \in \mathcal{N}(m-j-a_k)} C^1_1(j, m - a_k, n-1, j_0 - j) \right) \qquad (2.47d)$$

$$\bar{C}_2(j, m, j_0 - j)$$

$$= \sum_{a_k \in \mathcal{A}} f_{a_k}(m) \bar{C}_2(j, m - a_k, j_0 - j). \qquad (2.47e)$$

The starting point Eq. (2.47a) comes from the relation between the constrained combinatorics functions of the first and second kind as given by Eq. (2.21).

As we have already shown, for the case $m > j_0$ at hand, $n \geq 1$ and consequently the conditions for the validity of Theorem 1 are satisfied. Thus using expression (2.35) to replace $C^1_1(j, m, n, j_0 - j)$ leads to Eq. (2.47b).

The summations over $a_k \in \mathcal{A}$ and $n \in \mathcal{N}(m-j)$ are independent of each other and can therefore be exchanged. Furthermore, $f_{a_k}(m)$ is independent of n and can be taken outside the summation over n. We thus obtain Eq. (2.47c).

According to Lemma 4, $(n-1) \in \mathcal{N}(m-j-a_k)$ implies

that $n \in \mathcal{N}(m-j)$, for $a_k \in \mathcal{A}$. Thus

$$\{n;\ (n-1) \in \mathcal{N}(m-j, a_k)\} \subset \{n;\ n \in \mathcal{N}(m-j)\} \tag{2.48}$$

and consequently

$$\sum_{n \in \mathcal{N}(m-j)} C_1^1(j, m-a_k, n-1, j_0-j)$$

$$= \sum_{(n-1) \in \mathcal{N}(m-j, a_k)} C_1^1(j, m-a_k, n-1, j_0-j)$$

$$+ \sum_{\substack{(n-1) \notin \mathcal{N}(m-j-a_k) \\ n \in \mathcal{N}(m-j)}} C_1^1(j, m-a_k, n-1, j_0-j). \tag{2.49a}$$

According to Eq. (2.24b) the summation over $(n-1) \notin \mathcal{N}(m-j-a_k)$, $n \in \mathcal{N}(m-j)$ in Eq. (2.49a) adds up to zero. Hence

$$\sum_{n \in \mathcal{N}(m-j)} C_1^1(j, m-a_k, n-1, j_0-j)$$

$$= \sum_{(n-1) \in \mathcal{N}(m-j-a_k)} C_1^1(j, m-a_k, n-1, j_0-j), \tag{2.49b}$$

substituting Eq. (2.49b) in Eq. (2.47c) we obtain Eq. (2.47d).

Finally by using Eq. (2.21) another time we arrive at Eq. (2.47e). Thus Eq. (2.46) is satisfied. This completes the proof of Theorem 2.

III. LINEAR RECURSION RELATIONS

A. Notation

The general multiterm homogeneous linear recursion relation with nonconstant coefficients can be conveniently written as

$$b_m = \sum_{k=1}^{N} f_{a_k}(m) b_{m-a_k}, \quad m > j_0, \tag{3.1}$$

where the mth term, b_m, is related to N terms of lower rank, $b_{m-a_1}, \ldots, b_{m-a_N}$. The numbers $a_1, \ldots, a_N$ are positive integers assumed to be ordered:

$$0 < a_1 < \cdots < a_N. \tag{3.2}$$

The coefficients entering the recursion relation (3.1) are arbitrary functions of the level m. We have conveniently denoted the coefficient multiplying the term b_{m-a_k}, by $f_{a_k}(m)$.

In general, Eq. (3.1) is valid for $m > j_0$ for some integer j_0. Then the corresponding initial conditions are the values of $b_{j_0}, b_{j_0-1}, \ldots, b_{j_0+1-a_N}$ which we take to be

$$b_{j_0-i} = \lambda_i, \quad i = 0, 1, 2, \ldots, a_N - 1. \tag{3.3}$$

B. General solution

The problem at hand is to give an explicit expression for the coefficient b_m, in terms of the arbitrary functions $f_{a_k}(m)$, and an arbitrary set of parameters $\{\lambda_i;\ i = 0, 1, 2, \ldots, a_N - 1\}$ specifying arbitrary initial conditions. We will give the solution in the form of two theorems.

Theorem 3: The constrained combinatorics function of the second kind $\bar{C}_2(j, m, j_0-j)$ is a solution of the recursion relation (3.1) satisfying the initial conditions

$$b_{j_0-i} = \delta_{j_0-j, i}, \quad i = 0, 1, 2, \ldots, a_N - 1, \tag{3.4}$$

provided that the rank differences a_k are taken as partitioning subintervals and the coefficients $f_{a_k}(m)$ as the corresponding functions.

Proof: Let $\mathcal{A} = \{a_1, a_2, \ldots, a_N\}$. Then Eq. (2.46) can be rewritten as

$$\bar{C}_2(j, m, j_0-j) = \sum_{k=1}^{N} f_{a_k}(m) \bar{C}_2(j, m-a_k, j_0-j), \quad m > j_0. \tag{3.5}$$

If for a given j, included between j_0 and $j_0 - a_N + 1$, we set

$$b_m = \bar{C}_2(j, m, j_0-j), \quad j = j_0, j_0-1, \ldots, j_0-a_N+1 \tag{3.6}$$

in Eq. (3.5), we obtain Eq. (3.1), thus proving that $\bar{C}_2(j, m, j_0-j)$ is a solution of the recursion relation for $m > j_0$. Furthermore,

$$\bar{C}_2(j, j_0-i, j_0-j) = \delta_{j_0-i, j} = \delta_{j_0-j, i} \quad \text{for } i \geq 0. \tag{3.7}$$

For $i = j_0 - j$, this is true due to Eq. (2.22). On the other hand, for $i \neq j_0 - j$ the interval $(j_0 - i, j)$ can neither be partitioned into zero parts, since $j_0 - i - j \neq 0$, nor can it be partitioned into $n \geq 1$ parts having $\delta_1 > j_0 - j$, since $j_0 - i - j \leq j_0 - j$. Thus in this case $\mathcal{Q}_1(j_0 - i - j, n, j_0 - j) = \emptyset$ for all n, and consequently $\bar{C}_2(j, j_0-i, j_0-j) = 0$. Combining Eqs. (3.6) and (3.7), we obtain Eq. (3.4). This completes the proof of Theorem 3.

Theorem 4: The general solution of the recursion relation (3.1) satisfying the initial conditions (3.3) is given by

$$b_m = \sum_{i=0}^{a_N - 1} \lambda_i \bar{C}_2(j_0 - i, m, i), \quad m \geq j_0 - a_N + 1. \tag{3.8}$$

Proof: Since $\bar{C}_2(j, m, j_0-j)$ is a particular solution of the recursion relation, for $j_0 \geq j \geq j_0 - a_N + 1$, then by setting $i = j_0 - j$ we find that $\bar{C}_2(j_0 - i, m, i)$ is a solution of the recursion relation for $0 \leq i \leq a_N - 1$. Furthermore, the recursion relation is linear; hence b_m as given by Eq. (3.8) is also a solution. On the other hand, from Eq. (3.8), we have

$$b_{j_0-i} = \sum_{i'=0}^{a_N-1} \lambda_{i'} \bar{C}_2(j_0 - i', j_0 - i, i')$$

$$= \sum_{j'=j_0}^{j_0-a_N+1} \lambda_{j_0-j'} \bar{C}_2(j', j_0 - i, j_0 - j'), \tag{3.9}$$

where we have made the change of variable $j' = j_0 - i'$. From Eqs. (3.7) and (3.9) we obtain, for $0 \leq i \leq a_N - 1$,

$$b_{j_0-i} = \sum_{j'=j_0}^{j_0-a_N+1} \lambda_{j_0-j'} \delta_{j_0-j', i} = \lambda_i \quad \text{for } i = 0, 1, \ldots, a_N - 1. \tag{3.10}$$

This completes the proof of the theorem.

The notation used for the partitioning of the interval (j, m) is shown in Fig. 3.

IV. CONCLUSION

The general solution of homogeneous linear recursion relations has been presented in terms of functionals which we call constrained combinatorics functions of

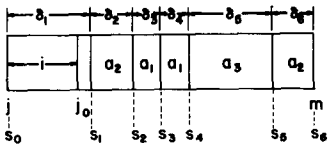


FIG. 3. Notation for the partition of an interval (j,m) into parts, subject to the condition $s_1 > j_0$, or equivalently $\delta_1 > i$.

the second kind. The practical usefulness of such solutions will increase as our understanding of the properties and interrelations of these functionals is improved.

The main results obtained can be summarized as follows: Given the linear recursion relation

$$b_m = \sum_{k=1}^{N} f_{a_k}(m) b_{m-a_k}, \quad m > j_0, \tag{4.1}$$

with the initial conditions

$$b_{j_0-i} = \lambda_i, \quad i = 0, 1, \ldots, a_N - 1. \tag{4.2}$$

The general solution is

$$b_m = \sum_{i=0}^{a_N-1} \lambda_i \bar{C}_2(j_0 - i, m, i), \tag{4.3}$$

where $\bar{C}_2(j_0 - i, m, i)$ is constructed as follows:

(i) Let $j = j_0 - i$ and construct the set $\mathcal{A} = \{a_1, a_2, \ldots, a_N\}$.

(ii) Consider all partitions (nq) of the interval (j, m) into parts δ_l belong to $\mathcal{A}$, subject to the condition $\delta_1 > i$.

(iii) Corresponding to each (nq) partition $(\delta_1, \delta_2, \ldots, \delta_n)$ construct the vector $\mathbf{S}_{nq}(j, m) = (s_0, s_1, s_2, \ldots, s_n)$ whose components are given by

$$s_0 = j, \quad s_p = j + \sum_{l=1}^{p} \delta_l, \quad p \geq 1. \tag{4.4}$$

(iv) Corresponding to each (nq) partition construct the functional

$$F_n^q(j, m) = \prod_{l=1}^{n} f_{\delta_l}(s_l). \tag{4.5}$$

(v) The constrained combinatorics function of the second kind is then given by the sum over all functions corresponding to partitions of (j, m) subject to the constraint $\delta_1 > i$ or equivalently $s_1 > j_0$,

$$\bar{C}_2(j, m, j_0 - j) = \sum_{\substack{n \\ s_1 > j_0}} \sum_{q} F_n^q(j, m). \tag{4.6}$$

Finally,

$$\bar{C}_2(j_0 - i, m, i) = \bar{C}_2(j, m, j_0 - j). \tag{4.7}$$

*Supported in part by the National Research Council of Canada.
†Present address: Department of Physics and Astronomy, University of Montana, Missoula, Montana 59801.
[1]A. F. Antippa and A. J. Phares, "The Linear Potential: An Exact Solution" (to be published).
[2]N. Ya. Vilenkin, *Combinatorics* (Academic, New York, 1971), pp. 74—75 (postage problem), and pp. 135—39 (characteristic equation).
[3]Wilfred Kaplan, *Ordinary Differential Equations* (Addison-Wesley, Reading, Mass., 1958), pp. 350—57.
[4]Henceforth we will use the word "parts" instead of the technical word "classes," used in combinatorics theory.
[5]"Partitioning subinterval" is more expressive for our purpose than the word summond.
[6]This is actually a functional rather than a function.

On integral relations for invariants constructed from three Riemann tensors and their applications in quantum gravity

P. van Nieuwenhuizen*

Institute for Theoretical Physics, State University of New York, Stony Brook, Long Island, New York 11794

C. C. Wu†

Physics Department, University of California, Santa Barbara, California 93106
(Received 28 July 1976)

The lowest order quantum corrections to pure gravitation are finite because there exists an integral relation between products of two Riemann tensors (the Gauss–Bonnet theorem). In this article several algebraic and integral relations are determined between products of three Riemann tensors in four- and six-dimensional spacetime. In both cases, one is left with only one invariant when $R_{\mu\nu}=0$, viz.,

$$\int(-g)^{1/2}(R_{\alpha\beta\mu\nu}R^{\mu\nu\rho\sigma}R_{\rho\sigma}{}^{\alpha\beta}).$$

It is explicitly shown that this invariant does not vanish, even when $R_{\mu\nu}=0$. Consequently, the two-loop quantum corrections to pure gravitation will only be finite if, due to miraculous cancellation, the coefficient of this invariant vanishes.

I. INTRODUCTION

Integral relations for products of Riemann tensors play an important role in quantum gravity because of the following theorem:

Theorem[1]: the leading k-loop divergences of the quantum S matrix for pure Einstein gravitation in n dimensions are of the form

$$S^{\text{div}}(k\text{-loop},\ n\text{-dim})=\frac{G^{k-1}}{(\nu-n)^k}\int d^n x(-g)^{1/2}A(x),$$

where G is Newton's constant, $(\nu-n)$ the parameter which regularizes divergences in quantum field theory according to the dimensional regularization scheme,[2] and

(i) $A(x)$ is a local scalar constructed from fields $g_{\mu\nu}(x)$.

(ii) $A(x)$ does not depend on G.

(iii) The fields $g_{\mu\nu}(x)$ in $A(x)$ satisfy $R_{\mu\nu}(g)=0$.

(iv) $A(x)$ is constructed from $(\frac{1}{2}n+k-l-1)$ Riemann tensors and $2l$ covariant derivatives D_α.

In particular, no expressions of the form R^{-1} or $(D_\mu D^\mu)^{-1}$ are allowed. The factor G^{k-1} is due to the facts that the Einstein action is proportional to G^{-1} and that the graviton field $h_{\mu\nu}$ is related to the metric $g_{\mu\nu}$ by $g_{\mu\nu}=\eta_{\mu\nu}+cG^{1/2}h_{\mu\nu}$ (where c is a constant) because any quantum boson field has the dimension of an inverse mass. From these two facts it follows that propagators are independent of G, while an m-point vertex is proportional to $G^{m/2-1}$. From these observations one easily proves that the leading divergence of a k-loop diagram, expressed as a function of $g_{\mu\nu}(x)$, depends on G as indicated. Finally, since the S matrix is dimensionless, A can only depend on the number of Riemann tensors indicated. The theorem says nothing about nonleading divergences, nor will we.

The simplest application of the theorem is the case of the one-loop diagrams (i.e., the first true quantum corrections) in pure Einstein gravitation in normal four-dimensional spacetime.[3,4] The most general form of $A(x)$ is according to (i) and (iv) given by

$$A=\alpha R^2+\beta R_{\mu\nu}R^{\mu\nu}+\gamma R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}. \tag{1}$$

The Gauss–Bonnet integral identity for arbitrary $g_{\mu\nu}$,

$$\delta\int d^4x(-g)^{1/2}(R_{\mu\nu}{}^{\alpha\beta}R_{\rho\sigma}{}^{\gamma\delta}\epsilon^{\mu\nu\rho\sigma}\epsilon_{\alpha\beta\gamma\delta})=0, \tag{2}$$

can be used to transform the last term in Eq. (1) into the other two as follows. For nonsingular $h_{\mu\nu}(x)$ which vanish sufficiently fast at infinity, one may omit the δ in Eq. (2) and one obtains, by expanding the product of the two ϵ symbols into Kronecker delta functions,

$$\int d^4x(-g)^{1/2}(R_{\mu\nu\alpha\beta}R^{\mu\nu\alpha\beta}-4R_{\mu\nu}R^{\mu\nu}+R^2)=0. \tag{3}$$

Finally, with $R_{\mu\nu}=0$ according to (iii), it follows that $A=0$. One concludes that the lowest order quantum corrections to the S matrix of pure Einstein gravitation are finite. Actually, the restriction $R_{\mu\nu}=0$ raises some problems about the validity of deducing Eq. (3) from Eq. (2), which we discuss in the next section.

In this article we consider the cases of two-loop divergences in four dimensions ($n=4,\ k=2$) and one-loop divergences in six dimensions ($n=6,\ k=1$). Elsewhere the results obtained here will be used to discuss the finiteness of the one but lowest order (two-loop) quantum corrections to the pure gravitational S matrix in four dimensions.

II. PRODUCTS OF THREE RIEMANN TENSORS

In both cases ($n=4,\ k=2$ and $n=6,\ k=1$) the scalar A in the theorem is obtained by contracting three Riemann tensors in all possible ways. Many contractions vanish or are equal due to the symmetries of the Riemann tensor

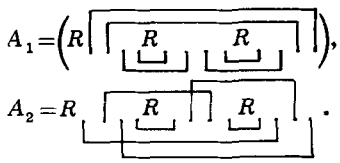
$$R_{\mu\nu\alpha\beta}=R_{\alpha\beta\mu\nu}=-R_{\beta\alpha\mu\nu}=-R_{\alpha\beta\nu\mu},$$
$$R_{[\alpha\beta\mu]\nu}=R_{\alpha\beta[\mu\nu;\delta]}=0. \tag{4}$$

For example, denoting a contraction graphically by a bar, one proves easily from the cyclic relation

$$R_{\gamma\delta}\,\lfloor\,\lfloor\; R \;\rfloor\,\alpha\,\rfloor\,\beta=\tfrac{1}{2}R_{\gamma\delta}\,\lfloor\,\lfloor\; R \;\rfloor\,\rfloor\,\alpha\beta. \tag{5}$$

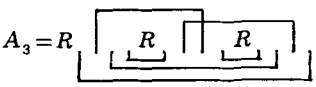
Using these relations, one finds that there are at most

two independent invariants which we display for clarity again in graphical notation



(6)

For example, the third invariant



(7)

is related to the previous two invariants by $A_3 = \frac{1}{4}A_1 + A_2$. Terms containing derivatives are of the form $RDDR$ and can be transformed into A_1 and A_2 after partial integration and use of the Bianchi identities when $R_{\mu\nu}=0$.

In six-dimensional spacetime there exists, however, an *integral relation* between A_1 and A_2, as we now discuss. The Gauss–Bonnet theorem in $2n$ dimensions for arbitrary $g_{\mu\nu}(x)$ reads[4]

$$\delta \int d^{2n}x(-g)^{1/2}(R_{\alpha_1\alpha_2}{}^{\mu_1\mu_2}\cdots R_{\alpha_{2n-1}\alpha_{2n}}{}^{\mu_{2n-1}\mu_{2n}}\epsilon^{\alpha_1\cdots\alpha_{2n}} \times \epsilon_{\mu_1\cdots\mu_{2n}}) = 0. \tag{8}$$

Upon expanding the ϵ tensors into δ tensors and putting $R_{\mu\nu}=0$ as in the introduction, one obtains the relation

$$\int d^6x(-g)^{1/2}(A_1 + 2A_2) = 0. \tag{9}$$

Actually, this relation should be read in the context of its applications to quantum field theory. When $R_{\mu\nu}=0$, one expects $g_{\mu\nu}(x)$ to develop singularities; moreover, it is conceivable that $g_{\mu\nu}(x)$ does not fall off in every direction fast enough for the integral in Eq. (9) to exist. In such cases, one cannot deduce Eq. (9) from Eq. (8) by considering successive variations which connect the given metric $g_{\mu\nu}(x)$ with the vacuum $\eta_{\mu\nu}$. However, in applications to quantum gravity one should interpret Eq. (9) in a perturbative sense. Each field $g_{\mu\nu}(x)$ represents the sum of all tree graphs of conventional Einstein gravitation with physical momenta ($p^2=0$) and physical polarizations ($\epsilon_\mu p^\mu = \epsilon_\mu{}^\mu = 0$) at the end points.[4] This "on-shell" condition is formally equivalent to $R_{\mu\nu}=0$; the tree graphs are in fact the perturbative solution of the differential equation $R_{\mu\nu}=0$. Considering a given tree diagram with N external legs leads to a divergent integral in Eq. (9) with the divergence of the form $\delta^4(\sum_{i=1}^N p_i)$, which (as needed) expresses energy–momentum conservation. The important result now is that the sum of all graphs with N external legs contained in Eq. (9) vanishes "on shell." One can go, formally at least, from Eq. (8) to Eq. (9) because in this perturbation expansion there are no singularities nor does the nonconvergence of Eq. (9) present a problem (it actually is needed as we discussed).

One concludes, therefore, that in six dimensions the most general form of the one-loop divergences of the S matrix is simply given by

$$S^{\text{div}}(1-\text{loop}, 6\,\text{dim}) = \alpha_6(\nu-6)^{-1}\int d^6x(-g)^{1/2}A_1(x), \tag{10}$$

where the constant α_6 can be determined by explicitly evaluating all tree diagrams with a given number of external legs and with one (and only one) vertex given by $A_1(x)$.

In four dimensions, the same symmetry arguments as before show that the leading two-loop divergences are a linear combination of A_1 and A_2 in Eq. (6). But here an *algebraic relation* exists between A_1 and A_2 when $R_{\mu\nu}=0$, which is best derived from spinor formalism.[5,6] Each Lorentz index μ can be decomposed into the direct product of two spinor indices ($SO_4 = SU_2 \times SU_2$), which transform independently. From now on we decompose Lorentz indices as $\alpha=(a,\underline{a})$, $\beta=(b,\underline{b})$, etc. Hence

$$\mu = m\otimes\underline{m}, \quad \mu = 1,4, \quad m=1,2, \quad \underline{m}=1,2. \tag{11}$$

Moreover, we choose a local inertial frame at the point considered. Any 2×2 antisymmetric tensor t^{ab} is proportional to ϵ^{ab} where $\epsilon^{11}=\epsilon^{22}=0$, and $\epsilon^{12}=-\epsilon^{21}=1$. With this observation, the antisymmetries of the Riemann tensor $R_{\alpha\beta\rho\sigma}$ in $(\alpha\beta)$ and $(\rho\sigma)$ lead to the decomposition

$$R_{\alpha\beta\rho\sigma} = A_{ab,rs}\epsilon^{\underline{a}\underline{b}}\epsilon^{\underline{r}\underline{s}} + B_{ab,\underline{r}\underline{s}}\epsilon^{ab}\epsilon^{\underline{r}\underline{s}} + C_{\underline{a}\underline{b},rs}\epsilon^{ab}\epsilon^{\underline{r}\underline{s}} + D_{\underline{a}\underline{b},\underline{r}\underline{s}}\epsilon^{ab}\epsilon^{rs}, \tag{12}$$

where A, B, C and D are symmetric in their first two and last two indices. From $R_{\alpha\beta\rho\sigma}=R_{\rho\sigma\alpha\beta}$ it follows that A and D are symmetric (in the sense that $A_{ab,rs}=A_{rs,ab}$) while $B^T=C$. The ten Einstein equations $\epsilon^{as}\epsilon^{\underline{q}\underline{s}}R_{\alpha\beta\rho\sigma}=0$ yield the ten relations[7]

$$B + C^T = 0, \quad (A_{11,22} - A_{12,12}) + (D_{11,22} - D_{12,12}) = 0$$

while the cyclic identity $R_{\alpha[\beta\rho\sigma]}$ provides one more relation

$$(A_{11,22} - A_{12,12}) - (D_{11,22} - D_{12,12}) = 0. \tag{13}$$

It follows that $B=C=0$ and that A and D are symmetric and traceless (in the sense that $\epsilon^{ar}A_{ab,rs}=0$) 3×3 matrices [combining the two indices (ab) into one index I with values 1, 2, 3]. As expected, one is left with ten independent components for the Riemann tensor under the restriction $R_{\mu\nu}=0$.

One concludes that the only nonvanishing scalars are given by $\text{tr}A^3$ and $\text{tr}D^3$. Terms like $\text{tr}A\,\text{tr}D^2$ vanish because $\text{tr}A$ does, while $\text{tr}(A^2D) = \text{tr}(AD^2) = 0$ as one may verify explicitly, using Eq. (12). In particular, one finds that A_2 vanishes, while $A_1 = \text{tr}(A^3 + D^3)$. The invariant $\text{tr}(A^3 - D^3)$ is proportional to $\epsilon^{\mu\nu\rho\sigma}R_{\rho\sigma\alpha\beta}R^{\alpha\beta\kappa\lambda}R_{\kappa\lambda\mu\nu}$ and cannot occur in the leading divergences since it breaks parity. For the leading divergences of the two-loop corrections of the gravitational S matrix in four dimensions, one finds thus

$$S^{\text{div}}(2\text{-loop}, 4\,\text{dim}) = \alpha_4 G(\nu-4)^{-2}\int d^4x(-g)^{1/2}A_1(x), \tag{14}$$

where the constant α_4 can be determined by calculating all tree graphs with a given number of external lines and one and only one vertex given by $A_1(x)$.

III. DO EXTRA INTEGRAL RELATIONS EXIST WHEN $R_{\mu\nu}=0$?

The question now arises whether the integrals in (10) and (14) themselves vanish (as opposed to their coefficients α_6 and α_4.) It is believed that the only integral identities for scalars build from arbitrary $g_{\mu\nu}(x)$ are

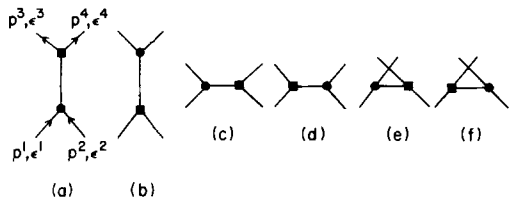


FIG. 1. Pole diagram contributing to the Born amplitude. ● denotes the vertex E_3 in Eq. (17) and ■ denotes the vertex V_3 in Eq. (18a).

the Gauss–Bonnet identities of Eq. (8). Not much seems to be known about similar integral relations under the restriction $R_{\mu\nu}=0$, presumably because, as noted before, such integrals do not exist in the classical sense. The question therefore is how to decide whether the integral of $A_1(x)$ vanishes. One possibility would be to take an exact, everywhere regular solution of Einstein's equations $R_{\mu\nu}=0$, to substitute it into $A_1(x)$ and to integrate (perhaps on a computer) the result—hoping that the integral exists. No such solution is known to us. Instead, we settled the question by doing precisely what the integrals in (10) and (14) are supposed to mean. That is, we replaced each field $g_{\mu\nu}(x)$ by a sum of tree graphs and considered all graphs with a particular number of external legs. The integrals are then trivial, each yielding the same factor $\delta^4(\sum p_i)$, and the question therefore is whether the sum of the finite remainders vanish.

Due to kinematical degeneracy in the case of three external legs, both integrals vanish, as the reader may easily verify. We consider therefore the next case: all tree diagrams with four external legs with (i) one vertex, obtained from $A_1(x)$ by expanding $g_{\mu\nu}=\eta_{\mu\nu}+\kappa h_{\mu\nu}$ (with $k^2=32\pi G$) and expanding $A_1(x)$ into powers of $h_{\mu\nu}$, (ii) all other vertices obtained by expanding the Einstein action in terms of $h_{\mu\nu}$, and (iii) the propagators being the usual propagators of quantum gravity. Denoting by E_N a N-point vertex in Einstein theory and by V_N the N-point vertices provided by $(-g)^{1/2}A_1(x)$, we have contracting all repeated indices with $\eta_{\mu\nu}$

$$\mathcal{L}^E=-2\kappa^{-2}(-g)^{1/2}g^{\mu\beta}g^{\nu\alpha}R_{\mu\nu\alpha\beta},\quad R^{\mu}{}_{\nu\alpha\beta}=\partial_\alpha\Gamma^{\mu}_{\nu\beta}-\cdots, \tag{15}$$

$$P_{\mu\nu,\rho\sigma}=\tfrac{1}{2}\left(\eta_{\mu\rho}\eta_{\nu\sigma}+\eta_{\mu\sigma}\eta_{\nu\rho}-\frac{2}{n-2}\eta_{\mu\nu}\eta_{\rho\sigma}\right)k^{-2},\quad n=\text{dim spacetime}, \tag{16}$$

$$E_3=\kappa(H_{\nu\alpha}-\tfrac{1}{2}H_{\lambda\lambda}\eta_{\nu\alpha})[-\tfrac{1}{4}h_{\beta\mu,\nu}h_{\beta\mu,\nu}+\tfrac{1}{2}h_{\gamma\nu,\beta}(h_{\alpha\beta,\gamma}-h_{\alpha\gamma,\beta})+\tfrac{1}{2}h_{\beta\gamma}(h_{\gamma\nu,\alpha\beta}+h_{\alpha\gamma,\nu\beta}-h_{\alpha\nu,\gamma\beta}-h_{\beta\gamma,\nu\alpha})]. \tag{17}$$

In this expression only the field H is to be contracted when E_3 is used to construct the diagrams of Fig. 1; the fields $h_{\alpha\beta}$ are already taken on-shell:

$$V_3=\kappa^3(h_{\mu\sigma,\nu\rho}+h_{\nu\rho,\mu\sigma}-h_{\mu\rho,\nu\sigma}-h_{\nu\sigma,\mu\rho})\times(h_{\sigma\beta,\rho\alpha}-h_{\sigma\alpha,\rho\beta})(h_{\beta\mu,\alpha\nu}), \tag{18a}$$

$$V_4=-\tfrac{3}{2}\kappa^4h_{\nu\tau}(h_{\beta\mu,\alpha\tau}-h_{\beta\tau,\alpha\mu})(h_{\sigma\beta,\rho\alpha}-h_{\sigma\alpha,\rho\beta})\times(h_{\mu\sigma,\nu\rho}-h_{\nu\sigma,\mu\rho}-h_{\mu\rho,\nu\sigma}+h_{\nu\rho,\mu\sigma})+3\kappa^4(h_{\beta\mu,\alpha\nu}-h_{\beta\nu,\alpha\mu})(h_{\sigma\beta,\alpha\rho}-h_{\rho\beta,\alpha\sigma}-h_{\alpha\sigma,\rho\beta}+h_{\alpha\rho,\sigma\beta})\times[-\tfrac{1}{2}h_{\mu\tau}(h_{\tau\sigma,\nu\rho}-h_{\nu\sigma,\tau\rho})+\tfrac{1}{4}(h_{\nu\tau,\sigma}+h_{\tau\sigma,\nu}-h_{\nu\sigma,\tau})\times(h_{\mu\rho,\tau}-h_{\tau\mu,\rho}-h_{\tau\rho,\mu})]. \tag{18b}$$

Since in tree graphs the dependence on the gauge drops out, one can choose the propagator in any gauge; the Feynman gauge in Eq. (16) is the simplest.

It is now a straightforward matter to calculate the diagrams of Figs. 1 and 2. The calculations were performed on a computer, using an algebraic manipulation program.[8] We start with the case of four-dimensional spacetime. The diagram in Fig. 2 is determined entirely by the contact term V_4 alone, and since all four external legs are on-shell, we have omitted in Eq. (18b) all terms which vanish on-shell. The diagrams in Fig. 1 contain one vertex V_3, one vertex E_3, and one propagator P. The polarization tensors $\epsilon^{\lambda}_{\mu\nu}(p)$ of external gravitons with helicities $\lambda=\pm2$ are given by

$$\epsilon^{\pm2}_{\mu\nu}(p)=\epsilon^{\pm1}_{\mu}(p)\epsilon^{\pm1}_{\nu}(p), \tag{19}$$

where $\epsilon^{\pm1}_{\mu}(p)$ are the polarization tensors for photons with helicities $\lambda=\pm1$.

The results are given in the table and we proceed to interpret them physically. Using parity invariance, time reversal invariance, crossing symmetry, and Bose symmetry, one finds that there are only three independent amplitudes $F(\lambda_3\lambda_4;\lambda_1\lambda_2)$ for the process $\lambda_1,\lambda_2\to\lambda_3,\lambda_4$

$$F(2,2;2,2),\quad F(2,2;2,-2),\quad F(2,2;-2,-2). \tag{20}$$

In particular $F(2,-2;2,-2)=F(2,2;2,2)$ due to (s,u) crossing. The results in I and II of Table I satisfy this criterion and provide a useful check. Note that only the total amplitude satisfies this result and not the seagull or pole graphs separately. This is due to the fact that only the total amplitude is gauge invariant. A second check is gravitational gauge invariance: Replacing any polarization tensor $\epsilon_{\mu\nu}$ by p_μ should give zero. The result in V satisfies also this criterion, which checks the relative signs and constants between V_3, E_3, and V_4. The amplitude $F(2,-2;2,2)$ in III is symmetric under all three crossing operations and should therefore be a symmetric function of s, t, and u. It is indeed. The same result holds for $F(2,2;-2,-2)$ in IV. Incidentally, the nonvanishing of III shows that in massless boson scattering, bosons need not retain their helicities. In Einstein theory, gravitons retain their helicities in Born approximation, but in Ref. 9 it was noted that this is an accident, and not true in general, as verified by III.

In Einstein gravitation kinematical singularities determine the Born amplitudes completely, due to the high spin of the gravitons. In our case the vertices V_3 and V_4 have two factors κ more than the vertices E_3 and E_4; hence we expect a less severe restriction. Using the general considerations of Ref. 9, one finds

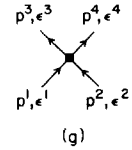


FIG. 2. Seagull diagram contributing to the Born amplitudes. ■ denotes the vertex V_4 in Eq. (18b).

TABLE I. Results for tree graphs in Figs. 1 and 2 in four dimension ($\kappa = 1$).

	Helicities considered	Nonzero pole graphs of Fig. 1	Seagull graph of Fig. 2	Total amplitude
I	$F(2,\ 2;\ 2, 2)$	$a+b=-48iE^8(1+3z^2)$	$g=12iE^6(1+3z^2)$	0
II	$F(2,\ -2;\ 2,\ -2)$	$e+f=-6iE^8(1+z)^5$	$g=-3iE^6(1+z)^4$	0
III	$F(2,\ 2;\ 2,\ -2)$	$\begin{cases} c+d=3iE^8(z-3)(1-z^2)^2 \\ e+f=3iE^8(-z-3)(1-z^2)^2 \end{cases}$	$g=-3iE^6(1-z^2)^2$	$\kappa^6 \frac{3}{8} istu$
IV	$F(2,\ 2;-2,\ -2)$	$\begin{cases} a+b=-48iE^8(1+3z^2) \\ c+d=-6iE^8(1-z)^5 \\ e+f=-6iE^8(1+z)^5 \end{cases}$	$g=6iE^6(11-10z^2-z^4)$	$\kappa^6 \frac{15}{4} istu$
V	$F(2,\ 2;-2,\ -2)$ with $\epsilon'_{\mu\nu}=p'_\mu p'_\nu$	$\begin{cases} a+b=48iE^{10}(z^2-1) \\ c+d=6iE^{10}(-1+3z-2z^2-2z^3+3z^4-z^5) \\ e+f=6iE^{10}(-1-3z-2z^2+2z^3+3z^4+z^5) \end{cases}$	$g=6iE^8(1-z^2)^2$	0
	$s=4E^2,\quad t=-2E^2(1-z),\quad u=-2E^2(1+z)$			
	Total amplitude $=\kappa^6\left(\frac{a+b}{s}+\frac{c+d}{t}+\frac{e+f}{u}+g\right)$			

that the helicity amplitudes of four massless bosons are given by

$$F(\lambda_3, \lambda_4; \lambda_1, \lambda_2) = (stu)^{-1} u^{|\lambda+\mu|/2} t^{|\lambda-\mu|/2} s^{|\Sigma\lambda_i|/2} \phi$$
$$\lambda = \lambda_1 - \lambda_2, \quad \mu = \lambda_3 - \lambda_4, \tag{21}$$

where s, t, and u are the Mandelstam invariants and $s+t+u=0$. The first factor is due to the three poles in the exchange graphs, the u and t factors give the singularities in the scattering angles, the s factors determine the singularities in the CM energy, while the remaining function ϕ is a polynomial in s, t, and u of that degree which makes F dimensionless. Since in our case each F is proportional to G^3, one has

$$\begin{aligned} &F(2, 2;2, 2) = (G^3/stu)s^4\phi_A, \quad \phi_A = a\Sigma^2, \\ &F(2, 2;2, -2) = (G^3/stu)(ut)^2 s^2 \quad \phi_B = b\Sigma^0, \\ &F(2, 2;-2, -2) = (G^3/stu)\phi_C, \\ &\phi_C = c\Sigma^6 + d(\Sigma^3)^2 + e(\Sigma^2)^3, \\ &\Sigma^n = s^n + t^n + u^n. \end{aligned} \tag{22}$$

We see from the table that these constraints are, indeed, satisfied. In particular $F(2, 2;2, -2)$ is completely determined up to the over-all constant $b = 3/8i$.

In the case of six dimensional spacetime the propagator acquires an *extra* term $\frac{1}{4}\eta_{\mu\nu}\eta_{\rho\sigma}$, which describes scalar exchange. However, we found by explicit calculation that the n-dependence of the propagator through $2/n-2$ is completely eliminated by the n-dependence of the contractions[2]

$$\eta_{\mu\nu}\eta_{\mu\nu} = n \tag{23}$$

so that the helicity amplitudes in four and six dimensions are equal. In fact, embedding four-dimensional spacetime into n-dimensional spacetime, the helicity amplitudes are n-independent for any n, since contractions of Eq. (23) at the E_3 vertex eliminate the n dependence of the propagator, while contractions at the V_3 vertex are n-independent.

One concludes that also in six-dimensional spacetime there are nonvanishing helicity amplitudes with one vertex $A_1(x)$.

IV. CONCLUSIONS

The leading divergences of the gravitational S matrix in the absence of matter are in both cases of one-loop corrections in six-dimensional spacetime and two-loop corrections in four-dimensional spacetime proportional to the spacetime integral of

$$\Delta\!\!\!/ = (-g)^{1/2} A_1(x) = (-g)^{1/2}(R_{\alpha\beta}{}^{\rho\sigma} R_{\rho\sigma}{}^{\mu\nu} R_{\mu\nu}{}^{\alpha\beta}).$$

This result was obtained from simple tensor algebra, together with the Gauss—Bonnet integral identity in six dimensions and an identity derived from spinor formalism in four dimensions.

By explicitly calculating the four-point on-shell Born graphs with one vertex given by $\Delta\!\!\!/$ and (in the case of one-pole diagrams) a second vertex given by the Einstein action, we showed that some helicity amplitudes vanish, while some do not. The results satisfy the checks of gauge invariance and kinematical singularities. This means that the integral of $\Delta\!\!\!/$ over spacetime does not vanish, so that no extra integral relation exists, even when $R_{\mu\nu}=0$.

These results mean that in quantum gravity calculations the coefficients of these nonrenormalizable divergences have to vanish if the S matrix is to be finite. For the one-loop divergences of the S matrix in six dimensions the coefficient $\Delta\!\!\!/$ [denoted in Eq. (10) by α_6] has been calculated and does not vanish.[8] The results of this paper mean that there is no easy way to conclude that the two-loop corrections in four dimensions are finite, but that instead direct hard work is needed to decide whether the coefficient of $\Delta\!\!\!/$ [denoted by α_4 Eq. (14)] vanishes.

ACKNOWLEDGMENTS

It is a pleasure to thank R. Geroch, G. 't Hooft, M. Veltman, and C. N. Yang for very useful discussions.

*Supported in part by National Science Foundation Grant #MPS-74-13208 A01.
†Supported in part by National Science Foundation.

[1]For details see P. van Nieuwenhuizen, in *Proceedings of the Marcel Grosman Conference* (North-Holland, Amsterdam, 1976).
[2]G. 't Hooft and M. Veltman, Nucl. Phys. B 44, 189 (1972).
[3]B.S. DeWitt, *Relativity, Groups and Topology* (Gordon and Breach, London, 1964); R. Bach, Math. Z. 9, 110 (1921); C. Lanczos, Ann. Math. 39, 842 (1938).
[4]G. 't Hooft and M. Veltman, Ann. Inst. H. Poincaré 20, 69 (1974).
[5]R. Penrose, in *Battelle Rencontres* 1967, *Lectures in Mathematics and Physics* (Benjamin, New York, 1968).
[6]F.A.E. Pirani, in *Lectures in General Relativity, Brandeis* 1964 *Summer Institute, Vol.* 1 (Prentice-Hall, Englewood Cliffs, New Jersey, 1965).
[7]Contraction proceeds as indicated since two spin-$\frac{1}{2}$ particles couple to a scalar as $|\frac{1}{2},m>|\frac{1}{2},s>\epsilon^{ms}$, while both spin must be coupled separately, since they transform independently.
[8]M. Levine, Ashmedai, Carnegie-Mellon Institute Report.
[9]M.T. Grisaru, P. van Nieuwenhuizen, and C.C. Wu, Phys. Rev. D 12, 397 (1975).
[10]P. van Nieuwenhuizen, Ann. Phys. (N.Y.) (to be published).

Comment on "New Jacobian theta functions and the evaluation of lattice sums" by I.J. Zucker [J. Math. Phys. 16, 2189 (1975)]

Murray Geller*

Jet Propulsion Laboratory, Pasadena, California
(Received 18 August 1976)

In a recent article by Zucker[1] pertaining to the evaluation of lattice sums, the q series

$$\theta_5 = 2q^{1/4}(1 - q^2 - q^6 + q^{12} + q^{20} \cdots), \tag{1}$$

was derived such that

$$\theta_5^2 = 2\theta_2(q^2)\,\theta_4(q^2), \tag{2}$$

where θ_2 and θ_4 are two of the Jacobian theta functions with zero argument, $\theta_2(0,q)$ and $\theta_4(0,q)$, respectively. The comment was made that the series θ_5 does not appear to have been considered by Jacobi. It is clear that, although θ_5 cannot be represented in terms of a linear combination of Jacobian theta functions of zero argument, it is easily seen that

$$\theta_5(q) = \sqrt{2}\,\theta_1(\pi/4, q) = \sqrt{2}\,\theta_2(\pi/4, q). \tag{3}$$

Hence, the full array of addition and multiplication formulas can be utilized to yield results involving θ_5 and the remaining theta functions, several examples of which are

$$\theta_5\,\theta_4(q^4) = \theta_2\,\theta_4(q^2) = \tfrac{1}{2}\theta_2(q^{1/2})\,\theta_5(q^{1/2}), \tag{4a}$$

$$\theta_5^4 = 2\theta_3\,\theta_4(\theta_3^2 - \theta_4^2) = 4\theta_3\,\theta_4\,\theta_2^2(q^2), \tag{4b}$$

$$\theta_5\,\theta_5(q^3) = \theta_3\,\theta_4(q^3) - \theta_4\,\theta_3(q^3), \tag{4c}$$

and

$$2^{-1/3}\,\theta_1'^{1/3}\,\theta_5^3 = 3\theta_1'(q^3)\,\theta_5(q^{1/3}) - \theta_1'(q^{1/3})\,\theta_5(q^3). \tag{4d}$$

We note that Tolke[2] had earlier recognized that of the ten products of theta functions, $\theta_i(z,q)\,\theta_j(z,q)$, for $i \leq j \leq 4$, only two were unable to be expressed in terms of a linear combination of theta functions with the argument nz, viz., $\theta_1(z,q)\,\theta_3(z,q)$ and $\theta_2(z,q)\theta_4(z,q)$. This led him to define the quantities

$$\theta_5(z,q) = 2\theta_1(z,q)\,\theta_3(z,q)/\theta_2^{1/2} \tag{5a}$$

$$\theta_6(z,q) = 2\theta_2(z,q)\,\theta_4(z,q)/\theta_2^{1/2}\,\theta_4^{1/2}\,\theta_4^{1/2}, \tag{5b}$$

from which we readily recognize that

$$\sqrt{2}\,\theta_{5_{\text{Zucker}}} = \theta_6(0,q^2)_{\text{Tolke}}. \tag{6}$$

In addition, Tolke presents many formulas relating θ_5 and θ_6 to the four Jacobian theta functions.

Finally, the series θ_5 of (1) can be summed analytically for selected values of the argument q:

$$q = \exp(-\pi/2),$$
$$\theta_5 = \Gamma(1/4)/2^{1/4}\,\pi^{3/4} = 1.291\,996\,007\cdots, \tag{7a}$$

$$q = \exp(-\pi\sqrt{2}/2), \tag{7b}$$
$$\theta_5 = (\sqrt{2}-1)^{3/8}\,\Gamma(1/8)/[2\pi\Gamma(1/4)]^{1/2} = 1.134\,229\,386\cdots$$

$$q = \exp(-\pi\sqrt{3}/2), \tag{7c}$$
$$\theta_5 = 3^{1/8}[\Gamma(1/3)]^{3/2}/2^{2/3}\,\pi = 1.008\,666\,841\cdots,$$

Many other analytical summations for θ_5 can be obtained by application of the multiplication formulas for theta functions.

*This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract #NAS7-100, sponsored by the National Aeronautics and Space Administration.

[1] I.J. Zucker, J. Math. Phys. 16, 2189 (1975).

[2] F. Tolke, *Praktische Funktionenlehre* (Springer-Verlag, Berlin, 1966), Vol. 2, Sec. 16.

ERRATA

Erratum: The spectrum of the Liouville–von Neumann operator [J. Math. Phys. 17, 57 (1976)]

Herbert Spohn

Fachbereich Physik der Universität München, München, Germany
(Received 4 August 1976)

(1) In Theorem 1, the first, the third, and the fourth set on the right side of the equalities should in each case be replaced by the closure of the same set.

(2) In the proof of Theorem 1, (a) should be replaced by: Since L_{ij} is multiplication by $x-y$ on $L^2(R^2, d\mu_i \times \mu_j)$

$$\sigma(L_{ij}) = \overline{\{x-y \mid x \in \sigma(H_i), y \in \sigma(H_j)\}}.$$

(3) I am grateful to J. Kyle for pointing out the error.

CUMULATIVE AUTHOR INDEX

All authors published so far in the current volume are listed alphabetically with the issue and page numbers following the dash. A cumulative author and subject index covering each volume is published annually. An (E) *after the page number indicates an Erratum.*

Antippa, Adel F.–(1) 173

Baumann, Klaus–(1) 171
Beckers, J.–(1) 72
Bricmont, J.–(1) 37

Christensen, Mark J.–(1) 113
Corones, James–(1) 163

Dassios, George–(1) 126
Debacker-Mathot, F.–(1) 37
Dobson, J.–(1) 116
Doebner, H.D.–(1) 138

Eimerl, David–(1) 106

Folland, N.O.–(1) 31
Fröman, Nanny–(1) 96
Fröman, Per Olof–(1) 96

Geller, Murray–(1) 187
Gilmore, R.–(1) 17

Hegerfeldt, Gerhard C.–(1) 171
Hietarinta, Jarmo–(1) 65
Hsu, J.P.–(1) 100

Jang, Pong Soo–(1) 41

Kalnins, E.G.–(1) 1
Ko, M.–(1) 58

Lehnigk, Siegfried H.–(1) 104
Lockwood, Loren A.–(1) 45
Lücke, W.–(1) 138

Mac, E.–(1) 100
Mansouri, Freydoon–(1) 52
Miller, W., Jr.–(1) 1

Nahm, W.–(1) 146, 155
Newman, C.M.–(1) 23
Newman, E.T.–(1) 58

Oliveira, Colber G.–(1) 120

Patera, J.–(1) 72
Penrose, R.–(1) 58
Perroud, M.–(1) 72
Phares, Alain J.–(1) 173
Pick, Štěpán–(1) 118

Rittenberg, V.–(1) 146, 155
Rund, Hanno–(1) 84

Scheunert, M.–(1) 146, 155
Schulman, L.S.–(1) 23
Schwartz, Charles–(1) 110
Spohn, Herbert–(1) 188 (E)

van Nieuwenhuizen, P.–(1) 182

Wald, Robert M.–(1) 41
Wang, A.P.–(1) 47
Weingarten, Don–(1) 165
Winternitz, P.–(1) 72
Wu, C.C.–(1) 182

