Riccati equations and perturbation expansions in quantum mechanics

George W. Rogers
Department of Physics and Astronomy, University of South Carolina, Columbia, South Carolina 29208 and Naval Surface Weapons Center, Dahlgren, Virginia 22448

(Received 17 July 1984; accepted for publication 21 September 1984)

General perturbation expansions, which allow corrections to any order to be written in quadrature, are presented for Riccati and other nonlinear first-order equations. These results are valid for eigenfunctions which are free of poles and zeros. A Riccati equation suitable for a Schrödinger or Klein–Gordon particle in a central field is expanded for a general state, with corrections to all orders expressed in quadrature. A general Riccati equation for a meromorphic eigenfunction leads to a similar expansion with corrections to all orders, including corrections to the zeros and simple poles, expressed in quadrature. This form is suitable for a Dirac particle in a central field but is more general. The general results are applied to specific examples from the literature.

I. INTRODUCTION

The Riccati equation formulation of central field problems in quantum mechanics has attracted increasing attention over the last two decades. Price[1] observed some thirty years ago that when the one-dimensional Schrödinger equation is reduced to Riccati form, the perturbative solution may be obtained in quadrature to any order in terms of the unperturbed solution and the perturbative potential. This was rediscovered later by Polikanov,[2] and later, independently by Aharonov and Au.[3] Polikanov[4] in another paper pointed out that in excited states, where the wave function has nodes, modifications become necessary. In this case, the expansion is straightforward but somewhat tedious,[3,4] so that orders higher than first have not been done for a general excited state. That is one of the topics of this paper. One way to circumvent this difficulty is to apply the first-order perturbation iteration method introduced by Hirschfelder[5] in connection with Rayleigh–Schrödinger perturbation theory, first applied to a Riccati equation in quantum mechanics by Au.[6] Au and Aharonov[7] have shown that, by considering the logarithmic derivative of the wave function of the Klein–Gordon equation in one-space and one-time, one can obtain an expansion of the same nature as the nonrelativistic (Schrödinger) case for perturbations of either scalar or fourth-component vector type. The Dirac equation may be reduced to Riccati form for central fields by considering the ratio of the radial wave function components. This was first done by Mikhailov and Polikanov,[8] who then obtained a perturbation expansion of the resulting Riccati equation. This was generalized to excited states by Au and Rogers,[9] to include scalar potentials by Au,[10] and to include anomalous magnetic moment interactions as well as simultaneous perturbations of more than one type by Rogers.[11] The Breit equation for two Dirac particles, in the absence of external fields, has been reduced to an independent Riccati equation for the $J=0$ states recently by Rogers,[12] which leads to a perturbation expansion similar to that obtained for the Dirac equation.

The perturbation expansion of a nonlinear first-order eigenvalue equation is considered in the first part of Sec. II for the set of solutions which are free of both zeros and simple poles. This is followed by a more specific treatment of the Riccati equation for the same set of restricted solutions. In Sec. III, a Riccati equation suitable for either a Schrödinger or a Klein–Gordon particle, but with a slightly more general form, is expanded for a set of solutions, including those for a general excited state; and the corrections to any order for a general state are expressed in quadrature. In Sec. IV, the general Riccati equation for a meromorphic function is expanded, with the corrections to any order expressed in quadrature. The problem of zeros in the coefficients is included in the treatment, which is applicable to both the Dirac equation and the Breit equation mentioned above. The results allow portions of the results of a number of papers[9–11] to be written out with minimal effort. More importantly, the final equations developed here allow one to write out the explicit higher-order corrections for excited states previously omitted in favor of the first-order perturbation iteration method. In Sec. V, several examples are presented and in Sec. VI, I make some concluding remarks.

II. PERTURBATION EXPANSION FOR NONLINEAR FIRST-ORDER EQUATIONS

A. The general problem

The problem considered here is that of a nonlinear ordinary first-order eigenvalue equation which can be written in the form

$$a(r,v,E)R(r)' + \sum_{n=0}^{N} b_n(r,v,E)R(r)^n = 0, \tag{2.1}$$

where v is a function of r, $R(r)$ is the eigenfunction, E is the eigenvalue, and N is the order of nonlinearity. The subsequent treatment assumes that an initial solution to this equation is known. That is, for a_0, b_{n0}, and v_0 there exists R_0 and E_0 such that

$$a(r,v_0E_0)R_0(r)' + \sum_{n=0}^{N} b_{n0}(r,v_0E_0)R_0(r)^n = 0. \tag{2.2}$$

The corrections to R_0 and E_0 are sought for a small change in the function $v(r)$ from v_0 to

$$v(r) = v_0(r) + \lambda v_1(r), \tag{2.3}$$

where v_1 will be referred to as the perturbation and λ is a number, to be referred to as the coupling constant, and plays an important role in the expansions to be developed.

The conventional method of obtaining a perturbation expansion for an eigenvalue problem is to expand both the

 0022-2488/85/040567-09$02.50

eigenfunction and the eigenvalue in power series of the coupling constant. Accordingly, I assume that for some range of λ about $\lambda = 0$, both R and E may be represented by the series expansions

$$R = \sum_{n=0}^{\infty} \lambda^n R_n, \tag{2.4}$$

and

$$E = \sum_{n=0}^{\infty} \lambda^n E_n. \tag{2.5}$$

If $a(r,v,E)$ and $b_n(r,v,E)$ can be represented by a polynomial expansion in v and E, this then leads to the expansions

$$a = \sum_{n=0}^{\infty} \lambda^n a_n, \tag{2.6}$$

and

$$b_n = \sum_{i=0}^{\infty} \lambda^i b_{ni}. \tag{2.7}$$

If Eqs. (2.3)–(2.7) are substituted into Eq. (2.1) the following expansion results:

$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \lambda^{n+m} a_n R'_m + \sum_{n=0}^{\infty} \lambda^n b_{0n} + \sum_{k=1}^{N}\sum_{n=0}^{\infty} \lambda^n b_{kn} \prod_{i=1}^{k} \left(\sum_{n_i=0}^{\infty} \lambda^{n_i} R_{n_i} \right) = 0. \tag{2.8}$$

By collecting terms of a given power of λ, one obtains a hierarchy of equations to any order in λ. In particular, each of these equations is linear and of the form

$$a_0 R'_n + \left[\sum_{k=0}^{N} k b_{k0} R_0^{(k-1)} \right] R_n + p_n + q_n = 0. \tag{2.9}$$

The term p_n may be chosen so that it contains the functions a_n and b_{kn} and hence E_n, whereas q_n would consist of only the lower-order functions, in both cases of overall order n. The exact form of p_n and q_n will be worked out only for the case $N = 2$, which is done below. These equations are in general inhomogeneous linear first-order differential equations, which allow the solutions to be written in quadrature. By starting with the first order or $n = 1$ corrections and performing the integrals at each succeeding order, one may work up the hierarchy to any desired order.

B. Expansion of the Riccati equation

The general Riccati equation is given by

$$a(r,v,E)R' + b(r,v,E)R^2 + c(r,v,E)R + d(r,v,E) = 0. \tag{2.10}$$

I assume that a,b,c and d can all be expanded in λ as was a in Eq. (2.6), and R can be expanded as in Eq. (2.4). The Riccati equation analog of Eq. (2.8) is

$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \lambda^{n+m}(a_n R'_m + c_n R_m) + \sum_{n=0}^{\infty} \lambda^n d_n + \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\sum_{k=0}^{\infty} \lambda^{n+m+k} b_n R_m R_k = 0. \tag{2.11}$$

By collecting all terms of a given order of λ, I obtain a hierarchy of equations with the nth-order equation given by

$$\sum_{m=0}^{n} (a_m R'_{n-m} + c_m R_{n-m}) + d_n + \sum_{m=0}^{n} b_m \left(\sum_{k=0}^{n-m} R_k R_{n-k-m} \right) = 0. \tag{2.12}$$

By arranging terms this can be written as

$$a_0 R'_n + (c_0 + 2b_0 R_0) R_n + p_n + q_n = 0, \tag{2.13}$$

where

$$p_n = d_n + a_n R'_0 + c_n R_0 + b_n R_0^2 \tag{2.14}$$

and

$$q_n = \sum_{m=1}^{n-1} (a_m R'_{n-m} + c_m R_{n-m} + b_0 R_m R_{n-m}) + \sum_{m=1}^{n-1} b_m \sum_{k=0}^{n-m} R_k R_{n-k-m}. \tag{2.15}$$

Thus, as can be seen from Eqs. (2.14) and (2.15), p_n contains E_n and q_n contains only quantities of order less than n.

C. Solution of the linear equations

I present in this section the solution to Eq. (2.13) which was derived from the Riccati equation (2.10). The generalization to Eq. (2.9) is easily made and will be indicated at the end of this section. The validity of the equations that follow is restricted to initial solutions with neither zeros nor poles, and $a_0(r)$ is assumed to be free of zeros also. In Secs. III and IV the results will be generalized to include these complications for Riccati equations which are encountered when dealing with excited states in quantum mechanics.

The solution to Eq. (2.13) involves the integrating factor defined by

$$\rho_0(r) = \exp\left[\int \frac{(c_0 + 2b_0 R_0)}{a_0} dr' \right]. \tag{2.16}$$

It is convenient to define an intermediate function G_0 by

$$G_0(r) = \exp\left[\int \frac{b_0}{a_0} R_0 \, dr' \right]. \tag{2.17}$$

In terms of G_0, the integrating factor is given by

$$\rho_0(r) = \exp\left[\int \frac{c_0}{a_0} dr' \right] G_0(r)^2. \tag{2.18}$$

The nth-order correction to the energy can then be obtained from the definite integral

$$\int_0^{\infty} \rho_0(r) \left[\frac{p_n(r) + q_n(r)}{a_0(r)} \right] dr = 0. \tag{2.19}$$

The nth-order correction to $R_0(r)$ is easily verified to be given by

$$R_n(r) = \frac{-1}{\rho_0(r)} \int_0^r \rho_0(r') \left[\frac{p_n(r') + q_n(r')}{a_0(r')} \right] dr'. \tag{2.20}$$

This gives the nth-order correction to any Riccati equation of the form of Eq. (2.10) for the restricted set of solutions under consideration.

For the more general equation (2.1), the corrections Eqs. (2.19) and (2.20) are correct if the following integrating factor is used instead of Eq. (2.16):

$$\rho_0(r) = \exp\left[\int_0^r \sum_{k=1}^{N} \frac{k b_{k0} R_0^{(k-1)}}{a_0} dr' \right]. \tag{2.21}$$

III. EXPANSION FOR SCHRÖDINGER AND KLEIN–GORDON PARTICLES

The Riccati equations derived from both the one-dimensional Schrödinger equation[1–3] and the one-space and one-time Klein–Gordon[7] equation are of the form

$$R(r)' - R(r)^2 + d(r,v,E) = 0, \tag{3.1}$$

where d is some function of r, $v(r)$, and the eigenvalue E. The functional form of d is left unspecified until the examples presented in Sec. V. The function R is the logarithmic derivative of the wave function for both types of particles. Thus, if the wave function has zeros, R has simple poles and is of the form

$$R = Q + \sum_{i=1}^{N} \frac{1}{(r-\alpha^i)}, \tag{3.2}$$

where Q is a function free of both zeros and poles, α^i is the position of the ith pole of R or zero of the wave function, and N is the number of poles of R. In what follows, I will make use of

$$f = \prod_{i=1}^{N} (r-\alpha^i), \tag{3.3}$$

so that

$$\frac{f'}{f} = \sum_{i=1}^{N} \frac{1}{(r-\alpha^i)}. \tag{3.4}$$

With Eqs. (3.2)–(3.4), Eq. (3.1) can be written in terms of Q as

$$Q' - Q^2 + 2(f'/f)Q - f''/f + d = 0. \tag{3.5}$$

I assume that Q can be expanded in power series in λ as was R in Sec. II. Each α^i is expanded in the same fashion and this allows the expansion of f to be carried out. I choose to express this expansion in terms of f_n and $\bar{f}_n$, where

$$f_0 = \prod_{i=1}^{N} (r-\alpha_0^i), \tag{3.6a}$$

$$f_n = \sum_{i=1}^{N} \frac{-\alpha_n^i}{(r-\alpha_0^i)} f_0, \quad n \geqslant 1, \tag{3.6b}$$

and $\bar{f}_n$ is uniquely defined through

$$f = f_0 + \lambda f_1 + \sum_{n=0}^{\infty} \lambda^n (f_n + \bar{f}_n), \tag{3.7}$$

with $\bar{f}_0$ and $\bar{f}_1$ both zero. Thus $\bar{f}_n$ contains only terms individually of order less than n. For example, $\bar{f}_2$ is given by

$$\bar{f}_2 = \sum_i \sum_{j\neq i} \frac{\alpha_1^i \alpha_1^j}{(r-\alpha_0^i)(r-\alpha_0^j)} f_0, \tag{3.8}$$

and higher orders are straightforward to write down.

To eliminate any poles from the Riccati equation (3.5) it is sufficient to multiply that equation by f. Once this has been done, Q, f, and d may all be expanded as indicated above and various powers of λ can be collected to obtain the linear equations. The zero-order (nonlinear) equation is just

$$f_0 Q_0' - f_0 Q_0^2 + 2f_0' Q_0 - f_0'' + d_0 f_0 = 0. \tag{3.9}$$

By defining

$$\Sigma_1(Q_n) = f_0 Q_n' - 2f_0 Q_0 Q_n + 2f_0' Q_n \tag{3.10}$$

and

$$\Sigma_2(f_n) = f_n Q_0' - f_n Q_0^2 + 2Q_0 f_n' - f_n'' + f_n d_0, \tag{3.11}$$

it is possible to write the nth-order equation as

$$\begin{aligned} &\Sigma_1(Q_n) + \Sigma_2(f_n) + d_n f_0 + \Sigma_2(\bar{f}_n) \\ &\quad + \sum_{i=1}^{n-1} (f_i + \bar{f}_i)[Q'_{(n-i)} + d_{(n-i)}] \\ &\quad + \sum_{i=1}^{n-1} 2(f_i' + \bar{f}_i')Q_{(n-i)} \\ &\quad - \sum_{i,j,k=0}^{n-1} (f_i + \bar{f}_i)Q_j Q_k \delta(i+j+k-n) = 0. \end{aligned} \tag{3.12}$$

Here I have used a δ function, which has the value $\delta(0) = 1$ and is zero otherwise. The integrating factor for this equation is

$$\rho_0(r) = f_0^2(r)\exp\Big[-\int_0^r 2Q_0\, dr'\Big]. \tag{3.13}$$

With the zero-order equation (3.9) and Eq. (3.13), one finds that

$$f_0\Sigma_1(Q_n) = \frac{f_0^2}{\rho_0}[Q_n\rho_0]', \tag{3.14}$$

$$f_0\Sigma_2(f_n) = \frac{f_0^2}{\rho_0}\Big[\sum_{i=1}^{N} \frac{-\alpha_n^i}{(r-\alpha_0^i)^2}\rho_0(r)\Big]', \tag{3.15}$$

and

$$f_0\Sigma_2(\bar{f}_n) = \frac{f_0^2}{\rho_0}\Big[(\bar{f}_n f_0' - \bar{f}_n' f_0)\frac{\rho_0}{f_0^2}\Big]'. \tag{3.16}$$

These relations allow the nth-order equation to be rewritten as

$$\begin{aligned} &[Q_n\rho_0]' - \sum_{i=1}^{N} \alpha_n^i \Big[\frac{\rho_0}{(r-\alpha_0^i)^2}\Big]' \\ &\quad + \Big[(\bar{f}_n f_0' - \bar{f}_n' f_0)\frac{\rho_0}{f_0^2}\Big]' + d_n\rho_0 \\ &= \frac{\rho_0}{f_0^2}\Big[-\sum_{i=1}^{n-1} f_0(f_i + \bar{f}_i)(Q'_{n-i} + d_{n-i}) \\ &\quad - \sum_{i=1}^{n-1} 2f_0(f_i' + \bar{f}_i')Q_{n-i} \\ &\quad + \sum_{i,j,k=0}^{n-1} f_0(f_i + \bar{f}_i)Q_j Q_k \delta(i+j+k-n)\Big]. \end{aligned} \tag{3.17}$$

This equation can be integrated from $r = 0$ to $r = \infty$, whence, on making use of the physical requirement that ρ_0 is zero at both limits, one has

$$\int_0^\infty d_n\,\rho_0\, dr = \int_0^\infty [\mathrm{rhs}]dr, \tag{3.18}$$

where [rhs] represents the right-hand side of Eq. (3.17). Once this integral has been performed, the value of E_n is determined.

The nth correction to the ith node of the wave function is given by

$$\begin{aligned} \alpha_n^i &= \frac{\bar{f}_n(\alpha_0^i)}{\Pi_{j\neq i}(\alpha_0^i - \alpha_0^j)} + \frac{\exp\big[\int_0^{\alpha_0^i} 2Q_0\, dr\big]}{\Pi_{j\neq i}(\alpha_0^i - \alpha_0^j)^2} \\ &\quad \times \Big\{\int_0^{\alpha_0^i} d_n\rho_0\, dr - \int_0^{\alpha_0^i} [\mathrm{rhs}]dr\Big\}. \end{aligned} \tag{3.19}$$

The nth-order correction to $Q(r)$ is found by integrating Eq. (3.17) from zero to r. This results in the expression

$$\rho_0(r)Q_n(r) = \sum_{i=1}^{N} \frac{\alpha_n^i \rho_0(r)}{(r-\alpha_0^i)^2} - [\bar{f}_n(r)f_0(r)' - \bar{f}_n(r)'f_0(r)]\,\frac{\rho_0(r)}{f_0(r)^2} - \int_0^r d_n\rho_0\,dr' + \int_0^r [\text{rhs}]\,dr'. \quad (3.20)$$

When $r=\alpha_0^i$, $f_0(\alpha_0^i)=0$, and hence $\rho_0(\alpha_0^i)=0$, which means that $Q_n(\alpha_0^i)$ cannot be determined directly from Eq. (3.20). This problem can be circumvented by referring back to Eq. (3.10) and noting that at $r=\alpha_0^i$

$$\Sigma_1[Q_n(\alpha_0^i)] = \prod_{\substack{j\neq i\\ j=1}}^{N} (\alpha_0^i - \alpha_0^j)Q_n(\alpha_0^i). \quad (3.21)$$

This is then the only term in the nth-order Eq. (3.12) which involves $Q_n(\alpha_0^i)$. That equation, along with Eq. (3.21) above, permits the evaluation of $Q_n(\alpha_0^i)$ once the other nth-order corrections have been determined from Eqs. (3.18) and (3.19). This observation has not previously appeared in the literature on the Schrödinger and Klein–Gordon Riccati equations. This seems to be a much simpler alternative to the method suggested by Privman.[13]

IV. THE RICCATI EQUATION AND EXPANSION FOR A MEROMORPHIC FUNCTION

The eigenfunction of the Riccati equation for a Dirac particle in a central field is the ratio of the two radial wave function components.[8,9] Thus, for a Dirac particle, the function R is a ratio of two functions, each of which has a finite number of zeros. This means that R is a meromorphic function with a finite number of zeros and also a finite number of simple poles. If there are N_1 simple poles and N_2 zeros in R, it can be written

$$R = \frac{\Pi_{i=1}^{N_2}(r-\beta^i)}{\Pi_{i=1}^{N_1}(r-\alpha^i)}\,Q(r), \quad (4.1)$$

where $Q(r)$ is free of both zeros and poles, β^i is the position of the ith zero, and α^i is the position of the ith simple pole on the real line. To simplify the expressions that follow, I define f and g by

$$f = \prod_{i=1}^{N_1} (r-\alpha^i), \quad (4.2a)$$

and

$$g = \prod_{i=1}^{N_2} (r-\beta^i). \quad (4.2b)$$

The generalization of the Riccati equation for a Dirac particle that I shall use is

$$a(r,v,E)R(r)' + b(r,v,E)R(r)^2 + c(r,v,E)R(r) + d(r,v,E) = 0. \quad (4.3)$$

This form is appropriate for not only a particle satisfying the central field Dirac equation, but also for two Dirac particles in a $J=0$ state of the Breit equation in the absence of external fields.[12] With this latter application in mind, I wish to include the possibility that $a(r,v,E)$ may be zero for some value of r.

I assume that the variable coefficients a, b, c, and d in Eq. (4.3) are rational functions of $v(r)$ and energy E. For a known solution R_0 satisfying

$$a_0(r,v_0,E_0)R_0' + b_0(r,v_0,E_0)R_0^2 + c_0(r,v_0,E_0)R_0 + d_0(r,v_0,E_0) = 0, \quad (4.4)$$

and a perturbation represented by λv_1 so that the new potential function $v(r)$ is given by

$$v(r) = v_0(r) + \lambda v_1(r). \quad (4.5)$$

I assume that Q, α^i, β^i, E, a, b, c, and d can all be represented by power series in λ. Specifically,

$$Q = \sum_{n=0}^{\infty} \lambda^n Q_n, \quad (4.6a)$$

$$\alpha^i = \sum_{n=0}^{\infty} \lambda^n \alpha_n^i, \quad (4.6b)$$

$$\beta^i = \sum_{n=0}^{\infty} \lambda^n \beta_n^i, \quad (4.6c)$$

$$E = \sum_{n=0}^{\infty} \lambda^n E_n, \quad (4.6d)$$

$$a(r,v,E) = \sum_{n=0}^{\infty} \lambda^n a_n(r, v_m, E_m;\ m\leqslant n), \quad (4.7a)$$

$$b(r,v,E) = \sum_{n=0}^{\infty} \lambda^n b_n(r, v_m, E_m;\ m\leqslant n), \quad (4.7b)$$

$$c(r,v,E) = \sum_{n=0}^{\infty} \lambda^n c_n(r, v_m, E_m;\ m\leqslant n), \quad (4.7c)$$

$$d(r,v,E) = \sum_{n=0}^{\infty} \lambda^n d_n(r, v_m, E_m;\ m\leqslant n). \quad (4.7d)$$

Here, the functions a_n, b_n, c_n, and d_n may, in general, be functions of not only v_n and E_n but also v_m and E_m, for $m\leqslant n$. I note that in normal expansions only v_0 and v_1 appear, but generalization to v_n is easily accomplished and will be assumed here. From the definitions of f and g and the expansion of α^i and β^i, the expansions of f and g are easily carried out and may be written in the form

$$f = \sum_{n=0}^{\infty} \lambda^n (f_n + \bar{f}_n), \quad (4.8a)$$

and

$$g = \sum_{n=0}^{\infty} \lambda^n (g_n + \bar{g}_n), \quad (4.8b)$$

where f_n and $\bar{f}_n$ are defined as in Sec. III and g_n and $\bar{g}_n$ are defined analogously.

If Eq. (4.3) is expanded directly as written, there is a problem in the neighborhood of each unperturbed pole of R. This is easily circumvented by multiplying the equation by f^2 before expanding. The result of this multiplication is

$$afgQ' + bg^2Q^2 + [a(fg' - f'g) + cfg]Q + df^2 = 0. \quad (4.9)$$

With the expansion (4.5)–(4.8) one easily obtains the different-order equations. The zero-order equation is

$$a_0 f_0 g_0 Q_0' + b_0 g_0^2 Q_0^2 + [a_0(f_0 g_0' - f_0' g_0) + c_0 f_0 g_0]Q_0 + d_0 f_0^2 = 0. \quad (4.10)$$

Before writing the nth-order terms, it is convenient to define a delta function such that $\delta(0)=1$ and $\delta(m)=0$ for $m\neq 0$. Then the nth-order terms arising from $afgQ'$ are

$$afgQ'|_{n\text{th-order}} = a_n f_0 g_0 Q_0' + a_0(f_n g_0 + f_0 g_n)Q_0' + a_0 f_0 g_0 Q_n' + a_0(\bar f_n g_0 + f_0 \bar g_n)Q_0' + \sum_{i,j,k,l=0}^{n-1} a_i(f_j + f_j)(g_k + \bar g_k) \times Q_l'\delta(i+j+k+l-n). \quad (4.11)$$

The nth-order terms arising from bg^2Q^2 are

$$bg^2Q^2|_{n\text{th-order}} = b_n g_0^2 Q_0^2 + 2b_0 g_0 g_n Q_0^2 + 2b_0 g_0^2 Q_{n0} Q_n + 2b_0 g_0 \bar g_n Q_0^2 + \sum_{i,j,k,l,m=0}^{n-1} b_i(g_j + \bar g_j)(g_k + \bar g_k)Q_l Q_m \times \delta(i+j+k+l+m-n). \quad (4.12)$$

The term linear in Q in Eq. (4.9) gives the terms

$$[a(fg'-f'g)+cfg]Q|_{n\text{th-order}} = Q_n[a_0(g_0' f_0 - g_0 f_0') + c_0 f_0 g_0] + Q_0[a_n(f_0 g_0' - f_0' g_0) + c_n f_0 g_0] + Q_0[a_0(f_0 g_n' - f_0' g_n) + c_0 f_0 g_n] + Q_0[a_0(f_n g_0' - f_n' g_0) + c_0 f_n g_0] + Q_0[a_0(f_0 \bar g_n - f_0' \bar g_n) + c_0 f_0 \bar g_n] + Q_0[a_0(\bar f_n g_0' - \bar f_n' g_0) + c_0 \bar f_n g_0] + \sum_{i,j,k,l=0}^{n-1} Q_i\{a_j[(f_k + \bar f_k)(g_l' + \bar g_l') - (f_k' + \bar f_k') \times (g_l + \bar g_l)] + c_j(f_k + \bar f_k)(g_l + \bar g_l)\} \times \delta(i+j+k+l-n). \quad (4.13)$$

The last term in Eq. (4.9) gives

$$df^2|_{n\text{th-order}} = d_n f_0^2 + 2d_0 f_0 f_n + 2d_0 f_0 \bar f_n + \sum_{i,j,k=0}^{n-1} d_i(f_j + \bar f_j)(f_k + \bar f_k)\delta(i+j+k-n). \quad (4.14)$$

I define the function X_0 through its logarithmic derivative

$$\frac{X_0'}{X_0} = \frac{c_0}{a_0} + \frac{2b_0 g_0 Q_0}{a_0 f_0} - \frac{2f_0'}{f_0}. \quad (4.15)$$

By using the zero-order Eq. (4.10), this can also be expressed as

$$\frac{X_0'}{X_0} = \frac{b_0 g_0}{a_0 f_0} - \frac{Q_0'}{Q_0} + \left(\frac{g_0'}{g_0} + \frac{f_0'}{f_0}\right) - \left(\frac{d_0}{a_0}\right)\left(\frac{f_0}{g_0}\right)\left(\frac{1}{Q_0}\right). \quad (4.16)$$

With the use of Eq. (4.15) the terms of Eqs. (4.11)–(4.14) involving Q_n are found to be equal to

$$a_0 f_0 g_0 Q_n' + 2b_0 g_0^2 Q_0 Q_n + Q_n[a_0(f_0 g_0' - f_0' g_0) + c_0 f_0 g_0] = (a_0/X_0)[X_0 f_0 g_0 Q_n]'. \quad (4.17)$$

By making use of Eq. (4.16), the terms of the nth-order equation involving f_n are found to be equal to

$$a_0 f_n g_0 Q_0' + 2d_0 f_0 f_n + Q_0[a_0(f_n g_0' - f_n' g_0) + c_0 f_n g_0] = (a_0/X_0)[-X_0 f_n g_0 Q_0]'. \quad (4.18)$$

Similarly, the terms with g_n are found to be equal to

$$a_0 f_0 g_n Q_0' + 2b_0 g_0 g_n Q_0^2 + c_0 f_0 g_n Q_0 + a_0(f_0 g_n' - f_0' g_n)Q_0 = (a_0/X_0)[X_0 f_0 g_n Q_0]'. \quad (4.19)$$

I next define

$$\Sigma_4 = a_n f_0 g_0 Q_0' + b_n g_0^2 Q_0^2 + d_n f_0^2 + Q_0[a_n(f_0 g_0' - f_0' g_0) + c_n f_0 g_0], \quad (4.20)$$

$$\Sigma_5 = a_0(\bar f_n g_0 + f_0 \bar g_n)Q_0' + 2b_0 g_0 \bar g_n Q_0^2 + 2d_0 f_0 \bar f_n + Q_0[a_0(f_0 \bar g_n' - f_0' \bar g_n) + c_0 f_0 \bar g_n] + Q_0[a_0(\bar f_n g_0 - \bar f_n' g_0) + c_0 \bar f_n g_0], \quad (4.21)$$

and

$$\Sigma_6 = \sum_{i,j,k,l=0}^{n-1} a_i(f_j + \bar f_j)(g_k + \bar g_k)Q_l'\delta(i+j+k+l-n) + \sum_{i,j,k,l,m=0}^{n-1} b_i(g_j + \bar g_j)(g_k + \bar g_k)Q_l Q_m \times \delta(i+j+k+l+m-n) + \sum_{i,j,k=0}^{n-1} d_i(f_j + \bar f_j)(f_k + \bar f_k)\delta(i+j+k-n) + \sum_{i,j,k,l=0}^{n-1} Q_i\{a_j[(f_l + \bar f_l)(g_k' + \bar g_k') - (f_l' + \bar f_l')(\bar g_k + \bar g_k)] + c_j(f_l + \bar f_l)(g_k + \bar g_k)\} \times \delta(i+j+k+l-n). \quad (4.22)$$

With the use of Eqs. (4.17)–(4.22), the nth-order equation can be written as

$$(a_0/X_0)[X_0 f_0 g_0 Q_n]' + (a_0/X_0)[X_0 f_0 g_n Q_0]' - \frac{a_0}{X_0}[X_0 f_n g_0 Q_0]' + \Sigma_4 + \Sigma_5 + \Sigma_6 = 0. \quad (4.23)$$

On multiplying this equation by (X_0/a_0) and integrating from 0 to r, I have

$$X_0 f_0 g_0 Q_n + (f_0 g_n - f_n g_0)X_0 Q_0 + \int_0^r \frac{X_0}{a_0}(\Sigma_4 + \Sigma_5 + \Sigma_6)dr' = 0. \quad (4.24)$$

If the integral in this equation is taken to $r=\infty$, then by making use of the fact that for physical solutions $X_0 = 0$ at $r=\infty$, I have

$$\int_0^r \frac{X_0}{a_0}(\Sigma_4 + \Sigma_5 + \Sigma_6)dr = 0. \quad (4.25)$$

If all the corrections for $m<n$ have been previously determined, then the only unknown appearing in this equation is E_n. Hence the nth-order correction to the energy is completely determined by this integral. Examples of the form of the integrand will be discussed in Sec. V. At $r=\alpha_0^i$, $f_0(\alpha_0^i)=0$ and the nth-order correction to α_0^i is given by

$$f_n(\alpha_0^i)g_0(\alpha_0^i)Q_0(\alpha_0^i) = \left(\frac{1}{X_0(\alpha_0^i)}\right)\int_0^{\alpha_0^i} \frac{X_0}{a_0}(\Sigma_4 + \Sigma_5 + \Sigma_6)dr. \quad (4.26)$$

At $r=\beta_0^i$, $g_0(\beta_0^i)=0$ and the nth-order correction to β_0^i is determined by the expression

$$g_n(\beta_0)f_0(\beta_0)Q_0(\beta_0) = \frac{-1}{X_0(\beta_0^i)}\int_0^{\beta_0^i}\frac{X_0}{a_0}(\Sigma_4+\Sigma_5+\Sigma_6)dr. \quad (4.27)$$

Once E_n, α_n^i, and β_n^i have been determined, Q_n is determined by

$$Q_n(r)f_0(r)g_0(r) + [f_0(r)g_n(r) - f_n(r)g_0(r)]Q_0(r) + \frac{1}{X_0(r)}\int_0^r \frac{X_0(r')}{a_0(r')}(\Sigma_4+\Sigma_5+\Sigma_6)dr' = 0, \quad (4.28)$$

except at the zeros of f_0 and g_0. At these points, specifically the α_0^i and the β_0^i, the value of $Q_n(\alpha_0^i)$ or $Q_n(\beta_0^i)$ is determined by the original nth-order equation. At any of these points, the terms involving Q_n' are multiplied by f_0 and g_0, one of which will be zero. Thus one is left with an algebraic equation for $Q_n(\alpha_0^i)$ or $Q_n(\beta_0^i)$ with all other variables of nth- or lower-order already determined. Hence this allows the evaluation of Q_n at these points. This method has been successfully employed in numerical work with the Dirac equation[11] and provides a simple alternative to the method suggested by Privman[13] for problems where the solution is of the form of Eq. (4.1)

If the coefficient a_0 has zeros for certain values of r, there are some restrictions on the solution for these methods to be assured to work. I begin by assuming that near one of the zeros of a_0, it can be factored as $a_0(r)=A_0(r)[r-\gamma_0]$, where the zero is at $r=\gamma_0$ and $A_0(r)$ is free of zeros in a neighborhood of γ_0. Then from the zero-order Eq. (4.4) I have the relation

$$b_0(\gamma_0)g_0[(\gamma_0)/f_0(\gamma_0)]Q_0(\gamma_0) + c_0(\gamma_0)Q_0(\gamma_0) + d_0(\gamma_0)[f_0(\gamma_0)/g_0(\gamma_0)] = 0. \quad (4.29)$$

This leads to a restriction of $Q_0(\gamma_0)$

$$Q_0(\gamma_0)[g_0(\gamma_0)/f_0(\gamma_0)] = \frac{-c_0(\gamma_0)}{2b_0(\gamma_0)}\left[1\pm\left(1-\frac{4b_0(\gamma_0)d_0(\gamma_0)}{c_0(\gamma_0)^2}\right)^{1/2}\right], \quad (4.30)$$

and a restriction on the coefficient functions

$$c_0(\gamma_0)^2 - 4b_0(\gamma_0)d_0(\gamma_0) \geqslant 0. \quad (4.31)$$

In addition, the integrand in Eqs. (4.26)–(4.28) must be free of poles. This is avoided if $X_0(\gamma_0)/a_0(\gamma_0)$ is finite. With a_0 written in terms of A_0 as given above, I require $\lim_{r\to\gamma_0}[X_0(r)/(r-\gamma_0)]$ to be finite. The function $X_0(r)$ can be written as

$$X_0(r) = (r-\gamma_0)\times\exp\left[\int_0^r\left(\frac{c_0+2b_0R_0}{A_0}-1\right)\frac{1}{(r-\gamma_0)}dr\right]. \quad (4.32)$$

In a neighborhood of γ_0, I can make the expansion

$$\frac{c_0+2b_0R_0}{A_0} - 1 = \sum_{n=0}^{\infty}\xi_n(r-\gamma_0)^n, \quad (4.33)$$

where the first coefficient of the expansion is given by

$$\xi_0 = [[c_0(\gamma_0)+2b_0(\gamma_0)R_0(\gamma_0)]/A_0(\gamma_0) - 1]. \quad (4.34)$$

With this expansion, near $r=\gamma_0$, $X_0(r)/(r-\gamma_0)$ is given by

$$\frac{X_0(r)}{(r-\gamma_0)} = (r-\gamma_0)^{\xi_0}r^{\xi_1}\exp\left[\int\sum_{n=2}^{\infty}\xi_n(r-\gamma_0)^{n-1}dr\right], \quad (4.35)$$

which leads to the requirement

$$\xi_0 = [[c_0(\gamma_0)+2b_0(\gamma_0)R_0(\gamma_0)]/A_0(\gamma)_0 - 1] \geqslant 0. \quad (4.36)$$

Therefore, if Eqs. (4.30), (4.31), and (4.36) are satisfied, the expressions for the nth-order corrections in this section are valid, even where a_0 has zeros.

V. EXAMPLES

A. The Klein–Gordon and Schrödinger equations

The single-particle central-field Klein–Gordon equation in the presence of both a fourth-component vector potential $V(r)$ and a scalar potential $S(r)$ can be written in natural units, $\hbar=c=1$, as[14]

$$\left[-\frac{1}{r}\frac{d^2}{dr^2}r + \frac{l(l+1)}{r^2} + m^2 + 2mS(r) - (E-V(r))^2\right]\psi(r) = 0. \quad (5.1)$$

The central-field Schrödinger equation can be written, again in natural units

$$\left[-\frac{1}{r}\frac{d^2}{dr^2}r + \frac{l(l+1)}{r^2} + 2m(V-E)\right]\psi(r) = 0. \quad (5.2)$$

In both cases, the radial part of the wave function can be written

$$\psi(r) = \exp[-G(r) - 1/r], \quad (5.3)$$

whence, with $Q=G'$, either equation can be written in the Riccati form

$$Q' - Q^2 + d = 0, \quad (5.4)$$

where d is

$$d^{\mathrm{KG}} = l(l+1)/r^2 + m^2 + 2mS(r) - [E-V(r)]^2, \quad (5.5a)$$

for the Klein–Gordon particle, and

$$d^{s} = l(l+1)/r^2 + 2m(V-E), \quad (5.5b)$$

for the Schrödinger particle. Letting the perturbation be given by λS_1 and λV_1, I have

$$S = S_0 + \lambda S_1, \quad (5.6)$$

$$V = V_0 + \lambda V_1, \quad (5.7)$$

with E and Q expanded as

$$E = \sum_{n=0}^{\infty}\lambda^n E_n \quad (5.8)$$

and

$$Q = \sum_{n=0}^{\infty}\lambda^n Q_n. \quad (5.9)$$

The expansion of d for the two cases follows easily from Eqs. (5.5)–(5.8). For the Klein–Gordon particle, I have

$$d_1^{\mathrm{KG}} = 2mS_1 - (E_1-V_1)(E_0-V_0), \quad (5.10a)$$

$$d_2^{\mathrm{KG}} = -(E_1-V_1)^2 - 2E_2(E_0-V_0), \quad (5.10b)$$

and

$$d_n^{KG} = 2E_n(V_0 - E_0) + 2E_{n-1}(V_1 - E_1), \quad n \geqslant 3. \tag{5.10c}$$

For the Schrödinger case, d_i^S is given by

$$d_1^S = 2m(V_1 - E_1), \tag{5.11a}$$

$$d_n^S = -2mE_n, \quad n \geqslant 2. \tag{5.11b}$$

For both equations, the integrating factor is

$$\rho_0 = \exp\left[-\int_0^r 2Q_0\, dr\right], \tag{5.12}$$

with

$$p_n = d_n, \tag{5.13}$$

and

$$q_n = -\sum_{m=1}^{n-1} R_m R_{n-m}, \tag{5.14}$$

with the nth-order corrections given by Eqs. (2.19) and (2.20).

This gives the ground state corrections for both types of particles. The expressions obtained this way differ from those of Au and Aharonov[7] for the Klein–Gordon particle in a ground state only in the limits of integration and the addition here of the central field angular momentum term. The lower limit of integration of ρ_0 is chosen by them such as to normalize ρ_0. The first excited state corrections for a Klein–Gordon particle have been given in Ref. 7 and are equivalent to those obtained by this procedure, except for the difference mentioned above for the nodeless states, and so will not be presented here. Instead, to demonstrate the form of the results with this notation for higher excited states, I will present the corrections for an excited state with two nodes, one at α_0^1 and the other at α_0^2. The Riccati equation for the case is

$$Q' - Q^2 + (2f'/f)Q - f''/f + d = 0, \tag{5.15}$$

with the d_i^{KG} and d_i^S given as before and

$$f_0 = (r - \alpha_0^1)(r - \alpha_0^2), \tag{5.16}$$

$$f_n = -\alpha_n^2(r - \alpha_0^1) - \alpha_n^1(r - \alpha_0^2), \quad n \geqslant 1, \tag{5.17}$$

$$\bar{f}_n = \sum_{i=1}^{n-1} \alpha_i^1 \alpha_{n-i}^2, \quad n \geqslant 2, \tag{5.18}$$

with $\bar{f}_1 = 0$, and

$$f_0' = (r - \alpha_0^1) + (r - \alpha_0^2), \tag{5.19}$$

$$f_n' = -(\alpha_n^1 + \alpha_n^2), \quad n \geqslant 1, \tag{5.20}$$

$$f_0'' = 2, \tag{5.21}$$

and

$$f_n'' = \bar{f}_n' = \bar{f}_n'' = 0. \tag{5.22}$$

The integrating factor is given by

$$\rho_0(r) = (r - \alpha_0^1)(r - \alpha_0^2)^2 e^{-\int_0^r 2Q_0\, dr'}. \tag{5.23}$$

The nth-order corrections are found from the equations

$$\rho_0(r)Q_n(r) = \left[\frac{\alpha_n^1}{(r - \alpha_0^1)^2} + \frac{\alpha_n^2}{(r - \alpha_0^2)^2}\right]\rho_0(r) - \sum_{i=1}^{n-1} \alpha_i^1 \alpha_{n-i}^2 (r - \alpha_0^1 + r - \alpha_0^1)\frac{\rho_0}{f_0^2} - \int_0^r d_n\, \rho_0\, dr' + \int_0^r [\text{rhs}]_{dr'}, \tag{5.24}$$

where the [rhs] of Eq. (3.12) can be written for this case as

$$\begin{aligned}[\text{rhs}] = \frac{\rho_0}{f_0}\Bigg\{ &\sum_{i=1}^{n-1} \Bigg[\alpha_n^2(r - \alpha_0^1) + \alpha_n^1(r - \alpha_0^2) - \sum_{j=1}^{n-1} \alpha_j^1 \alpha_{i-j}^2\Bigg] \\ &\times (Q'_{n-i} + d_{n-i}) + \sum_{i=1}^{n-1} 2[\alpha_n^1 + \alpha_n^2] Q_{n-i} \\ &+ \sum_{i,j,k=0}^{n-1} \Bigg[-\alpha_i^2(r - \alpha_0^1) - \alpha_i^1(r - \alpha_0^2) \\ &+ \sum_{m=1}^{i-1} \alpha_m^1 \alpha_{i-m}^2\Bigg] Q_j Q_k \delta(i + j + k - n)\Bigg\}.\end{aligned} \tag{5.25}$$

From Eqs. (5.24) and (5.25) all corrections may be determined as discussed earlier.

B. The Dirac equation for nodeless states

The Dirac equation for a combination of scalar potential S, fourth-component vector potential V, and anomolous magnetic moment term ϵ, when all three terms have spherical symmetry, leads to the Riccati equation for nodeless states[11]

$$R' - 2kR/r - 2\epsilon R + R^2(E - V + m + S) + (E - V - m - S) = 0. \tag{5.26}$$

From this equation, the coefficients are identified as

$$a = 1, \tag{5.27a}$$

$$b = E - V + m + S, \tag{5.27b}$$

$$c = -2k/r - 2\epsilon, \tag{5.27c}$$

$$d = E - V - m - S. \tag{5.27d}$$

Then, with

$$V = V_0 + \lambda V_1 \tag{5.28a}$$

$$\epsilon = \epsilon_0 + \lambda \epsilon_1, \tag{5.28b}$$

and

$$S = S_0 + \lambda S_1, \tag{5.28c}$$

the nth-order correction to the energy is, from Eq. (2.19), found to be given by

$$\begin{aligned} E_n \int_0^\infty (1 + R_0^2)\rho_0\, dr &= \int_0^\infty 2\epsilon_1 R_{n-1}\rho_0\, dr \\ &- \int_0^\infty \Bigg[(E_0 - V_0 + m + S_0)\sum_{m=1}^{n-1} R_m R_{n-m} \\ &+ (E_1 - V_1 + S_1)\sum_{k=0}^{n-1} R_k R_{n-k-m} \\ &+ \sum_{m=2}^{n-1} E_m \sum_{k=0}^{n-m} R_k R_{n-k-m}\Bigg]\rho_0\, dr. \end{aligned} \tag{5.29}$$

This duplicates the combined results of Refs. 9–11 for nodeless states.

For the same problem, the first-order corrections to excited states with nodes and poles are easily written down from Sec. IV. The sigmas are easily found to be given by

$$\Sigma_4 = (E - V_1 + S_1)g_0^2 Q_0^2 + (E_1 - V_1 - S_1)f_0^2 - 2\epsilon_1 f_0 g_0 Q_0, \tag{5.30}$$

and

$$\Sigma_5 = \Sigma_6 = 0, \tag{5.31}$$

so that the first-order corrections are determined from

$$Q_1(r)f_0(r)g_0(r)X_0(r) + [f_0(r)g_1(r) - f_1(r)g_0(r)] \times Q_0(r)X_0(r) + \int_0^r X_0(r')[(E_1 - V_1)(f_0^2 + g_0^2 Q_0^2) + S_1(g_0^2 Q_0^2 - f_0^2) - 2\epsilon_1 f_0 g_0 Q_0]\,dr' = 0. \tag{5.32}$$

From Eq. (4.16) X_0 is given by

$$X_0(r) = r^{-2k}\exp\left[-\int 2\epsilon_0\,dr\right] \times \exp\left\{2\int [(E_0 - V_0 + m + S_0)g_0 Q_0 - f_0'] \frac{1}{f_0}\,dr\right\}, \tag{5.33}$$

where the constants of integration are unimportant so long as X_0 is zero at $r = 0$ and $r = \infty$. This duplicates the first-order results of Ref. 11.

VI. CONCLUDING REMARKS

The formulation of perturbation theory presented in this paper allows the computation of the corrections to the eigenvalue and eigenfunction to any order in the coupling constant. The only drawback is the hierarchical nature of the quadratures. That is, before the nth-order corrections can be evaluated, the corrections to the $(n-1)$ lower orders must be found first. This means a total of n integrations are required to determine the nth-order corrections. Although in certain cases such as the anharmonic oscillator it is possible to perform the integrals analytically,[15,16] it is in general necessary to perform the integrals numerically. The advantage of this formulation is that only the initial solution to the specific state under consideration needs to be known, whereas in the usual Rayleigh–Schrödinger perturbation theory the complete set of unperturbed states must be known. In certain problems this is a tremendous advantage.

I would next like to comment on the conditions under which the Riccati form can be obtained. In general, it is easy to show that a set of n linear homogeneous coupled first-order equations for the functions F_i, for $i = 1, 2,..., n$, can be transformed into $(n-1)$ coupled first-order nonlinear inhomogeneous equations and one dependent first-order equation, whose solution can be written in quadrature in terms of the other functions. That is, the transformation $F_1 = G$, $F_i = R_i G$, $i > 1$, results in equations of the functional form

$$G'/G = f_1(R_i,\ i \neq 1), \tag{6.1}$$

and

$$R_i' = f_i(R_j,\ j \neq 1). \tag{6.2}$$

Thus, in the case of the central field Dirac equation, which can be written as two coupled first-order equations,[17] the above transformation results in a single independent equation for R as in Eq. (5.26) (see Refs. 8 and 9) and a dependent equation for which the solution may be written in quadrature.

If there are more than two original coupled equations, the transformation still reduces the number of simultaneous equations of the form of Eq. (6.2) by one with the other equation of the form (6.1). This procedure cannot generally be continued because the coupled equations (6.2) are no longer homogeneous. Thus the expansions of this paper are in general applicable to problems involving one or two first-order homogeneous or one second-order homogeneous equation. If more equations are involved, the transformation given above reduces the number of equations to be solved by one, but the price is nonlinearity and inhomogeneity of the resulting equations.

This expansion is not directly applicable to the problem of computing radiative corrections. These corrections involve operator expansions, which involve normal ordering and commutation relations and thus are different from this function expansion.

Once the operator expansion has been carried out to some order, however, it may be possible to utilize the perturbation expansion of this paper for the parts of the operator expansion that can be expressed in terms of effective potentials. As an example, part of the first-order corrections to the Breit equation in the absence of external fields can be calculated in terms of the Breit operator given by[18]

$$V_1 = -(e^2/r)[\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2 + (\boldsymbol{\alpha}_1 \cdot \mathbf{r})(\boldsymbol{\alpha}_2 \cdot \mathbf{r})/r^2], \tag{6.3}$$

with

$$E_1 = \langle 0|V_1|0\rangle, \tag{6.4}$$

where the standard bracket notation has been used and $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are the spin matrices for particles one and two, respectively. The first-order corrections due to the effective potential V_1 are expressible in quadrature. If V_1 is used as a basis for computing second-order corrections, one gets quite wrong results.[18] If, however, the second-order corrections are worked out from the operator expansion, Bethe and Salpeter[18] have pointed out that part of this correction can be expressed in terms of an effective potential V_2 given by

$$V_2(r) = \frac{e^2}{4\pi^2 mc}\int \frac{d^3k}{k} e^{-\mathbf{k}\cdot\mathbf{r}} \times [\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2 - [(\boldsymbol{\alpha}_1 \cdot \mathbf{k})(\boldsymbol{\alpha}_2 \cdot \mathbf{k})/k^2]], \tag{6.5}$$

with E_2 determined from

$$E_2 = \sum_{l \neq 0} \frac{\langle 0|V_2|l\rangle\langle l|V_2|0\rangle}{(E_0 - E_l)}, \tag{6.6}$$

where the sum is over the entire basis of intermediate states. In order to evaluate the corrections due to V_2 without the sum over intermediate states, $E_1(V_2)$ and $Q_1(V_2)$ must be calculated, but have no physical significance. Instead, they are necessary to compute $E_2 = E_2[V_2, E_1(V_2), Q_1(V_2)]$ because of the hierarchical nature of the expansion. Thus it may be possible to use this formulation indirectly in the calculation of parts of the radiative corrections in systems of physical interest.

In this paper I have shown that for a restricted set of solutions it is possible to obtain a perturbation expansion with all corrections expressed in quadrature for any first-order nonlinear eigenvalue equation. Expansions for Riccati equations have been developed in detail for important cases.

The development of Sec. II is appropriate for the set of solutions which possess neither zeros nor poles, and has applications to the Schrödinger, Klein–Gordon, and Dirac equations for central fields. In Sec. III, the set of solutions corresponding to excited states of the Schrödinger and Klein–Gordon equations has been developed with the nth-order corrections to an arbitrary excited state presented. In Sec. V, this has been applied to a Klein–Gordon particle in an excited state with two nodes. In Sec. IV a general Riccati equation for the set of meromorphic solutions has been expanded with all results written in quadrature. This is appropriate for the Dirac equation for a general central field and allows a majority of the first-order results of several papers[9–11] to be easily written down. An example is provided in Sec. V. The results of Sec. IV allow the nth-order correction to an arbitrary excited state to be written in quadrature. This section is also applicable to more complicated Riccati equations, such as one obtained from the Breit equation for two particles in the absence of an external field, which the author is studying and will hopefully report on elsewhere.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge C. K. Au for introducing the author to logarithmic perturbation theory and both Y. Aharonov and C. K. Au for direction of previous work out of which this work is an outgrowth. The author is also happy to thank J. L. Safko for his encouragement and comments.

[1]R. J. Price, Proc. Phys. Soc. London **67**, 383 (1954).
[2]V. S. Polikanov, Zh. Eksp. Teor. Fiz. **52**, 1326 (1967) [Sov. Phys. JETP **25**, 882 (1967)].
[3]Y. Aharonov and C. K. Au, Phys. Rev. Lett. **42**, 1582 (1979); **43**, 176 (E) (1979); **44**, 619(E) (1980).
[4]V. S. Polikanov, Theoret. Math. Phys. (USSR) **24**, 230 (1975).
[5]J. O. Hirschfelder, J. Chem. Phys. **39**, 2009 (1963).
[6]C. K. Au, Phys. Lett. A **77**, 221 (1980).
[7]C. K. Au and Y. Aharonov, J. Math. Phys. **22**, 1429 (1981).
[8]A. I. Mikhailov and V. S. Polikanov, Zh. Eksp. Teor. Fiz. **54**, 175 (1968). [Sov. Phys. JETP **27**, 95 (1968)].
[9]C. K. Au and G. W. Rogers, Phys. Rev. A **22**, 1820 (1980).
[10]C. K. Au, Phys. Rev. A **23**, 2068 (1981).
[11]G. W. Rogers, Phys. Rev. A **30**, 35 (1984).
[12]G. W. Rogers, currently unpublished.
[13]V. Privman, Phys. Rev. A **24**, 2362 (1981).
[14]G. Baym, *Lectures on Quantum Mechanics* (Benjamin, New York, 1969).
[15]C. K. Au, G. W. Rogers, and Y. Aharonov, Phys. Lett. A **95**, 287 (1983).
[16]A. D. Dolgov, V. L. Eletskii, and V. S. Popov, Zh. Eksp. Teor. Fiz. **79**, 1704 (1980) [Sov. Phys. JETP **52**, 861 (1980)].
[17]H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Springer, Berlin, 1957), Sec. 14.
[18]Reference 17, Sec. 38.

Closed formula for the product of n Dirac matrices

Matías Moreno[a)]

Centro de Investigación y de Estudios Avanzados del IPN, Departamento de Física, Apdo. Postal 14-740 07000, México D. F., Mexico

(Received 25 July 1984; accepted for publication 19 October 1984)

The product of n Dirac γ matrices is evaluated in terms of traces of at most $(n+3)$ γ matrices. This leads to a method that can be used to generalize identities for a correlated product of γ matrices. The expression for the n γ matrix product also provides a method for the evaluation of a scalar, a pseudoscalar, a vector, an axial vector, and a second rank antisymmetric tensor associated with a one-line fermion amplitude that drastically simplifies the squaring of fermionic amplitudes. The general spin and particle–antiparticle dependence of the squared amplitude is given.

I. INTRODUCTION

In the evaluation of cross sections or lifetimes involving spin $\frac{1}{2}$ particles, one of the main practical difficulties is the evaluation of traces of products of Dirac γ matrices. Although the problem is a purely algebraic one, the nature of the difficulty can be easily realized by noting that the trace of a product of $2n$ γ matrices has $(2n-1)!!$ terms. Thus one is limited in practice to consider those cases in which the number of γ matrices is small. Even if a symbolic manipulation program is available the rapid growth of the number of terms will sooner or later saturate the computer facilities.

It is therefore important to develop methods of computation and identities that maintain or reduce the number of γ matrices to a minimum.

Important steps have been taken in the past to solve these practical difficulties. In particular, Caianiello and Fubini[1] showed that the trace of an arbitrary product of γ matrices is related to a determinant and, using this property, obtained that many terms vanish when the trace of more than 12 γ matrices is taken. As a by-product, they got rules for the simplification of expressions of the form $\gamma^{\mu}\cdots\gamma_{\mu}$, where the dots stand for a given product of N γ matrices, i.e., an N γ string. This last formula was also obtained independently by Chisholm,[2] who in turn generalized it to the case in which the contracted indices appeared in two different traces or one in a trace and the other in a γ string. An elegant algorithm was developed by Kahane[3] to deal with the problem of reducing the product of a γ string when a subset of Dirac matrices in it is contracted by pairs. This in turn was generalized by Chisholm.[4] Recently, Sirlin[5] has found that additional simplifications result when the chiral projectors appear in these products.

In this work a closed formula for the product of n γ matrices is found. The coefficients of the expansion of this product in the covariant basis are given in terms of at most a trace of $(n+3)$ γ matrices. Thus a general element of the Dirac algebra, i.e., a sum of γ strings with given coefficients, can be *explicitly expressed* in the covariant basis. The usefulness of this procedure is clear once one realizes that, in practice, one only knows the coefficients of the general Dirac algebra element in terms of the γ strings.

The plan of the article is as follows. In Sec. II we define Dirac algebra, set up our conventions (that follow those of Bjorken and Drell[6]), give some useful known results, and obtain the multiplication table of the covariant basis. Section III contains the central result of this work, namely a closed expression for the coefficients of the expansion in the covariant basis of the product of n γ matrices. Section IV is devoted to the development of identities in which a general element of Dirac algebra is left and right multiplied by two basis elements with one or two Lorentz indices contracted. These identities, when combined with the expression for the product of n γ matrices, lead to a method that can be used to generalize the Caianello–Fubini–Chisholm–Sirlin identities for a general Dirac algebra element. Seen in the light of this generalization, the algebraic content of these identities is clear: The tensor coefficient just vanishes when a Dirac algebra element is sandwiched between γ^{μ} and γ_{μ}. In Sec. V, a general method for the evaluation of squares of matrix elements of processes that contain distinguishable fermionic lines is presented. The importance for this purpose of the expression for the product of n γ matrices obtained in Sec. III is then made clear: If one begins with a matrix element $\mathscr{M}$ with a product of n γ matrices, call it Γ, the squaring of it gives the trace of $(2n+2)$ γ matrices, $|\mathscr{M}|^2 = \mathrm{Tr}\,\mathscr{P}_i\Gamma\mathscr{P}_f\bar{\Gamma}$, where $\mathscr{P}_i$ and $\mathscr{P}_f$ are particle projectors. This means $(2n+1)!!$ terms in the spin-averaged transition probability. This is to be compared with $(n+2)!!$ typical terms that are obtained using the method proposed in this work. In the final section several applications are given. These include the computation of a general vector and axial-vector matrix elements and how to get the spin dependence of a matrix element out of the unpolarized case.

II. DIRAC ALGEBRA IN THE COVARIANT BASIS

Dirac algebra is generated by products and sums of four basic elements γ^0, γ^1, γ^2, and γ^3 that satisfy the anticommutation relation

$$\{\gamma^{\mu}, \gamma^{\nu}\} = 2g^{\mu\nu}\mathbf{1}, \tag{2.1}$$

where $g^{\mu\nu}$ are the matrix elements of $g = \mathrm{diag}\,(1, -1, -1, -1)$; and $\mathbf{1}$ is the identity element of the algebra. The simplest effective realization of this algebra is given in terms of 4×4 matrices. This implies that any element of the algebra is a linear superposition of 16 elements that define a basis. An appropriate basis for field theory computations is given by the covariant basis.[7] This basis is defined by the unit element $\mathbf{1}$, and γ^{μ} themselves, the six commutators of the basic elements

[a)] On sabbatical leave from Instituto de Física, Universidad Nacional Autónoma de México, Apdo. Postal 20-364, 01000 México D.F., Mexico.

 0022-2488/85/040576-09$02.50

$$\sigma^{\mu\nu} := (i/2)[\gamma^\mu, \gamma^\nu], \tag{2.2}$$

the γ^5,

$$\gamma^5 := i\gamma^0\gamma^1\gamma^2\gamma^3 = (i/4!)\varepsilon_{\alpha\beta\mu\nu}\gamma^\alpha\gamma^\beta\gamma^\mu\gamma^\nu, \tag{2.3}$$

where $\varepsilon_{\alpha\beta\mu\nu}$ is the Levi-Civita completely antisymmetric symbol, with $\varepsilon_{0123} = -\varepsilon^{0123} = 1$, and, following Bjorken and Drell conventions,[6] we sum from 0 to 3 when repeated Greek indices appear. We complete the definition of the covariant basis with the four "axial-vector" elements

$$\gamma^5\gamma^\mu.$$

In terms of the 16 elements a general element of Dirac algebra is given by

$$\Gamma = S + iP\gamma^5 + V_\mu\gamma^\mu + iA_\mu\gamma^5\gamma^\mu - (i/2)T_{\mu\nu}\sigma^{\mu\nu}, \tag{2.4}$$

with the expansion coefficients S, P, V_μ, A_μ, and $T_{\mu\nu} = -T_{\nu\mu}$ arbitrary complex numbers. Phases and signs in these coefficients have been introduced for later convenience, but they are mostly there to compensate the i factors in the γ^5 and $\sigma^{\mu\nu}$ definitions. The name covariant associated with this basis stems from field theory because when a matrix element of Γ between spinors $\bar{\chi} = \chi^\dagger\gamma^0$ and ψ is taken, the resulting object,

$$\bar{\chi}\Gamma\psi, \tag{2.5}$$

will be a Lorentz scalar *if* S, P, V_μ, A_μ, and $T_{\mu\nu}$ transform as a scalar, pseudoscalar, vector, axial vector, and tensor of rank 2, respectively.

The most widely used relations among elements of this basis are

$$\gamma^5 \cdot \gamma^5 = \mathbf{1}, \tag{2.6}$$

$$\{\gamma^\mu, \gamma^5\} = 0, \tag{2.7}$$

and the so-called Gordon relation

$$\begin{aligned} \gamma^\mu\gamma^\nu &= \tfrac{1}{2}\{\gamma^\mu, \gamma^\nu\} + \tfrac{1}{2}[\gamma^\mu, \gamma^\nu] \\ &= g^{\mu\nu} - i\sigma^{\mu\nu}. \end{aligned} \tag{2.8}$$

A less common identity is given by the product of three γ matrices

$$\gamma^\mu\gamma^\alpha\gamma^\nu = g^{\mu\alpha}\gamma^\nu - g^{\mu\nu}\gamma^\alpha + g^{\alpha\nu}\gamma^\mu + i\varepsilon^{\mu\alpha\nu\beta}\gamma^5\gamma_\beta. \tag{2.9}$$

Another useful relation is

$$\sigma^{\mu\nu}\gamma^5 = -(i/2)\varepsilon^{\mu\nu\rho\sigma}\sigma_{\rho\sigma} = -i\tilde{\sigma}^{\mu\nu}, \tag{2.10}$$

where the dual of a second-rank tensor $D_{\mu\nu}$ is defined by

$$\tilde{D}_{\mu\nu} := \tfrac{1}{2}\varepsilon_{\mu\nu\rho\sigma}D^{\rho\sigma}. \tag{2.11}$$

Using these identities one can construct the multiplication table of the basis; this is given in Table I. The algebraic content of this table can be reexpressed through the product of two elements of the algebra. Denoting by subindices the expansion coefficients of both elements

$$\begin{aligned} \Gamma_1\Gamma_2 \equiv{}& (S_1 + iP_1\gamma^5 + V_{1\alpha}\gamma^\alpha + iA_{1\alpha}\gamma^5\gamma^\alpha - (i/2)T_{1\alpha\beta}\sigma^{\alpha\beta})(S_2 + iP_2\gamma^5 + V_{2\mu}\gamma^\mu + iA_{2\mu}\gamma^5\gamma^\mu - (i/2)T_{2\mu\nu}\sigma^{\mu\nu}) \\ ={}& (S_1S_2 - P_1P_2 + V_1 \cdot V_2 + A_1 \cdot A_2 - \tfrac{1}{2}T_{1\alpha\beta}T_2^{\alpha\beta}) + i(S_1P_2 + P_1S_2 + A_1 \cdot V_2 - V_1 \cdot A_2 + \tfrac{1}{2}\tilde{T}_{1\alpha\beta}T_2^{\alpha\beta})\gamma^5 \\ &+ (S_1V_{2\alpha} + V_{1\alpha}S_2 + A_{1\alpha}P_2 - P_1A_{2\alpha} + T_{1\alpha\mu}V_2^\mu + V_1^\mu T_{2\mu\alpha} + \tilde{T}_{1\alpha\mu}A_2^\mu - A_1^\mu\tilde{T}_{2\mu\alpha})\gamma^\alpha \\ &+ i(P_1V_{2\alpha} - V_{1\alpha}P_2 + S_1A_{2\alpha} + A_{1\alpha}S_2 - \tilde{T}_{1\alpha\mu}V_2^\mu + V_1^\mu\tilde{T}_{2\mu\alpha} + T_{1\alpha\mu}A_2^\mu + A_1^\mu T_{2\mu\alpha})\gamma^5\gamma^\alpha \\ &- (i/2)(2V_{1\alpha}V_{2\beta} - \widetilde{2V_{1\alpha}A_{2\beta}} + 2A_{1\alpha}V_{2\beta} + 2A_{1\alpha}A_{2\beta} + T_{1\alpha\beta}S_2 + S_1T_{2\alpha\beta} + \tilde{T}_{1\alpha}P_2 + P_1\tilde{T}_{2\alpha\beta} + 2T_{1\alpha\mu}T_2{}^{\mu}{}_{\beta})\sigma^{\alpha\beta}. \end{aligned} \tag{2.12}$$

Let us close this section defining the even and odd parts of a Dirac algebra element with respect to γ^5. We call even or odd that part of a Dirac algebra element that commutes or anticommutes with γ^5, respectively. Thus, we can decompose

$$\Gamma = \Gamma_{\mathscr{E}} + \Gamma_{\mathscr{O}}, \tag{2.13}$$

with

$$[\Gamma_{\mathscr{E}}, \gamma^5] = 0 \tag{2.14}$$

and

$$\{\Gamma_{\mathscr{O}}, \gamma^5\} = 0. \tag{2.15}$$

In terms of the covariant basis

$$\Gamma_{\mathscr{E}} = S + iP\gamma^5 - (i/2)T_{\mu\nu}\sigma^{\mu\nu} \tag{2.16}$$

and

$$\Gamma_{\mathscr{O}} = V_\mu\gamma^\mu + iA_\mu\gamma^5\gamma^\mu. \tag{2.17}$$

III. PRODUCT OF n γ MATRICES IN TERMS OF S-EXPANSION COEFFICIENTS

Because every element of Dirac algebra stems from a sum of products of γ matrices, it is important to study the

TABLE I. Multiplication table of the covariant basis of Dirac algebra. The symbol $\delta^{\alpha\beta,\mu\nu} := g^{\alpha\mu}g^{\beta\nu} - g^{\alpha\nu}g^{\beta\mu}$. Elements in the first column multiply on the right elements of the first row. The multiplication by the unit element has been omitted for brevity.

1	γ^5	γ^μ	$\gamma^5\gamma^\mu$	$\sigma^{\mu\nu}$
γ^5	1	$\gamma^5\gamma^\mu$	γ^μ	$-i\tilde{\sigma}^{\mu\nu}$
γ^α	$-\gamma^5\gamma^\alpha$	$g^{\alpha\mu} - i\sigma^{\alpha\mu}$	$-g^{\alpha\mu}\gamma^5 + \tilde{\sigma}^{\alpha\mu}$	$i\delta^{\alpha\beta,\mu\nu}\gamma_\beta - \varepsilon^{\alpha\mu\nu\beta}\gamma^5\gamma_\beta$
$\gamma^5\gamma^\alpha$	$-\gamma^\alpha$	$+g^{\alpha\mu}\gamma^5 - \tilde{\sigma}^{\alpha\mu}$	$-g^{\alpha\mu} + i\sigma^{\alpha\mu}$	$i\delta^{\alpha\beta,\mu\nu}\gamma^5\gamma_\beta - \varepsilon^{\alpha\mu\beta}\gamma_\beta$
$\sigma^{\alpha\beta}$	$-i\tilde{\sigma}^{\alpha\beta}$	$i\delta^{\alpha\beta,\nu\mu}\gamma_\nu - \varepsilon^{\alpha\beta\mu\nu}\gamma^5\gamma_\nu$	$i\delta^{\alpha\beta,\nu\mu}\gamma^5\gamma_\nu - \varepsilon^{\alpha\beta\mu\nu}\gamma_\nu$	$\delta^{\alpha\beta,\mu\nu} - i\varepsilon^{\alpha\beta\mu\nu}\gamma^5 - i[g^{\alpha\mu}\sigma^{\beta\nu} - g^{\alpha\nu}\sigma^{\beta\mu} + g^{\beta\nu}\sigma^{\alpha\mu} - g^{\beta\mu}\sigma^{\alpha\nu}]$

relation among elements of this algebra that are related through products of γ matrices. In this respect, it is useful to define

$$\Gamma^{(\mu)} = \Gamma\gamma^{\mu}, \tag{3.1}$$

where in the covariant basis

$$\Gamma^{(\mu)} = S^{(\mu)} + iP^{(\mu)}\gamma^5 + V_\alpha^{(\mu)}\gamma^\alpha + iA_\alpha^{(\mu)}\gamma^5\gamma^\alpha - (i/2)T_{\alpha\beta}^{(\mu)}\sigma^{\alpha\beta}. \tag{3.2}$$

From the multiplication table or from Eq. (2.12), one immediately gets the relation between the expansion coefficients of $\Gamma^{(\mu)}$ and Γ. Explicitly, for the even $\Gamma^{(\mu)}$ coefficients,

$$S^{(\mu)} = V^{\mu}, \tag{3.3a}$$

$$P^{(\mu)} = A^{\mu}, \tag{3.3b}$$

$$T_{\alpha\beta}^{(\mu)} = V_\alpha g_\beta^\mu - V_\beta g_\alpha^\mu + A_\theta \varepsilon^{\theta\mu}_{\cdot\cdot\alpha\beta}, \tag{3.3c}$$

and, for the odd $\Gamma^{(\mu)}$ coefficients,

$$V_\alpha^{(\mu)} = Sg_\alpha^\mu + T_{\cdot\alpha}^{\cdot\ \mu}, \tag{3.4a}$$

$$A_\alpha^{(\mu)} = \widetilde{T}^{\mu}_{\cdot\alpha} + g_\alpha^\mu P. \tag{3.4b}$$

Here we remark that even (odd) $\Gamma^{(\mu)}$ coefficients are given solely in terms of odd (even) Γ coefficients.

Consider next, the element of Dirac algebra formed by the product of $2n$ γ matrices, Γ^{2n}. This element naturally is purely even and therefore

$$\Gamma^{2n} = S^{2n} + iP^{2n}\gamma^5 - (i/2)T_{\alpha\beta}^{2n}\sigma^{\alpha\beta}. \tag{3.5}$$

Multiplication of Γ^{2n} by a γ matrix gives a purely odd element

$$\Gamma^{2n+1} = V_\alpha^{2n+1}\gamma^\alpha + iA_\alpha^{2n+1}\gamma^5\gamma^\alpha; \tag{3.6}$$

calling γ^μ the matrix that multiplies Γ^{2n}, we get

$$\Gamma^{2n+1} = \Gamma^{2n}\gamma^\mu = \Gamma^{2n(\mu)}. \tag{3.7}$$

If we multiply once more, this time by γ^ν, we obtain

$$\Gamma^{2n+2} = \Gamma^{2n+1}\gamma^\nu = \Gamma^{2n+1(\nu)} = \Gamma^{2n(\mu\nu)}. \tag{3.8}$$

Using Eq. (2.12), it follows that

$$V_\alpha^{2n+1} = V_\alpha^{2n(\mu)} = S^{2n}g_\alpha^\mu + T^{2n\cdot\mu}_{\ \ \ \alpha}, \tag{3.9a}$$

$$A_\alpha^{2n+1} = A_\alpha^{2n(\mu)} = \widetilde{T}^{2n\mu}_{\ \ \ \cdot\alpha} + g_\alpha^\mu P^{2n}, \tag{3.9b}$$

and

$$S^{2n+2} = S^{2n(\mu\nu)} = V^{2n(\mu)\nu}, \tag{3.10a}$$

$$P^{2n+2} = P^{2n(\mu\nu)} = A^{2n(\mu)\nu}, \tag{3.10b}$$

$$T_{\alpha\beta}^{2n+2} = T_{\alpha\beta}^{2n(\mu\nu)} = V_\alpha^{2n(\mu)}g_\beta^\nu - V_\beta^{2n(\mu)}g_\alpha^\nu + A_\theta^{2n(\mu)}\varepsilon^{\theta\nu}_{\cdot\cdot\alpha\beta}. \tag{3.10c}$$

After substitution of the first set of equations in the second, we obtain

$$S^{2n(\mu\nu)} = S^{2n}g^{\mu\nu} - T^{2n\mu\nu}, \tag{3.11a}$$

$$P^{2n(\mu\nu)} = \widetilde{T}^{2n\mu\nu} + P^{2n}g^{\mu\nu}, \tag{3.11b}$$

$$T_{\alpha\beta}^{2n(\mu\nu)} = T_{\alpha\beta}^{2n}g^{\mu\nu} + S^{2n}\delta_{\alpha\beta}^{\mu\nu} + T^{2n\overline{\theta}}_{\theta\,\cdot}\delta^{\mu\nu}_{\overline{\theta}\tau}\delta^{\theta\tau}_{\alpha\beta}. \tag{3.11c}$$

In the last formula, use has been made of Eq. (A7c) to explicitly show the antisymmetry in α, β.

From Eqs. (3.10) and (3.11a) we can formulate the following theorem.

Theorem: The coefficients in the covariant expansion of the product of n γ matrices are a linear combination of S coefficients of at most a product of $(n+3)$ γ matrices.

In fact, from Eq. (3.11a)

$$T^{2n\alpha\beta} = S^{2n}g^{\alpha\beta} - S^{2n(\alpha\beta)} = \tfrac{1}{2}(S^{2n(\beta\alpha)} - S^{2n(\alpha\beta)}), \tag{3.12a}$$

and from Eq. (3.10a)

$$V^{2n(\mu)\alpha} = S^{2n(\mu\alpha)}. \tag{3.12b}$$

Combining this formula with Eq. (3.11c) and (A7), we get

$$A^{2n(\mu)\alpha} = -\tfrac{1}{6}T^{2n(\mu\rho)\theta\tau}\varepsilon^{\alpha}_{\cdot\rho\theta\tau} = \tfrac{1}{6}S^{2n(\mu\rho\theta\tau)}\varepsilon^{\alpha}_{\cdot\rho\theta\tau}. \tag{3.12c}$$

Finally, from this and Eq. (3.10b), we get

$$P^{2n(\mu\nu)} = A^{2n(\mu)\nu} = \tfrac{1}{6}S^{2n(\mu\rho\theta\tau)}\varepsilon^{\nu}_{\cdot\rho\theta\tau}. \tag{3.12d}$$

The relevance of this theorem is clear if one observes that the S coefficients are directly related, in the matricial representation of the algebra elements, to the traces (invariants) of the algebra elements. From the well-known trace theorems,[8] we obtain

$$S^{2n} = S^{(\alpha_1\alpha_2\cdots\alpha_{2n})} = \tfrac{1}{4}\mathrm{Tr}(\gamma^{\alpha_1}\cdots\gamma^{\alpha_{2n}}) =: [\alpha_1\alpha_2\cdots\alpha_{2n}], \tag{3.13}$$

where a convenient notation has been introduced in the last line. In terms of it, the recursive relation between S^{2n} and S^{2n-2} reads

$$\begin{aligned} S^{2n} &= [\alpha_1\alpha_2\cdots\alpha_{2n}] \\ &= \sum_{j=2}^{2n}(-1)^P g^{\alpha_1\alpha_j}[\alpha_2\alpha_3\cdots\alpha_{j-1}\alpha_{j+1}\cdots\alpha_{2n}] \\ &= \sum_{j=2}^{2n}(-1)^j g^{\alpha_1\alpha_j}[\alpha_2\alpha_3\cdots\alpha_{j-1}\alpha_{j+1}\cdots\alpha_{2n}], \end{aligned} \tag{3.14}$$

with $(-1)^P = +1$ or -1 if the permutation

$$\begin{pmatrix} 1\,2\,3\cdots 2n \\ 1\,j\,2\cdots 2n \end{pmatrix}$$

is even or odd. Thus, when fully expanded, S^{2n} has $(2n-1)!!$ terms. If no confusion arises, we will also use the square bracket notation for "contracted" γ products. For example,

$$\begin{aligned} \mathrm{Tr}\, \not{a}\, \not{b}\, \gamma^\mu \not{d} &= a_\alpha b_\beta d_\nu\, \mathrm{Tr}\, \gamma^\alpha\gamma_\beta\gamma^\mu\gamma^\nu \\ &= 4a_\alpha b_\beta d_\nu [\alpha\beta\mu\nu] \\ &=: 4[a\, b\, \mu\, d\,]. \end{aligned} \tag{3.15}$$

Using (3.14) we can obtain explicit formulas for the relations (3.12). From (3.12a) and (3.14) we get

$$T^{2n\alpha\beta} = \sum_i (-1)^i g^{\alpha\alpha_i}[\beta\alpha_1\cdots\alpha_{i-1}\alpha_{i+1}\cdots\alpha_{2n}]. \tag{3.16a}$$

This expression can be still simplified to

$$T^{2n\alpha\beta} = -2\sum_{i<j=1}^{2n}(-1)^{i-j}g^{\alpha\alpha_i} \times g^{\beta\alpha_j}[\alpha_1\cdots\alpha_{i-1}\alpha_{i+1}\cdots\alpha_{j-1}\alpha_{j+1}\cdots\alpha_{2n}], \tag{3.16b}$$

which is easier to use than the more compact looking form of Eq. (3.12a) because the antisymmetry of the α,β indices has been taken into account. Equation (3.16b) has $n[(2n-1)!!]$ different terms.

The V_α^{2n+1} coefficient in Eq. (3.12b) cannot, in general, be simplified any further. An explicit formula for it is

$$V_\alpha^{2n+1}=\sum_{i=1}^{2n+1}(-1)^{i-1}g_\alpha^{\alpha_i}[\alpha_1\cdots\alpha_{i-1}\alpha_{i+1}\cdots\alpha_{2n+1}], \tag{3.17}$$

which has $(2n+1)!!$ different terms.

In analogous form, we can find explicit formulas for $P^{2n(\mu\alpha)}$ and A_α^{2n+1}. From Eqs. (3.12b) and (3.12c), we obtain

$$\begin{aligned}A^{2n+1\alpha}&=A^{2n(\alpha_{2n+1})\alpha}=P^{2n(\alpha_{2n+1}\alpha)}\\&=\sum_{i<j<k=3}^{2n+1}(-1)^{i+j+k}[\alpha_1\cdots\alpha_{i-1}\alpha_{i+1}\cdots\alpha_{j-1}\alpha_{j+1}\\&\quad\cdots\alpha_{k-1}\alpha_{k+1}\cdots\alpha_{2n+1}]\varepsilon^{\alpha_i\alpha_j\alpha_k\alpha}.\end{aligned} \tag{3.18}$$

As we will see in the example below it is possible to reduce the number of terms in Eq. (3.18) using the identity

$$\begin{aligned}g^{\mu\nu}\varepsilon^{\alpha\beta\gamma\delta}=&-g^{\mu\alpha}\varepsilon^{\beta\gamma\delta\nu}+g^{\mu\beta}\varepsilon^{\alpha\gamma\delta\nu}\\&-g^{\mu\gamma}\varepsilon^{\alpha\beta\gamma\nu}+g^{\mu\delta}\varepsilon^{\alpha\beta\gamma\nu}.\end{aligned} \tag{3.19}$$

The number of terms in (3.18) is $(2n+1)!!(n/3)$, for $n\geqslant 1$. Use of (3.19) reduces this a bit, in any case this is smaller than the $(2n+3)!!$ originally expected.

We are now prepared to generalize Eqs. (2.8) and (2.9). Thus Γ^4 is

$$\begin{aligned}&\gamma^{\alpha_1}\gamma^{\alpha_2}\gamma^{\alpha_3}\gamma^{\alpha_4}\\&=(g^{\alpha_1\alpha_2}g^{\alpha_3\alpha_4}-g^{\alpha_1\alpha_3}g^{\alpha_2\alpha_4}+g^{\alpha_1\alpha_4}g^{\alpha_2\alpha_3})\\&\quad-i[g^{\alpha_1\alpha_2}\sigma^{\alpha_3\alpha_4}-{}^{\alpha_1\alpha_3}\sigma^{\alpha_2\alpha_4}+g^{\alpha_1\alpha_4}\sigma^{\alpha_2\alpha_3}\\&\quad+g^{\alpha_2\alpha_3}\sigma^{\alpha_1\alpha_4}-g^{\alpha_2\alpha_4}\sigma^{\alpha_1\alpha_3}+g^{\alpha_3\alpha_4}\sigma^{\alpha_1\alpha_2}]\\&\quad+i\varepsilon^{\alpha_1\alpha_2\alpha_3\alpha_4}\gamma^5.\end{aligned} \tag{3.20}$$

For Γ^5, we get

$$\begin{aligned}&\gamma^{\alpha_1}\gamma^{\alpha_2}\gamma^{\alpha_3}\gamma^{\alpha_4}\gamma^{\alpha_5}\\&=[\alpha_1\alpha_2\alpha_3\alpha_4]\gamma^{\alpha_5}-[\alpha_1\alpha_2\alpha_3\alpha_5]\gamma^{\alpha_4}\\&\quad+[\alpha_1\alpha_2\alpha_4\alpha_5]\gamma^{\alpha_3}-[\alpha_1\alpha_3\alpha_4\alpha_5]\gamma^{\alpha_2}+[\alpha_2\alpha_3\alpha_4\alpha_5]\gamma^{\alpha_1}\\&\quad+i\{g^{\alpha_1\alpha_2}\varepsilon^{\alpha_3\alpha_4\alpha_5\nu}-g^{\alpha_1\alpha_3}\varepsilon^{\alpha_2\alpha_4\alpha_5\nu}\\&\quad+g^{\alpha_1\alpha_4}\varepsilon^{\alpha_2\alpha_3\alpha_5\nu}-g^{\alpha_1\alpha_5}\varepsilon^{\alpha_1\alpha_2\alpha_3\nu}\\&\quad+g^{\alpha_2\alpha_3}\varepsilon^{\alpha_1\alpha_4\alpha_5\nu}-g^{\alpha_2\alpha_4}\varepsilon^{\alpha_1\alpha_3\alpha_5\nu}\\&\quad+g^{\alpha_2\alpha_5}\varepsilon^{\alpha_1\alpha_3\alpha_4\nu}+g^{\alpha_3\alpha_4}\varepsilon^{\alpha_1\alpha_2\alpha_5\nu}\\&\quad-g^{\alpha_3\alpha_5}\varepsilon^{\alpha_1\alpha_2\alpha_4\nu}+g^{\alpha_4\alpha_5}\varepsilon^{\alpha_1\alpha_2\alpha_3\nu}\}\gamma^5\gamma_\nu.\end{aligned} \tag{3.21a}$$

Using (3.19) simplifies the term $\gamma^5\gamma_\nu$. We get, using it twice,

$$\begin{aligned}&\gamma^{\alpha_1}\gamma^{\alpha_2}\gamma^{\alpha_3}\gamma^{\alpha_4}\gamma^{\alpha_5}\\&=[\alpha_1\alpha_2\alpha_3\alpha_4]\gamma^{\alpha_5}-[\alpha_1\alpha_2\alpha_3\alpha_5]\gamma^{\alpha_4}\\&\quad+[\alpha_1\alpha_2\alpha_4\alpha_5]\gamma^{\alpha_3}-[\alpha_1\alpha_3\alpha_4\alpha_5]\gamma^{\alpha_2}+[\alpha_2\alpha_3\alpha_4\alpha_5]\gamma^{\alpha_1}\\&\quad+i(g^{\alpha_1\alpha_2}\varepsilon^{\alpha_3\alpha_4\alpha_5\nu}-g^{\alpha_1\alpha_3}\varepsilon^{\alpha_2\alpha_4\alpha_5\nu}\\&\quad+g^{\alpha_2\alpha_3}\varepsilon^{\alpha_1\alpha_4\alpha_5\nu}+g^{\alpha_4\alpha_5}\varepsilon^{\alpha_1\alpha_2\alpha_3\nu}\\&\quad-g^{\alpha_4\nu}\varepsilon^{\alpha_1\alpha_2\alpha_3\alpha_5}+g^{\alpha_5\nu}\varepsilon^{\alpha_1\alpha_2\alpha_3\alpha_4})\gamma^5\gamma_\nu\end{aligned} \tag{3.21b}$$

which has six, instead of ten, factors in the $\gamma^5\gamma^\nu$ term but in which the symmetry among the indices has been lost.

IV. IDENTITIES FOR CORRELATED γ MATRICES' PRODUCTS

One of the possibilities that the theorem of the previous section brings out is the systematic development of identities for correlated products of γ matrices. Examples of such identities are

$$\gamma^\mu\gamma_\mu=4, \tag{4.1a}$$

$$\gamma^\mu\gamma^\alpha\gamma_\mu=-2\gamma^\alpha. \tag{4.1b}$$

These identities find their use in higher-order calculations in field theory and in the simplification of expressions for the crossed terms of the square of identical particle amplitudes. In particular, Sirlin[5] has found recently several identities of this kind and has used them in higher-order computations of gauge theories.

Perhaps the most clear way to understand why the contraction of two γ matrices leads to simplifications is to observe that

$$\begin{aligned}\gamma^\mu\Gamma\gamma_\mu&=\gamma^\mu(S+iP\gamma^5+V_\alpha\gamma^\alpha\\&\quad+iA_\alpha\gamma^5\gamma^\alpha-(i/2)T_{\alpha\beta}\sigma^{\alpha\beta})\gamma_\mu\\&=2(2S-2iP\gamma^5-V_\alpha\gamma^\alpha+iA_\alpha\gamma^5\gamma^\alpha).\end{aligned} \tag{4.2}$$

This means that the resulting coefficient of the covariant expansion is simply the original coefficient times a factor. Besides this, the tensor coefficient is annihilated by the $\gamma^\mu\Gamma\gamma_\mu$ operation. From Eq. (4.2) and the formulas (3.13), (3.16b), (3.17), and (3.18) we can, by a simple substitution, obtain any relation of the form

$$\begin{aligned}\gamma^\mu\Gamma^N\gamma_\mu&=\gamma^\mu\gamma^{\alpha_1}\gamma^{\alpha_2}\cdots\gamma^{\alpha_N}\gamma_\mu\\&=\begin{cases}4(S^{2n}-iP^{2n}\gamma^5), & \text{if } N=2n,\\ 2(-V_\alpha^{2n+1}\gamma^\alpha+iA_\alpha^{2n+1}\gamma^5\gamma^\alpha), & \text{if } N=2n+1.\end{cases}\end{aligned} \tag{4.3}$$

We can now explore relations of the form (4.2) for other basic element. From the γ^5 dual properties (2.14) and (2.15), it is clear that

$$\gamma^5\Gamma\gamma_5=S+iP\gamma^5-(i/2)T_{\alpha\beta}\sigma^{\alpha\beta}-(V_\alpha\gamma^\alpha+iA_\alpha\gamma^5\gamma^\alpha). \tag{4.4}$$

From this and (4.2) we get

$$\gamma^5\gamma^\mu\Gamma\gamma^5\gamma_\mu=4(-S+iP\gamma^5)+2(-V_\alpha\gamma^\alpha+iA_\alpha\gamma^5\gamma^\alpha). \tag{4.5}$$

These formulas form the basis for a set of relations of the type (4.3).

Less direct relations emerge from the doubly contracted products. For this case we have

$$\gamma^\mu\gamma^\nu\Gamma\gamma_\nu\gamma_\mu=16(S+iP\gamma^5)+4(V_\alpha\gamma^\alpha+iA_\alpha\gamma^\alpha\gamma^5), \tag{4.6}$$

and

$$\gamma^\mu\gamma^\nu\Gamma\gamma_\mu\gamma_\nu=-8(S+iP\gamma^5)+4(V_\alpha\gamma^\alpha+iA_\alpha\gamma^5\gamma^\alpha). \tag{4.7}$$

Subtracting the two last formulas, we get

$$\gamma^\mu\gamma^\nu\Gamma\sigma_{\mu\nu}=12(-iS+P\gamma^5) \tag{4.8}$$

and

$$\sigma^{\mu\nu}\Gamma\sigma_{\mu\nu} = 12(S + iP\gamma^5). \qquad (4.9)$$

The last relation shows that the $\sigma^{\mu\nu}$ sandwich of an arbitrary algebra element has annihilated all but the S and P expansion coefficients.

Let us finally quote two mixed basis elements contracted products

$$\begin{aligned}\gamma_\mu\Gamma\sigma^{\mu\nu} = & (- 3iV^\nu) + i(3iA^\nu)\gamma^5 \\ & + (3iSg^\nu_\alpha + iT^{\nu}_{\cdot\alpha})\gamma^\alpha \\ & + i(3iPg^\nu_\alpha + i\tilde{T}^{\nu}_{\cdot\alpha})\gamma^5\gamma^\alpha \\ & - (i/2)(- iV_\alpha\delta^{\alpha\nu}_{\theta\tau} + iA_\alpha\varepsilon^{\alpha\nu}_{\cdot\cdot\theta\tau})\sigma^{\theta\tau},\end{aligned}$$

and

$$\begin{aligned}\sigma^{\nu\mu}\Gamma\gamma_\mu = & (2iV^\nu) + i(3iA^\nu)\gamma^5 + (- 3iSg^\nu_\alpha + iT^{\nu}_{\cdot\alpha})\gamma^\alpha \\ & + i(- 3iPg^\nu_\alpha - iT^{\nu}_{\cdot\alpha})\gamma^5\gamma^\alpha \\ & - (i/2)(- iV_\alpha\delta^{\alpha\nu}_{\theta\tau} + iA_\alpha\varepsilon^{\alpha\nu}_{\cdot\cdot\theta\tau})\sigma^{\theta\tau}.\end{aligned}$$

From which a new set of identities can be derived using the relations of the previous section.

We can therefore induce from these considerations a simple rule to obtain identities among correlated products of γ matrices: First, derive the identity for a general element of Dirac algebra. Then use the formulas for the expansion coefficients of products of γ matrices.

V. USING THE COVARIANT BASIS

A. Squaring of matrix elements: Spin and particle–antiparticle dependence

Let us consider a one-line spin $\frac{1}{2}$ fermion amplitude, such as the one in Fig. 1. In terms of the incoming, i, and outgoing, f, spinors the amplitude can be written as

$$\mathscr{M}_{fi} = \bar{\omega}_f\Gamma\omega_i, \qquad (5.1)$$

where Γ is a 4×4 matrix that depends on external parameters (polarizations, momenta, etc.) of any particle or source that participates in the process. Plane wave spinors ω_f and ω_i may correspond to particle, u, or antiparticle, v, states. In general to make contact with the experiment one must take the square of $\mathscr{M}_{fi}$. The standard[8] procedure leads to

$$|\mathscr{M}_{fi}|^2 = \mathrm{Tr}[\Gamma\Lambda(p_i)\Sigma(s_i)\bar{\Gamma}\Lambda(p_f)\Sigma(s_f)], \qquad (5.2)$$

where

$$\bar{\Gamma} = \gamma^0\Gamma^\dagger\gamma^0, \qquad (5.3)$$

and the energy Λ and spin Σ projectors have the form

$$\Lambda(p) = (\not{p} \pm m)/2m, \qquad (5.4)$$

$$\Sigma(s) = (1 + \gamma^5\not{s})/2; \qquad (5.5)$$

with $p\cdot s = 0$ and the signs in the energy projector correspond to particle (+) or antiparticle (−) states.

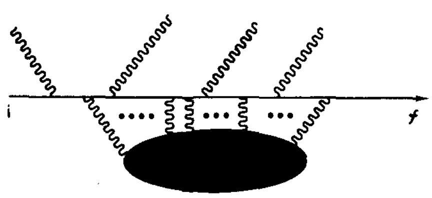


FIG. 1. One fermion line amplitude considered in Eq. (5.1). The ball might contain within itself other fermion lines provided they are distinguishable. Wavy lines denote arbitrary bosonic interactions.

When expressed in terms of the covariant basis, Γ has the form of Eq. (2.4) with the expansion coefficients depending on external parameters. We can also expand $\bar{\Gamma}$ in the covariant basis. We get

$$\bar{\Gamma} = \bar{S} + i\bar{P}\gamma^5 + \bar{V}_\alpha\gamma^\alpha + i\bar{A}_\alpha\gamma^5\gamma^\alpha - (i/2)\bar{T}_{\alpha\beta}\sigma^{\alpha\beta}, \qquad (5.6)$$

with

$$\bar{S} = S^*, \quad \bar{P} = P^*, \quad \bar{V}_\alpha = V^*_\alpha,$$
$$\bar{A}_\alpha = -A^*_\alpha, \quad \bar{T}_{\alpha\beta} = -T^*_{\alpha\beta}. \qquad (5.7)$$

Using the multiplication table, we will obtain expressions for

$$\Gamma_1 = \Gamma\cdot\Lambda(p_i)\cdot\Sigma(s_i) \qquad (5.8a)$$

and for

$$\Gamma_2 = \bar{\Gamma}\cdot\Lambda(p_f)\cdot\Sigma(s_f). \qquad (5.8b)$$

In terms of Γ_1 and Γ_2, the trace in Eq. (5.2) can be evaluated immediately. Since the only matrix with nonvanishing trace of the basis is $\mathbf{1}$, we get from Eq. (2.12)

$$|\mathscr{M}_{fi}|^2 = 4(S_1S_2 - P_1P_2 + V_1\cdot V_2 + A_1\cdot A_2 - \tfrac{1}{2}T_{1\alpha\beta}T_2^{\alpha\beta}). \qquad (5.9)$$

In the expression (5.8) the order of $\Lambda(p)$ and $\Sigma(s)$ projectors is irrelevant as long as $p\cdot s = 0$. Because one is frequently interested in spin-summed probabilities we consider first

$$\Gamma(p,m) = \Gamma\cdot\Lambda(p).$$

Performing this product, we get for particles

$$S(p,m) = \tfrac{1}{2}(S + (V\cdot p)/m), \qquad (5.10a)$$

$$P(p,m) = \tfrac{1}{2}(P + (A\cdot p)/m), \qquad (5.10b)$$

$$V_\alpha(p,m) = \tfrac{1}{2}(V_\alpha + T_{\alpha\mu}(p^\mu/m) + S(p_\alpha/m)) \qquad (5.10c)$$

$$A_\alpha(p,m) = \tfrac{1}{2}(A_\alpha + (p^\mu/m)\tilde{T}_{\mu\alpha} + P(p_\alpha/m)) \qquad (5.10d)$$

$$T_{\alpha\beta}(p,m) = \frac{1}{2}\left[T_{\alpha\beta} + \frac{1}{m}(V_\alpha p_\beta - V_\beta p_\alpha) + \frac{1}{m}\varepsilon_{\alpha\beta\mu\nu}A^\mu p^\nu\right]. \qquad (5.10e)$$

If, instead of a particle projector, one needs the antiparticle projector, one simply changes the sign of p in this formula, without changing the sign of p in S, P, V, A, or T. From this relation for $\Gamma(p,m)$, we can get the general spin dependence. Computing

$$\Gamma(p,m;s) = \Gamma(p,m)\Sigma(s) = \Gamma\Lambda(p,m)\Sigma(s),$$

one obtains

$$S(p,m;s) = \frac{1}{4}\left[S + \frac{V\cdot p}{m} - i\left(A\cdot s + \frac{p^\mu}{m}\tilde{T}_{\mu\alpha}s^\alpha\right)\right], \qquad (5.11a)$$

$$P(p,m;s) = \frac{1}{4}\left[P + \frac{A\cdot p}{m} + i\left(V\cdot s + s^\alpha T_{\alpha\beta}\frac{p^\beta}{m}\right)\right], \qquad (5.11b)$$

$$V_\alpha(p,m;s)=\frac{1}{4}\left[V_\alpha+\frac{i}{m}\varepsilon_{\alpha\mu\nu\rho}p^\mu V^\nu s^\rho+S\frac{p_\alpha}{m}+iPs_\alpha-\frac{i}{m}(A\cdot s\,p_\alpha-A\cdot p s_\alpha)+T_{\alpha\mu}\frac{p^\mu}{m}+is^\mu\widetilde{T}_{\mu\alpha}\right], \tag{5.11c}$$

$$A_\alpha(p,m;s)=\frac{1}{4}\left[A_\alpha+\frac{i}{m}\varepsilon_{\alpha\mu\nu\rho}p^\mu A^\nu s^\rho-iSs_\alpha+P\frac{p_\alpha}{m}+\frac{i}{m}(s\cdot Vp_\alpha-p\cdot Vs_\alpha)+\frac{p^\mu}{m}\widetilde{T}_{\mu\alpha}+is^\mu T_{\mu\alpha}\right], \tag{5.11d}$$

$$T_{\alpha\beta}(p,m;s)=\frac{1}{4}\Big[T_{\alpha\beta}+\frac{i}{m}(p^\mu\widetilde{T}_{\mu\beta}s_\alpha-p^\mu\widetilde{T}_{\mu\alpha}s_\beta)+i\varepsilon_{\alpha\beta\mu\nu}\left(T^{\mu\rho}\frac{p_\rho}{m}+\frac{p^\mu}{m}S+V^\mu\right)s^\nu$$
$$+\frac{i}{m}(s_\alpha p_\beta-p_\alpha s_\beta)P+\frac{1}{m}(V_\alpha p_\beta-p_\alpha V_\beta)+i(s_\alpha A_\beta-A_\alpha s_\beta)+\frac{1}{m}\varepsilon_{\alpha\beta\mu\nu}A^\mu p^\nu\Big]. \tag{5.11e}$$

The evaluation of the square of the amplitude becomes a matter of simple substitution of Eq. (5.10) or Eq. (5.11) into Eq. (5.9) taking into account Eqs. (5.7) and (5.8). Arbitrary polarization of incoming or outgoing states can be obtained using Eq. (5.11). If antiparticles instead of particles are involved it is enough to change the sign of p in Eqs. (5.10) or Eqs. (5.11) without changing this sign within S, P, V, A, or T.

One can now appreciate better the importance of the theorem in Sec. III because using it we can write any amplitude of the form (5.1) in terms of the covariant basis. Thus, the method of squaring amplitudes presented above becomes practical.

B. Two cases: The vector plus axial-vector amplitude and the vector plus scalar case

In the actual substitution of the general expression for Γ we do not expect a great simplification of the final result. However, in many applications only some of the coefficients are nonzero. We consider first the unpolarized case in which Γ has only nonzero V and A, i.e.,

$$\Gamma=V_\alpha\gamma^\alpha+iA_\alpha\gamma^5\gamma^\alpha, \tag{5.12}$$

which appears frequently in low-order perturbative calculations of electroweak theories. Substituting Eq. (5.10) into Eq. (5.9) one finds for the spin-summed probabilities

$$\sum_{s_i s_f}|\mathscr{M}_{fi}|^2=V_\alpha V^*_\beta\left[g^{\alpha\beta}\left(1-\frac{p_i\cdot p_f}{m^2}\right)+\frac{p_i^\alpha p_f^\beta+p_i^\beta p_f^\alpha}{m^2}\right]$$
$$+A_\alpha A^*_\beta\left[-g^{\alpha\beta}\left(1+\frac{p_i\cdot p_f}{m^2}\right)+\frac{p_i^\alpha p_f^\beta+p_i^\beta p_f^\alpha}{m^2}\right]$$
$$+(V_\alpha A^*_\beta+V^*_\alpha A_\beta)\frac{p_{f_\mu}p_{i_\nu}}{m^2}\varepsilon^{\alpha\beta\mu\nu}. \tag{5.13}$$

We note that the expression is identical for particles or antiparticles. Furthermore, if V and A are relatively real, as in the example below, we get

$$\sum_{s_i s_f}|\mathscr{M}_{fi}|^2=V^2\left(1-\frac{p_i\cdot p_f}{m^2}\right)-A^2\left(1+\frac{p_i\cdot p_f}{m^2}\right)$$
$$+(2/m^2)(p_i\cdot Vp_f\cdot V+p_i\cdot Ap_f\cdot A$$
$$+\varepsilon^{\alpha\beta\mu\nu}V_\alpha A_\beta p_{f\mu}p_{i\nu}). \tag{5.14}$$

The second case that we want to analyze is the amplitude with nonvanishing vector and scalar components. We will give the square of the amplitude with and without spin for this case. The particle–particle amplitude is

$$\mathscr{M}=\bar{u}_f(S+\not{V})u_i. \tag{5.15}$$

Squaring it gives, using Eqs. (5.11) in (5.9),

$$|\mathscr{M}|^2=\mathrm{Tr}(\Gamma\Lambda_i\Sigma_i\overline{\Gamma}\Lambda_f\Sigma_f)$$
$$=\frac{1}{4}\Big\{\left(S+\frac{p_i}{m}\cdot V\right)\left(S^*+\frac{p_f}{m}\cdot V^*\right)-(iV\cdot s_i)(iV^*\cdot s_f)+\left(V_\alpha+\frac{p_{i\alpha}}{m}S+\frac{i}{m}\varepsilon_{\alpha\beta\mu\nu}p_i^\beta V^\mu s_i^\nu\right)$$
$$\times\left(V^{*\alpha}+\frac{p_f^\alpha}{m}S^*+\frac{i}{m}\varepsilon^{\alpha\rho\sigma\tau}p_{f\rho}V^*_\sigma s_{f\tau}\right)+\left(-iSs_{i\alpha}-\frac{i}{m}V\cdot p_is_{i\alpha}+\frac{i}{m}V\cdot s_ip_{i\alpha}\right)\Big(-iS^*s_f^\alpha-\frac{i}{m}V^*\cdot p_fs_f^\alpha$$
$$+\frac{i}{m}V^*\cdot s_fp_f^\alpha\Big)-\frac{1}{2}\left[i\varepsilon_{\alpha\beta\theta\tau}\left(\frac{1}{m}Sp_i^\theta+V^\theta\right)s_i^\tau+\frac{1}{m}(V_\alpha p_{i\beta}-V_\beta p_{i\alpha})\right]$$
$$\times\left[i\varepsilon^{\alpha\beta}{}_{\cdot\cdot\rho\sigma}\left(\frac{1}{m}S^*p_f^\rho+V^{*\rho}\right)s_f^\sigma+\frac{1}{m}(V^{\alpha*}p_f^\beta-V^{\beta*}p_f^\alpha)\right]\Big\}, \tag{5.16}$$

simplification of this equation with the aid of the Appendix gives

$$|\mathscr{M}|^2=\frac{1}{4}\Big\{\left[SS^*\left(1+\frac{p_i\cdot p_f}{m^2}\right)+V\cdot V^*\left(1-\frac{p_i\cdot p_f}{m^2}\right)+\frac{1}{m}(SV^*+S^*V)\cdot(p_i+p_f)+V_\alpha V^*_\beta\frac{p_i^\alpha p_f^\beta+p_i^\beta p_f^\alpha}{m}\right](1-s_i\cdot s_f)$$
$$+\frac{1}{m^2}\Big[-(SS^*+V\cdot V^*)p_i\cdot s_fp_f\cdot s_i+\left(\frac{S}{m}p_i+V\right)\cdot s_f\left(\frac{S^*}{m}p_f+V^*\right)\cdot s_im^2$$
$$+\left(\frac{S^*}{m}p_i+V^*\right)\cdot s_f\left(\frac{S}{m}p_f+V\right)\cdot s_im^2+s_f\cdot p_i(V^*\cdot s_iVp_f+V\cdot s_iV^*\cdot p_f)+s_i\cdot p_f(V^*\cdot s_fV\cdot p_i+V\cdot s_fV^*\cdot p_i)$$
$$-p_i\cdot p_f(V^*\cdot s_iV\cdot s_f+V\cdot s_fV^*\cdot s_i)\Big]+\frac{i}{m}\varepsilon_{\alpha\beta\mu\nu}\left[\frac{p_f^\alpha p_i^\beta}{m}(S^*V^\mu-SV^{\mu*})+V^\alpha V^{\beta*}(p_i-p_f)^\mu\right]+(s_f+s_i)^\nu\Big\}. \tag{5.17}$$

This expression reduces in the case of a vanishing s_i and s_f to

$$|\mathscr{M}|^2=\frac{1}{4}\left[SS^*\left(1+\frac{p_i\cdot p_f}{m^2}\right)+V\cdot V^*\left(1-\frac{p_i\cdot p_f}{m^2}\right)+\frac{1}{m}(SV^*+S^*V)\cdot(p_i+p_f)+V_\alpha V^*_\beta\frac{p_i^\alpha p_f^\beta+p_i^\beta p_f^\alpha}{m^2}\right]. \tag{5.18}$$

Let us now see how these rather general formulas work in a more specific example.

C. An example: One photon emission of an electron in an external field

Let us see now how the method outlined above works in an example. Consider the lowest-order amplitude for emission (or absorption) of a photon by an electron in an arbitrary external electromagnetic field. If the external field is a Coulomb potential this will lead to the Bethe-Heitler formula. If the photon energy is small the emitted radiation is, of course, bremsstrahlung. Two Feynman diagrams contribute to this process, they are depicted in Fig. 2. Calling k the photon momentum, ϵ its polarization vector, and $\mathscr{A}$ the Fourier transform of the external field, we get

$$\mathscr{M}_{fi}=\bar u(p_f)\left[-\left(\frac{\epsilon\cdot p_i}{k\cdot p_i}-\frac{\epsilon\cdot p_f}{k\cdot p_f}\right)\not{\mathscr{A}}+\frac{1}{2k\cdot p_f}\not{k}\not{\epsilon}\not{\mathscr{A}}+\frac{1}{2kp_i}\not{\mathscr{A}}\not{\epsilon}\not{k}\right]u(p_i), \tag{5.19}$$

where the usual manipulation[4] of the amplitude has been done to separate different powers of k. Our method now requires us to express (5.19) in the covariant basis. This can be always done using the theorem in Sec. III. For the problem at hand it is enough to use Eq. (2.9) for the triple product of γ's, which is a particular case of the product of n γ matrices given at the end of Sec. III. We have, using Eq. (2.9),

$$\not{k}\not{\epsilon}\not{\mathscr{A}}=\not{k}\epsilon\cdot\mathscr{A}-\not{\epsilon}k\cdot\mathscr{A}+i(\varepsilon^{\alpha\beta\mu\nu}k_\alpha\epsilon_\beta\mathscr{A}_\mu)\gamma^5\gamma_\nu,\quad \not{\mathscr{A}}\not{\epsilon}\not{k}=\not{k}\epsilon\cdot\mathscr{A}-\not{\epsilon}k\cdot\mathscr{A}+i(\varepsilon^{\alpha\beta\mu\nu}\mathscr{A}_\alpha\epsilon_\beta k_\mu)\gamma^5\gamma_\nu.$$

Substituting this in Eq. (5.16),

$$\mathscr{M}_{fi}=\bar u(p_f)\left\{\left[-\left(\frac{\epsilon\cdot p_i}{k\cdot p_i}-\frac{\epsilon\cdot p_f}{k\cdot p_f}\right)\mathscr{A}_\nu+\frac{1}{2}\left(\frac{1}{k\cdot p_f}+\frac{1}{k\cdot p_i}\right)(\epsilon\cdot\mathscr{A}k_\nu-k\cdot\mathscr{A}\epsilon_\nu)\right]\gamma^\nu\right.$$
$$\left.+i\left[\frac{1}{2}\left(\frac{1}{k\cdot p_f}-\frac{1}{k\cdot p_i}\right)k^\alpha\epsilon^\beta\mathscr{A}^\mu\varepsilon_{\alpha\beta\mu\nu}\right]\gamma^5\gamma^\nu\right\}u(p_i). \tag{5.20}$$

From this we can identify the gauge invariant expressions for V and A:

$$V_\alpha=-\left(\frac{\epsilon\cdot p_i}{k\cdot p_i}-\frac{\epsilon\cdot p_f}{k\cdot p_f}\right)\mathscr{A}_\alpha+\frac{1}{2}\left(\frac{1}{k\cdot p_f}+\frac{1}{k\cdot p_i}\right)(\epsilon\cdot\mathscr{A}k_\alpha-k\cdot\mathscr{A}\epsilon_\alpha),$$
$$A_\alpha=\frac{1}{2}\left(\frac{1}{k\cdot p_f}-\frac{1}{k\cdot p_i}\right)k^\mu\epsilon^\nu\mathscr{A}^\beta\varepsilon_{\mu\nu\beta\alpha}. \tag{5.21}$$

We can now substitute this into Eq. (5.15) and integrate over the required phase space in order to compute a cross section. We remark that the evaluation of the square of the matrix element can always end when the coefficients of Γ in the covariant basis are identified. In fact, this will be the most compact form of expressing this square in most cases. The vectors V and A are the natural vector and axial vector of the amplitude.

If one needs to make connection with other expressions for $|\mathscr{M}_{fi}|^2$ the formulas in the Appendix for products of the Levi–Civita tensor can be very helpful. Using them, we have obtained the following form of the spin-summed $|\mathscr{M}_{fi}|^2$ containing only scalar products of $\mathscr{A}$, ϵ, k, p_f, and p_i:

$$\sum_{s_i,s_f}|\mathscr{M}_{fi}|^2=\left(\frac{\epsilon\cdot p_i}{k\cdot p_i}-\frac{\epsilon\cdot p_f}{k\cdot p_f}\right)^2\left[\mathscr{A}^2\left(1-\frac{p_i\cdot p_f}{m^2}\right)+2\frac{p_i\cdot\mathscr{A}p_f\cdot\mathscr{A}}{m^2}\right]+\frac{(\mathscr{A}\cdot k)^2}{k\cdot p_ik\cdot p_f}\left[\epsilon^2\left(1-\frac{p_i\cdot p_f}{m^2}\right)+\frac{2p_i\cdot\epsilon p_f\cdot\epsilon}{m^2}\right]$$
$$-\frac{2k\cdot\mathscr{A}\epsilon\cdot\mathscr{A}}{m^2}\frac{1}{k\cdot p_ik\cdot p_f}(p_i\cdot k\epsilon\cdot p_f+p_f\cdot k\epsilon\cdot p_i)-\frac{k\cdot\mathscr{A}\epsilon^2}{m^2}\left[(p_f\cdot\mathscr{A}-p_i\cdot\mathscr{A})\left(\frac{1}{p_i\cdot k}-\frac{1}{p_f\cdot k}\right)\right]$$
$$+\frac{p_f\cdot kp_i\cdot k}{2m^2}\left[\epsilon^2\mathscr{A}^2\left(\frac{1}{k\cdot p_i}-\frac{1}{k\cdot p_f}\right)^2+4(\epsilon\cdot\mathscr{A})^2\frac{1}{p_i\cdot kp_f\cdot k}\right]-\frac{1}{m^2}\left(\frac{\epsilon\cdot p_i}{k\cdot p_i}-\frac{\epsilon\cdot p_f}{k\cdot p_f}\right)$$
$$\times\left[2\epsilon\cdot\mathscr{A}(p_f\cdot\mathscr{A}+p_i\cdot\mathscr{A})-2k\cdot\mathscr{A}\left(\frac{\epsilon\cdot p_ip_f\cdot\mathscr{A}}{k\cdot p_i}+\frac{\epsilon\cdot p_fp_i\cdot\mathscr{A}}{k\cdot p_f}\right)+\mathscr{A}^2\left(\frac{\epsilon\cdot p_f}{k\cdot p_f}-\frac{\epsilon\cdot p_i}{k\cdot p_i}\right)(k\cdot p_i-k\cdot p_f)\right], \tag{5.22}$$

which leads to the Bethe-Heitler formula. The first term of it is just the bremsstrahlung soft photon amplitude.

We observe that, in the method of evaluation of $|\mathscr{M}_{fi}|^2$ that we have followed, the largest S coefficient that has been needed is S^4 (in the development of the triple γ product), which contains three independent terms. This can be compared with the S^8 that appears in the usual evaluation of the spinless form of Eq. (5.2). Here, S^8 has $7!!=105$ terms. Equa-

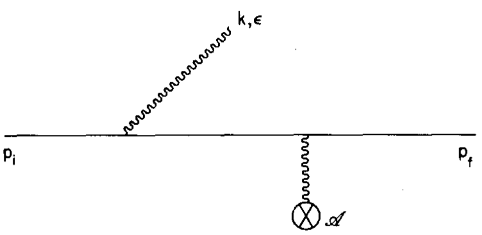


FIG. 2. Lowest-order Feynman amplitudes of photon emission. Wavy lines are photons, $\mathscr{A}$ is a general source. The emitted photon has momentum k and polarization ϵ.

tion (5.19) has been, in fact, verified with this method[9] or with more recent methods.[10]

VI. FINAL REMARKS

We have shown how the use of the covariant basis of the algebra of Dirac leads to a general procedure for the evaluation of squares of one-line spin $\frac{1}{2}$ amplitudes. This method saves, in general, a lot of effort in this evaluation. The method is based on the theorem of Sec. III in which the multiple product of n γ matrices was computed in terms of the S coefficients (traces) of at most $(n+3)$ γ matrices. Furthermore, this theorem brings out additionally a systematic procedure for the generation of generalized identities for correlated products of γ matrices of Dirac algebra elements, as was pointed out in Sec. IV.

A by-product of this method is the observation that, if parity is conserved, each physical amplitude naturally defines a scalar, a pseudoscalar, a vector, an axial vector, and an antisymmetric tensor of rank 2 in terms of which the probabilities of the physical process can be evaluated, as was done in Sec. V. The spin dependence of this probability for a general one-line process was expressed in terms of these five objects. These five objects (S, P, V, A, T) are themselves spin independent and are also independent on the exchange of particle by antiparticle.

The method of squaring amplitudes can be extended to the case of identical particle scattering. The only complication is the presence of the crossed term in the square of the amplitude. Such a term implies a trace of the form

$$\mathrm{Tr}[(\Gamma_1\mathscr{P}(p_1,s_1)\bar{\Gamma}_1\mathscr{P}(p_4,s_4)\Gamma_2\mathscr{P}(p_2,s_2)\bar{\Gamma}_2\mathscr{P}(p_3,s_3) \pm \Gamma_1\mathscr{P}(p_1,s_1)\bar{\Gamma}_1\mathscr{P}(p_3,s_3)\Gamma_2\mathscr{P}(p_2,s_2)\bar{\Gamma}_2(p_4,s_2)],$$

with $\mathscr{P}(p,s) = \Lambda(p)\Sigma(s)$. This trace is, in general, more difficult to evaluate than the one in Eq. (5.2). However, some simplification arising from correlated products of γ matrices usually occurs.

Let us finally comment on the squaring of amplitudes method presented in this work as compared to other approaches to the problem. We have in mind the Jacob and Wick[11] helicity formalism and the recently proposed projector or covariant polarization method of Caffo and Remiddi.[2] Both of these two formalisms exploit the properties of the initial and final wave functions. The work here presented concentrates instead on the "transition operator" Γ. We therefore regard the two approaches as complementary, the use of one does not prevent taking the other. The main limitation of our approach is its present restriction to spin $\frac{1}{2}$ particles. This, however, might be overcome in the future using a formalism of Dirac–Fierz–Bargmann–Wigner.[13] This extension has been done for the Caffo–Remidii work by Passarino.[14]

On the other hand, the main advantage of the method here proposed is that one does not foresee the technical troubles faced when helicity-type amplitudes are used. These problems stem from the kinematical singularities that often appear in the helicity states.

ACKNOWLEDGMENTS

Its a pleasure to thank G. Cocho, A. García, P. Kielanowski, M. Moshinsky, M. A. Pérez, and J. Plebański for many enlightening discussions on this work. The warm hospitality that the author has received from the CINVESTAV, where this work was done, is greatly appreciated. I want also thank Mr. F. Barrios and Mr. S. Saavedra for helping to check the formulas and correct some algebraic mistakes in the last section. I would also like to thank M. Veltman for remarks about the connection of this paper to previous works on the subject.

APPENDIX: AUXILIARY RELATIONS

Let us first define the generalization Kronecker delta of $2n$ indices, each of which can take p values, as

$$\delta\begin{pmatrix}\alpha_1\alpha_2\cdots\alpha_n\\ \beta_1\beta_2\cdots\beta_n\end{pmatrix} = \begin{cases} 1, & \text{if } P\begin{pmatrix}\alpha_1\cdots\alpha_n\\ \beta_1\cdots\beta_n\end{pmatrix} \text{ is even and all } \alpha \text{ indices are different,}\\ -1, & \text{if } P\begin{pmatrix}\alpha_1\cdots\alpha_n\\ \beta_1\cdots\beta_n\end{pmatrix} \text{ is odd and all } \alpha \text{ indices are different,}\\ 0, & \text{other cases,}\end{cases} \tag{A1}$$

where P denotes a permutation of the lower indices with respect to the upper indices and all indices α must be different among themselves. In terms of this delta, we have

$$\varepsilon_{\mu\nu\alpha\beta} = \delta\begin{pmatrix}0 & 1 & 2 & 3\\ \mu & \nu & \alpha & \beta\end{pmatrix}. \tag{A2}$$

We remark that the position of the indices of the generalized Kronecker δ bears no relation with covariant or contravariant indices. Taking this remark into account and raising indices with $g^{\alpha\rho} = \mathrm{diag}(1, -1, -1, -1)$, we have

$$\varepsilon^{\mu\nu\alpha\beta} = \delta\begin{pmatrix}0 & 1 & 2 & 3\\ \mu & \nu & \alpha & \beta\end{pmatrix}. \tag{A3}$$

From the last two formulas it follows that

$$\varepsilon^{\alpha_1\alpha_2\alpha_3\alpha_4}\varepsilon_{\beta_1\beta_2\beta_3\beta_4} = -\delta\begin{pmatrix}\alpha_1 & \alpha_2 & \alpha_3 & \alpha_4\\ \beta_1 & \beta_2 & \beta_3 & \beta_4\end{pmatrix}. \tag{A4}$$

On the other hand the general relation between δ of successive orders is

$$\delta\begin{pmatrix}\alpha_1 & \alpha_2\cdots\alpha_n\\ \beta_1 & \beta_2\cdots\beta_n\end{pmatrix} = \delta\begin{pmatrix}\alpha_1\\ \beta_1\end{pmatrix}\delta\begin{pmatrix}\alpha_2 & \alpha_3\cdots\alpha_n\\ \beta_2 & \beta_3\cdots\beta_n\end{pmatrix}$$

$$- \delta \begin{pmatrix} \alpha_1 \\ \beta_2 \end{pmatrix} \delta \begin{pmatrix} \alpha_2 & \alpha_3 \cdots \alpha_n \\ \beta_1 & \beta_3 \cdots \beta_n \end{pmatrix}$$

$$- \delta \begin{pmatrix} \alpha_1 \\ \beta_3 \end{pmatrix} \delta \begin{pmatrix} \alpha_2 & \alpha_3 \cdots \alpha_n \\ \beta_2 & \beta_1 \cdots \beta_n \end{pmatrix}$$

$$\vdots$$

$$- \delta \begin{pmatrix} \alpha_1 \\ \beta_n \end{pmatrix} \delta \begin{pmatrix} \alpha_2 & \alpha_3 \cdots \alpha_n \\ \beta_2 & \beta_3 \cdots \beta_1 \end{pmatrix}. \tag{A5}$$

Therefore

$$\sum_{\alpha=1}^{p} \delta \begin{pmatrix} \alpha & \alpha_2 \cdots \alpha_n \\ \alpha & \beta_2 \cdots \beta_n \end{pmatrix} = (p - n + 1) \delta \begin{pmatrix} \alpha_2 \cdots \alpha_n \\ \beta_2 \cdots \beta_n \end{pmatrix}. \tag{A6}$$

In the special case of Eq. (A4), we have

$$\varepsilon^{\alpha\,\alpha_2\alpha_3\alpha_4}\,\varepsilon_{\alpha\beta_2\,\beta_3\,\beta_4} = -\,\delta \begin{pmatrix} \alpha_2\alpha_3\alpha_4 \\ \beta_2\,\beta_3\,\beta_4 \end{pmatrix}, \tag{A7a}$$

$$\varepsilon^{\alpha\,\beta\alpha_3\alpha_4}\,\varepsilon_{\alpha\beta\,\beta_3\,\beta_4} = -\,2\delta \begin{pmatrix} \alpha_3\alpha_4 \\ \beta_3\,\beta_4 \end{pmatrix}, \tag{A7b}$$

$$\varepsilon^{\alpha\beta\mu\alpha_4}\,\varepsilon_{\alpha\beta\mu\beta_4} = -\,6\delta \begin{pmatrix} \alpha_4 \\ \beta_4 \end{pmatrix}, \tag{A7c}$$

and

$$\varepsilon^{\alpha\beta\mu\nu}\,\varepsilon_{\alpha\beta\mu\nu} = -\,4! = -\,24. \tag{A7d}$$

[1]E. R. Caianiello and S. Fubini, Nuovo Cimento **9**, 1218 (1952).

[2]J. S. R. Chisholm, Proc. Cambridge Philos. Soc. **48**, 300 (1952); Nuovo Cimento **30**, 426 (1963).

[3]J. Kahane, J. Math. Phys. **9**, 1732 (1968).

[4]J. S. R. Chisholm, Comput. Phys. Comm. **4**, 205 (1972).

[5]A. Sirlin, Nucl. Phys. B **192**, 93 (1982).

[6]J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw–Hill, New York, 1964).

[7]An interesting modification of this basis has been discussed by J. Plebański in which the elements of the covariant basis are multiplied by $\frac{1}{2}(1 + \gamma^5)$, making the new basis covariant and dual: J. Plebański, "On the Generators of the N dimensional Pseudo-Unitary and Pseudo-Orthogonal Groups," Lecture notes delivered at the Physics Department, Centro de Investigación y de Estudios Avanzados del IPN, México, 1966.

[8]Reference 6, Chap. 7.

[9]F. Barrios, B. Sc. thesis, Universidad Nacional Autonoma de México, 1984 (unpublished).

[10]G. Passarino, Phys. Rev. D **28**, 2867 (1983).

[11]M. Jacob and G. C. Wick, Ann. Phys. **7**, 404 (1959).

[12]M. Caffo and E. Remiddi, Helv. Phys. Acta **55**, 339 (1982).

[13]P. A. M. Dirac, Proc. R. Soc. London Ser. **A155**, 447 (1936); M. Fierz, *ibid.* **173**, 221 (1939); V. Bargmann and E. P. Wigner, Proc. Nat. Acad. Sci. **34**, 211 (1982); E. M. Corson, *Introduction to Tensors, Spinors and Relativistic Wave-Equations* (Blackie, Glasgow, 1955); D. Lurié, *Particles and Fields* (Interscience, New York, 1968).

[14]G. Passarino, Nucl. Phys. B **237**, 249 (1984).

The relation of a theory of countable sets to the field equations of physics

D. J. BenDaniel
Symmetrics, Inc., Indian Springs Way, Wellesley, Massachusetts 02181

(Received 19 March 1984; accepted for publication 14 September 1984)

The systematic development of mathematics is based on the theory of sets. We present an axiom system in which, in contradistinction to the usual theories, it seems possible to define formal provability yet some useful mathematics can be derived. Some features of the theory suggest that these axioms can provide a possible new foundation for mathematical physics.

The usual foundations of the theory of sets are the axioms of Zermelo, Fraenkel, and Skolem (ZFS). The ZFS system includes the axioms of extensionality, schema of separation, pairing, sums, infinity, power set, schema of replacement, choice and regularity.[1] This system has been extensively studied and is considered by most experts to be consistent.

In particular, the axiom schema of separation (AS) has been the subject of attention because of its role in certain paradoxes. It is well known that AS can be derived from the axiom schema of replacement (AR).[2]

We begin our discussion by deleting AR (and also AS) from ZFS and designate this system as ZFS − AR. This is our point of departure. To this system, we adjoin an axiom, which we shall call the "axiom of bijective replacement" (ABR), as follows,

$$\text{(ABR)}\quad \forall x\in A\ \exists y\,(\phi(x,y)\wedge \forall x'\ \forall y'(\phi(x',y')\rightarrow(x\neq x'\leftrightarrow y\neq y')))\rightarrow\exists U\,[\forall s\ s\in U\leftrightarrow\exists t\in A\phi(t,s)],$$

where $\phi(x,y)$ is any formula in which x and y are free and U is not free. ABR says that a set is defined if there is a prescription replacing one-for-one all its elements by the elements of another set. From this replaceability, we can show, using the axiom of choice (AC),[3] that AS and ABR are independent and that AS + ABR = AR. Thus ABR restores as much of AR as we can without reintroducing AS.

We now adjoin another axiom, "all sets are countable" (ASC), which can be generally written as

$$\text{(ASC)}\quad \forall U\ \exists M\ \forall x\in U\ \exists n\ [(x,n)\in M\wedge\forall Z\ n\in Z\wedge\forall y((y,n)\in M\rightarrow y=x)],$$

where Z is any set defined by the axiom of infinity (AI).[4] Here, the set of all natural numbers N is the minimal set defined by AI. ASC, with AC and ABR, says that from any set there is a one-to-one mapping onto a set of natural numbers.

We designate the resultant theory ZFS − AR + ABR−ASC or ZFS − AS + ASC as T (see Ref. 5) and shall attempt to show that T contains all sets of physical relevance.

We consider first some important foundational differences between ZFS and T. The statement "A subset S of U exists with the property ψ" is written

$$\exists S\,[\forall x\in U\ x\in S\leftrightarrow\psi(x)].$$

In ZFS, this statement is always valid by AS. In T, on the other hand, the validity of this statement or the validity of its negation must be shown. For example, the well-known Cantor proof, which in ZFS, through the use of a diagonal set, shows there is no one-to-one mapping between an infinite set and its power set, is instead a proof in T of the nonexistence of that diagonal set. Also, transfinite recursion, often used in ZFS, is not available in T. To show this, we write down the general statement of transfinite recursion for sets of natural numbers

$$\forall n\in N\ \exists u\in N(\phi(n,u)\wedge\forall m\in N\ \forall v\in N(\phi(m,v)\rightarrow m\neq n\leftrightarrow v\neq u)\wedge((\phi(0,w)\wedge\psi(w))\wedge\forall l\in N\ \forall p\in l\wedge\phi(p,x)\wedge\psi(x)\rightarrow\phi(l,y)\wedge\psi(y)))\rightarrow\forall q\in N\ \exists z\in N\ \phi(q,z)\psi(z).$$

The proof of this statement as a theorem of ZFS is well known. However, what is actually proven is the conditional

$$\exists S'[\forall x\in U\ x\in S'\leftrightarrow\sim\psi(x)]\rightarrow S'=0,$$

where U is the set defined by ϕ, using ABR. In ZFS, the antecedent in this statement is always valid by AS and, hence, by implication, the set S is the null set and the transfinite recursion theorem is proven. In T, on the other hand, the set S is the null set if and only if it exists and thus the proof fails.

As a result, only those parts of mathematics not requiring transfinite recursion are derivable in T. For example, the existence of the set of all prime numbers cannot be shown. We can show, nevertheless, the basic operations of addition and multiplication of all positive and negative integers.[6] Furthermore, we can establish, using ABR and the axiom of sums ($A\Sigma$),[7] the existence of the set of all fractional numbers, sequences of distinct fractionals as bijective maps from the natural numbers and the real numbers as limits of sequences of distinct fractional numbers. The set of real numbers exists and is countable.[8]

The investigation of analysis in the space of countable reals is underway and proofs are still naive. It is, however, a complete metric space. One can show in T that every bounded set of reals has a least upper (and greatest lower) bound. Whether the space is compact and in what sense it is a continuum are open questions.

Let us now construct sets which have physical relevance, that is, functions having a range and domain of continuous real or complex variables. We first state the simple result that the range of a constant function is given by AC. To show the range of nonconstant functions of a real variable, we write down the general statement that a mapping $\phi(x,u)$ of real variables is bijective, letting the domain (the set of all x) be the open set $(0,1)$ and restricting u to the same set,

 0022-2488/85/040585-03$02.50

$$[\forall x\in(0,1)\ \exists u\in(0,1)[\phi(x,u)\wedge x'\in(0,1)\ \forall u'\in(0,1)\exists C_{u,x},C_{x,u}(\phi(x',u')\rightarrow(x'-x')\leqslant C_{x,u}(u-u')\wedge(u-u')\leqslant C_{u,x}(x-x'))].$$

The range (the set of all u) exists by ABR. We now introduce the postulate that $C_{x,u}C_{u,x}=1\wedge C_{u,x}>0\wedge C_{x,u}>0$. Hence $C_{u,x}$ and $C_{x,u}$ each have a least upper bound given by the reciprocal of the greatest lower bound of the other, and we have

$$[\forall x\in(0,1)\ \exists u\in(0,1)[\phi(x,u)\wedge\exists C_1C_2\ \forall x'\in(0,1)\ \forall u'\in(0,1)\ (\phi(x',u')\rightarrow(x-x'|\leqslant C_1|u-u'|\wedge|u-u'|\leqslant C_2|x-x'|)],$$

where C_1 and C_2 are upper bounds. This result may be termed a bi-Lipschitz condition. We can show that u, being bounded, is continuous and differentiable, from the fact that bounded monotonic sequences converge. An identical result is obtained with $C_{x,u}C_{u,x}=1\wedge C_{u,x}<0\wedge C_{x,u}<0$. Furthermore, simply by scaling the variable u, the range can be made equal to the domain and the quantifiers can be equivalently $\forall u\exists x$. We recognize that, repetitively applied, our postulate means "There is a set of equivalent real varaibles, x_1, x_2, ... x_p, in which any change in one is accompanied by some change in all the others" and may be interpreted physically as universal coupling.

From these monotonic pieces, one can construct, us using $A\ \Sigma$, functions which are not in general monotonic but yet are bi-Lipschitz everywhere *except for isolated missing points*. These functions are thus semicontinuous and can be shown to be of bounded variation. If all of the relevant results of analysis based on ZFS can also be derived in this theory, we obtain a Hilbert space. The fundamental equations of physics describing field phenomena in space-time are well known to have general solutions which form such Hilbert spaces.

To show functions of a complex variable, an analogous proof uses bijective mappings of real pairs (x,y) and (u,v). Conformal mappings $w=u+iv=f(z)+f(x+iy)$ are obtained from the postulate $(C_{u,x}=C_{v,y})\wedge(C_{v,x}=-C_{u,y})\wedge C_{x,u}=C_{y,v})\wedge(C_{x,v}=-C_{y,u})$. The result from $(C_{u,x}=-C_{v,y})\wedge(C_{v,x}=C_{u,y})\wedge(C_{x,u}=-C_{y,v})\wedge(C_{x,v}=-C_{y,u})$ refers to mappings where the orientation is reversed. These derivations may be extend to higher dimensions.

We are left with the question, however, as to why physics would have a special preference for the theory T, since the same results could be achieved in ZFS, albeit with an *ad hoc* assumption of bijectivity. Is there any link between mathematical foundations and natural events? An answer may lie in the general treatment of provability due to Tarski. Tarski has shown that, in any consistent theory, one cannot define both the set G of all formulas $\psi(x)$ and the set V of all such formulas which are provable.[9] In set theories rich enough to contain recursive arithmetic and also in which G and V are sets, such as ZFS, it can be shown that the set G is definable and hence the set V cannot be, leading to his result that provability cannot be defined in those theories. *When viewed within T*, even though basic arithmetic is interpretable, without recursion one cannot show G to be a set, since there is no formula generating just those strings of symbols that are syntactically correct, i.e., there can be no prescription for meaning in a system which is rich. This allows a possible extension of T, call it T*, obtained by now adjoining to T an axiom that the set V is definable. This axiom can be written

$$(\mathrm{AV})\quad \exists V[\forall n\in N\ n\in V\leftrightarrow\exists p\in N\ \mathrm{Dem}\ (p,n)],$$

where Dem (p,n) is that arithmetical formula in T* which represents the statement, "The entire preceding sequence of provable formulas, whose aggregate Gödel number is p, is a proof in T* of the formula, which is equivalent to none of those preceding, whose Gödel number is n." With this, we can now show the existence of undecidable formulas in T*. We first note that there is a one-to-one mapping from the set V onto a set $\tilde{V}$ which contains just the Gödel numbers of the negations of the formulas whose Gödel numbers are in V. The set V is definable in T*, hence $\tilde{V}$ is definable and their union $V\cup\tilde{V}$ as well. If this union were equal to G, then G would be definable but that contradicts the Tarski proof. Thus $V\cup\tilde{V}\neq G$. It follows that there are undecidable formulas in T*. The existence of undecidable formulas can be considered the price we pay for provability. It seems reasonable to require provability in the description of natural events yet not to be deterred by undecidable formulas. We would thus prefer T*, if it is consistent, as a foundation for mathematical physics.

ACKNOWLEDGMENTS

My profound thanks to Professor Jan Mycielski of the Department of Mathematics of the University of Colorado who is helping me grasp some of the rudiments of mathematical foundations, to the late Dr. S. Ulam of Los Alamos National Laboratory who put me in contact with him, and to Dr. R. G. Segers of Exxon and Professor W. V. Quine of the Department of Philosophy of Harvard University who have made very helpful and constructive comments.

[1] A review of ZFS and several alternative systems can be found in A. A. Fraenkel, Y. Bar-Hillel, and A. Levy, *Foundations of Set Theory* (North-Holland, Amsterdam, 1958). A suitable general development of the ZFS theory is P. Suppes, *Axiomatic Set Theory* (Van Nostrand, Princeton, 1960).

[2] AS determines a set U from a condition ψ by asserting that any statement of the form $\exists U[\forall x\in B\ x\in U\leftrightarrow\psi(x)]$ is valid, where $\psi(x)$ is any formula in which x is free and U is not free. References (1) give a history of earlier concepts of this axiom.

[3] We use AC only in the form "For each one-to-many mapping there is at least one one-to-one mapping".

[4] AI can be written $\exists Z[0\in Z\wedge\forall x x\in Z\rightarrow(\{x\}\cup x)\in Z]$. In ZFS, more than one set is given by AI with the set of all natural numbers, N, being minimal: $\exists N(\forall n\ n\in N\leftrightarrow\forall Z\ n\in Z)$.

[5] The consistency of T − ABR is proven (R. M. Solovay, private communication) by the existence of a countable model. The consistency of T is made plausible by the facts that any set defined by ABR must be equipollent with an existing set and that transfinite recursion is not available in T.

[6] A discussion of the axiomatic subtheories of arithmetic occurs in the well-known review, by A. Mostowski, R. M. Robinson, and A. Tarski "Undecidability and Essential Undecidability in Arithmetic," in *Undecidable Theories* (North-Holland, Amsterdam, 1953). Defining the numerals Δ_n from N by a bijective mapping, we can interpret in T the fragment of arithmetic which contains all but IV below, whose derivation requires recursion,

I	$\forall n, p \in N$	$\Delta_n + \Delta_p = \Delta_{n+p}$,
II	$\forall n, p \in N$	$\Delta_n \cdot \Delta_p = \Delta_{n \cdot p}$,
III	$\forall n\ p \in N$	$\Delta_n \neq \Delta_p \leftrightarrow n \neq p$,
IV	$\forall n \in N$	$x = \Delta_n \rightarrow x = \Delta_0 \vee x = \Delta_n \vee \ldots x = \Delta_n$,
V	$\forall n, p \in N$	$\Delta_n < \Delta_p \vee \Delta_p < \Delta_n$.

These five axioms are the minimal subtheory for which all recursive sets are definable. Thus some recursive concepts are not definable as sets in T. Furthermore, all the Peano axioms except the induction axiom can also be obtained in T. Peano arithmetic without induction is not essentially undecidable.

[7]AΣ says that a set exists which contain exactly all the elements of the sets included in a set of sets, written as $\forall A\ \exists B\ \forall x\ [x \in B \leftrightarrow \exists C (x \in C \wedge C \in A)]$.

[8]Derivation of the reals is given in D. J. BenDaniel, "A Theory of Countable Sets," to be submitted to Symbolic Logic.

[9]G and V are sets of natural numbers which are obtained by any suitable Gödelization procedure for expressions. The Tarski proof is discussed in detail in the reference of footnote 6 where a diagonal function D is utilized instead of the set G.

Spinors as fundamental objects

Paolo Budinich and Krystyna Bugajska[a)]
International School for Advanced Studies, Trieste, Italy

(Received 17 August 1984; accepted for publication 9 November 1984)

We define, on the algebraic Dirac spinor space Ψ, some operators $D^\pm$ and $T^\pm$. By means of them we show how the fundamental operations of Hermitian conjugation, complex conjugation, bar conjugation, and so on may be introduced in the Clifford algebra approach. These definitions depend on the geometrical property of the pure imaginary unit i of the Dirac algebra $D_{s,t}$ and are the same only for mod 4 dimensions of vector space-times $R^{s,t}$. Furthermore, on the set $\Psi \times \Psi$ we introduce equivalence relations $R_\pm$ and define bijections $\chi_\pm$ between $\Psi \times \Psi / R_\pm$ and $D_{s,t}$. We investigate some properties of $\chi_\pm$ and give the necessary and sufficient conditions for $u \in D_{s,t}$ to belong to some minimal left ideal of $D_{s,t}$. Next we use the decomposition of the Dirac algebra $D_{s,t}$ onto the Dirac spinor spaces to demonstrate two different ways of an action of any element $s \in \mathrm{Spin}(x,t)$. These considerations throw a new light onto the problem of the covariant derivative on the bundle of algebraic spinors over a space-time manifold.

I.

Let us assume that we have an n-dimensional real vector space-time of signature (s,t) which we shall denote by $R^{s,t}$. Let $R_{s,t}$ be its corresponding universal Clifford algebra.[1] It is known that we can find the faithful matrix representation of any Clifford algebra $R_{s,t}$ which exhibits its character as the real algebra of endomorphisms of some F-linear space $S = S(s,t)$. (Here F denotes a division ring given by entries of matrices in the matrix representation of $R_{s,t}$.) Furthermore we know that S can be realized by any minimal left ideal of $R_{s,t}$. But to determine a minimal left ideal of $R_{s,t}$ we have to fix some primitive idempotent $f = f^2$ (see Ref. 2). Let $\{e_i\}$, $i \in (1,\dots,s+t)$ be an orthonormal base of our vector space $R^{s,t}$ and let $\{e_K\}$ be the corresponding canonical basis of $R_{s,t}$ (here K denotes a multi-index), i.e.,

$$e_K = e_{i_1} \cdots e_{i_k}, \quad 1 \leqslant i_1 < i_2 \cdots < i_k \leqslant n. \tag{1.1}$$

It can be shown[3] that for any Clifford algebra $R_{s,t}$ there exist $\chi = t - H(t-s)$ pairwise commuting, nonannihilating idempotents of the form

$$\tfrac{1}{2}(1 + e_K), \tag{1.2}$$

such that their product

$$f^1 = (1/2^\chi)(1 + e_{K_1}) \cdots (1 + e_{K_\chi}) \tag{1.3}$$

is a primitive idempotent.[4] Now, when we vary independently the signs ϵ_i in the product

$$(1/2^\chi)(1 + \epsilon_1 e_{K_1}) \cdots (1 + \epsilon_\chi e_{K_\chi}), \tag{1.4}$$

we obtain 2^χ orthogonal primitve idempotents which we shall denote by f^α, $\alpha = 1,\dots,2^\chi$. They determine the decomposition of $R_{s,t}$ in the corresponding to f_1^α minimal left ideals, i.e.,

$$R_{s,t} = \bigoplus_{\alpha/1}^{2_\chi} R_{s,t} f^\alpha = \bigoplus_{\alpha/1}^{2^\chi} S^\alpha. \tag{1.5}$$

For the simple Clifford algebra $R_{s,t}$ we can always take the matrix representation of $R_{s,t}$ in which the αth column represents S^α (see Refs. 5 and 6). Thus the decomposition (1.5) is equivalent to the decomposition of the matrix algebra in its columns. In a general case, entries of a matrix realization of a given Clifford algebra $R_{s,t}$ belong to one of the rings: $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, ${}^2\mathbb{R}$, or ${}^2\mathbb{H}$, respectively, depending on the concrete dimension $\eta = s + t$ and the concrete signature (s,t). For this reason we are interested rather in the corresponding Dirac algebra $D_{s,t}$ instead of the Clifford algebra $R_{s,t}$ itself. By definition,[7] $D_{s,t}$ is algebraically equivalent to the complexification of the corresponding Clifford algebra $R_{s,t}$. More precisely, for even-dimensional vector space-time $R^{s,t}$, $D_{s,t}$ is determined as a real Clifford algebra of appropriate enlarged vector space $R^{s',t'}$, $s' + t' = s + t + 1$, such that

$$R_{s,t} \subset D_{s,t} := R_{s',t'} \cong R^{\mathbb{C}}_{s,t}. \tag{1.6}$$

Again, Dirac spinor space $\Psi(s,t)$ is, by definition, given by a left minimal ideal of $D_{s,t}$, i.e.,

$$\Psi(s,t) = S(s',t'). \tag{1.7}$$

When $s - t = 0, 2 \bmod 8$, $R_{s,t}$ is realized by the real matrices and in addition we have $\chi(s,t) = \chi(s',t')$. This means that any primitive idempotent f of $R_{s,t}$ can be taken as a primitive idempotent of $R_{s',t'} = D_{s,t}$. Hence in this case

$$\Psi(s,t) = R_{s',t'} f \cong (R_{s,t})^{\mathbb{C}} = S(s,t)^{\mathbb{C}}. \tag{1.8}$$

The above isomorphism can be obtained by taking as the basis elements of the Dirac spinor space $\Psi(s,)$ the basis elements of $S(s,t)$.

II.

Let us fix some even-dimensional vector space $R^{s,t}$, $s + t = 2r$. Its Dirac algebra $D_{s,t}$ is realized by a $2^r \times 2^r$ complex matrix algebra $\mathbb{C}(2^r)$ (see Ref. 8). Let f be a primitive idempotent of $R_{s',t'} = D_{s,t}$ of the form (1.3). Let $\{\rho_1,\dots,\rho_N\}$, $N = 2^r$ be the basis of the corresponding spinor space $S(s',t')$. We can assume that

$$\rho_i = u_i f, \quad i = 1,\dots,N, \tag{2.1}$$

with $u_1 = 1$ and $u_i = e_{S_i}$. [Here, S_i is a multi-index given by (1.1).] Thus any spinor $\psi \in \Psi(s,t) = S(s',t')$ can be written as $\psi = uf$, where

[a)] Present address: Institut für Theoretische Physik der Technischen Universität Clausthal, 3392 Clausthal-Zellerfeld, West Germany.

 0022-2488/85/040588-05$02.50

$$u = \psi_i u_i, \quad \psi_i \in \mathbb{C}. \tag{2.2}$$

Any primitive idempotent f determines not only the left minimal ideal $R_{s',t'} f$ but also the right minimal ideal $fR_{s',t'}$. Hence it is obvious that any element $\tilde{\psi} \in \tilde{\Psi}(s,t) := fD_{s,t}$ can be written as

$$\tilde{\psi} = fu, \tag{2.3}$$

with u given as above, i.e., $u = \psi_i u_i$.

Let β_+ and β_- be the anti-involutions of $R_{s',t'}$ induced by the identity transformation and reflection of $R^{s',t'}$, respectively. These maps allow us to construct operators

$$D^{\pm} : \Psi \to \tilde{\Psi} \tag{2.4}$$

in the following way.

First, $\forall \psi \in \Psi$, we define

$$D^+(\psi) = \omega_+ \beta_+(\psi) = f\omega_+ \beta_+(u), \tag{2.5}$$
$$D^-(\psi) = \omega_- \beta_-(\psi) = f\omega_- \beta_-(u).$$

Similarly we can define operators $\tilde{D}^{\pm} : \tilde{\Psi} \to \Psi$

$$\tilde{D}^+(\tilde{\psi}) = \beta_+(\tilde{\psi})\omega_+ = \beta_+(u)\omega_+ f, \tag{2.6}$$
$$\tilde{D}^-(\tilde{\psi}) = \beta_-(\tilde{\psi})\omega_- = \beta_-(u)\omega_- f.$$

The elements $\omega_\pm$ have to be taken in such a way that

$$\omega_\pm \beta_\pm(f)\omega_\pm^{-1} = f. \tag{2.7}$$

Such elements $\omega_\pm \in R_{s',t'}$ always exist although the property (2.7) does not determine them uniquely.

We also introduce the T operators by

$$T^+(\psi) = D^+(\psi)\omega_+^{-1} = \omega_+ \beta_+(\psi)\omega_+^{-1},$$
$$= f\omega_+ \beta_+(u)\omega_+^{-1}, \tag{2.8}$$
$$T^-(\psi) = D^-(\psi)\omega_-^{-1} = \omega_- \beta_-(\psi)\omega_-^{-1} = f\omega_- \beta_-(u)\omega_-^{-1},$$

and similarly

$$\begin{aligned} \tilde{T}^+(\tilde{\psi}) &= \omega_+^{-1}\beta_+(\tilde{\psi})\omega_+ = \omega_+^{-1}\beta_+(u)\omega_+ f, \\ \tilde{T}-(\tilde{\psi}) &= \omega_-^{-1}\beta_-(\tilde{\psi})\omega_- = \omega_-^{-1}\beta_-(u)\omega_- f. \end{aligned} \tag{2.9}$$

We can check that

$$\tilde{T}^+T^+(\psi) = \psi, \qquad \tilde{D}^+D^+(\psi) = \epsilon_+\psi,$$
$$\text{whereas} \tag{2.10}$$
$$\tilde{T}^-T^-(\psi) = \psi, \qquad \tilde{D}^-D^-(\psi) = \epsilon_-\psi,$$

with ϵ_+ given by

$$\beta_\pm(\omega_\pm) = \epsilon_\pm \omega_\pm^{-1}. \tag{2.11}$$

Let us recall that $R_{s',t'} = D_{s,t}$ is a real Clifford algebra which is realized by the algebra of $2^r \times 2^r$ complex matrices. The complex nature of $R_{s',t'}$ is introduced by the product

$$e_0 e_1 e_2 \cdots e_{s+t}, \tag{2.12}$$

which plays the role of the pure imaginary unit i. [Here, e_0 is determined by an additional dimension to our starting vector space $R^{s,t}$ (see Ref. 9)].

Thus we always have that either $\beta_+(i)$ or $\beta_-(i)$ is equal to $-i$ and $\beta_+(i) = -\beta_-(i)$. Hence

$$\tilde{T}^-T^+(\psi) = \psi_i^* \mathscr{C}^{-1}\beta_{-+}(u_i)\mathscr{C}f, \tag{2.13}$$

with

$$\mathscr{C} = \beta_-(\omega_+)\omega_-; \quad \beta_{-+} = \beta_- \circ \beta_+ \tag{2.14}$$

and

$$\tilde{D}^-D^+(\psi) = \psi_i^* \beta_{-+}(u_i)\mathscr{C}f. \tag{2.15}$$

Similarly, we obtain that

$$\tilde{T}^+T^-(\psi) = \psi_i^* B^{-1}\beta_{+-}(u_i)Bf \tag{2.16}$$

and

$$\tilde{D}^+D^-(\psi) = \psi_i^* \beta_{+-}(u_i)Bf,$$

with

$$B = \beta_+(\omega_-)\omega_+; \quad \beta_{+-} = \beta_+ \circ \beta_-. \tag{2.17}$$

Because we can easily check that $\beta_{+-}(u_i) = \beta_{-+}(u_i)$ as well as $\mathscr{C} = \pm B$, we obtain that

$$\tilde{T}^-T^+ = \tilde{T}^+T^-. \tag{2.18}$$

Now we can introduce the operations $+$, T, and $*$. However, the definitions of these operations will depend on the sign ϵ in the formula

$$\beta_-(i) = \epsilon i, \tag{2.19}$$

i.e., will depend on the dimension $n = s + t$ of our vector space-time $R^{s,t}$. We shall define

$$\psi^+ = T^+(\psi), \quad \psi^T = T^-(\psi), \quad \psi^* = \tilde{T}^+T^-(\psi), \tag{2.20}$$

when $\epsilon = +1$, and

$$\psi^+ = T^-(\psi), \quad \psi^T = T^+(\psi), \quad \psi^* = \tilde{T}^+T^-(\psi), \tag{2.21}$$

when $\epsilon = -1$.

We see that in the general case we can write

$$\psi^+ = T^\epsilon(\psi), \quad \psi^T = T^{-\epsilon}(\psi),$$

and

$$\psi^* = \tilde{T}^\epsilon T^{-\epsilon}(\psi) = \tilde{T}^{-\epsilon}T^\epsilon(\psi), \tag{2.22}$$

where ϵ is given by (2.19). This is a generalization of some operations introduced by Budinich.[10] We can easily check that $\epsilon = +1$ for $n = 2 \bmod 4$, whereas $\epsilon = -1$ for $n = 0 \bmod 4$.

However, any element $\psi \in \Psi(s,t) = R_{s',t'} f \subset R_{s',t'}$ has its matrix representation as well as any element $\tilde{\psi} \in \tilde{\Psi}(s,t) = fR_{s',t'} \subset R_{s',t'}$ has one. Besides, in the matrix algebra $R_{s',t'} \cong \mathbb{C}(2^r)$ we have very well-defined operations of the Hermitian conjugation, transposition, and complex conjugation. Thus it is quite natural that we want to represent the spinors ψ^+, ψ^T, and ψ^* given by (2.20) or (2.21) by the Hermitian conjugated, transposed, and complex conjugated to ψ matrices, respectively. But this requirement puts some conditions on ω_+ additional to (2.7).

To see that we can always fix such elements ω_+ and ω_- let us do this for lower-dimensional space-times $R^{s,t}$ and then use the isomorphism[11]

$$R_{s'+2k,t'+2k} \cong R_{s',t'} \otimes (R_{1,1})^{2k}. \tag{2.23}$$

Let us take, for example, the vector space $R^{2,0}$, i.e., $\epsilon = +1$. Then for $f = \frac{1}{2}(1 + e_1)$; $\rho_1 = f$, $\rho_2 = e_2 f$ we can take $\omega_+ = 1$, $\omega_- = e_{12}$. Now, any spinor ψ^+, ψ^T, and ψ^* is represented by the matrix approximately related to the matrix which represents spinor ψ. Let us take the vector space $R^{3,1}$, i.e., $\epsilon = -1$. Let us fix the Dirac spinor space $\Psi(3,1)$ by taking as an element f

$$f = \tfrac{1}{4}(1 + e_1)(1 + e_{34}).$$

Here, $e_1^2 = e_2^2 = e_3^2 = 1 = -e_4^2$.

Now, for $(\rho_1, \rho_2, \rho_3, \rho_4) = (f, e_2 f, e_3 f, e_{32} f)$ elements $\omega_+ = e_{132}$ and $\omega_- = ie_4$ will satisfy our requirements.

Thus, in the further considerations we shall assume that elements ω_+ and ω_- not only fulfill the condition (2.7) but also allow us to represent elements ψ^+, ψ^T, and ψ^* given by (2.20) and (2.21) by Hermitian conjugated, transposed, or complex conjugated to ψ matrices, respectively.

From (2.18) we have that

$$\psi^* = (\psi^+)^T = (\psi^T)^+. \tag{2.24}$$

Thus we see that the operations $+$, T, $*$, which we use in the quantum theories, can be expressed in the language of the Clifford algebra approach by (2.20) or (2.21), depending on the goemetrical property of the pure imaginary unit $e_0 e_1 \cdots e_{s,t} = i$. Using above operators we also can define the scalar products of spinors. Namely, for any ψ, $\varphi \in \Psi(s,t)$, $\psi = uf$, $\varphi = vf$ we put

$$(\psi,\varphi)_\pm = D^\pm(\psi)\varphi = T^\pm(\psi)\omega_+\varphi = f\omega_\pm \beta_\pm(u)vf. \tag{2.25}$$

Let us define $\bar{\psi}$ as given by

$$\bar{\psi} = \psi^+\omega_\epsilon = T(\psi)\omega_\epsilon = D^\epsilon(\psi). \tag{2.26}$$

Now we can write

$$(\psi,\varphi)_\epsilon = \bar{\psi}\varphi. \tag{2.27}$$

The map

$$\psi \rightsquigarrow \bar{\psi}, \tag{2.28}$$

given by (2.26) corresponds to the operation of bar conjugation well known in quantum theories. For example, for $n = 4$, $(s,t) = (3,1)$, and for the Dirac spinor space and its base introduced above we have

$$\bar{\psi} = T^-(\psi)\omega_- = i\psi^+ t\gamma_4 = \psi^+\gamma_{0123}. \tag{2.29}$$

III.

Let f have the form given by (1.3). Similarly, as previously, we shall denote it by f^1. Let Ψ^1 denote the Dirac spinor space determined by f^1, i.e., $\Psi^1 = R_{s',t'} f^1$. Similarly, let $\{f^\alpha, \alpha = 1\cdots 2^r\}$ denote the family of mutually orthogonal primitive idempotents introduced by (1.4), and let $\{\Psi^\alpha, \alpha = 1\cdots 2^\chi\}$ be the set of Dirac spinor spaces corresponding to $\{f^\alpha\}$.

We introduce the maps

$$\chi_\pm: \quad \Psi^1 \times \Psi^1 \to R_{s',t'} \tag{3.1}$$

in the following way: for any ψ, $\varphi \in \Psi^1$,

$$\chi_\pm(\psi,\varphi) = \psi\omega_\pm \beta_\pm(\varphi). \tag{3.2}$$

First of all, it is easy to see that the image of $\chi_+(\chi_-)$ is given as a left and a right ideal of $R_{s',t'}$ simultaneously. It means that the map $\chi_+(\chi_-)$ has to be onto the Dirac–Clifford algebra $D_{s,t} = R_{s',t'}$. Because $\beta_\delta(i) = +i$ for some δ equal to $+1$ or -1 we obtain from (3.2) that

$$\chi_\delta(\lambda\chi, \mu\varphi) = \lambda\mu\chi_\delta(\psi,\varphi) \tag{3.3}$$

and

$$\chi_{-\delta}(\lambda\chi, \mu\varphi) = \lambda\,\mu^*\chi_{-\delta}(\psi,\varphi), \tag{3.4}$$

for any $\lambda, \mu \in \mathbb{C}$.

Hence we can introduce on the set $\Psi^1 \times \Psi^1$ the following relations of equivalence:

$$R_\delta: \quad (\lambda\psi, \mu\varphi) \sim \lambda\,\mu(\psi,\varphi), \tag{3.5}$$

$$(\psi_1 + \psi_2, \varphi) \sim (\psi_\alpha, \varphi) + (\psi_2, \varphi), \tag{3.5'}$$

$$(\psi, \varphi_1 + \varphi_2) \sim (\psi_1\varphi_1) + (\psi_1\varphi_2), \tag{3.5''}$$

and

$$R_{-\delta}: \quad (\lambda\psi, \mu\varphi) \sim \lambda\mu^*(\psi,\varphi), \tag{3.6}$$

with the rest of the identifications equal to (3.5′) and (3.5″). Now it is obvious that the following lemma is true.

Lemma 1: The maps $\chi_\pm$ are bijections between the sets

$$\Psi^1 \times \Psi^1 / R_\pm \quad \text{and} \quad R_{s',t'}. \tag{3.7}$$

■

Let us notice that

$$\Psi^1 \times \Psi^1 / R_\delta = \Psi^1 \otimes \Psi^1. \tag{3.8}$$

Now let $\{\rho_i\}$, $i = 1\cdots 2^r$ be the basis of Ψ^1 of the form given by (2.1), i.e., $\rho_i = u_i f^1$. It is easy to see that for any $\alpha \in (1, \ldots, 2^\chi)$ and for any $i \in (1, \ldots, 2^r)$,

$$f^\alpha u_i = u_i f^\beta, \tag{3.9}$$

with β depending on (i,α), i.e., $\beta = \beta(i,\alpha)$.

Lemma 2: Let us fix some $\lambda \in (1\cdots 2^r)$. Then the set of elements $\chi_\pm(\rho_i, \rho_\lambda)$, $i = 1\cdots 2^r$ forms the basis of some appropriate Dirac spinor space $\Psi^\beta = R_{s',t'} f^\beta$.

Proof:

$$\begin{aligned}\chi_\pm(\rho_i, \rho_\lambda) &= u_i f^1 \omega_\pm \beta_\pm(f^1)\beta_\pm(u_\lambda)\\ &= \delta u_i \omega_\pm \beta_\pm(f^1)\beta_\pm(u_\lambda)\\ &= \tilde{\delta} u_i \omega_\pm u_\lambda f^\beta,\end{aligned}$$

where $\delta, \tilde{\delta}^+ = \pm 1$, and β depends on the sign of the map χ and on λ $[\beta_\pm(f^1) = f^\alpha$, with $\alpha = \alpha(\pm)]$.

It is easy to see that for any Dirac–Clifford algebra $D_{s,t}$ we have

$$2^{2r+1-1-\chi} = 2^r, \tag{3.10}$$

i.e., $r = \chi$.

Now, because $\chi_\pm$ are bijections and because $\Psi^{\beta_1} \cap \Psi^{\beta_2} = 0$ for $\beta_1 \neq \beta_2$, we obtain immediately that

$$\chi_\pm(\Psi^1 \times \{\rho_{\lambda_1}\}) \cap (\Psi^1 \times \{\rho_{\lambda_2}\}) = 0, \quad \text{for } \lambda_1 \neq \lambda_2.$$

Again by Lemma 1 we obtain that $\chi_\pm(\rho_i, \rho_\lambda) i = 1, \ldots, 2^r$ has to form a base of Ψ^β. ■

The above consideration can be very easily generalized to the following statement.

Theorem 1: For any ψ, $\varphi \in \Psi^1$ the image $\chi_+(\psi,\varphi)$ [as well as $\chi_-(\psi,\varphi)$] belongs to some minimal left ideal of the Clifford algebra $R_{s',t'}$. And conversely let $W \in R_{s',t'}$ be an element of some minimal left ideal of $R_{s',t'}$. Then there exist spinors $\psi_\pm$, $\varphi_\pm \in \Psi^1$ such that $W = \chi_\pm(\psi_\pm, \varphi_\pm)$.

Proof: the first part is simple. For the second, let W belong to some minimal left ideal $R_{s',t'} h$, where h is a primitive idempotent, i.e.,

$$W = ph, \quad \text{with} \quad p \in R_{s',t'}.$$

But

$$h = h^\alpha f^\alpha,$$

and for any f^α we can find invertible elements $v^\alpha_\pm$ such that $v^\alpha_\pm f^\alpha = \beta_\pm(f^1)v^\alpha_\pm$.

Taking into account the fact that for any $\psi = uf^1$, $\varphi = vf^1 \in \Psi^1$,

$$\chi_\pm(\psi,\varphi) = \delta u\omega_\pm \beta_\pm(f^1)\beta_\pm(v), \quad \text{with } \delta = \pm 1,$$

we obtain immediately the second assertion. ■

The above theorem provides us with a simple criterion which tells us which elements of a given Clifford algebra $R_{s',t'}$ belong to some minimal left ideals.

IV.

By construction, we have that

$$R^{s,t} \subset R^{s,t} \oplus Re_0 = R^{s',t'}$$
$$D_{s,t} \equiv R_{s',t'}. \tag{4.1}$$

Thus, the vector space $R^{s,t}$ can be obtained as the image of the maps $\chi_\pm$ of some elements belonging to $\Psi^1 \times \Psi^1$. Let $s \in \mathrm{Spin}(s,t) \subset \mathrm{Spin}(s',t')$. Of course s belongs to the automorphism group of the spinor space Ψ^1. We can easily check that

$$\chi_\pm(s\psi, s\varphi) = s\chi_\pm(\psi,\varphi)s^{-1}, \quad \forall \psi,\varphi \in \Psi^1. \tag{4.2}$$

Thus, to any transformation $s \in \mathrm{Spin}(s,t)$ of the Dirac spinor space Ψ^1 we can associate a unique transformation $g = \tau(s) \in \mathrm{SO}(s,t)$ of our starting vector space $R^{s,t}$. [Here τ is the covering map τ: $\mathrm{Spin}(s,t) \rightarrowtail \mathrm{SO}(s,t)$.]

Let $(e_1,\ldots,e_{s+t})$ be an orthogonal base of $R^{s,t}$. Let g be some orthogonal transformation of $R^{s,t}$. We can understand this transformation as the result of the transformation $\pm s \in \mathrm{Spin}(s,t)$ of the spinor space Ψ^1, such that $\tau(\pm s) = g$. However, there also exists another, completely unequivalent interpretation of the transformation g of the vector space $R^{s,t}$. It is related with the decomposition of $R_{s',t'}$ onto its minimal left ideals according to (1.5), i.e.,

$$R_{s',t'} = \bigoplus_{\alpha/1}^{N} \Psi^\alpha. \tag{4.3}$$

This approach allows us to treat the vector space $R^{s,t}$ as the vector space spanned by elements $\{e_\mu\}$ which are compositions of spinors belonging to different concrete spinor spaces Ψ^α. More precisely, let $\gamma_\mu : \mu = 1,\ldots,s+t$ be the matrix representation of e_u in the Dirac–Clifford algebra $D_{s,t}$. Then we can write that

$$e_\mu = \psi_\mu^1 + \psi_\mu^2 + \cdots + \psi_\mu^N, \tag{4.4}$$

or in the matrix representation

$$\gamma_\mu = \psi_\mu^1 + \psi_\mu^2 + \cdots + \psi_\mu^N. \tag{4.4'}$$

Here $\psi_\mu^\alpha \in \Psi^\alpha$ and any spinor ψ_μ^α has to be unequal to zero.

Let $\{\rho_i^\alpha\}$, $i = 1,\ldots,N$, be the base of Ψ^α. Then

$$\psi_\mu^\alpha = \sum_{i/1}^{N} \psi_{\mu i}^\alpha \rho_i^\alpha, \tag{4.5}$$

and the coefficient matrix $A_i^\alpha = ||\psi_{\mu i}^\alpha||$ is exactly equal to our nonsingular matrix γ_μ. Let

$$g: e_\mu \to e'_\mu, \tag{4.6}$$

i.e.,

$$g: \gamma_\mu \to \gamma'_\mu, \tag{4.6'}$$

and let $s \in \mathrm{Spin}(s,t)$ be such that $\tau(s) = g$. Let $\chi_\delta(\Psi^1, \{\rho_\lambda^1\}) = \Psi^\beta$ with $\beta = \beta(\delta,\lambda)$.

In a general case we have that

$$\chi_\delta(s\Psi^1, \{s\rho_\lambda^1\}) \neq \Psi^\beta. \tag{4.7}$$

Thus we can say that the transformation g breaks the decomposition (4.3) of $R_{s',t'}$. We can see this also in another way. Namely, any $s \in \mathrm{Spin}(s,t)$ determines the following maps:

(a) $s: R_{s',t'} \to R_{s',t'}$, (4.8)

given by

$$u \xrightarrow{s} su, \quad \forall u \in R_{s',t'}, \tag{4.8'}$$

and

(b) $s: R_{s',t'} \to R_{s',t'}$, (4.9)

given by

$$u \to sus^{-1} : g \circ u, \quad \forall u \in R_{s',t'}. \tag{4.9'}$$

The former map is related with the action of s on the spinor spaces Ψ^α and of course preserves the decomposition (4.3), whereas the latter is related with the bijections

$$\chi_\delta : \Psi^1 \times \Psi^1 / R_\delta \leftrightarrow R_{s',t'},$$

and obviously does not preserve the decomposition (4.3).

However, if we fix some matrix realization of $R_{s',t'}$ then it is equivalent to the fact that we fix some concrete primitive idempotent $f = f^1$ as well as a related decomposition (4.3). By (4.3), (4.4'), (4.6), and (4.6') we see that

$$\gamma'_\mu = \psi'^1_\mu + \psi'^2_{r\mu} + \cdots + \psi'^n_\mu$$

and

$$\psi'^\alpha_\mu = g_{\nu\mu}\psi_\nu^\alpha, \quad \forall_\mu = 1,\ldots,s+t, \quad \forall \alpha = 1,\ldots,N, \tag{4.10}$$

where $g_{\nu\mu}$ is the matrix element of the transformation g (4.6), i.e., $e'_\mu = g_{\nu\mu} e_\nu$.

Thus we can look at the map g as preserving the decomposition (4.3), but transforming spinors which build the basis of $R^{s,t}$ according to (4.4).

The relation (4.10) tells us that we cannot relate with g any transformation of the spinor space Ψ^α. The formula (4.10) only means that the subspace of Ψ^α spanned by $\{\psi_\mu^\alpha\}_{,\mu=1,\ldots,s+t}$ is transformed onto itself. Let us notice that the set of spinors $\{\psi_\mu^\alpha\}$ can be linearly dependent. Then, according to (4.10) g transforms the same element $\psi_\mu^\alpha = \psi_\gamma^\alpha$, $\mu \neq \gamma'$, of Ψ^α in different ways. We can summarize this in the following diagram:

$$\begin{tikzcd}
K_\delta(\Psi^1 \times \Psi^1/R_\delta) \arrow[r, "s\times s"] \arrow[d] & K_\delta(\Psi^1 \times \Psi^1/R_\delta) \arrow[d] \\
R^{s,t} \arrow[r, "g=\tau(s)"] \arrow[d, "\zeta"'] & R^{s,t} \arrow[d, "\zeta"] \\
\left\{\sum_\alpha \psi_\mu^\alpha\right\} \arrow[r, squiggly, "\kappa"] & \left\{\sum_\alpha \psi_\mu^\alpha\right\}.
\end{tikzcd}$$

Here $K_\delta(\Psi^1 \times \Psi^1/R_\delta)$ denotes the coimage of $R^{s,t}$ of the bijection χ_δ, the map ζ is uniquely determined by (4.4) and (4.4'), and κ by (4.10). By (4.2) the upper diagram commutes.

[1] I. Porteous, *Topological Geometry* (Van Nostrand–Reinhold, London, 1969).
[2] P. Lounesto, "Spinors and Brauer–Wall groups," Rep. HTKK-MAT-A124 (1978).
[3] See Ref. 2.
[4] H is a Radon–Hurwitz number.

[5]I. Benn and R. Tucker, Commun. Math. Phys. **89**, 341 (1983); P. Becher and H. Joos, Z. Phys. C. **15**, 343 (1982); T. Banks, Y. Dothan, and D. Horn, "Geometric Fermions," TAUP 1037-82, Tel Aviv University, 1982.
[6]K. Bugajska, "Spinors and space-times," (submitted for publication); Geometrical properties of the algebraic spinors for $R^{3,1}$" (submitted for publication).
[7]See Refs. 2 and 6.
[8]P. Budinich, "On pure conformal spinors and internal symmetry," Report ISAS 40/83/E.P. Trieste, 1983.
[9]See Ref. 6.
[10]See Ref. 8.
[11]D. Husemoller, *Fibre Bundles* (McGraw-Hill, New York, 1966).

Infinitesimal symmetry transformations. II. Some one-dimensional nonlinear systems

M. Aguirre
Instituto de Física, Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile

J. Krause
Facultad de Física, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile

(Received 18 May 1984, accepted for publication 12 October 1984)

The converse problem of similarity analysis is discussed for the infinitesimal symmetry transformations of ordinary second-order differential equations which are nonlinear in $\dot{x}$ (and may be linear or nonlinear in x). A natural classification of the problem arises, according to the highest order N of nonlinearity in $\dot{x}$. The completely general maximal Lie algebra is obtained for the case $N \leqslant 3$. In the case $N \geqslant 4$ one has, besides the system of differential equations for the infinitesimal generators, an extra set of anholonomic constraints, which operates as a symmetry-breaking mechanism producing a strong reduction in the number of surviving parameters. Miscellaneous examples are given, which illustrate some features of similarity analysis of nonlinear systems. The infinitesimal point transformation symmetries of the Van der Pol oscillator are also briefly discussed.

I. INTRODUCTION

In a previous paper[1] we have obtained the completely general Lie algebra associated with the point symmetry transformations of a given inhomogeneous ordinary *linear* differential equation of the second order. Here we present a generalization of our previous work. Indeed, using the same mathematical approach introduced in paper I, in this article we tackle the converse problem of similarity analysis[2] for the infinitesimal symmetry groups of some *nonlinear* ordinary second-order differential equations which are relevant in mechanics. As we shall see, it turns out that our previous results (for linear systems) correspond to special cases of the similarity Lie algebras we present in this work.

Since the motivation underlying this article is the same already formulated in paper I, we would like to refer the reader to the Introduction of that paper. It should be mentioned, however, that the issue discussed in this paper bears a particular interest from the standpoint of contemporary classical mechanics. Recently, a large amount of research has been related to nonlinear systems having a single degree of freedom,[3] or multidegrees of freedom.[4] Although in this paper we do not touch on the physical aspects of nonlinear systems (and, moreover, we do only enface a very restricted area of the enormous field of nonlinear analysis), we wish to remark that similarity methods may be called to play an interesting role in nonlinear mechanics. For instance, in a recent article[5] the complete nonlinear problem of water waves is investigated, according to the perfect-fluid model, using general methods of infinitesimal transformation theory[6] for finding the symmetry groups of free-boundary problems.

It appears, therefore, to be an attractive endeavor to obtain the Lie algebras associated with the point transformation invariance of a nonlinear system having a single degree of freedom.[7]

The organization of this paper is as follows. In Sec. II we present a general discussion of the infinitesimal symmetries of an ordinary differential equation of the second order, which is nonlinear in $\dot{x}$. Thus, a natural classification of the problem arises, according as the highest order N of nonlinearity in $\dot{x}$ is $N \leqslant 3$ or $N \geqslant 4$. We examine these two cases in Secs. III and IV, respectively. Then we briefly present some miscellaneous examples (Sec. V) of the previous formalism for the converse problem of infinitesimal similarity analysis is nonlinear mechanics. Given its importance, in Sec. V we also discuss the point transformation symmetries of the Van der Pol oscillator, as an example of a differential equation which is linear in $\dot{x}$ and nonlinear in x.

II. INFINITESIMAL SYMMETRIES OF A SECOND-ORDER NONLINEAR DIFFERENTIAL EQUATION: GENERAL DISCUSSION

Let us briefly analyze the symmetries of an ordinary second-order nonlinear differential equation

$$\ddot{x} = f_N(t,x,\dot{x}), \tag{2.1}$$

in the neighborhood of the identity. We recall that if the infinitesimal point transformation

$$t' = t + \epsilon\eta(t,x), \tag{2.2a}$$

$$x' = x + \epsilon\theta(t,x) \tag{2.2b}$$

(with $0 \leqslant \epsilon < 1$) corresponds to a symmetry of Eq. (2.1), then the generating functions $\eta(t,x)$ and $\theta(t,x)$ satisfy the well-known identity

$$\begin{aligned}\theta_{tt} &+ (2\theta_{xt} - \eta_{tt})\dot{x} + (\theta_{xx} - 2\eta_{xt})\dot{x}^2 - \eta_{xx}\dot{x}^3 \\ &+ ((\theta_x - 2\eta_t) - 3\eta_x\dot{x})f_N(t,x,\dot{x}) \\ &- \eta f_{Nt}(t,x,\dot{x}) - \theta f_{Nx}(t,x,\dot{x}) \\ &+ (\theta_t + (\theta_x - \eta_t)\dot{x} - \eta_x\dot{x}^2)f_{N\dot{x}}(t,x,\dot{x}) \equiv 0\end{aligned} \tag{2.3}$$

[cf. Eq. (I.2.5)]. Thus, it is evident that the power series of $f_N(t,x,\dot{x})$ in terms of $\dot{x}$ shall give us a set of differential equations (and possible subsidiary anholonomic constraints) for the generators. Furthermore, it is also clear that a Laurent series in $\dot{x}$ for $f_N(t,x\dot{x})$ would entail a dynamical extravagance. Hence, for the purposes of nonlinear mechanics, it is

 0022-2488/85/040593-08$02.50

general enough to assume a positive power series in $\dot{x}$; i.e.,

$$\ddot{x} = \sum_{k=0}^{N} \alpha_k(t,x)\dot{x}^k, \quad N = 0,1,2,..., \tag{2.4}$$

which covers, once and for all, the (linear and nonlinear) cases of mechanical interest. Usually, in the applications, $f_N(t,x,\dot{x})$ is a polynomial in $\dot{x}$; not so in relativistic mechanics, however.

Now, if we substitute from Eq. (2.4) into the identity (2.3), after separating the null coefficients of the different powers of $\dot{x}$, we get:

$$\eta_{xx} = -\alpha_2\eta_x + \alpha_3(\eta_t - 2\theta_x) - \alpha_{3t}\eta - \alpha_{3x}\theta - 4\alpha_4\theta_t, \tag{2.5}$$

$$\theta_{xx} - 2\eta_{xt} - 2\alpha_1\eta_x = \alpha_2\theta_x + \alpha_{2t}\eta + \alpha_{2x}\theta + 3\alpha_3\theta_t, \tag{2.6}$$

$$2\theta_{xt} - \eta_{tt} - 3\alpha_0\eta_x - \alpha_1\eta_t - \alpha_{1t}\eta - \alpha_{1x}\theta = 2\alpha_2\theta_t, \tag{2.7}$$

$$\theta_{tt} - \alpha_0(2\eta_t - \theta_x) - \alpha_{0t}\eta - \alpha_{0x}\theta - \alpha_1\theta_t = 0; \tag{2.8}$$

and also

$$(k+1)\alpha_{k+1}\theta_t = (k-4)\alpha_{k-1}\eta_x - \alpha_{kt}\eta - \alpha_{kx}\theta + \alpha_k((k-2)\eta_t - (k-1)\theta_x), \tag{2.9}$$

which holds for $k = 4,5,...,N-1$;and moreover, if N is finite ($N \geqslant 4$):

$$((N-2)\eta_t - (N-1)\theta_x)\alpha_N - \eta\alpha_{Nt} - \theta\alpha_{Nx} + (N-4)\eta_x\alpha_{N-1} = 0, \tag{2.10}$$

$$(N-3)\alpha_N\eta_x = 0. \tag{2.11}$$

This brings into the fore a natural classification of the problem in three cases: (A) $0 \leqslant N \leqslant 3$, (B) $4 \leqslant N < \infty$, and (C) $N = \infty$.

In Eqs. (2.5)–(2.8) one has a set of four linear homogeneous differential equations of the second order of the two unknown generating functions (η,θ) in two independent variables (t,x). [These equations are the direct generalization of Eqs. (I.2.7)–(I.2.10) to the nonlinear case in $\dot{x}$.] In Eqs. (2.9), (2.10), and (2.11) one has a set of linear homogeneous and anholonomic constraints, of the first order, for (η,θ), which must be satisfied in case B. Clearly, case A has no constraints; while, in case C, Eq. (2.9) holds for $k = 4,5,...$, and Eqs. (2.10) and (2.11) become meaningless.

Therefore, in handling the infinitesimal symmetries of a system which is nonlinear in $\dot{x}$ (and which may be linear or nonlinear in x) we may operate on a *linear basis* of independent solutions of Eqs. (2.5)–(2.8), according to the same approach used in paper I.[8] Indeed, given $\alpha_k(t,x)$, for $k = 0,1,2,3,4$, in Eq. (2.4), Eqs. (2.5)–(2.8) can be solved for $\eta(t,x)$ and $\theta(t,x)$ (at least, in principle). The general solution of this linear homogeneous system can be formally written as a superposition of linearly independent solutions and contains *at most* eight constants of integration[9] [as a glance at Eqs. (2.5)–(2.8) neatly shows]. Therefore, according to the principle of superposition, these integration constants appear as the arbitrary numerical coefficients of the linear combination adopted as general solution for $\eta(t,x)$ and $\theta(t,x)$ [cf., Eqs. (3.6) and (3.7), below]. Clearly, this fact motivates that the set of required initial data for solving Eqs. (2.5)–(2.8) is (at most) eight-dimensional [cf. Eqs. (3.12), *infra*].

In case A these constants of integration can be considered as essential parameters of the Lie group which describes the invariance of the differential equation. In cases B and C, however, the subsidiary constraints [Eqs. (2.9)–(2.11)] may produce a strong reduction in the number of allowable nonzeroth constants of integration which can be used as essential parameters of the symmetry group. Thus, if the polynomial series in $\dot{x}$ is "long enough" one arrives almost certainly at the trivial solution,

$$\eta(t,x) = \theta(t,x) = 0, \tag{2.12}$$

unless the coefficients $\alpha_k(t,x)$ meet very "fortunate" conditions. [Of course, this issue is just a matter of principle. In practice, when $N \geqslant 4$, one takes advantage of the simplification introduced by Eqs. (2.9)–(2.11), and one solves these equations before tackling Eqs. (2.5)–(2.8).]

Hence, the conclusion follows that differential equations of order higher than one only exceptionally admit continuous groups of point symmetry transformations. (This is in contrast with the well-known general result concerning first-order differential equations.[10]) In particular, if a second-order differential equation admits a continuous symmetry group generated by (2.2), one concludes that these infinitesimal symmetry transformations correspond to a Lie group having no more than eight essential parameters.[9]

It is not out of place to illustrate the facts we present with an example. Hence, in Sec. V we present a *compared* similarity analysis of two nonlinear differential equations, in order to exhibit some peculiar relations between symmetry and nonlinearity.

We wish to end this section with a brief remark concerning the "conditions" under which the method presently discussed in this paper (as well as in paper I) is meaningful for obtaining the Lie algebra associated with an ordinary second-order differential equation. We first plainly observe that $\eta = \theta = 0$ is always a (trivial) solution of Eq. (2.3) [or, for that matter, of Eqs. (2.5)–(2.8) and Eqs. (2.9)–(2.11)]. Furthermore, there are ordinary differential equations of the second order such that they admit the trivial solution (2.12) as the *only* solution of Eq. (2.3). [To recall an extreme (and well known) example of this fact, the equation $\ddot{x} = x^2 + t^2$ gives $\eta = \theta = 0$ as the only solution of Eq. (2.3).] Clearly, in such cases, the only point symmetry admitted by the differential equation is the identity. Simple as it is, this fact is telling us that our approach to the Lie algebra of a differential equation of the second order is completely general, since it stems from the converse problem of infinitesimal similarity analysis. In other words, in order to use the proposed method it is not necessary to assume (*a priori*) that the differential equation admits certain one-parameter symmetries that belong to a finite dimensional Lie algebra; whether it does or not, is an outcoming (i.e., *a posteriori*) result of the similarity analysis itself.

III. LIE ALGEBRAS OF $\ddot{x} = f_N(t,x,\dot{x})$: CASE $N \leqslant 3$

According to our previous remarks, we consider first the infinitesimal symmetry problem set by a differential equation of the following form:

$$\ddot{x} = \alpha_0(t,x) + \alpha_1(t,x)\dot{x} + \alpha_2(t,x)\dot{x}^2 + \alpha_3(t,x)\dot{x}^3, \tag{3.1}$$

where some of the α's may be constant and eventually zero. Therefore, Eqs. (2.5)–(2.8) become in the following system:

$$\eta_{xx} + \alpha_2 \eta_x + \alpha_3(2\theta_x - \eta_t) + \alpha_{3t}\eta + \alpha_{3x}\theta = 0, \tag{3.2}$$

$$\theta_{xx} - 2\eta_{xt} - 2\alpha_1 \eta_x - \alpha_2 \theta_x - \alpha_{2t}\eta - \alpha_{2x}\theta - 3\alpha_3\theta_t = 0, \tag{3.3}$$

$$2\theta_{xt} - \eta_{tt} - 3\alpha_0\eta_x - \alpha_1\eta_t - \alpha_{1t}\eta - \alpha_{1x}\theta - 2\alpha_2\theta_t = 0, \tag{3.4}$$

$$\theta_{tt} - \alpha_0(2\eta_t - \theta_x) - \alpha_{0t}\eta - \alpha_{0x}\theta - \alpha_1\theta_t = 0, \tag{3.5}$$

with no constraints at all. As we have already remarked in Sec. II, the general solution of this homogeneous linear system can be formally written as a superposition of linearly independent basis solutions $\eta_a(t,x)$ and $\theta_a(t,x)$, $a = 1,2,...,r<8$; thus, using Einstein's "dummy index" convention, we write

$$\eta(t,x) = q^a \eta_a(t,x), \tag{3.6}$$

$$\theta(t,x) = q^a \theta_a(t,x). \tag{3.7}$$

Henceforth we assume $r = 8$ and follow the same method introduced in paper I in order to obtain the structure constants, without recourse to the detailed knowledge of the basis functions $\eta_a(t,x)$ and $\theta_a(t,x)$. Since the calculations are much more involved in the present case, we shall state them in a sketchy manner, for the sake of brevity.

It is well known that, because of the Lie–Cartan integrability conditions, the formal infinitesimal operators

$$X_a(t,x) = \eta_a(t,x)\partial_t + \theta_a(t,x)\partial_x, \tag{3.8}$$

subject to the Killing equations (3.2)–(3.5), satisfy a Lie algebra

$$[X_a(t,x), X_b(t,x)] = f^c_{ab} X_c(t,x), \tag{3.9}$$

where f^c_{ab} denotes the structure constants. Thus, we observe that the full symmetry group of the differential equation (3.1) formally admits the following identities [cf. Eqs. (I.2.20) and (I.2.21)]:

$$f^c_{ab}\eta_c = [\eta_a, \eta_{bt}] + [\theta_a, \eta_{bx}], \tag{3.10}$$

$$f^c_{ab}\theta_c = [\eta_a, \theta_{bt}] + [\theta_a, \theta_{bx}]. \tag{3.11}$$

Of course, these commutation relations are particularly critical in the sense that they hold only for those ranges of t and x on which the given functions $\alpha_k(t,x)$, k $= 0,1,2,3$, are regular.[11] Hence, let us assume "initial data" at some regular point $(t,x) = (t_0,x_0)$ (say) to evaluate the structure constants. So, in order to represent the algebra, we introduce the following parametrization:

$$\begin{aligned} &q^1 = \eta(t_0,x_0), && q^2 = \theta(t_0,x_0),\\ &q^3 = \eta_t(t_0,x_0), && q^4 = \theta_x(t_0,x_0),\\ &q^5 = \eta_x(t_0,x_0), && q^6 = \theta_t(t_0,x_0),\\ &q^7 = \tfrac{1}{2}\eta_{tt}(t_0,x_0), && q^8 = \tfrac{1}{2}\theta_{xx}(t_0,x_0); \end{aligned} \tag{3.12}$$

and thus, according to Eqs. (3.6) and (3.7), we adopt the following set of "initial data":

$$\begin{aligned} &\eta_a(t_0,x_0) = \delta_{a1}, && \theta_a(t_0,x_0) = \delta_{a2},\\ &\eta_{at}(t_0,x_0) = \delta_{a3}, && \theta_{ax}(t_0,x_0) = \delta_{a4},\\ &\eta_{ax}(t_0,x_0) = \delta_{a5}, && \theta_{at}(t_0,x_0) = \delta_{a6},\\ &\eta_{att}(t_0,x_0) = 2\delta_{a7}, && \theta_{axx}(t_0,x_0) = 2\delta_{a8}. \end{aligned} \tag{3.13}$$

As we shall see presently, these nonsingular "initial data" uniquely determine the general form of the structure constants (up to equivalence). However, as we have already remarked in paper I, the main point in this matter is that, if we use a different set of "initial conditions," defined at the same or at a different regular point, we obtain a different parametrization of the group, which corresponds to a mere change of the basis of the algebra.

In this manner, Eqs. (3.10) and (3.11) give us immediately the following subsets of structure constants:

$$f^1_{ab} = [\delta_{a1}, \delta_{b3}] + [\delta_{a2}, \delta_{b5}], \tag{3.14}$$

$$f^2_{ab} = [\delta_{a1}, \delta_{b6}] + [\delta_{a2}, \delta_{b4}], \tag{3.15}$$

of which the meaning is clear. Of course, in order to obtain the structure constants $f^3_{ab}, f^4_{ab}, f^5_{ab}$, and f^6_{ab}, we have to take the derivatives $f^c_{ab}\eta_{ct}, f^c_{ab}\theta_{cx}, f^c_{ab}\eta_{cx}$, and $f^c_{ab}\theta_{ct}$, respectively, in Eqs. (3.10) and (3.11), use the fact that η_a and θ_a $(a = 1,2,...,8)$ satisfy Eqs. (3.2)–(3.5), and evaluate these derivatives at $(t,x) = (t_0,x_0)$. Thus we get

$$f^3_{ab} = [\delta_{a1}, \tfrac{1}{2}\hat{\alpha}_{2t}\delta_{b2} + 2\delta_{b7}] - [\delta_{a2}, \tfrac{1}{2}\hat{\alpha}_2\delta_{b4} + \hat{\alpha}_1\delta_{b5} + \tfrac{3}{2}\hat{\alpha}_3\delta_{b6} - \delta_{b8}] - [\delta_{a5}, \delta_{b6}], \tag{3.16}$$

$$f^4_{ab} = [\delta_{a1}, \tfrac{1}{2}\hat{\alpha}_{1x}\delta_{b2} + \tfrac{1}{2}\hat{\alpha}_1\delta_{b3} + \tfrac{3}{2}\hat{\alpha}_0\delta_{b5} + \hat{\alpha}_2\delta_{b6} + \delta_{b7}] + 2[\delta_{a2}, \delta_{b8}] + [\delta_{a5}, \delta_{b6}], \tag{3.17}$$

$$\begin{aligned} f^5_{ab} = {}& [\delta_{a1}, (\hat{\alpha}_{3t} - \tfrac{1}{2}\hat{\alpha}_{2x})\delta_{b2} - \tfrac{1}{2}\hat{\alpha}_2\delta_{b4} - \hat{\alpha}_1\delta_{b5} - \tfrac{3}{2}\hat{\alpha}_3\delta_{b6} + \delta_{b8}]\\ &+ [\delta_{a2}, \hat{\alpha}_3(\delta_{b3} - 2\delta_{b4}) - \hat{\alpha}_2\delta_{b5}] - [\delta_{a3}, \delta_{b5}] + [\delta_{a4}, \delta_{b5}], \end{aligned} \tag{3.18}$$

$$\begin{aligned} f^6_{ab} = {}& [\delta_{a1}, (\hat{\alpha}_{0x} - \tfrac{1}{2}\hat{\alpha}_{1t})\delta_{b2} + 2\hat{\alpha}_0\delta_{b3} - \hat{\alpha}_0\delta_{b4} + \hat{\alpha}_1\delta_{b6}]\\ &+ [\delta_{a2}, \tfrac{1}{2}\hat{\alpha}_1\delta_{b3} + \tfrac{3}{2}\hat{\alpha}_0\delta_{b5} + \hat{\alpha}_2\delta_{b6} + \delta_{b7}] + [\delta_{a3}, \delta_{b6}] - [\delta_{a4}, \delta_{b6}], \end{aligned} \tag{3.19}$$

where $\hat{\alpha}$ denotes $\alpha(t_0,x_0)$, quite generally. By the same token, to calculate f^7_{ab} and f^8_{ab} we need the derivatives $f^c_{ab}\eta_{ctt}$ and $f^c_{ab}\theta_{cxx}$, respectively, in Eqs. (3.10) and (3.11); then we perform various substitutions in the right-hand members thereof, and evaluate them at (t_0,x_0). So we obtain, finally,

$$\begin{aligned} f^7_{ab} = {}& [\delta_{a1}, -\{\tfrac{1}{2}\hat{\alpha}_0(\tfrac{1}{2}\hat{\alpha}_{2x} - \hat{\alpha}_{3t}) - \tfrac{1}{2}\hat{\alpha}_1(\hat{\alpha}_{1x} - \tfrac{1}{2}\hat{\alpha}_{2t}) + \hat{\alpha}_2\hat{\alpha}_{0x} - \hat{\alpha}_3\hat{\alpha}_{0t} - \hat{\alpha}_{0xx} + \tfrac{1}{3}(2\hat{\alpha}_{1tx} - \hat{\alpha}_{2tt})\}\delta_{b2}\\ &- (2\hat{\alpha}_0\hat{\alpha}_2 - \tfrac{1}{2}\hat{\alpha}_1^2 - 2\hat{\alpha}_{0x} + \hat{\alpha}_{1t})\delta_{b3} + \tfrac{3}{4}\hat{\alpha}_0\hat{\alpha}_2\delta_{b4} + \hat{\alpha}_0\hat{\alpha}_1\delta_{b5} + \tfrac{1}{2}(\hat{\alpha}_{1x} - \hat{\alpha}_{2t} - \tfrac{3}{2}\hat{\alpha}_0\hat{\alpha}_3)\delta_{b6} - \tfrac{3}{2}\hat{\alpha}_0\delta_{b8}]\\ &- [\delta_{a2}, \{\tfrac{3}{2}\hat{\alpha}_0\hat{\alpha}_3 - \tfrac{1}{3}(\tfrac{1}{2}\hat{\alpha}_{1x} - \hat{\alpha}_{2t})\}\delta_{b3} - \tfrac{1}{4}\{\hat{\alpha}_1\hat{\alpha}_2 + \tfrac{2}{3}(\hat{\alpha}_{1x} - \hat{\alpha}_{2t})\}\delta_{b4}\\ &+ \tfrac{1}{2}(\hat{\alpha}_0\hat{\alpha}_2 - \hat{\alpha}_1^2 - 2\hat{\alpha}_{0x} + \hat{\alpha}_{1t})\delta_{b5} + \{\tfrac{1}{4}\hat{\alpha}_1\hat{\alpha}_3 - \hat{\alpha}_{2x} + 2\hat{\alpha}_{3t}\}\delta_{b6} + \tfrac{1}{2}\hat{\alpha}_1\delta_{b8}] + [\delta_{a3}, \hat{\alpha}_0\delta_{b5} + \delta_{b7}] \end{aligned}$$

$$+ [\delta_{a4}, -\tfrac{1}{2}\hat{\alpha}_0\delta_{b5} + \tfrac{1}{2}\hat{\alpha}_2\delta_{b6}] + \tfrac{1}{2}\hat{\alpha}_1[\delta_{a5},\delta_{b6}] + [\delta_{a6},\delta_{b8}], \tag{3.20}$$

$$\begin{aligned} f^8_{ab} = {} & [\delta_{a1}, \{-\hat{\alpha}_0\hat{\alpha}_{3x} + \hat{\alpha}_1\hat{\alpha}_{3t} + \tfrac{1}{2}\hat{\alpha}_2(\tfrac{1}{2}\hat{\alpha}_{1x} - \hat{\alpha}_{2t}) - \tfrac{1}{2}\hat{\alpha}_3(\hat{\alpha}_{0x} - \tfrac{1}{2}\hat{\alpha}_{1t}) + \tfrac{1}{3}(\hat{\alpha}_{1xx} - 2\hat{\alpha}_{2tx}) + \hat{\alpha}_{3tt}\}\delta_{b2} \\ & + \{\tfrac{1}{4}\hat{\alpha}_1\hat{\alpha}_2 + \tfrac{1}{3}(\hat{\alpha}_{1x} - \tfrac{1}{2}\hat{\alpha}_{2t})\}\delta_{b3} + \{-\tfrac{3}{2}\hat{\alpha}_0\hat{\alpha}_3 + \tfrac{1}{3}(\hat{\alpha}_{1x} - \tfrac{1}{2}\hat{\alpha}_{2t})\delta_{b4} + \{-\tfrac{1}{4}\hat{\alpha}_0\hat{\alpha}_2 + \hat{\alpha}_{0x} - \tfrac{1}{2}\hat{\alpha}_{1t}\}\delta_{b5} \\ & + \tfrac{1}{2}\{-\hat{\alpha}_1\hat{\alpha}_3 + \hat{\alpha}_2^2 + \hat{\alpha}_{2x} - 2\hat{\alpha}_{3t}\}\delta_{b6} + \tfrac{1}{2}\hat{\alpha}_2\delta_{b7}] + [\delta_{a2}, \tfrac{3}{4}\hat{\alpha}_1\hat{\alpha}_3\delta_{b3} + \{-2\hat{\alpha}_1\hat{\alpha}_3 + \tfrac{1}{2}\hat{\alpha}_2^2 + \hat{\alpha}_{2x} - 2\hat{\alpha}_{3t}\}\delta_{b4} \\ & + \tfrac{1}{2}\{-\tfrac{3}{2}\hat{\alpha}_0\hat{\alpha}_3 + \hat{\alpha}_{1x} - \hat{\alpha}_{2t}\}\delta_{b5} + \hat{\alpha}_2\hat{\alpha}_3\delta_{b6} + \tfrac{3}{2}\hat{\alpha}_3\delta_{b7}] + [\delta_{a3}, -\tfrac{1}{2}\hat{\alpha}_1\delta_{b5} + \tfrac{1}{2}\hat{\alpha}_3\delta_{b6}] \\ & + [\delta_{a4}, -\hat{\alpha}_3\delta_{b6} + \delta_{b8}] + [\delta_{a5}, \tfrac{1}{2}\hat{\alpha}_2\delta_{b6} + \delta_{b7}]. \end{aligned} \tag{3.21}$$

We present the results of these calculations in Table I, wherefrom the rather formidable Lie algebra follows for the yet unknown infinitesimal operators $X_a(t,x)$. Clearly, we recover from Table I the Lie algebra for the linear system $\ddot{x} + f_2(t)\dot{x} + f_1(t)x = f_0(t)$ considered in paper I (cf. Tables I and II of that paper).

It must be borne in mind that the general algebra, whose structure constants have been obtained above, corresponds to the *maximal* possible Lie algebra (i.e., $r = 8$) associated with a differential equation of the form presented in Eq. (3.1). Indeed, it may happen that the fundamental system of Eqs. (3.2)–(3.5) admits less than eight linearly independent solutions ($r < 8$), in which case some of the q's are zero. For instance, one may get after solving Eqs. (3.2)–(3.5), in a particular case: $\eta = \eta(t)$ and $\theta = \theta(x)$; which means, according to the parametrization adopted in Eqs. (3.12): $q^5 = q^6 = 0$. But then one may use the same rule (i.e., $q^5 = q^6 = 0 \Rightarrow X_5 = X_6 = 0$) one uses for having the subalgebras associated with the subgroups one obtains by the elimination of some of the parameters. Of course, in such a case one must substitute the new structure constants $f^b_{5a} = f^b_{6a} = f^5_{ab} = f^6_{ab} = 0$ everywhere in Table I. It is only in this very special sense that all the Lie algebras associated with Eq. (3.1) are contained in Table I. The situation we comment presents itself very frequently for nonlinear differential equations. We have already recalled the extreme example of the equation $\ddot{x} = x^2 + t^2$, which belongs in the class of Eq. (3.1), and which, however, gives $\eta = \theta = 0$ as the only solution of Eqs. (3.2)–(3.5).

IV. LIE ALGEBRAS OF $\ddot{x} = f_N(t,x,\dot{x})$: CASE $N>4$

Now let us briefly consider the infinitesimal symmetries of a differential equation of the following type:

$$\ddot{x} = \alpha_0(t,x) + \cdots + \alpha_3(t,x)\dot{x}^3 + \cdots + \alpha_N(t,x)\dot{x}^N, \tag{4.1}$$

where $\alpha_N \neq 0$ for some $N = 4,5,\ldots < \infty$. In this case one has the holonomic constraint $\eta = \eta(t)$ [cf. Eq. (2.11)] and, therefore, taking into account this constraint, the generators have to satisfy the following system of differential equations:

$$2\theta_{xt} - \ddot{\eta} - \alpha_1\dot{\eta} - \alpha_{1t}\eta - \alpha_{1x}\theta - 2\alpha_2\theta_t = 0, \tag{4.2}$$

$$\theta_{tt} - \alpha_0(2\dot{\eta} - \theta_x) - \alpha_{0t}\eta - \alpha_{0x}\theta - \alpha_1\theta_t = 0, \tag{4.3}$$

TABLE I. The nonzeroth structure constants of the maximal Lie algebra associated with the differential equation $\ddot{x} = \alpha_0(t,x) + \alpha_1(t,x)\dot{x} + \alpha_2(t,x)\dot{x}^2 + \alpha_3(t,x)\dot{x}^3$; here $\hat{\alpha}$ denotes $\alpha(t_0,x_0)$ quite generally.

$$f^1_{13} = 1, \quad f^1_{25} = 1.$$

$$f^2_{16} = 1, \quad f^2_{24} = 1.$$

$$\begin{aligned} & f^3_{12} = \tfrac{1}{2}\hat{\alpha}_{2t}, \quad f^3_{17} = 2, \quad f^3_{24} = -\tfrac{1}{2}\hat{\alpha}_2, \quad f^3_{25} = -\hat{\alpha}_1, \\ & f^3_{26} = -\tfrac{3}{2}\hat{\alpha}_3, \quad f^3_{28} = 1, \quad f^3_{56} = -1. \end{aligned}$$

$$\begin{aligned} & f^4_{12} = \tfrac{1}{2}\hat{\alpha}_{1x}, \quad f^4_{13} = \tfrac{1}{2}\hat{\alpha}_1, \quad f^4_{15} = \tfrac{3}{2}\hat{\alpha}_0, \\ & f^4_{16} = \hat{\alpha}_2, \quad f^4_{17} = 1, \quad f^4_{28} = 2, \quad f^4_{56} = 1. \end{aligned}$$

$$\begin{aligned} & f^5_{12} = \hat{\alpha}_{3t} - \tfrac{1}{2}\hat{\alpha}_{2x}, \quad f^5_{14} = -\tfrac{1}{2}\hat{\alpha}_2, \quad f^5_{15} = -\hat{\alpha}_1, \\ & f^5_{16} = -\tfrac{3}{2}\hat{\alpha}_3, \quad f^5_{18} = 1, \quad f^5_{23} = \hat{\alpha}_3, \quad f^5_{24} = -2\hat{\alpha}_3, \\ & f^5_{25} = -\hat{\alpha}_2, \quad f^5_{35} = -1, \quad f^5_{45} = 1. \end{aligned}$$

$$\begin{aligned} & f^6_{12} = \hat{\alpha}_{0x} - \tfrac{1}{2}\hat{\alpha}_{1t}, \quad f^6_{13} = 2\hat{\alpha}_0, \quad f^6_{14} = -\hat{\alpha}_0, \quad f^6_{16} = \hat{\alpha}_1, \quad f^6_{23} = \tfrac{1}{2}\hat{\alpha}_1, \\ & f^6_{25} = \tfrac{3}{2}\hat{\alpha}_0, \quad f^6_{26} = \hat{\alpha}_2, \quad f^6_{27} = 1, \quad f^6_{36} = 1, \quad f^6_{46} = -1. \end{aligned}$$

$$\begin{aligned} & f^7_{12} = -\tfrac{1}{2}\hat{\alpha}_0(\tfrac{1}{2}\hat{\alpha}_{2x} - \hat{\alpha}_{3t}) + \tfrac{1}{2}\hat{\alpha}_1(\hat{\alpha}_{1x} - \tfrac{1}{2}\hat{\alpha}_{2t}) - \hat{\alpha}_2\hat{\alpha}_{0x} + \hat{\alpha}_3\hat{\alpha}_{0t} + \hat{\alpha}_{0xx} - \tfrac{1}{3}(2\hat{\alpha}_{1tx} - \hat{\alpha}_{2tt}), \\ & f^7_{13} = -2\hat{\alpha}_0\hat{\alpha}_2 + \tfrac{1}{2}\hat{\alpha}_1^2 + 2\hat{\alpha}_{0x} - \hat{\alpha}_{1t}, \quad f^7_{14} = \tfrac{3}{4}\hat{\alpha}_0\hat{\alpha}_2, \quad f^7_{15} = \hat{\alpha}_0\hat{\alpha}_1, \\ & f^7_{16} = \tfrac{1}{2}(\hat{\alpha}_{1x} - \hat{\alpha}_{2t} - \tfrac{3}{2}\hat{\alpha}_0\hat{\alpha}_3), \quad f^7_{18} = -\tfrac{3}{2}\hat{\alpha}_0, \quad f^7_{23} = -\tfrac{3}{2}\hat{\alpha}_0\hat{\alpha}_3 + \tfrac{1}{3}(\tfrac{1}{2}\hat{\alpha}_{1x} - \hat{\alpha}_{2t}), \\ & f^7_{24} = \tfrac{1}{4}\hat{\alpha}_1\hat{\alpha}_2 + \tfrac{1}{6}(\hat{\alpha}_{1x} - \hat{\alpha}_{2t}), \quad f^7_{25} = -\tfrac{1}{2}(\hat{\alpha}_0\hat{\alpha}_2 - \hat{\alpha}_1^2 - 2\hat{\alpha}_{0x} + \hat{\alpha}_{1t}), \\ & f^7_{26} = -\tfrac{1}{4}\hat{\alpha}_1\hat{\alpha}_3 + \hat{\alpha}_{2x} - 2\hat{\alpha}_{3t}, \quad f^7_{28} = -\tfrac{1}{2}\hat{\alpha}_1, \quad f^7_{35} = \hat{\alpha}_0, \quad f^7_{37} = 1, \\ & f^7_{45} = -\tfrac{1}{2}\hat{\alpha}_0, \quad f^7_{46} = \tfrac{1}{2}\hat{\alpha}_2, \quad f^7_{56} = \tfrac{1}{2}\hat{\alpha}_1, \quad f^7_{68} = 1. \end{aligned}$$

$$\begin{aligned} & f^8_{12} = -\hat{\alpha}_0\hat{\alpha}_{3x} + \hat{\alpha}_1\hat{\alpha}_{3t} + \tfrac{1}{2}\hat{\alpha}_2(\tfrac{1}{2}\hat{\alpha}_{1x} - \hat{\alpha}_{2t}) - \tfrac{1}{2}\hat{\alpha}_3(\hat{\alpha}_{0x} - \tfrac{1}{2}\hat{\alpha}_{1t}) + \tfrac{1}{3}(\hat{\alpha}_{1xx} - 2\hat{\alpha}_{2tx}) + \hat{\alpha}_{3tt}, \\ & f^8_{13} = \tfrac{1}{4}\hat{\alpha}_1\hat{\alpha}_2 + \tfrac{1}{3}(\hat{\alpha}_{1x} - \tfrac{1}{2}\hat{\alpha}_{2t}), \quad f^8_{14} = -\tfrac{3}{2}\hat{\alpha}_0\hat{\alpha}_3 + \tfrac{1}{3}(\hat{\alpha}_{1x} - \tfrac{1}{2}\hat{\alpha}_{2t}), \\ & f^8_{15} = -\tfrac{1}{4}\hat{\alpha}_0\hat{\alpha}_2 + \hat{\alpha}_{0x} - \tfrac{1}{2}\hat{\alpha}_{1t}, \quad f^8_{16} = \tfrac{1}{2}(-\hat{\alpha}_1\hat{\alpha}_3 + \hat{\alpha}_2^2 + \hat{\alpha}_{2x} - 2\hat{\alpha}_{3t}), \quad f^8_{17} = \tfrac{1}{2}\hat{\alpha}_2, \\ & f^8_{23} = \tfrac{3}{4}\hat{\alpha}_1\hat{\alpha}_3, \quad f^8_{24} = -2\hat{\alpha}_1\hat{\alpha}_3 + \tfrac{1}{2}\hat{\alpha}_2^2 + \hat{\alpha}_{2x} - 2\hat{\alpha}_{3t}, \\ & f^8_{25} = \tfrac{1}{2}(-\tfrac{3}{2}\hat{\alpha}_0\hat{\alpha}_3 + \hat{\alpha}_{1x} - \hat{\alpha}_{2t}), \quad f^8_{26} = \hat{\alpha}_2\hat{\alpha}_3, \quad f^8_{27} = \tfrac{3}{2}\hat{\alpha}_3, \quad f^8_{35} = -\tfrac{1}{2}\hat{\alpha}_1, \\ & \mathrm{f}^8_{36} = \tfrac{1}{2}\hat{\alpha}_3, \quad \mathrm{f}^8_{46} = -\hat{\alpha}_3, \quad \mathrm{f}^8_{48} = 1, \quad \mathrm{f}^8_{56} = \tfrac{1}{2}\hat{\alpha}_2, \quad \mathrm{f}^8_{57} = 1. \end{aligned}$$

as well as the following set of anholonomic constraints:

$$(k+1)\alpha_{k+1}\theta_t+(k-1)\alpha_k\theta_x+\alpha_{kx}\theta - (k-2)\alpha_k\dot{\eta}+\alpha_{kt}\eta-\delta_{k2}\theta_{xx}=0, \qquad (4.4)$$

for $2<k\leqslant(N-1)$ (i.e., when $N=4,5,...<\infty$), and

$$(N-1)\alpha_N\theta_x+\alpha_{Nx}\theta-(N-2)\alpha_N\dot{\eta}+\alpha_{Nt}\eta=0, \qquad (4.5)$$

for $N=4,5,...<\infty$. Let us introduce the same parametrization adopted before [cf. Eqs. (3.12)]. Obviously, now we have $q^5=\eta_x(t=t_0)=0$ (i.e., $X_5=0$), and thus the Lie group has seven parameters at most. As a matter of principle, Eqs. (4.2) and (4.3) can be solved, to yield:

$$\eta(t)=q^a\eta_a(t), \qquad \theta(t,x)=q^a\theta_a(t,x), \qquad (4.6)$$

where, clearly, $\eta_5=\theta_5=0$. But then, if we substitute from Eqs. (4.6) into Eqs. (4.4) and (4.5), and evaluate them at $(t,x)=(t_0,x_0)$, we get a set of $N-1$ linear homogeneous relations (for the following six constants of integration: q^1,q^2,q^3,q^4,q^6 and q^8); i.e.,

$$(k+1)\hat{\alpha}_{k+1}q^6+(k-1)\hat{\alpha}_kq^4+\hat{\alpha}_{kx}q^2 - (k-2)\hat{\alpha}_kq^3+\hat{\alpha}_{kt}q^1-2\delta_{k2}q^8=0, \qquad (4.7)$$

$$(N-1)\hat{\alpha}_Nq^4+\hat{\alpha}_{Nx}q^2-(N-2)\,\hat{\alpha}_Nq^3+\hat{\alpha}_{Nt}q^1=0, \qquad (4.8)$$

with $2<k<N-1$ and $N\geqslant4$. Thus, in any given instance, the rank of the matrix [six columns $\times(N-1)$ rows] which figures in this linear scheme shall give us the number of essential parameters of the *maximal* Lie group associated with Eq. (4.1). Again, the symmetry-breaking mechanism afforded by the anholonomic constraints becomes apparent.

Of course, in any attempt at a *general* discussion of the generators (4.6) the mathematics becomes bewildering. It is interesting, however, to consider at least the most simple case when all the coefficients α_j, $j=1,2,...,N$, in Eq. (4.1) are constants. We discuss this example in the next Section (Sec. V, Example 5b). Nevertheless, it must be understood that the almost unsurmountable difficulties one faces for having a general formal discussion of the issue when the α_j's are not constants (and $N\geqslant4$) does *not* mean that the *method* fails in these instances.

As for the case when $N=\infty$, the severe restrictions imposed by the first-order constraints become even stronger. Notwithstanding this fact, this does not mean that one should expect the identity to be the only allowable symmetry of a differential equation of the form

$$\ddot{x}=\sum_{k=0}^{\infty}\alpha_k(t,x)\dot{x}^k \qquad (4.9)$$

in every instance. In the next section we also present an example of the method (Example 5d) for this extreme case.

V. SOME MISCELLANEOUS EXAMPLES

In this section we present some instances, which serve to illustrate some particular points of the previous method for obtaining the Lie algebra associated with the point symmetries of one-dimensional systems with a nonlinear behavior in $\dot{x}$. The examples appended in this section also illustrate some interesting features of the converse problem of similarity analysis for nonlinear systems. For the sake of brevity, our discussion is very sketchy.

We shall consider the infinitesimal symmetry properties of the following systems in one dimension: (a) compared similarity analysis of two nonlinear differential equations (with $N=3$ and $N=5$); (b) nonlinear differential equation, with constant coefficients and $N\geqslant4$; (c) the Van der Pol oscillator; and (d) relativistic uniformly accelerated motion (i.e., $N=\infty$).

Example 5a: Here we present a compared similarity analysis of two nonlinear differential equations [i.e., Eq. (5.1) *versus* Eq. (5.6) below], in order to exhibit some peculiar relations between symmetry and nonlinearity. Thus, let us consider the differential equation

$$\ddot{x}=\alpha_3\dot{x}^3, \qquad (5.1)$$

where α_3 is a constant. In this case Eqs. (2.6)–(2.8) become

$$\begin{aligned}&\theta_{tt}=0,\\&2\theta_{xt}-\eta_{tt}=0,\\&\theta_{xx}-2\eta_{xt}=3\alpha_3\theta_t,\\&\eta_{xx}=\alpha_3(\eta_t-2\theta_x),\end{aligned} \qquad (5.2)$$

which we integrate immediately. Using the parametrization shown in Eqs. (3.12) with $(t_0,x_0)=(0,0)$, we obtain

$$\eta(t,x)=q^1+q^3t+q^5x+q^7t^2+(q^8-\tfrac{3}{2}\alpha_3q^6)tx + \tfrac{1}{2}\alpha_3(q^3-2q^4)x^2-\tfrac{1}{2}\alpha_3(q^8+\tfrac{1}{2}\alpha_3q^6)x^3-\tfrac{1}{4}\alpha_3^2q^7x^4, \qquad (5.3)$$

$$\theta(t,x)=q^2+q^4x+q^6t+q^8x^2+q^7tx+\tfrac{1}{2}\alpha_3q^7x^3, \qquad (5.4)$$

and therefore the infinitesimal operators come out as follows:

$$\begin{aligned}&X_1=\partial_t,\quad X_2=\partial_x,\quad X_3=(t+\tfrac{1}{2}\alpha_3x^2)\partial_t,\\&X_4=-\alpha_3x^2\partial_t+x\partial_x,\quad X_5=x\partial_t,\\&X_6=-\tfrac{1}{4}\alpha_3x(6t+\alpha_3x^2)\partial_t+t\partial_x,\\&X_7=(t^2-\tfrac{1}{4}\alpha_3^2x^4)\partial_t+x(t+\tfrac{1}{2}\alpha_3x^2)\partial_x,\\&X_8=x(t-\tfrac{1}{2}\alpha_3x^2)\partial_t+x^2\partial_x.\end{aligned} \qquad (5.5)$$

The Lie algebra is given in Table II (and, of course, it is a particular case of the maximal algebra whose nonzeroth structure constants are given in Table I, as the reader can easily check).

Had we considered, instead of Eq. (5.1), the (perturbed) differential equation

$$\ddot{x}=\alpha_3\dot{x}^3+\alpha_5(t,x)\dot{x}^5 \qquad (5.6)$$

(with α_3 constant), we would equally arrive at Eqs. (5.2), and therefore Eqs. (5.3) and (5.4) would follow after integration. However, the anholonomic constraints presented in Eqs. (2.9)–(2.11) must be taken into account in this case, which yield the following subsidiary conditions:

$$\begin{aligned}&\alpha_5(t,x)\theta_t=0,\\&\alpha_5(t,x)(3\eta_t-4\theta_x)-\alpha_{5t}(t,x)\eta-\alpha_{5x}(t,x)\theta=0,\\&\alpha_5(t,x)\eta_x=0.\end{aligned} \qquad (5.7)$$

Hence, we obtain

TABLE II. The Lie algebra associated with the differential equation $\ddot{x} = a_3\dot{x}^3, a_3 = \text{constant}$. One gets the commutator $[X_a, X_b]$ at the intersection of the ath row with the bth column.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	0	0	X_1	0	0	$X_2 - \frac{3}{2}\alpha_3 X_5$	$2X_3 + X_4$	X_5
X_2	0	0	$\alpha_3 X_5$	$X_2 - 2\alpha_3 X_5$	X_1	$-\frac{3}{2}\alpha_3 X_3$	$X_6 + \frac{3}{2}\alpha_3 X_8$	$X_3 + 2X_4$
X_3	$-X_1$	$-\alpha_3 X_5$	0	0	$-X_5$	$X_6 + \frac{1}{2}\alpha_3 X_8$	X_7	0
X_4	0	$-X_2 + 2\alpha_3 X_5$	0	0	X_5	$-X_6 - \alpha_3 X_8$	0	X_8
X_5	0	$-X_1$	X_5	$-X_5$	0	$-X_3 + X_4$	X_8	0
X_6	$-X_2 + \frac{3}{2}\alpha_3 X_5$	$\frac{3}{2}\alpha_3 X_3$	$-X_6 - \frac{1}{2}\alpha_3 X_8$	$X_6 + \alpha_3 X_8$	$X_3 - X_4$	0	0	X_7
X_7	$-2X_3 - X_4$	$-X_6 - \frac{3}{2}\alpha_3 X_8$	$-X_7$	0	$-X_8$	0	0	0
X_8	$-X_5$	$-X_3 - 2X_4$	0	$-X_8$	0	$-X_7$	0	0

$$\eta(t) = q^1 + q^3 t, \quad \theta(x) = q^2 + \tfrac{1}{2}q^3 x, \tag{5.8}$$

and, moreover,

$$q^1\alpha_{5t}(t,x) + q^2\alpha_{5x}(t,x) + q^3(t\alpha_{5t}(t,x) + \tfrac{1}{2}x\alpha_{5x}(t,x) - \alpha_5(t,x)) = 0, \tag{5.9}$$

which, in general, yields $q^1 = q^2 = q^3 = 0$, unless the function $\alpha_5(t,x)$ meets very "fortunate" conditions. We present some of these "fortunate" cases in Table III. The only exceptional "case" missing in Table III (i.e., $\alpha_5 = t\alpha_{5t} + \frac{1}{2}x\alpha_{5x}$ *and* $\alpha_{5t} = \lambda\alpha_{5x}$) is impossible. Hence we conclude that, in general, no matter how small is the nonlinear perturbative term $\alpha_5(t,x)\dot{x}^5$ in Eq. (5.6), as far as it is not zero, it operates as a very strong symmetry breakdown mechanism for the unperturbed system defined in Eq. (5.1), because of the subsidiary constraints which must be satisfied.

Example 5b: Let us consider the (nonlinear in $\dot{x}$) differential equation of the second order, with constant coefficients,

$$\ddot{x} = \alpha_0 + \cdots + \alpha_3\dot{x}^3 + \cdots + \alpha_N\dot{x}^N, \tag{5.10}$$

where $\alpha_N \neq 0$ for some $N = 4,5,... < \infty$. Then Eqs. (4.4) (for $k = N-1$) and (4.5) give us, immediately,

$$\eta(t) = q^1 + q^3 t, \tag{5.11}$$

$$\theta(t,x) = q^2 - \frac{\alpha_{N-1}}{N(N-1)\alpha_N}q^3 t + \frac{N-2}{N-1}q^3 x; \tag{5.12}$$

which correspond to a chosen parametrization at the regular point $(t_0,x_0) = (0,0)$, and which, upon substitution into Eqs. (4.2), (4.3), and (4.4) (for $2<k<N-2$) yield

$$(N(N-k)\alpha_k\alpha_N - (k+1)\alpha_{k+1}\alpha_{N-1})q^3 = 0,$$
$$k = 0,1,...,N-2. \tag{5.13}$$

TABLE III. Exceptional cases for the generators associated with $\ddot{x} = \alpha_3\dot{x}^3 + \alpha_5(t,x)\dot{x}^5, \alpha_3 = \text{constant}$.

$\alpha_5(t,x)$	$\eta(t,x)$	$\theta(t,x)$
At	$q^3 t$	q^2
$At + C$	$q^1(1-(A/c)t)$	$q^2 - \frac{1}{2}(A/C)q^1 x$
$At + at^2$	0	q^2
Bx^2	$q^1 + q^3 t$	$\frac{1}{2}q^3 x$
$Bx^2 + C$	q^1	0
$Bx^2 + bx^4$	q^1	0
$At + Bx^2$	$q^3 t$	$\frac{1}{2}q^3 x$
$At + Bx^2 + C$	$q^1(1-(A/C)t)$	$-\frac{1}{2}(A/C)q^1 x$

In consequence, if

$$\alpha_{k+1} \neq \frac{N(N-k)\alpha_N}{(k+1)\alpha_{N-1}}\alpha_k, \tag{5.14}$$

for *some* k in the range $0<k<N-2$, we must set $q^3 = 0$; and thus the generators are simply

$$\eta = q^1 \quad \text{(time displacement)},$$
$$\theta = q^2 \quad \text{(space displacement)}, \tag{5.15}$$

as expected. On the other hand, if

$$\alpha_{k+1} = \frac{N(N-k)\alpha_N}{(k+1)\alpha_{N-1}}\alpha_k \tag{5.16}$$

holds for *all* $k = 0,1,...,N-2$, that is, if (and only if) the differential equation (5.10) is of the form

$$\ddot{x} = \alpha_N\left(\dot{x} + \frac{\alpha_{N-1}}{N\alpha_N}\right)^N = \alpha(\dot{x}+\beta)^N, \quad (N>4), \tag{5.17}$$

the following exceptional case obtains:

$$\eta(t) = q^1 + q^3 t,$$
$$\theta(t,x) = q^2 + q^3\frac{(N-2)x - \beta t}{N-1}. \tag{5.18}$$

Hence, the Lie algebra associated with a second-order differential equation like (5.17) is

$$[X_1,X_2] = 0,$$
$$[X_1,X_3] = X_1 - \frac{\beta}{N-1}X_2, \tag{5.19}$$
$$[X_2,X_3] = \frac{N-2}{N-1}X_2.$$

Of course (since $N \geqslant 4$, *ex hypothesis*), the reader can easily check that this algebra does *not* correspond to a particular case of the maximal Lie algebra presented in Table I (which holds only for $N \leqslant 3$).

Example 5c (the Van der Pol oscillator): There are many examples of bizarre nonlinear phenomena that can be modeled by a system with a viscous reaction linear in $\dot{x}$, although with a variable viscosity which is nonlinear in x, and under the action of a time independent applied force. Hence, the equation of motion of such a nonuniform linear viscous system is of the general form

$$\ddot{x} = \alpha_0(x) + \alpha_1(x)\dot{x}, \tag{5.20}$$

which was studied by Levinson and Smith,[12] some 40 years ago. Well known instances of systems that can be represented by this general type of equation are the onset of coherent radiation in lasers and masers,[13] self-excitations in electric circuits,[14] and many others. Indeed, differential equations of this general type are fundamental in the dynamic theory dealing with nonlinear dissipative oscillators.[15]

From Eqs. (3.2)–(3.5) we obtain, in this case,

$$\eta_{xx} = 0, \tag{5.21}$$

$$\theta_{xx} - 2\eta_{xt} - 2\alpha_1(x)\eta_x = 0, \tag{5.22}$$

$$2\theta_{xt} - \eta_{tt} - 3\alpha_0(x)\eta_x - \alpha_1(x)\eta_t - \alpha_1'(x)\theta = 0, \tag{5.23}$$

$$\theta_{tt} - \alpha_0(x)(2\eta_t - \theta_x) - \alpha_0'(x)\theta - \alpha_1(x)\theta_t = 0, \tag{5.24}$$

as the fundamental equations for the symmetry generators of Eq. (5.20). For the sake of concreteness, let us consider in particular the Van der Pol dissipative oscillator,[16]

$$\ddot{x} + \omega_0^2 x - \gamma(1 - x^2)\dot{x} = 0, \tag{5.25}$$

$\gamma > 0$, which is the canonical instance of the general differential equation stated in Eq. (5.20). (Here we have a concrete example of a differential equation which is linear in $\dot{x}$ and nonlinear in x.) In this case, Eqs. (5.21) and (5.22) can be easily integrated; after some manipulations, one gets:

$$\eta(t,x) = \phi_1(t)x + \phi_2(t), \tag{5.26}$$

$$\theta(t,x) = (\dot{\phi}_1(t) + \gamma\phi_1(t))x^2 + \phi_3(t)x + \phi_4(t) - \tfrac{1}{6}\gamma\phi_1(t)x^4, \tag{5.27}$$

which are analogous (but not equal) to Eqs. (I.2.11) and (I.2.12), respectively. If one substitutes from Eqs. (5.26) and (5.27) into Eqs. (5.23) and (5.24), one obtains that the time dependent coefficients $\phi_j(t)$, $j = 1,2,3,4$, have to satisfy:

$$2\dot{\phi}_3 - \ddot{\phi}_2 - \gamma\dot{\phi}_2 = 0, \tag{5.28a}$$

$$3(\ddot{\phi}_1 + \gamma\dot{\phi}_1) + 3\omega_0^2\phi_1 + 2\gamma\phi_4 = 0, \tag{5.28b}$$

$$\gamma(\phi_2 + 2\phi_3) = 0, \tag{5.29a}$$

$$\gamma(5\dot{\phi}_1 + 6\gamma\phi_1) = 0 \tag{5.29b}$$

$$\gamma^2\phi_1 = 0, \tag{5.29c}$$

and

$$\ddot{\phi}_4 - \gamma\dot{\phi}_4 + \omega_0^2\phi_4 = 0, \tag{5.30a}$$

$$\ddot{\phi}_3 + 2\omega_0^2\dot{\phi}_2 - \gamma\dot{\phi}_3 = 0, \tag{5.30b}$$

$$\dddot{\phi}_1 + (\omega_0^2 - \gamma^2)\dot{\phi}_1 - \gamma\omega_0^2\phi_1 + \gamma\dot{\phi}_4 = 0, \tag{5.30c}$$

$$\gamma\phi_3 = 0, \tag{5.31a}$$

$$\gamma(5\ddot{\phi}_1 + 7\gamma\dot{\phi}_1 - 5\omega_0^2\phi_1) = 0, \tag{5.31b}$$

$$\gamma^2\dot{\phi}_1 = 0, \tag{5.31c}$$

respectively.

Thus, clearly, the solution is

$$\eta = q^1, \quad \theta = 0. \tag{5.32}$$

Hence, the Van der Pol oscillator has only the (almost trivial) point symmetry of time translation invariance. We observe that in the limit of the simple harmonic oscillator ($\gamma = 0$), Eqs. (5.29) and (5.31) collapse into six useless identities, and Eqs. (5.28) and (5.30) give us the generators presented in Eqs. (I.3.10) and (I.3.11) for the linear oscillator. Interesting enough, notwithstanding the fact that the simple harmonic oscillator is a special case of the Van der Pol oscillator, the symmetry group of point transformations of the nonlinear oscillator is just a trivial subgroup (i.e., $q^1 \neq 0$, and $q^2 = \cdots = q^8 = 0$) of the group of the linear oscillator (even in the limit when $\gamma \to 0$, since Eqs. (5.32) do *not* depend on γ, indeed).

Furthermore, for the forced Van der Pol oscillator, with a guiding force $f_0(t)$, one has certainly the identity ($\eta = \theta = 0$) as the only surviving symmetry [unless $\dot{f}_0 = 0$, where Eqs. (5.32) still hold]. Also, since the Rayleigh oscillator,[17]

$$\ddot{x} + \omega_0^2 x - \gamma(1 - \dot{x}^2)\dot{x} = 0, \tag{5.33}$$

can be put in the Van der Pol form by means of the substitution $y = \sqrt{3}\dot{x}$, it is obvious that Eqs. (5.32) represent the only symmetry of this equation too.

In summary, for a Van der Pol oscillator one does not have an active point transformation (besides time translation) which transforms one solution into another by means of a continuous adjustment of a symmetry group parameters. This makes a great contrast with the simple harmonic oscillator, whose classical states are all continuously connected by means of its eight-parameter symmetry group of point transformations.[18]

Hence, the nonlinear behavior in x may also determine a lack of full symmetry in a broken symmetry model, in which the original symmetry of the unperturbed model can not be recovered, no matter how small the perturbative coupling constant $\gamma \neq 0$ may be.

Example 5d (hyperbolic motion) [19]: Finally, for the case when $N = \infty$, we consider uniformly accelerated motion in relativistic mechanics, for a one-dimensional system. The definition of uniform acceleration in relativistic mechanics can be stated as

$$\frac{da^\mu}{d\tau} = a^\nu a_\nu v^\mu, \tag{5.34}$$

where a^μ and v^μ are the four-acceleration and four-velocity of the particle, respectively, and τ denotes proper time (we set $c = 1$). It can be shown easily that the equation of motion

$$\dddot{x} = -3(1 - \dot{x}^2)^{-1}\dot{x}\ddot{x}^2 \tag{5.35}$$

is equivalent to Eq. (5.34). (For details, see Rohrlich, Ref. 19, pp. 115–117.) Clearly, Eq. (5.35) admits the following first integral:

$$\ddot{x} = (1 - \dot{x}^2)^{3/2} g, \tag{5.36}$$

where g is a constant. This equation is interesting by itself since, in one-dimensional space, it corresponds to the following relativistic equation of motion

$$\frac{d}{dt}((1 - \dot{x}^2)^{-1/2}\dot{x}) = g, \tag{5.37}$$

which meaning is clear. [In three-dimensional space, however, the equivalence of (5.36) and (5.37) requires some special assumptions on collinearity.] Thus, one may consider Eq. (5.36) as the differential equation one has to solve for hyperbolic motion in this case.

If one considers the infinitesimal point symmetries of Eq. (5.36), as the reader can easily check, one gets

$$t' = t + \epsilon q^1 \quad \text{(time translation)}$$
$$x' = x + \epsilon q^2 \quad \text{(space translation)} \tag{5.38}$$

as the only infinitesimal point transformation leaving invariant Eq. (5.36).

ACKNOWLEDGMENTS

Work was supported in part (M.A.) by Dirección General de Investigación de la Universidad Católica de Valparaíso, through DGI Grant No. 123.726/82, and in part (J.K.) by Dirección de Investigación de la Pontificia Universidad Católica de Chile, through Grant No. 46/83.

[1]M. Aguirre and J. Krause, J. Math. Phys. **25**, 210 (1984). We refer to this paper as paper I, and to equations in it by equation numbers like Eq. (I.2.1).

[2]Cf. G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer, New York, 1974).

[3]See, for instance, N. N. Krylov and N. N. Bogoliubov, *Introduction to Nonlinear Mechanics* (Princeton U.P., Princeton, 1947). See, also, F. Brauer and J. A. Nohel, *The Qualitative Theory of Ordinary Differential Equations* (Benjamin, New York, 1969).

[4]The current literature on nonlinear systems having many degrees of freedom is not as developed as for one-dimensional systems. Exceptions are the books by R. M. Evan-Ivanowski, *Resonance Oscillators in Mechanical Systems* (Elsevier, New York, 1976), and by A. H. Nayfeh and D. T. Mook, *Nonlinear Oscillations* (Wiley, New York, 1979). We thank María C. Depassier for bringing Nayfeh's book to our attention.

[5]T. Brooke Benjamin and P. J. Olver, J. Fluid Mech. **125**, 137 (1982). This paper includes an extensive bibliography on both issues, water waves and similarity analysis.

[6]L. V. Ovsiannikov, *Group Analysis of Differential Equations* (Academic, New York, 1982).

[7]This is a report of work in progress. The identification of the Lie algebras obtained in this paper and in paper I is already under study. Results will be published elsewhere.

[8]This is contrary to an erroneous statement we make in the last paragraph of Sec. II in paper I.

[9]L. Bianchi, *Lezioni sulla teoria dei gruppi continui finiti di transformazioni* (Zanichelli, Bologna, 1928), Chap. X.

[10]E. E. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), Chap. IV.

[11]We do appreciate the comments of the referee of a previous version of this paper (and also of the referee of paper I), concerning this important point of rigor. In paper I we have tacitly assumed that the point $(t,x) = (0,0)$ is such a regular point, for the sake of concreteness. Our general (i.e., formal) results are independent of this detail.

[12]N. Levinson and O. K. Smith, Duke Math. J. **9**, 382 (1942).

[13]H. Haken, Rev. Mod. Phys. **47**, 67 (1975).

[14]R. S. Stratonovich, *Topics in the Theory of Random Noise* (Gordon and Breach, New York, 1967), Vol. II.

[15]A. H. Nayfeh and D. T. Mook, Ref. 4.

[16]B. Van der Pol, Philos. Mag. **2**, (7), 978 (1926).

[17]Lord Rayleigh, Philos. Mag. **15**, (1), 229 (1883).

[18]C. E. Wulfman and B. G. Wybourne, J. Phys. A **9**, 507 (1976).

[19]F. Rohrlich, *Classical Charged Particles* (Addison–Wesley, Reading, MA, 1965), Sec. 5-3.

Decomposition formulas of exponential operators and Lie exponentials with some applications to quantum mechanics and statistical physics

Masuo Suzuki
Department of Physics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

(Received 17 July 1984; accepted for publication 19 October 1984)

Decomposition formulas of general exponential operators in a Banach algebra and in a Lie algebra are presented that yield a basis of Monte Carlo simulation of quantum systems. They are applied to study the relaxation and fluctuation from the initially unstable point and to confirm algebraically the scaling theory of transient phenomena. A global approximation method of transient phenomena is also formulated on the basis of decomposition formulas. It is applied to the laser model as a simple example.

I. INTRODUCTION

Exponential operators appear very often in statistical physics and quantum mechanics. In particular, Feynman's formula

$$\left(\frac{d}{d\lambda} e^{A+\lambda B}\right)_{\lambda=0} = \int_0^1 e^{(1-s)A} B\, e^{sA}\, ds, \tag{1.1}$$

has been used very frequently in perturbational calculations. This has been extended to the following formula[1–3]:

$$\frac{d}{d\lambda} e^{A(\lambda)} = \int_0^1 e^{(1-s)A(\lambda)} A'(\lambda)\, e^{sA(\lambda)}\, ds$$
$$= \int_0^1 e^{sA(\lambda)} A'(\lambda)\, e^{(1-s)A(\lambda)}\, ds. \tag{1.2}$$

This is furthermore extended to ordered exponentials[4] in the succeeding section.

Some generalized decomposition formulas of exponential operators are derived in Sec. III. One of the simplest examples is the following formula[5]:

$$\|e^{A+B} - (e^{(1/n)A} e^{(1/n)B})^n\|$$
$$\leq (\|[A,B]\|/2n)\exp(\|A\| + \|B\|). \tag{1.3}$$

This yields Trotter's formula,[6]

$$\lim_{n\to\infty} (e^{(1/n)A} e^{(1/n)B})^n = e^{A+B}, \tag{1.4}$$

for bounded operators A and B. This has been used in performing Monte Carlo simulations of quantum spin systems.[7] The above inequality (1.3) is extended to some more general forms in Sec. III.

In Sec. IV, we derive some decomposition formulas for exponential operators[8] satisfying the Lie algebra. It is well known[1,8] that the exponential operator $\exp(A+B)$ of the two-component Lie algebra (A,B) is decomposed as

$$e^{A+B} = e^A e^{f(\alpha)B};\quad f(\alpha) = (1-e^{-\alpha})/\alpha, \tag{1.5}$$

where $[A,B] = \alpha B$. This decomposition formula is conveniently used[8,9] in solving the linear Fokker–Planck equation. More general formulas on the Lie algebra including an infinite-component one are also given in Sec. IV. These formulas are applied in Sec. V to derive the scaling theory[10,11] for transient phenomena near the instability point and to formulate a global approximation method of transient phenomena.

II. DERIVATIVES OF GENERALIZED EXPONENTIAL OPERATORS

In many systems far from equilibrium, a time-dependent operator $\mathscr{H}(t)$ appears usually and consequently the following ordered exponentials $V(t)$ and $V^{-1}(t)$ are used[12,13]:

$$V(t) = \exp_+ \int_0^t \mathscr{H}(s)ds$$
$$= 1 + \sum_{n=1}^{\infty} \int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n\, \mathscr{H}(t_1)\cdots\mathscr{H}(t_n), \tag{2.1}$$

and

$$V^{-1}(t) = \exp_-\left(-\int_0^t \mathscr{H}(s)ds\right)$$
$$= 1 + \sum_{n=1}^{\infty} (-1)^n \int_0^t dt_1 \int_0^{t_1} dt_2$$
$$\cdots \int_0^{t_{n-1}} dt_n\, \mathscr{H}(t_n)\cdots\mathscr{H}(t_1). \tag{2.2}$$

From the definitions (2.1) and (2.2), we have that $V^{-1}(t)V(t) = V(t)V^{-1}(t) = 1$ and that

$$\frac{d}{dt} V(t) = \mathscr{H}(t)V(t)$$

and

$$\frac{d}{dt} V^{-1}(t) = -V^{-1}(t)\mathscr{H}(t). \tag{2.3}$$

Now it is easy to prove[14] the formulas on the differentiation with respect to the parameter ξ appearing in $\mathscr{H}(t)$,

$$\frac{\partial}{\partial\xi} V(t) = V(t) \int_0^t V^{-1}(s) \frac{\partial\mathscr{H}(s)}{\partial\xi} V(s)ds \tag{2.4}$$

and

$$\frac{\partial}{\partial\xi} V^{-1}(t) = -\int_0^t V^{-1}(s) \frac{\partial\mathscr{H}(s)}{\partial\xi} V(s)ds\, V^{-1}(t), \tag{2.5}$$

by noting that

$$\frac{d}{dt}\left(V^{-1}(t)\frac{\partial}{\partial\xi} V(t)\right) = V^{-1}(t)\frac{\partial\mathscr{H}(t)}{\partial\xi} V(t). \tag{2.6}$$

These formulas will be used in the following sections.

 0022-2488/85/040601-12$02.50

III. GENERALIZED DECOMPOSITION FORMULAS OF EXPONENTIAL OPERATORS

As was exemplified in Sec. I, the decomposition of exponential operators is very useful in quantum mechanics and statistical physics.

The inequality (1.3) is easily extended to the following theorem.

Theorem 1: For any set of operators $\{A_j\}$ in a Banach algebra (i.e., normed space),

$$\left\|\exp \sum_{j=1}^{p} A_j - \left(\prod_{j=1}^{p} e^{(1/n)A_j}\right)^n\right\| \leq \frac{1}{2n}\left(\sum_{j>k}\|[A_j,A_k]\|\right)\exp \sum_{j=1}^{p}\|A_j\| \tag{3.1}$$

with an arbitrary positive integer p.

Proof: If we put, as in Ref. 18,

$$g=\exp\left(\frac{1}{n}\sum_{j=1}^{p}A_j\right) \quad \text{and} \quad h=\prod_{j=1}^{p}\exp\left(\frac{1}{n}A_j\right), \tag{3.2}$$

then the left-hand side of the inequality (3.1) is bounded as

$$\|g^n - h^n\| \leq n\|g-h\|\exp\left(\frac{n-1}{n}\sum_{j=1}^{p}\|A_j\|\right). \tag{3.3}$$

In order to find an upper bound of the norm $\|g-h\|$, we note first the following identity.[5]

Identity 1:

$$e^{\lambda(A+B)} = e^{\lambda A}e^{\lambda B} - \int_0^\lambda dt\int_0^t ds \times e^{tA}e^{(t-s)B}[A,B]e^{sB}e^{(\lambda-t)(A+B)}. \tag{3.4}$$

This is easily derived by noting the relation

$$1-e^{\lambda A}e^{\lambda B}e^{-\lambda(A+B)} = \int_0^\lambda dt\, e^{tA}[e^{tB},A]e^{-t(A+B)} \tag{3.5}$$

and Kubo's identity[15]

$$[A,e^{t\mathscr{H}}] = \int_0^t ds\, e^{(t-s)\mathscr{H}}[A,\mathscr{H}]e^{s\mathscr{H}}. \tag{3.6}$$

By applying Identity 1 repeatedly, we obtain the following identity.

Identity 2: With the notation $A_0 \equiv 0$,

$$\begin{aligned}\exp\left(\lambda\sum_{j=1}^{p}A_j\right) - e^{\lambda A_1}e^{\lambda A_2}\cdots e^{\lambda A_p} &= \int_0^\lambda dt\int_0^t ds\sum_{k=1}^{p-1}e^{\lambda A_1}e^{\lambda A_2}\cdots e^{\lambda A_{k-1}}e^{tA_k} \\ &\times\exp\left[(t-s)\sum_{j=k+1}^{p}A_j\right]\left[\sum_{j=k+1}^{p}A_j,A_k\right] \\ &\times\exp\left(s\sum_{j=k+1}^{p}A_j\right)\exp\left[(\lambda-t)\sum_{j=k}^{p}A_j\right].\end{aligned} \tag{3.7}$$

Here we have also used generalized Kubo's identity

$$\begin{aligned}[A,e^{\lambda\mathscr{H}_1}e^{\lambda\mathscr{H}_2}\cdots e^{\lambda\mathscr{H}_p}] &= \sum_{k=1}^{p}\int_0^\lambda ds\, e^{\lambda\mathscr{H}_1}e^{\lambda\mathscr{H}_2}\cdots e^{\lambda\mathscr{H}_{k-1}}e^{(\lambda-s)\mathscr{H}_k} \\ &\times[A,\mathscr{H}_k]e^{s\mathscr{H}_k}e^{\lambda\mathscr{H}_{k+1}}\cdots e^{\lambda\mathscr{H}_p}.\end{aligned} \tag{3.8}$$

This is easily derived from the relation

$$[A,B_1B_2\cdots B_p] = \sum_{k=1}^{p}B_1B_2\cdots B_{k-1}[A,B_k]B_{k+1}\cdots B_p \tag{3.9}$$

and from Kubo's identity (3.6).

Now we take the norm of (3.7). Then we obtain the following inequality.

Inequality 1: For any set of operators $\{A_j\}$ in a Banach algebra,

$$\left\|\exp\left(\lambda\sum_{j=1}^{p}A_j\right) - e^{\lambda A_1}\cdots e^{\lambda A_p}\right\| \leq \frac{\lambda^2}{2}\sum_{j>k}\|[A_j,A_k]\|\exp\left(\lambda\sum_{j=1}^{p}\|A_j\|\right). \tag{3.10}$$

That is, we have an upper bound of the norm $\|g-h\|$ by putting $\lambda = 1/n$ in (3.10). Thus, we arrive finally at Theorem 1.

By the way it may be worthwhile to note the following identities.

Identity 3 (Kubo's dual identity):

$$[A,e^{t\mathscr{H}}] = \int_0^t ds\, e^{s\mathscr{H}}[A,\mathscr{H}]e^{(t-s)\mathscr{H}}. \tag{3.11}$$

Identity 4:

$$[e^{\lambda A},e^{\mu B}] = \int_0^\lambda dt\int_0^\mu ds\, e^{(\lambda-t)A}e^{(\mu-s)B}[A,B]e^{sB}e^{tA}. \tag{3.12}$$

For an ordered exponential, we have the following generalized Kubo's identity.

Identity 5:

$$\begin{aligned}\left[A(t),\exp_+\int_0^t\mathscr{H}(s)ds\right] &= V(t)\int_0^t ds\, V^{-1}(s)[A(t),\mathscr{H}(s)]V(s) \\ &= \int_0^t ds[V(s)A(t)V^{-1}(s),\mathscr{H}(s)]V(t),\end{aligned} \tag{3.13}$$

and

$$\begin{aligned}\left[A(t),\exp_- -\int_0^t\mathscr{H}(s)ds\right] &= \int_0^t ds\, V^{-1}(s)[\mathscr{H}(s),A(t)]V(s)V^{-1}(t) \\ &= V^{-1}(t)\int_0^t ds\,[\mathscr{H}(s),V(s)A(t)V^{-1}(s)]\end{aligned} \tag{3.14}$$

with $V(t)$ defined by (2.1).

The proof of Identity 5 is easily given by differentiating the function $f(t,t') = V^{-1}(t)A(t')V(t) - A(t')$ with respect to t and by putting $t'=t$ after integrating the differential equation thus obtained.

Now we discuss some symmetrized product approximants of exponential operators, which was introduced by Hirsch *et al.*[16] and De Laedt *et al.*[17] by modifying generalized approximants by the present author.[18] First we discuss the symmetrization of (1.3) with respect to the operators A and B. We have the following inequality.

Theorem 2: For any operators A and B in a Banach algebra,

$$\left|\left|e^{A+B}-(e^{(1/2n)A}e^{(1/n)B}e^{(1/2n)A})^n\right|\right| \le \frac{1}{n^2}\Delta_2(A,B), \tag{3.15}$$

where

$$\Delta_2(A,B)=\tilde{\Delta}_2(A,B)\exp(\|A\|+\|B\|),$$

and

$$\tilde{\Delta}_2(A,B)=\tfrac{1}{12}\{\|[[A,B],B]\|+\tfrac{1}{2}\|[[A,B],A]\|\}. \tag{3.16}$$

It should be noted that the upper bound of the difference between the two operators [namely, exp $(A+B)$ and the symmetrized approximant] decreases[5,17] proportionally to n^{-2} as n increases. Thus, the above symmetrized product is a much better approximant.

Proof of Theorem 2: First we prove the following inequality.

Inequality 2: For any operators A and B in a Banach algebra,

$$\|e^{\lambda(A+B)}-e^{(\lambda/2)A}e^{\lambda B}e^{(\lambda/2)A}\|\le\Delta_2(\lambda A,\lambda B), \tag{3.17}$$

where $\Delta_2(A,B)$ is defined by (3.16).

The proof of Inequality 2: If we put

$$F(\lambda)=1-e^{(\lambda/2)A}e^{\lambda B}e^{(\lambda/2)A}e^{-\lambda(A+B)}, \tag{3.18}$$

we obtain $F(0)=0$ and

$$\frac{d}{d\lambda}F(\lambda)=\frac{1}{2}e^{(\lambda/2)A}e^{\lambda B}\int_0^\lambda G(s)ds \times e^{(\lambda/2)A}e^{-\lambda(A+B)}, \tag{3.19}$$

using Kubo's identity (3.6) and his dual identity (3.11), where

$$G(s)=e^{(s/2)A}[A,B]e^{-(s/2)A}-e^{-sB}[A,B]e^{sB}. \tag{3.20}$$

By noting that $G(0)=0$, we have

$$G(s)=\int_0^s du\left\{\frac{1}{2}e^{(u/2)A}[A,[A,B]]e^{-(u/2)A}+e^{-uB}[B,[A,B]]e^{uB}\right\}. \tag{3.21}$$

Thus, we obtain the following formula.

Identity 6:

$$e^{\lambda(A+B)}=e^{(\lambda/2)A}e^{\lambda B}e^{(\lambda/2)A}+\frac{1}{2}\int_0^\lambda dt\int_0^t ds \times e^{(t/2)A}e^{tB}G(s)e^{(t/2)A}e^{(\lambda-t)(A+B)}, \tag{3.22}$$

with $G(s)$ defined by (3.21).

Therefore, we get the inequality (3.17). By using an inequality similar to (3.3) and Inequality 2 for $\lambda=1/n$, we arrive finally at Theorem 2.

For a general set of operators $\{A_j\}$, the following weak bound has been proven[5] already.

Inequality 3: For any set of operators $\{A_j\}$ in a Banach algebra,

$$\left|\left|\exp\left(\sum_{j=1}^p A_j\right)-(e^{(1/2n)A_1}\cdots e^{(1/2n)A_{p-1}}e^{(1/n)A_p}e^{(1/2n)A_{p-1}}\cdots e^{(1/2n)A_1})^n\right|\right|\le\frac{1}{3n^2}\left(\sum_{j=1}^p\|A_j\|\right)^2\exp\left(\sum_{j=1}^p\|A_j\|\right). \tag{3.23}$$

However, an upper bound $\Delta_p(A_1,A_2,\ldots,A_p)/n^2$ stronger than (3.23) is obtained recursively by applying Inequality 2 as

$$\begin{aligned}&\left|\left|\exp\left(\lambda\sum_{j=1}^p A_j\right)-(e^{(\lambda/2)A_1}\cdots e^{(\lambda/2)A_{p-1}}e^{\lambda A_p}e^{(\lambda/2)A_{p-1}}\cdots e^{(\lambda/2)A_1})\right|\right|\\ &\le\left|\left|\exp\left(\lambda\sum_{j=1}^p A_j\right)-\exp\left(\frac{\lambda}{2}A_1\right)\exp\left(\lambda\sum_{j=2}^p A_j\right)\exp\left(\frac{\lambda}{2}A_1\right)\right|\right|\\ &\quad+\left|\left|e^{(\lambda/2)A_1}\left\{\exp\left(\lambda\sum_{j=2}^p A_j\right)-e^{(\lambda/2)A_2}\cdots e^{\lambda A_p}\cdots e^{(\lambda/2)A_2}\right\}e^{(\lambda/2)A_1}\right|\right|\le\Delta_2\left(\lambda A_1,\lambda\sum_{j=2}^p A_j\right)+e^{\|\lambda A_1\|}\Delta_{p-1}(\lambda A_2,\ldots,\lambda A_p).\end{aligned} \tag{3.24}$$

Therefore, we obtain the following recursion relation:

$$\Delta_p(A_1,\ldots,A_p)=e^{\|A_1\|}\Delta_{p-1}(A_2,\ldots,A_p)+\Delta_2\left(A_1,\sum_{j=2}^p A_j\right). \tag{3.25}$$

Thus, we arrive at the following inequality.

Inequality 4: For any set of operators $\{A_j\}$ in a Banach algebra,

$$\left|\left|\exp\left(\lambda\sum_{j=1}^p A_j\right)-e^{(\lambda/2)A_1}\cdots e^{(\lambda/2)A_{p-1}}e^{\lambda A_p}e^{(\lambda/2)A_{p-1}}\cdots e^{(\lambda/2)A_1}\right|\right|\le\Delta_p(\lambda A_1,\ldots,\lambda A_p), \tag{3.26}$$

where $\Delta_p(A_1,\ldots,A_p)$ is given by

$$\Delta_p(A_1,\ldots,A_p)=\sum_{k=1}^{p-1}\exp\left(\sum_{j=1}^{k-1}\|A_j\|\right)\Delta_2(A_k,A_{k+1}+\cdots+A_p)\le\tilde{\Delta}_p(A_1,\ldots,A_p)\exp\left(\sum_{j=1}^p\|A_j\|\right),$$

and

$$\tilde{\Delta}_p(A_1,\ldots,A_p)=\sum_{k=1}^{p-1}\tilde{\Delta}_2(A_k,A_{k+1}+\cdots+A_p). \tag{3.27}$$

This inequality yields the following theorem in a way similar to the proof of Theorem 1 through (3.3).

Theorem 3: For any set of operators $\{A_j\}$ in a Banach algebra

$$\left|\left|\exp\left(\sum_{j=1}^p A_j\right)-(e^{(1/2n)A_1}\cdots e^{(1/2n)A_{p-1}}e^{(1/n)A_p}e^{(1/2n)A_{p-1}}\cdots e^{(1/2n)A_1})^n\right|\right|\le\frac{1}{n^2}\tilde{\Delta}_p(A_1,\ldots,A_p)\exp\left(\sum_{j=1}^p\|A_j\|\right), \tag{3.28}$$

where $\tilde{\Delta}_p(A_1,...,A_p)$ is given by (3.27)

The above symmetrized decomposition of exponential operators is extended as follows. Hirsch *et al.*[16] and De Raedt *et al.*[17] introduced a symmetrization of the form

$$e^{\lambda(A+B)} = e^{(\lambda/2)A} e^{(\lambda/2)B} e^{(\lambda^3/4)C_3} e^{(\lambda/2)B} e^{(\lambda/2)A} + O(\lambda^5) \quad (3.29)$$

for two operators A and B, where

$$C_3 = \tfrac{1}{6}[[B,A],A+2B] \equiv 4R_2(A,B). \quad (3.30)$$

This is easily extended to any set of operators $\{A_1,A_2,...,A_p\}$ as

$$\exp\left(\lambda \sum_{j=1}^{p} A_j\right) = e^{(\lambda/2)A_1}...e^{(\lambda/2)A_p} e^{\lambda^3 R_p} e^{(\lambda/2)A_p}...e^{(\lambda/2)A_1} + O(\lambda^5), \quad (3.31)$$

where

$$R_p = \frac{1}{6}\left[\frac{\partial^3}{\partial\lambda^3}\left(e^{-(\lambda/2)A_p}...e^{-(\lambda/2)A_1} \times \exp\left(\lambda \sum_{j=1}^{p} A_j\right) e^{-(\lambda/2)A_1}...e^{-(\lambda/2)A_p}\right)\right]_{\lambda=0}. \quad (3.32)$$

In order to find $R_p = R_p(A_1,...A_p)$ explicitly in terms of commutators of $\{A_j\}$, we apply repeatedly the following property:

$$e^{\lambda(A,B)} - e^{(\lambda/2)A} e^{\lambda B} e^{(\lambda/2)A} = \lambda^3 R_2(A,B) + O(\lambda^4), \quad (3.33)$$

being equivalent to (3.29). The recursion formula for R_p thus obtained is given by

$$R_p(A_1,...,A_p) = R_2(A_1,A_2 + \cdots + A_p) + R_{p-1}(A_2,A_3,...,A_p) \quad (3.34)$$

through the relation

$$\exp\left(\lambda \sum_{j=1}^{p} A_j\right) - e^{(\lambda/2)A_1} \exp\left(\lambda \sum_{j=2}^{p} A_j\right) e^{(\lambda/2)A_1} = \lambda^3 R_2\left(A_1, \sum_{j=2}^{p} A_j\right) + O(\lambda^4). \quad (3.35)$$

Therefore, we obtain

$$R_p(A_1,...,A_p) = \sum_{k=1}^{p-1} R_2(A_k,A_{k+1} + \cdots + A_p). \quad (3.36)$$

The above arguments can be generalized in the following.

Symmetrized Decomposition Formula: For any set of operators $\{A_j\}$,

$$\exp\left(2\lambda \sum_{j=1}^{p} A_j\right) = e^{\lambda A_1}...e^{\lambda A_p} e^{\lambda^3 S_3}...e^{\lambda^{2n+1} S_{2n+1}} ...e^{\lambda^{2n+1} S_{2n+1}}... e^{\lambda^3 S_3} e^{\lambda A_p}...e^{\lambda A_1}, \quad (3.37)$$

where with $S_1 = 0$ we have

$$S_{2n+1} = \frac{1}{2(2n+1)!}\left[\frac{\partial^{2n+1}}{\partial\lambda^{2n+1}}\left(e^{-\lambda^{2n-1}S_{2n-1}} ...e^{-\lambda^3 S_3} e^{-\lambda A_p}...e^{-\lambda A_1} \times \exp\left(2\lambda \sum_{j=1}^{p} A_j\right) e^{-\lambda A_1} ...e^{-\lambda A_p} e^{-\lambda^3 S_3}...e^{-\lambda^{2n-1}S_{2n-1}}\right)\right]_{\lambda=0}. \quad (3.38)$$

One of the remarkable points in the above expansion formula is that only odd terms in λ appear in (3.37). That is,

$$\exp\left(\lambda \sum_{j=1}^{p} A_j\right) = e^{(\lambda/2)A_1}...e^{(\lambda/2)A_p} e^{(\lambda/2)^3 S_3} ...e^{(\lambda/2)^{2n+1} S_{2n+1}}...e^{(\lambda/2)^3 S_3} e^{(\lambda/2)A_p} ...e^{(\lambda/2)A_1} + O(\lambda^{2n+3}), \quad (3.39)$$

namely there appears no correction of the order of λ^{2n+2} in the above expansion (3.37) or (3.39).

The proof of this formula is given by mathematical induction. For this purpose, we put

$$F(\lambda) \equiv \exp\left(2\lambda \sum_{j=1}^{p} A_j\right) = e^{\lambda A_1}...e^{\lambda A_p} e^{\lambda^3 S_3} ...e^{\lambda^{2n+1}S_{2n+1}} F_{2n+1}(\lambda) e^{\lambda^{2n+1}S_{2n+1}} ...e^{\lambda^3 S_3} e^{\lambda A_p}...e^{\lambda A_1}. \quad (3.40)$$

It is easy to show that

$$S_3 = 4R_p(A_1,A_2,...,A_p)$$

and

$$F_3(\lambda) = 1 + O(\lambda^5) \quad (3.41)$$

with (3.36). Now we assume that S_{2k+1} is given by the form (3.38) for $k \le n$ and that $F_{2k+1}(\lambda) = 1 + O(\lambda^{2k+3})$ for $k \le n$. First it is shown from the above assumption that

$$F_{2n+1}(\lambda) = e^{\lambda^{2n+3}S_{2n+3}} F_{2n+3}(\lambda) e^{\lambda^{2n+3}S_{2n+3}}, \quad (3.42)$$

where $F_{2n+3}(\lambda) = 1 + O(\lambda^{2n+4})$, at least, from the definition of S_{2n+3}. Next, from the property that $F(\lambda)F(-\lambda) = 1$, we have $F_{2n+3}(\lambda)F_{2n+3}(-\lambda) = 1$. Thus we obtain that $F_{2n+3}(\lambda) = 1 + O(\lambda^{2n+5})$. Therefore, we arrive at the conclusion that the above assumption holds also in the case of $n+1$. This completes the proof.

The convergence of the above general expansion formulas (3.37) and (3.39) can be shown in a way quite similar to that[18] for the ordinary Zassenhaus formula for a certain range of λ.

The above symmetrized decomposition formula gives the following approximation method.

Generalized Symmetrized Approximants of Exponential Operators: For any set of operators in a Banach algebra,

$$\exp\left(\sum_{j=1}^{p} A_j\right) = F_{n,m}(A_1,...,A_p) + O(1/n^{2m+2}), \quad (3.43)$$

where

$$F_{n,m}(A_1,...,A_p) = e^{(1/2n)A_1}...e^{(1/2n)A_p} e^{(1/8n^3)S_3} ...e^{2(1/2n)^{2m+1}S_{2m+1}}...e^{(1/8n^3)S_3} \times e^{(1/2n)A_p}...e^{(1/2n)A_1})^n, \quad (3.44)$$

with $\{S_{2k+1}\}$ defined by (3.38).

By the way, it may be also instructive to note the following identity.

Identity 7: With Kubo's notation $A^\times B \equiv [A,B] = AB - BA$,

$$e^{t(A+B)} = e^{tA} e^{tB} \exp_+ \left[\int_0^t ds\, e^{-sB^\times}(e^{-sA^\times} - 1)B \right]$$

$$= \left\{ \exp\left[\int_0^t ds\, e^{sB^\times}(e^{sA^\times} - 1)B \right] \right\} e^{tB} e^{tA}. \qquad (3.45)$$

The proof of this identity is straightforward. This will be used to derive some identities for the two-component Lie algebra such as (1.5) for $A^\times B = \alpha B$, as will be discussed in the succeeding section.

The above identity is easily extended in the following.

Identity 8:

$$\exp\left(t \sum_{j=1}^{p} A_j \right) = e^{tA_1} \cdots e^{tA_p} \exp_+ \int_0^t D(s) ds$$

$$= \exp_- \int_0^t D(-s) ds\, e^{tA_p} \cdots e^{tA_1}, \qquad (3.46)$$

where

$$D(s) = \sum_{k=2}^{p} (e^{-sA_p^\times} \cdots e^{-sA_k^\times})(e^{-sA_{k-1}^\times} \cdots e^{-sA_1^\times} - 1) A_k. \qquad (3.47)$$

The proof is straightforward.

The above identity 8 is also extended to ordered exponentials in the following.

Identity 9:

$$\exp_+ \int_0^t ds \{A(s) + B(s)\}$$

$$= \exp_+ \int_0^t A(s) ds \exp_+ \int_0^t B(s) ds$$

$$\times \exp_+ \int_0^t C_+(s) ds,$$

$$\exp_- \int_0^t ds \{A(s) + B(s)\}$$

$$= \exp_- \int_0^t C_-(s) ds \exp_- \int_0^t B(s) ds$$

$$\times \exp_- \int_0^t A(s) ds, \qquad (3.48)$$

where

$$C_\pm(t) = \exp_\mp \left(\mp \int_0^t B^\times(s) ds \right)$$

$$\times \left\{ \exp_\mp \left(\mp \int_0^t A^\times(s) ds \right) - 1 \right\} B(t). \qquad (3.49)$$

The proof of Identity 9 is given as follows. First we put

$$F_+(t) = \exp_- \left(-\int_0^t B(s) ds \right)$$

$$\times \exp_- \left(-\int_0^t A(s) ds \right) \exp_+ \int_0^t \{A(s) + B(s)\} ds,$$

and

$$F_-(t) = \exp_- \int_0^t ds \{A(s) + B(s)\}$$

$$\times \exp_+ \left(-\int_0^t A(s) ds \right) \exp_+ \left(-\int_0^t B(s) ds \right). \qquad (3.50)$$

Then, we have

$$\frac{d}{dt} F_\pm(t) = C_\pm(t) F_\pm(t) \text{ and } F_\pm(0) = 1, \qquad (3.51)$$

where we have used the relations (2.3) and the following identity.

Identity 10:

$$\left(\exp_\pm \int_0^t A^\times(s) ds \right) B(t)$$

$$= \left(\exp_\pm \int_0^t A(s) ds \right) B(t) \left(\exp_\mp - \int_0^t A(s) ds \right). \qquad (3.52)$$

In order to obtain expressions of $\{S_{2n+1}\}$ explicitly, the following alternative formulation of the symmetrized decomposition formula will be more useful. First, from Identity 8, we have

$$\exp\left(2\lambda \sum_{j=1}^{p} A_j \right) = e^{\lambda A_1} \cdots e^{\lambda A_p} F_1(\lambda) e^{\lambda A_p} \cdots e^{\lambda A_1}, \qquad (3.53)$$

where

$$F_1(\lambda) = \exp_+ \int_0^\lambda D(s) ds \exp_- \int_0^\lambda D(-s) ds \qquad (3.54)$$

with (3.47). Thus, the expansion coefficient S_{2n+1} in (3.37) is given by

$$S_{2n+1} = \frac{1}{2(2n+1)!} \left[\frac{\partial^{\partial n+1}}{\partial \lambda^{2n+1}} e^{-\lambda^{2n-1} S_{2n-1}} \cdots e^{-\lambda^3 S_3} F_1(\lambda) e^{-\lambda^3 S_3} \cdots e^{-\lambda^{2n-1} S_{2n-1}} \right]_{\lambda=0}. \qquad (3.55)$$

This is very convenient, because the right-hand side of (3.55) is expressed already in terms of commutators of $A_1, \ldots, A_p$. For example, we have

$$S_3 = \frac{1}{12} \left(\frac{\partial^3}{\partial \lambda^3} F_1(\lambda) \right)_{\lambda=0} = \frac{1}{6} \left(\frac{\partial^2}{\partial \lambda^2} D(\lambda) \right)_{\lambda=0} = \frac{1}{6} \left(\frac{\partial^2}{\partial \lambda^2} D(-\lambda) \right)_{\lambda=0}$$

$$= \frac{1}{6} \sum_{k=2}^{p} \left[\frac{d^2}{d\lambda^2} (e^{\lambda A_p^\times} \cdots e^{\lambda A_{k+1}^\times})(e^{\lambda A_k^\times} \cdots e^{\lambda A_1^\times} - 1) A_k \right]_{\lambda=0} = \frac{1}{6} \sum_{k=2}^{p} \left\{ 2 \sum_{i \neq j} \sum_{j \leq k-1} A_i^\times A_j^\times + \sum_{i \leq k} (A_i^\times)^2 \right\} A_k. \qquad (3.56)$$

This agrees with the result (3.36), as it should be.

In particular, for $p=2$ we have

$$S_3=\frac{Q_2}{3}=\frac{1}{6}\{\mathscr{N}(A^\times+B^\times)^2\}B=\frac{1}{6}(A^\times+2B^\times)A^\times B,$$

$$S_5=\frac{Q_4}{5}+\frac{[Q_2,Q_1]}{5\cdot 3}=\frac{1}{5}[S_3,A^\times B]+\frac{1}{5!}\{\mathscr{N}(A^\times+B^\times)^4\}B,$$

$$S_7=\frac{Q_6}{7}+\frac{[Q_2,Q_3]}{7\cdot 3}+\frac{[Q_4,Q_1]}{7\cdot 5}+\frac{[Q_1,[Q_1,Q_2]]}{7\cdot 5\cdot 3},$$

$$=\frac{1}{7}[S_5,A^\times B]+\frac{1}{7\cdot 6}[S_3,\{\mathscr{N}(A^\times+B^\times)^3\}B]+\frac{1}{7!}\{\mathscr{N}(A^\times+B^\times)^6\}B,..., \tag{3.57}$$

where the symbol $\mathscr{N}$ denotes the ordering that $B^\times$ should be on the left-hand side of $A^\times$, and the $\{Q_n\}$ are defined by

$$Q_n=(1/n!)\{\mathscr{N}(A^\times+B^\times)^n\}B. \tag{3.58}$$

Any higher-order term S_{2n+1} can be obtained similarly from (3.55) with the use of the following general expansion formula on $F_1(\lambda)$:

$$F_1(\lambda)=F_1^+(\lambda)F_1^-(\lambda), \tag{3.59}$$

where

$$F_1^+(\lambda)=\exp_+\int_0^\lambda(e^{-sB^\times}e^{-sA^\times}-1)B\,ds=\exp_+\int_0^\lambda\{\mathscr{N}(e^{-s(A^\times+B^\times)}-1)\}B\,ds$$

$$=1+\sum_{n=1}^\infty\frac{(-1)^n}{n+1}\lambda^{n+1}Q_n+\sum_{n=1}^\infty\sum_{m=1}^\infty\frac{(-1)^n(-1)^m\lambda^{m+n+2}}{(m+n+2)(m+1)}Q_nQ_m$$

$$+\cdots\sum_{n_1=1}^\infty\cdots\sum_{n_k=1}^\infty\frac{(-1)^{n_1}\cdots(-1)^{n_k}\lambda^{n_1+\cdots+n_k+k}}{(n_1+\cdots+n_k+k)\cdots(n_1+n_2+2)(n_1+1)}Q_{n_k}\cdots Q_{n_1}+\cdots, \tag{3.60}$$

and

$$F_1^-(\lambda)=\exp_-\int_0^\lambda(e^{sB^\times}e^{sA^\times}-1)B\,ds=1+\sum_{k=1}^\infty\sum_{n_1=1}^\infty\cdots\sum_{n_k=1}^\infty\frac{(-1)^{n_1}\cdots(-1)^{n_k}\lambda^{n_1+\cdots+n_k+k}}{(n_1+\cdots+n_k+k)\cdots(n_1+n_2+2)(n_1+1)}Q_{n_1}\cdots Q_{n_k}. \tag{3.61}$$

IV. DECOMPOSITION FORMULAS OF LIE EXPONENTIALS

In the present section, we derive some decomposition formulas of exponential oprators composed of generalized Lie algebra.

The simplest example[8,9] of such exponential oprators may be $\exp(A+B)$ for the two-dimensional Lie algebra (A,B), namely $[A,B]=\alpha B$. As was mentioned in Sec. I, it is decomposed as (1.5). It is easily extended in the following.

Formula 1: If $[A,B]=\alpha B$, then

$$\exp(A+B)=\exp\{\lambda\tilde{f}(\alpha)B\}e^A\exp\{(1-\lambda)f(\alpha)B\} \tag{4.1}$$

for an arbitrary value of λ, where

$$f(\alpha)=\frac{1-e^{-\alpha}}{\alpha}\quad\text{and}\quad\tilde{f}(\alpha)=e^\alpha f(\alpha)=\frac{e^\alpha-1}{\alpha}. \tag{4.2}$$

The proof is given by using the formula (1.5) as follows:

$$e^{A+B}=e^{(A+\lambda B)+(1-\lambda)B}=e^{A+\lambda B}e^{f(\alpha)(1-\lambda)B}$$

$$=e^{\lambda\tilde{f}(\alpha)B}e^Ae^{(1-\lambda)f(\alpha)B}, \tag{4.3}$$

where we have used the fact that if $[A,B]=\alpha B$, then $[A,B']=\alpha B'$ for $B'=\lambda B$, and have also used the simple relation

$$e^Ae^{f(\alpha)B}e^{-A}=e^{\tilde{f}(\alpha)B}. \tag{4.4}$$

The above formula will be applied in Sec. V to the relaxation near the instability point.

It should be remarked that the above decomposition formula (4.1) holds for any value of λ. This freedom will be used in the succeeding section in order to find a global solution of nonlinear relaxation from or near the unstable point of the relevant system.

The above result is easily extended to a more general case.

Formula 2: If $[A,B]=\alpha B$, then

$$\exp_+\int_0^t\{a(s)A+b(s)B\}ds$$

$$=\exp\left\{\int_0^t b(s)\lambda(s,t)\exp\left(\alpha\int_s^t a(u)du\right)dsB\right\}$$

$$\times\exp\left(\int_0^t a(s)ds\,A\right)\exp\left\{\int_0^t b(s)(1-\lambda(s,t))\right.$$

$$\left.\times\exp\left(-\alpha\int_0^s a(u)du\right)ds\,B\right\} \tag{4.5}$$

for an arbitrary function $\lambda(s,t)$.

Proof: First we prove the following lemmas.

Lemma 1:

$$\exp_+\int_0^t\{a(s)A+b(s)B\}ds$$

$$=e^{\mu(t,0)A}\exp\left\{\int_0^t b(s)e^{-\alpha\mu(s,0)}ds\,B\right\}, \tag{4.6}$$

for $[A,B]=\alpha B$, where

$$\mu(t,s) = \int_s^t a(s) ds. \quad (4.7)$$

Lemma 2: For $[A,B] = \alpha B$,

$$\exp_+ \int_0^t \{a(s)A + b(s)B\} ds = \exp\left\{\int_0^t b(s)\, e^{\alpha\mu(t,s)}\, ds\, B\right\} e^{\mu(t,0)A}. \quad (4.8)$$

The proofs of Lemmas 1 and 2 are given in the same way as for Lemma 3.

Now we extend the above arguments to the following infinite-dimensional Lie algebra

$$[A(t), B(s)] = \alpha(t,s) B(s). \quad (4.9)$$

This appears when we treat Fokker–Planck equations with time-dependent drift and diffusion terms. We have the following formula.

Formula 3: For the Lie algebra (4.9) we have

$$\exp_+ \int_0^t \{A(s) + B(s)\} ds = \exp_+ \left(\int_0^t A(s) ds\right) \times \exp_+ \left(\int_0^t B(s) \exp\left(-\int_0^s \alpha(u,s) du\right) ds\right), \quad (4.10)$$

and

$$\exp_- \int_0^t \{A(s) + B(s)\} ds = \exp_- \left(\int_0^t B(s) \exp\left(\int_0^s \alpha(u,s) du\right) ds\right) \exp_- \int_0^t A(s) ds. \quad (4.11)$$

Proof: If we put

$$G(t) = \exp_- \left(-\int_0^t A(s) ds\right) \exp_+ \int_0^t \{A(s) + B(s)\} ds, \quad (4.12)$$

then we obtain

$$\frac{d}{dt} G(t) = \left\{\exp_- \left(-\int_0^t A^\times(s) ds\right) B(t)\right\} G(t) = \left\{\exp\left(-\int_0^t \alpha(s,t) ds\right) B(t)\right\} G(t), \quad (4.13)$$

where we have used Identity 10 in (3.52) and the commutation relation (4.9). The integration of (4.13) yields (4.10). Equation (4.11) is also derived in a quite similar way.

Formula 3 is transformed into the following form.

Formula 4: For the Lie algebra (4.9),

$$\exp_+ \int_0^t \{A(s) + B(s)\} ds = \exp_+ \left\{\int_0^t B(s) \exp\left(\int_s^t \alpha(u,s) du\right) ds\right\} \times \exp_+ \int_0^t A(s) ds, \quad (4.14)$$

and

$$\exp_- \int_0^t \{A(s) + B(s)\} ds = \exp_- \int_0^t A(s) ds \times \exp_- \int_0^t \left\{B(s) \exp\left(-\int_s^t \alpha(u,s) du\right)\right\} ds. \quad (4.15)$$

The proof of Formula 4 is given by putting $\lambda(s,t) = \exp(-\int_0^s \alpha(u,s) du)$ and $\lambda(s,t) = \exp(-\int_s^t \alpha(u,s) du)$ in the following lemma.

Lemma 3: For the Lie algebra (4.9), we have

$$\exp_\pm \int_0^t A(s) ds \exp_\pm \int_0^t \lambda(s,t) B(s) ds = \exp_\pm \left\{\int_0^t \lambda(s,t) B(s) \exp\left(\int_0^t \alpha(u,s) du\right) ds\right\} \times \exp_\pm \int_0^t A(s) ds \quad (4.16)$$

for an arbitrary function of $\lambda(s,t)$.

Proof of Lemma 3: It is easily shown by noting that

$$\exp_\pm \left(\int_0^t A(s) ds\right) \exp_\pm \left(\int_0^t \lambda(s,t) B(s) ds\right) \times \exp_\pm \left(-\int_0^t A(s,) ds\right) = \exp_\pm \left\{\int_0^t \lambda(s,t) \exp\left(\int_0^t \alpha(u,s) du\right) B(s) ds\right\}. \quad (4.17)$$

In particular, when $[B(t), B(s)] = 0$ for any values of t and s, we have the following decomposition formula.

Formula 5: For the Lie algebra (4.9) and for $[B(t), B(s)] = 0$,

$$\exp_\pm \int_0^t \{A(s) + B(s)\} ds = \exp_\pm \int_0^t \{(1 - \lambda(s,t)) B(s) \exp \beta_\pm(s,t)\} ds \times \exp_\pm \left(\int_0^t A(s) ds\right) \times \exp_\pm \int_0^t \{\lambda(s,t) B(s) \exp(-\beta_\pm(s,t))\} ds, \quad (4.18)$$

where

$$\beta_+(s,t) = \int_s^t \alpha(u,s) du, \quad \beta_-(s,t) = \int_0^s \alpha(u,s) du. \quad (4.19)$$

This is easily derived by applying Formulas 3 and 4 to an exponential operator of the form $\exp_\pm \int_0^t \{[A(s) + (1 - \lambda(s,t)) B(s)] + \lambda(s,t) B(s)\} ds$ by noting that

$$[A(s) + (1 - \lambda(s,t)) B(s), B(s')] = \alpha(s,s') B(s') \quad (4.20)$$

under the assumption in Formula 5.

Formulas 1 and 2 are some special cases of Formula 5. These formulas will be applied to solve rigorously or asymptotically Fokker–Planck equations with time-dependent coefficients in the succeeding section.

V. APPLICATIONS TO TRANSIENT PHENOMENA NEAR THE INSTABILITY POINT

A. Linear case

Decomposition formulas in Sec. III have been already applied to Monte Carlo simulations of quantum systems.[7,16,17,19–23] The purpose of the present section is to apply the decomposition formulas presented in Sec. IV to relaxation phenomena.

First we discuss here the following simple linear Fokker–Planck equation:

$$\frac{\partial}{\partial t} P(x,t) = \left(-\frac{\partial}{\partial x} \gamma x + \epsilon \frac{\partial^2}{\partial x^2} \right) P(x,t). \qquad (5.1)$$

This is a well-known exactly soluble equation. As a simple instructive example, we try to apply[8] Formula 1 or Eq. (1.5) to the following formal solution of (5.1):

$$P(x,t) = \exp\left(-t \frac{\partial}{\partial x} \gamma x + t\epsilon \frac{\partial^2}{\partial x^2} \right) \cdot P(x,0). \qquad (5.2)$$

Then, by noting that $[A,B] = \alpha B$ with $\alpha = 2\gamma t$ and with

$$A = -t \frac{\partial}{\partial x} \gamma x \quad \text{and} \quad B = t\epsilon \frac{\partial^2}{\partial x^2}, \qquad (5.3)$$

we obtain

$$\begin{aligned} P(x,t) &= \exp\left(-t \frac{\partial}{\partial x} \gamma x \right) \\ &\quad \times \exp\left\{ (1 - e^{-2\gamma t}) \left(\frac{\epsilon}{2\gamma} \right) \frac{\partial^2}{\partial x^2} \right\} P(x,0) \\ &= \left\{ \frac{2\pi\epsilon(e^{2\gamma t} - 1)}{\gamma} \right\}^{-1/2} \\ &\quad \times \exp\left\{ -\frac{(y - e^{-\gamma t} x)^2}{2\epsilon(1 - e^{-2\gamma t})/\gamma} \right\} P(y,0)dy, \end{aligned} \qquad (5.4)$$

where we have used the following formulas:

$$\exp\left(-\gamma(t) \frac{\partial}{\partial x} x \right) P(x) = e^{-\gamma(t)} P(xe^{-\gamma(t)}), \qquad (5.5)$$

and

$$\begin{aligned} &\exp\left(\epsilon(t) \frac{\partial^2}{\partial x^2} \right) P(x) \\ &= \{4\pi\epsilon(t)\}^{-1/2} \int_{-\infty}^{\infty} \exp\left\{ -\frac{(x-y)^2}{4\epsilon(t)} \right\} P(y)dy. \end{aligned} \qquad (5.6)$$

These are well-known expressions.

Now we extend the above formulation to the following time-dependent Fokker–Planck equation:

$$\frac{\partial}{\partial t} P(x,t) = \left(-\frac{\partial}{\partial x} (a(t)x + b(t)) + \epsilon(t) \frac{\partial^2}{\partial x^2} \right) P(x,t). \qquad (5.7)$$

The formal solution of (5.7) is given by the following time-ordered exponential:

$$\begin{aligned} P(x,t) = \exp_+ \Big\{ \int_0^t \Big(&-\frac{\partial}{\partial x}(a(s)x + b(s)) \\ &+ \epsilon(s) \frac{\partial^2}{\partial x^2} \Big) ds \Big\} P(x,0). \end{aligned} \qquad (5.8)$$

If we put

$$A(t) = -\frac{\partial}{\partial x} (a(t)x + b(t)) \quad \text{and} \quad B(t) = \epsilon(t) \frac{\partial^2}{\partial x^2}, \qquad (5.9)$$

then we find again the following Lie algebra:

$$[A(t), B(s)] = 2a(t)B(s). \qquad (5.10)$$

Thus, we can apply Formula 3 to (5.8). Then we obtain

$$\begin{aligned} P(x,t) &= \exp_+ \left(-\int_0^t \frac{\partial}{\partial x} (a(s)x + b(s)ds) \right) \\ &\quad \times \exp\left(\sigma(t) \frac{\partial^2}{\partial x^2} \right) P(x,0) \end{aligned} \qquad (5.11)$$

with

$$\sigma(t) = \int_0^t \epsilon(s) \exp\left(-2 \int_0^s a(u)du \right) ds. \qquad (5.12)$$

This is again expressed by the following integral:

$$\begin{aligned} P(x,t) &= \frac{\lambda(t)}{\sqrt{4\pi\sigma(t)}} \\ &\quad \times \int_{-\infty}^{\infty} \exp\left\{ -\frac{(y - \lambda(t)x + \int_0^t b(s)\lambda(s)ds)^2}{4\sigma(t)} \right\} \\ &\quad \times P(y,0)dy, \end{aligned} \qquad (5.13)$$

where

$$\lambda(t) = \exp\left(-\int_0^t a(s)ds \right). \qquad (5.14)$$

Here we have used the following transformation formula:

$$\begin{aligned} &\exp_+ \left\{ -\int_0^t \frac{\partial}{\partial x} (a(s)x + b(s))ds \right\} \cdot P(x) \\ &= \lambda(t) P\left(\lambda(t)x - \int_0^t b(s)\lambda(s)ds \right). \end{aligned} \qquad (5.15)$$

The above solution (5.13) has been already derived in different methods.[24,25]

B. Scaling limit

Next, we discuss some asymptotic applications of our decomposition formulas given in Sec. IV. We study here the following Fokker–Planck equation:

$$\frac{\partial}{\partial t} P(x,t) = \left(-\frac{\partial}{\partial x} \alpha(x) + \epsilon \frac{\partial^2}{\partial x^2} \right) P(x,t) \qquad (5.16)$$

with a nonlinear drift velocity $\alpha(x,)$. The formal solution of (5.16) is given by

$$P(x,t) = \exp(t\mathscr{L}_{\text{drift}} + t\mathscr{L}_{\text{diff}}) P(x,t), \qquad (5.17)$$

where

$$\mathscr{L}_{\text{drift}} = -\frac{\partial}{\partial x} \alpha(x) \quad \text{and} \quad \mathscr{L}_{\text{diff}} = \epsilon \frac{\partial^2}{\partial x^2}. \qquad (5.18)$$

It should be noted here[8,26,27] that $\mathscr{L}_{\text{drift}}$ and $\mathscr{L}_{\text{diff}}$ do not necessarily in general satisfy the Lie algebra. However, there is the possibility[8] that the previous decomposition formulas can be used in some asymptotic limit, namely in the scaling limit.[10,11]

As was discussed by the present author[10,11] in the scaling theory of transient phenomena, one of the most interesting nonequilibrium phenomena is the relaxation from or near the unstable equilibrium point. Such a situation is described by (5.16) with $\alpha(x)$ of the form

$$\alpha(x) = \gamma x + (\text{nonlinear term}) \qquad (5.19)$$

for $\gamma > 0$. We are now satisfied[10,11] in asymptotic evaluation of relaxation in the small diffusion limit $\epsilon \to 0$, namely the limit of small $\mathscr{L}_{\text{diff}}$. However, if we neglect the diffusion term $\mathscr{L}_{\text{diff}}$ from the beginning in our unstable situation, then nothing happens.[10,11] The diffusion term is essential in the early state (or initial regime) of relaxation as was emphasized in the scaling theory.[10,11]

On the other hand, the nonlinear part of the drift term $\mathscr{L}_{\text{drift}}$ becomes more and more important, as time t increases, because it gives a correct saturation effect. Thus, both the diffusion term and the nonlinear part of the drift term play essential roles in the relaxation from or near the instability point. From the above arguments,[10,11] however, there is a possibility to treat separately the above two terms in some appropriate limit.

In order to find what is our desired limit,[10,11] we first discuss the *linear* case (5.1) by neglecting the nonlinear terms in (5.16) or (5.19). Then, the solution of (5.1) is given by (5.4) or

$$P_{(l)}(x,t) = \exp\left\{(e^{2\gamma t}-1)\left(\frac{\epsilon}{2\gamma}\right)\frac{\partial^2}{\partial x^2}\right\} \times \exp\left(-\gamma t\frac{\partial}{\partial x}x\right)P_{(l)}(x,0) \tag{5.20}$$

from Formula 1 in (4.1). If the initial system is just at the instability point $x=0$, namely $P(x,0)=\delta(x)$, then the first drift operator in (5.20) has no effect. Thus, we obtain

$$P_{(l)}(x,t) = \exp\left\{\epsilon e^{2\gamma t}(1-e^{-2\gamma t})\left(\frac{1}{2\gamma}\right)\frac{\partial^2}{\partial x^2}\right\}\delta(x). \tag{5.21}$$

Therefore, in the limit of large t, we have

$$P_{(l)}(x,t) \simeq P_{(l)}^{(\mathrm{sc})}(x,\epsilon e^{2\gamma t}) \equiv \exp\left(\epsilon e^{2\gamma t}\left(\frac{1}{2\gamma}\right)\frac{\partial^2}{\partial x^2}\right)\delta(x). \tag{5.22}$$

This form is also found directly from the explicit solution (5.4) with $P(y,0)=\delta(y)$. It should be remarked here that (5.21) is the exact solution of (5.1) and consequently that it contains completely the linear drift effect through the decomposition formula 1. That is, the second modified diffusion factor in (5.20) plays a role of *renormalized diffusion effect* due to the linear drift term. The function $\tilde{f}(2\gamma t)\equiv(e^{2\gamma t}-1)/(2\gamma t)$ yields the renormalization of time in the decomposition formula 1.

One of the most remarkable features of the above solution is that it has a scaling property in the sense that it is invariant asymptotically for the following transformation

$$\epsilon\to\epsilon' = b\epsilon \quad\text{and}\quad t\to t' = t-(1/2\gamma)\log b. \tag{5.23}$$

From this consideration, it is expected that our desired limit is the *scaling limit*[10,11,28] that

$$\text{sc}-\lim \equiv \lim_{\epsilon\to 0}\lim_{t\to\infty} \quad \text{for } \tau \text{ fixed}, \tag{5.24}$$

where

$$\tau = \epsilon e^{2\gamma t}. \tag{5.25}$$

Even in the general nonlinear case, this scaling property is expected to hold in the sense that the solution $P(x,t)$ of (5.16) becomes a *finite* function only of the scaling variable τ in the above scaling limit (5.25).

Then, we discuss here the nonlinear case from our viewpoint of scaling on the basis of our algebraic method, namely the decomposition formulas presented in Sec. IV. The formal solution of (5.16) is also expressed as

$$P(x,t) = e^{A+B+C}P(x,0), \tag{5.26}$$

where

$$A = -t\frac{\partial}{\partial x}\gamma x, \quad B = t\epsilon\frac{\partial^2}{\partial x^2},$$
$$C = -t\frac{\partial}{\partial x}(\alpha(x)-\gamma x). \tag{5.27}$$

As is seen from the above scaling argument in the linear case, one substantial aspect of the scaling theory is the separation of procedure to solve nonlinear systems with random force. This separation is performed algebraically in the present formulation. The above consideration on the linear case suggests the following asymptotic separation[8]:

$$\text{sc}-\lim e^{A+B+C} = e^{A+C}e^{f(\alpha)B}, \tag{5.28}$$

where

$$f(\alpha) = (1-e^{-\alpha})/\alpha \quad\text{and}\quad [A,B] = \alpha B. \tag{5.29}$$

Here $\alpha = 2\gamma t$ in the present problem.

In order to confirm the validity of this asymptotic separation (5.28), we give first the following systematic expansion formula.

Formula 6: When $[A,B]=\alpha B$, we have

$$\exp(A+C+\lambda B) = H(\lambda)e^{A+C}e^{\lambda f(\alpha)B}, \tag{5.30}$$

where

$$H(\lambda) = \exp_+\int_0^\lambda G(\mu)d\mu, \tag{5.31}$$

and

$$G(\lambda) = \int_0^1 e^{s(A+C+\lambda B)}Be^{-s(A+C+\lambda B)} \times\{1-2f(\alpha)\delta(s-1)\}ds. \tag{5.32}$$

In particular, the first-order term of $H(\lambda)$ is given by

$$G(0) = \epsilon\int_0^1 e^{s(A+C)}\frac{\partial^2}{\partial x^2}e^{-s(A+C)} \times\{1-2f(\alpha)\delta(s-1)\}ds. \tag{5.33}$$

Here, the exponential factor $\exp[s(A+C)]$ is a drift operator of the form

$$\exp\{s(A+C)\}\cdot P(x) = \exp\left\{-st\frac{\partial}{\partial x}\alpha(x)\right\}P(x) = \{\alpha(\xi(x,t))/\alpha(x)\}P(\xi(x,t)), \tag{5.34}$$

where[11]

$$\xi(x,t) = F^{-1}(e^{-\gamma t}F(x))$$

and

$$F(x) = \exp\int_{a_0}^x \frac{\gamma}{\alpha(y)}dy. \tag{5.35}$$

Thus, we find that $[G(0)/\epsilon]$ is bounded if $\alpha(x)$ contains a nonlinear term by which the system approaches a *finite* stationary state. In general, $[G(\lambda)/\epsilon]$ is shown to be bounded in the same situation. Thus, we arrive at the conclusion that the asymptotic separation (5.28) is valid in the scaling limit (5.24). That is, the desired scaling solution for (5.16) is given by[10,11]

$$P^{(\mathrm{sc})}(x,t)=e^{t\mathscr{L}_{\mathrm{drift}}}e^{tf(2\gamma t)\mathscr{L}_{\mathrm{diff}}}P(x,0). \tag{5.36}$$

Each operator in (5.36) can be manipulated analytically and consequently the scaling solution $P^{(\mathrm{sc})}(x,t)$ is explicitly found for a specific drift function $\alpha(x)$ and for any unstable initial distribution such as

$$P(x,0)=\frac{1}{\sqrt{2\pi\epsilon\sigma_0}}\exp\Big(-\frac{(x-\delta)^2}{2\epsilon\sigma_0}\Big). \tag{5.37}$$

According to the expression (5.36), the renormalized diffusion effect is essential in the initial regime. The nonlinear drift term becomes dominant in the second nonlinear regime, as was already discussed in a different method.[10,11]

The validity condition of the scaling theory on the initial condition is given by

$$\delta^2\lesssim\epsilon \quad \text{and} \quad \sigma_0=O(1), \tag{5.38}$$

as is seen from the above arguments. The condition (5.38) assures that the initial system is located essentially at the unstable point. When the above instability condition (5.38) is not satisfied, the ordinary Ω-expansion method,[28–30] namely the perturbational expansion around the deterministic solution, is valid.

C. Scaling theory for Fokker–Planck equations with time-dependent coefficients

Next we try to extend the above arguments to a time-dependent Fokker–Planck equation of the form

$$\frac{\partial}{\partial t}P(x,t)=\Big(-\frac{\partial}{\partial x}\alpha(x,t)+\epsilon(t)\frac{\partial^2}{\partial x^2}\Big)P(x,t). \tag{5.39}$$

In fact, any realistic nonequilibrium situation will be described by (5.39) more appropriately than (5.16), because it takes a finite time to make the relevant system quenched, for example, from above the critical point to below the critical point. This situation has been discussed already by many authors.[31–33]

Now the formal scaling solution of (5.39) is given by

$$P^{(\mathrm{sc})}(x,t)=\exp_+\int_0^t\mathscr{L}_{\mathrm{drift}}(s)ds \times\exp\Big(D(t)\frac{\partial^2}{\partial x^2}\Big)\cdot P(x,0) \tag{5.40}$$

from Formula 3 in (4.10) similarly to (5.36), where

$$\mathscr{L}_{\mathrm{drift}}(t)=-\frac{\partial}{\partial x}\alpha(x,t), \tag{5.41}$$

and

$$D(t)=\int_0^t\epsilon(s)\exp\Big(-2\int_0^s\gamma(u)du\Big)ds. \tag{5.42}$$

Here $\gamma(t)$ denotes a time-dependent growing rate defined by

$$\alpha(x,t)=\gamma(t)x+(\text{nonlinear term}). \tag{5.43}$$

Therefore, when the initial distribution is given by (5.37), the scaling solution of (5.39) is expressed by

$$P^{(\mathrm{sc})}(x,t)=\exp_+\Big(-\int_0^t\frac{\partial}{\partial x}\alpha(x,s)ds\Big)P_G(x,t), \tag{5.44}$$

where

$$P_G(x,t)=\{2\pi(\epsilon\sigma_0+2D(t))\}^{-1/2}\times\exp\Big(-\frac{x^2}{2(\epsilon\sigma_0+2D(t))}\Big). \tag{5.45}$$

In particular, for the case of the laser model[10,11,30] described by

$$\alpha(x,t)=\gamma(t)x-g(t)x^3, \tag{5.46}$$

the scaling solution is given by

$$P^{(\mathrm{sc})}(x,t)=\frac{1}{\sqrt{2\pi\tau(t)}}\Big\{1-2x^2e^{-\beta(t)}\int_0^t g(s)e^{\beta(s)}ds\Big\}^{-3/2}\times\exp\Big\{-\frac{x^2}{2\tau(t)}\Big(1-2x^2\int_0^t g(s)e^{\beta(s)-\beta(t)}ds\Big)^{-1}\Big\}, \tag{5.47}$$

where

$$\tau(t)=(\epsilon\sigma_0+2D(t))\exp(\beta(t)) \tag{5.48}$$

and

$$\beta(t)=2\int_0^t\gamma(s)ds. \tag{5.49}$$

Now, one of the most interesting physical quantities in the present problem is the *onset time* t_0 at which a macroscopic order or structure begins to appear, and which is determined, in the present problem, by

$$(\epsilon\sigma_0+2D(t_0))\exp\beta(t_0)\simeq1. \tag{5.50}$$

In the special limit, we consider first the time-independent case that $\gamma(t)=\gamma$, $g(t)=g$, and $\epsilon(t)=\epsilon$, namely sudden quenching. Then, the onset time t_0 is given, from (5.50), by

$$t_0\simeq(1/2\gamma)\log[\epsilon(\sigma_0+1/\gamma)]^{-1}, \tag{5.51}$$

as was already discussed in the previous papers by the present author.[10,11] Namely, the onset time becomes larger and larger, as the strength of the random force ϵ and the intial fluctuation σ_0 become smaller and smaller. This was called[10,11] synergism of the initial fluctuation and the random force.

Next we consider the shift of the onset time Δt_0 due to the time-dependent growing rate $\gamma(t)$. As was discussed by Weidlich and Haag,[31] by Nozieres and Saint-James,[32] and by Wong,[33] we consider the case of a linear dependence of $\gamma(t)$ on time t as

$$\begin{aligned}\gamma(t)&=-\gamma_0, && \text{for} \quad t\le0,\\ \gamma(t)&=vt-\gamma_0, && \text{for} \quad 0\le t\le T,\\ \gamma(t)&=\gamma_e, && \text{for} \quad T\le t,\end{aligned} \tag{5.52}$$

with $v=(\gamma_0+\gamma_e)/T$. Then, we obtain the shift of onset time, Δt_0, in the form

$$\Delta t_0=t_0-t_0(T=0)=[(\gamma_0+\gamma_e)/(2\gamma_e)]\ T, \tag{5.53}$$

for $t_0\gg T$. Thus, the shift of onset time Δt_0 becomes larger, as γ_0 and T increase, as it should be. These results agree well with those obtained by other authors.[31–33]

D. Global approximation method (GAM)

Finally we try to extend the scaling theory to find a global solution[8] of the system which is valid in the whole region of time. Our idea is to make use of decomposition formulas 1,2, and 5.

We consider Eq. (5.16). Formula 1 suggests the following approximation:

$$P^{(\mathrm{GAM})}(x,t) = \exp\left(\lambda(t)\tilde{f}(2\gamma t)t\epsilon\frac{\partial^2}{\partial x^2}\right)\exp\left(-t\frac{\partial}{\partial x}\alpha(x)\right)\times\exp\left((1-\lambda(t))f(2\gamma t)t\epsilon\frac{\partial^2}{\partial x^2}\right)P(x,0), \tag{5.54}$$

for the nonlinear drift $\alpha(x)$. The validity of this asymptotic evaluation is confirmed and is shown to agree with the scaling solution in the scaling regime. Our problem is to show that the above solution is optimized by adjusting the arbitrary function $\lambda(t)$ in (5.54).

The criteria for determining $\lambda(t)$ are the following.

(A) The fluctuation around the stationary point $x = x_{\mathrm{st}}$ is given correctly up to the order of ϵ in the limit $t\to\infty$.

(B) The function $\lambda(t)$ is as simple as possible.

First we note that the distribution function $\tilde{P}(x,t)$ defined by

$$\tilde{P}(x,t) = \exp\left(-t\frac{\partial}{\partial x}\alpha(x)\right)\times\exp\left\{(1-\lambda(t))f(2\gamma t)t\epsilon\frac{\partial^2}{\partial x^2}\right\}P(x,0) \tag{5.55}$$

approaches a delta function (or sum of delta functions) around $x = x_{\mathrm{st}}$ (if there are several stationary points). Therefore, the fluctuation $x = x_{\mathrm{st}}$ of the solution $P^{(\mathrm{GAM})}(x,t)$ in (5.54) for $t\to\infty$ is given by

$$\langle(x-x_{\mathrm{st}})^2\rangle = \lim_{t\to\infty}\{2\lambda(t)\tilde{f}(2\gamma t)t\epsilon\} \tag{5.56}$$

from the expression (5.54), by using the formula (5.6). Now we assume that $\langle(x-x_{\mathrm{st}})^2\rangle = \epsilon\sigma_e$. Then, we have such a condition on $\lambda(t)$ as

$$\lambda(t)(e^{2\gamma t}-1)/\gamma \simeq \sigma_e \tag{5.57}$$

for large t. The simplest form of $\lambda(t)$ to satisfy (5.57) is

$$\lambda(t) = \sigma_e\gamma e^{-2\gamma t}. \tag{5.58}$$

Thus, we obtain the global approximation solution $P^{(\mathrm{GAM})}(x,t)$ given by (5.54) with (5.58), namely

$$P^{(\mathrm{GAM})}(x,t) = \exp\left(\tilde{\epsilon}(t)\frac{\partial^2}{\partial x^2}\right)\exp\left(-t\frac{\partial}{\partial x}\alpha(x)\right)\times\exp\left(\epsilon(t)\frac{\partial^2}{\partial x^2}\right)P(x,0), \tag{5.59}$$

where

$$\epsilon(t) = (\epsilon/2\gamma)(1-e^{-2\gamma t})(1-\sigma_e\gamma e^{-2\gamma t}) \tag{5.60}$$

and

$$\tilde{\epsilon}(t) = \tfrac{1}{2}\epsilon\sigma_e(1-e^{-2\gamma t}). \tag{5.61}$$

As a simple application of our global approximation method, we discuss the laser model $\alpha(x) = \gamma x - gx^3$. This system has two stationary states $x_{\mathrm{st}} = \pm(\gamma/g)^{1/2}$ and the fluctuation around them is given by $\epsilon\sigma_e = \epsilon/(2\gamma)$. Now the global approximate solution of this system to give correctly the stationary fluctuation for the initial Gaussian distribution (5.37) with $\delta = 0$ is given by[8]

$$P^{(\mathrm{GAM})}(x,t) = \{2\pi\epsilon\sigma(t)\}^{-1/2}\int_{-\Delta}^{\Delta}\exp\left(-\frac{(\xi-x)^2}{2\epsilon\sigma(t)}\right)\times[1/(2\pi\tau)^{1/2}](1-2g\sigma(t)\xi^2)^{-3/2}\times\exp\left(-\frac{\xi^2}{2\tau(1-2g\sigma(t)\xi^2)}\right)d\xi, \tag{5.62}$$

where $\sigma(t) = (1-e^{-2\gamma t})\sigma_e$, $\Delta = \{2g\sigma(t)\}^{-1/2}$, and

$$\tau = \tau(t)e^{2\gamma t} \equiv \epsilon\{\sigma_0 + (2-e^{-2\gamma t})\sigma(t)\}\exp(2\gamma t). \tag{5.63}$$

The above result (5.62) is rewritten as

$$P^{(\mathrm{GAM})}(x,t) = \frac{1}{2\pi(\epsilon\sigma(t)\tau(t))^{1/2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{s^2}{2\tau(t)} - \frac{1}{2\epsilon\sigma(t)}\left(x-\frac{s}{(e^{-2\gamma t}+n(t)s^2)^{1/2}}\right)^2\right\}ds, \tag{5.64}$$

where $n(t) = 2g\sigma(t)$. This is our desired global solution. In order to see how it approaches the correct stationary state, we separate it into the following two parts:

$$P^{(\mathrm{GAM})}(x,t) = P_{+}^{(\mathrm{GAM})}(x,t) + P_{-}^{(\mathrm{GAM})}(x,t), \tag{5.65}$$

where each part is defined by the positive and negative regions of the integral (5.64) with respect to s, respectively. In the limit of large t, $P_{+}^{(\mathrm{GAM})}(x,t)$ is, for example, calculated explicitly as

$$P_{+}^{(\mathrm{GAM})}(x,t) \simeq \frac{\exp\{-(x-x(t))^2/2\epsilon\sigma(t)\}}{2\pi(\epsilon\sigma(t)\tau(t))^{1/2}}\times\int_0^{\infty}\exp\left\{-\frac{s^2}{2\tau(t)} - \frac{(x-x(t))x^3(t)}{2\epsilon\sigma(t)\exp(2\gamma t)}\frac{1}{s^2}\right\}ds$$

$$= \frac{1}{2(2\pi\epsilon\sigma(t))^{1/2}}\exp\left\{-\frac{(x-x(t))^2}{2\epsilon\sigma(t)} - \left[\frac{(x-x(t))x^3(t)}{\epsilon\sigma(t)\tau(t)\exp(2\gamma t)}\right]^{1/2}\right\}$$

$$\Rightarrow P_{\mathrm{eq}}^{+}(x) \equiv \frac{1}{2}\frac{1}{\sqrt{2\pi\epsilon\sigma_e}}\exp\left\{-\frac{(x-x_{\mathrm{st}})^2}{2\epsilon\sigma_e}\right\}, \tag{5.66}$$

for $x > x(t)$, where $x(t) = n(t)^{-1/2}$ and we have used the following integral formula:

$$\int_0^{\infty}e^{-ax^2-b/x^2}dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}e^{-2(ab)^{1/2}} \tag{5.67}$$

for $a > 0$ and $b > 0$. Similarly, $P_{-}^{(\mathrm{GAM})}(x,t)$ is also shown to approach $P_{\mathrm{eq}}^{(-)}(x)$ correctly for $t\to\infty$.

The present result (5.64) obtained by GAM agrees with Weiss' result[34] derived by using the path integral formulation.

The above formulation can be extended to Fokker–Planck equations with time-dependent coefficients, (5.39), as

$$P^{(\mathrm{GAM})}(x,t) = \exp\left\{\int_0^t(1-\lambda(s,t))\epsilon(s)\exp\left(2\int_s^t\gamma(u)du\right)ds\,\frac{\partial^2}{\partial x^2}\right\}$$

$$\times \exp_+ \left\{ -\int_0^t \frac{\partial}{\partial x} \alpha(x,s) ds \right\}$$
$$\times \exp \left\{ \int_0^t \lambda(s,t) \epsilon(s) \exp\left(-2 \int_0^s \gamma(u) du \right) ds \frac{\partial^2}{\partial x^2} \right\}, \tag{5.68}$$

from Formula 5.

VI. SUMMARY AND DISCUSSION

In the present paper, we have given some decomposition formulas of exponential operators and Lie exponentials and they have been applied to study the relaxation and fluctuation from or near the instability point. A global approximation method of transient phenomena near the instability point has been formulated on the basis of decomposition formulas into three parts. An application to the laser model has been presented in detail.

Here it should be remarked that spin operators belong clearly to a Banach algebra, but that differential operators such as $\mathscr{L}_{\text{diff}}$ in Sec. V do not belong to a Banach algebra in a strict sense. However, if we confine operands (namely the distribution function in the case of the Fokker–Planck equation) into functions $\{P(x)\}$ which decrease rapidly for $x = \pm\infty$, then our decomposition formulas given in the present paper are still valid.

Some applications to other phenomena such as combustion[35,36] and to nonuniform systems will be reported elsewhere.

ACKNOWLEDGMENTS

The present author would like to thank Professor R. Kubo, Professor J. R. Klauder, Professor G. Nicolis, Mrs. H. Tasaki, T. Ikegami, and M. Katori for useful discussion.

This work is partially financed by the Scientific Research Fund of the Ministry of Education.

[1] R. M. Wilcox, J. Math. Phys. **8**, 962 (1967).
[2] R. F. Snider, J. Math. Phys. **5**, 1586 (1964), Appendix B.
[3] M. Suzuki, Prog. Theor. Phys. **53**, 1657 (1975).
[4] R. Kubo, J. Phys. Soc. Jpn. **17**, 1100 (1962).
[5] M. Suzuki, Phys. Rev. B (in press).
[6] H. F. Trotter, Proc. Am. Math. Soc. **10**, 545 (1959).
[7] M. Suzuki, S. Myashita, and A. Kuroda, Prog. Theor. Phys. **58**, 1377 (1977).
[8] M. Suzuki, Physica A **117** 103 (1983). This is a brief report of the idea of the present study. See also M. Suzuki, in *Statistical Physics and Chaos in Fusion Plasmas*, edited by C. W. Horton and L. E. Reichl (Wiley, New York, 1984), p. 311.
[9] W. Miller, Jr., "*Symmetry and Separation of Variables*," in *Encylopedia of Mathematics and its Applications*, Vol. 4, (Addison–Wesley, Reading, MA, 1977).
[10] M. Suzuki, Prog. Theor. Phys. **56**, 77, 477 (1976); J. Stat. Phys. **16**, 11, 477 (1977).
[11] M. Suzuki, Adv. Chem. Phys. **46**, 195 (1981).
[12] R. Kubo, J. Math. Phys. **4**, 174 (1963).
[13] M. Suzuki, Prog. Theor. Phys. **69**, 160 (1980).
[14] M. Suzuki, J. Stat. Phys. **20**, 163 (1979). The second equation of Eq. (30) in this reference would be corrected as in the text. By the way, Eq. (38) in this reference is completely wrong, and consequently an explicit condition for the boundedness of the temporal evolution operator $\Gamma(r,t)$ is still an open question.
[15] R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957).
[16] J. E. Hirsch, R. L. Sugar, D. J. Scalapino, and R. Blankenbecler, Phys. Rev. B **26**, 5033 (1982) (Footnote 9).
[17] H. De Raedt and B. De Raedt, Phys. Rev. A **28** 3575 (1983); H. De Raedt and A. Lagendijk, Phys. Rev. Lett. **46**, 77 (1981) and **49** 1552 (1982); Phys. Rev. B **24** 463 (1981).
[18] M. Suzuki, Commun. Math. Phys. **51**, 183 (1976) and **57** 193 (1977); Prog. Theor. Phys. **56**, 1454 (1976).
[19] J. J. Cullen and D. P. Landau, Phys. Rev. B **27**, 297 (1983).
[20] A. Wiesler, Phys. Lett. A **89**, 359 (1982); J. W. Lyklema, Phys. Rev. Lett. **49**, 88 (1982); Phys. Rev. B **27**, 3108 (1983).
[21] M. Barma and B. S. Shastry, Phys. Rev. B **18**, 3351 (1978).
[22] J. E. Hirsch, D. J. Scalapino, R. L. Sugar, and R. Blankenbecler, Phys. Rev. Lett. **47**, 1628 (1981); Phys. Rev. B **26**, 5033 (1982).
[23] B. DeRaedt and H. De Raedt, Phys. Rev. B **10**, 5325 (1984).
[24] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
[25] P. Hänggi and H. Thomas, Z. Phys. B **22**, 295 (1975).
[26] F. Fu-kang and J. Lu, Commun. Theor. Phys. (Beijing) **2**, 1481 (1983).
[27] I. An, S. Chen, and H-Y. Guo, Physica A **128**, 520 (1984).
[28] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1983).
[29] N. G. van Kampen, Can. J. Phys. **39**, 551 (1961).
[30] R. Kubo, K. Matsuo, and K. Kitahara, J. Stat. Phys. **9**, 51 (1973).
[31] W. Weidlich and G. Haag, Z. Phys. B **39**, 81 (1980).
[32] P. Nozieres and D. Saint-James, Prog. Theor. Phys. Suppl. **69**, 122 (1980).
[33] Y. M. Wong and P. H. E. Meijer, Phys. Rev. A **26**, 611 (1982).
[34] U. Weiss, *Chaos and Order in Nature*, edited by H. Haken (Springer, Berlin, 1981), p. 77; Phys. Rev. A **25**, 2444 (1982). See also B. Caroli, C. Caroli, and B. Roulet, J. Stat. Phys. **26**, 83 (1981).
[35] M. Frankowicz, M. Marek Mansour, and G. Nicolis, Physica A **125**, 237 (1984).
[36] G. Nicolis, F. Baras and M. Malek Mansour, in *Chemical Instability*, edited by G. Nicolis and F. Baras (Reidel, Dordrecht, 1984).

N-dimensional spinors: Their properties in terms of finite groups

H. W. Braden

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, United Kingdom

(Received 13 March 1984; accepted for publication 10 September 1984)

A classification scheme is presented for the finite multiplicative group generated by the gamma matrices associated with a given Clifford algebra. This group reflects the periodicities observed in n-dimensional spinors, and its representations and other properties are studied, thus highlighting the dependence on the number of spacelike and timelike vectors. The reality of the representations is examined and tabulated; application is made to the imposition of Majorana and Weyl conditions.

I. INTRODUCTION

In the study of the Dirac and wave equations on a given space-time, Clifford algebras[1,2] play a fundamental role. Accordingly, they (and the associated study of spinors) have been the subject of many investigations (Good[3] gives an early review). The purpose of this paper is to study these algebras and spinors by the study of an associated finite group. Such a line of investigation is not new, but before reviewing this work, a description of this group (made precise in Sec. II) is in order.

Associated with a Clifford algebra is a natural finite group,[4] the multiplicative group, generated by the gamma matrices representing the algebra. For example, the Pauli σ matrices generate a group, abstractly the quaternion group. This group, though finite, is large enough to reflect the representation structure of the underlying group of metric-preserving transformations and the related spin groups we are interested in. Indeed, the nontrivial representations of this group generate the fundamental spin representations of the underlying orthogonal group. The connection is as follows. A rotation can always be expressed as a product of reflections. The group we are considering is the double group of the (abelian) group of reflections in the coordinate axes. Such finite groups have a very specific structure; they are extraspecial groups or their central extensions (described in Sec. II). These finite groups provide a useful and simple method for studying n-dimensional spinors and properties of Clifford algebras.

This underlying finite group has been mentioned or used in passing by a variety of authors. Eddington[5] noted the four-dimensional case's relation to the group of collineations of Kummer's quartic surface. Boerner[6] made use of them when describing the spin representations of the orthogonal groups.[7] Remakrishnan encountered them in a program he calls "L-Matrix theory," and tried generalizing the group to describe para-Fermi statistics.[8] More recently, Salingaros encountered these groups when putting a group structure on the differential forms of a given space-time,[9] and he later identified their connection with extraspecial groups.[10] (This contains some errors, noted later.)

The same group we are describing also arises in a variety of other contexts. First, a uniform description of the double groups of the (finite) reflection groups has been developed.[11] Here the double group is generated directly in terms of a Cartan matrix and these extraspecial groups. This generalizes Schur's[12] work on the double group of the symmetric group and has significance for those interested in lattice symmetries.[13] Second and more recently, this group has arisen when constructing vertex operators associated with a Kac–Moody algebra.[14] In this regard, the method of classification developed later is particularly useful. These groups also appeared in the classification of the finite simple groups.[15]

Despite their ubiquity, it appears that a detailed description of these groups is lacking in the mathematical physics literature. The purpose of this account is to provide a new direct classification of the group associated with an algebra of given dimension and signature. In so doing we shall also review many of the known properties of these groups and their representations. These results will be applied to the known representation theory of Clifford algebras and spinors, so obtaining results which are frequently used in supersymmetric calculations today.[16]

An outline of the paper is as follows. In Sec. II we associate a group to a Clifford algebra with a given dimension and signature. The elementary properties of this group are given, and it is shown the group is one of five types; it is extraspecial or a central product of one of these. This section is mostly review. Section III answers the question: which group is to be associated with a specific algebra. The techniques of this section are new and improve the existing enumerative method of classification. It is this section that is particularly relevant to the discussion of vertex operators. Having now associated a particular group to a given signature and dimension, Sec. IV discusses the representations of these groups. Many known results are drawn upon in this section; some old results are slightly extended so the reality properties and orders of the group elements may be discussed; the tensor products of representations are described. Section V gives some application of these results, showing how the Majorana and Weyl restrictions on spinors may be imposed. Section VI is a brief conclusion.

II. CONSTRUCTION AND PROPERTIES OF THE GROUP ASSOCIATED WITH THE ALGEBRA

In this section we shall associate a group with a Clifford algebra over a space-time with given signature. We shall identify the group as one of five types and in the next section determine which of these groups is associated with a particular signature and dimension. First, we define an algebra over the field F by the anticommutation relations

 0022-2488/85/040613-08$02.50

$$e_i^2 = \eta_{ii}, \quad \eta_{ii} \in F, \tag{1a}$$

$$e_i e_j + e_j e_i = 0, \quad i \neq j, \quad 1 \leq i, \quad j \leq n. \tag{1b}$$

If the field is of characteristic different from (2), (char $F \neq 2$) these may be expressed equivalently by

$$e_i e_j + e_j e_i = \{e_i, e_j\} = 2\eta_{ii}\delta_{ij} \equiv 2g_{ij}. \tag{2}$$

When all the g_{ij} are zero we have the Grassmann, or exterior, algebra; with $\eta_{ii} = \pm 1$, we have a Clifford algebra. Let

$$\eta_{ii} = +1, \quad i \leq r, \tag{3a}$$

$$\eta_{ii} = -1, \quad r < i \leq n. \tag{3b}$$

We label this algebra $C^{r,s}$, with $s = n - r$. We may think of n here as the dimension of the underlying space-time, and s the number of spacelike dimensions.

Associated with this Clifford algebra is a finite group, $G^{r,n-r}$; loosely, this is the group of products of the units of the algebra. The abstract definition of this group is

$$\begin{aligned} G^{r,n-r} &\equiv \langle \Gamma_1, \Gamma_2, \ldots, \Gamma_n, -1 \,|\, 1 = (-1)^2 = \Gamma_j^2, \quad j \leq r; \\ &-1 = \Gamma_k^2, \quad r < k \leq n, \\ &[\Gamma_i, \Gamma_j] = -1, \quad i \neq j, \quad [\Gamma_i, -1] = 1 \rangle. \end{aligned} \tag{4}$$

The group is written multiplicatively with identity 1 and group commutator $[x,y] = xyx^{-1}y^{-1}$. (The same group is obtained via a differential geometric construction by Salingaros.[9,10]) When no confusion arises, we will drop the superscripts from G. The group G is realized as the multiplicative group of products (and their negatives) of the gamma matrices which represent a given algebra.

We proceed now to identify the class of groups G we are dealing with. This is done by determining the conjugacy classes of G, its center, and commutator. Let $g = \Gamma_{i_1}\Gamma_{i_2}\cdots\Gamma_{i_d} \equiv \Gamma_{i_1 i_2, \ldots, i_d}$, with $i_1 < i_2 < \cdots < i_d$. There are $\binom{n}{d}$ such possible g; including their negatives, there are $2\binom{n}{d}$. Thus the order of G is $|G| = 2^{1+n}$. The conjugacy classes of G may be seen from the following identities:

$$\Gamma_i g \Gamma_i^{-1} = (-1)^d g \quad \text{if } i \neq i_j, \quad j = 1, \ldots, d; \tag{5a}$$

$$\Gamma_i g \Gamma_i^{-1} = (-1)^{d-1} g \quad \text{if } i = i_j, \quad \text{some } j = 1, \ldots, d. \tag{5b}$$

The elements g and $-g$ are therefore in the same conjugacy class unless n is both odd and $d = n$. Only in this latter case will g (and also $-g$) be self-conjugate (like -1 and 1), and so in the center of $G, Z(G)$. As there are $\binom{n}{d}$ distinct elements $\Gamma_{i_1 i_2, \ldots, i_d}$, there are then $2^n + 1$ conjugacy classes of G for n even, and $2^n + 2 = 2(2^{n-1} + 1)$ conjugacy classes for n odd.

To identify the structure of the center of G in the case n odd define

$$\Delta = \Gamma_1 \Gamma_2 \cdots \Gamma_n. \tag{6}$$

Then

$$\Delta^2 = (-1)^{n(n+1)/2 - r}. \tag{7}$$

We have, then, the following cases for the center of G:

$$Z(G) = \begin{cases} \langle -1 \rangle \cong C_2, & n \text{ even}, \\ \langle -1, \Delta \rangle \cong & \begin{cases} \langle -1 \rangle \times \langle \Delta \rangle \cong C_2 \times C_2, \\ \quad \text{if } n \text{ odd and} \\ \quad n(n+1)/2 - r \text{ even}, \\ = \langle \Delta \rangle \cong C_4, \text{ if } n \text{ odd and} \\ \quad n(n+1)/2 - r \text{ odd}. \end{cases} \end{cases}$$

Here C_k denotes the cyclic group of k elements.

Using the results of Eq. (5) we find the following:

$$G' = [G, G] = \langle -1 \rangle, G/G' = C_2 \times C_2 \times \cdots \times C_2, \quad n \text{ times}.$$

Thus G/G' is elementary abelian and hence $G' = \Phi(G)$, where $\Phi(G)$ is the Frattini subgroup[17] of G, the intersection of all the maximal subgroups of G.

This structure to G is quite restrictive and specifies G. We recall the following definition and lemma.[17,18]

Definition 1: A group is a p group if every element of the group except the identity has order a power of the prime p. A p group E, such that $|E| = p^{1+2k}$ is known as extraspecial if $E' = \Phi(G) = Z(E) = C_p$.

Lemma 2: Let E be an extraspecial subgroup of the p group P, such that $[P,E] \leq Z(E)$. Then $P = E \circ C_p(E)$.

Here (and throughout) $\circ$ means the central product of two groups: that is, the direct product with centers identified. (This is distinct from the Kronecker product, which is mistakenly described in Salingaros.[9,10] Also, $C_G(E)$ denotes the centralizer of E in G. In the case when $n = 2m$, our group G is extraspecial. Further, we have for $n = 2m + 1$,

$$[G,E] \leq [G,G] = G' = Z(E), \quad |E| = 2^{2m+1},$$

and so the conditions of the lemma are satisfied. Thus when n is odd the group G is a central extension of an extraspecial group. In particular

$$G = E \circ Z(G), \quad |G| = 2^{2m+2}, \quad |E| = 2^{2m+1}. \tag{8}$$

To complete the specification of G we need to know more about the extraspecial groups E. We recall[19,20] that there are two nonabelian p groups of order p^3, each being extraspecial. In our case, these are the dihedral and quaternion groups of order 8, denoted D and Q, respectively, throughout this paper. Further, every extraspecial p group is the central product of these two nonabelian p groups. If $P_1, \ldots, P_m$ are extraspecial p groups of order p^3 then, up to isomorphism, there is only one central product of $P_1, \ldots, P_m$, with center of order p. This is extraspecial of order p^{2m+1} and denoted $P_1 \circ P_2 \circ \cdots \circ P_m$. We can talk therefore about the central product of $P_1, \ldots, P_m$. How many different extraspecial groups of order p^{2m+1} are there? For the case $p = 2$ we are considering, the answer to this is given by the following theorem.[20]

Theorem 3: Let E_m be an extraspecial two-group of order 2^{2m+1}. Then there are two types of isomorphism classes of such groups, namely the following.

(a) E_{m+}, the central product of m dihedral groups D. This possesses maximal abelian normal subgroups of type $(\underbrace{4,2,\ldots,2}_{m-1})$ and $(\underbrace{2,2,\ldots,2}_{m+1})$:

(b) E_{m-}, the central product of $(m-1)$ dihedral groups D and one quaternion group Q. This possesses maximal abelian normal subgroups of type $(\underbrace{4,2,\ldots,2}_{m-1})$.

This theorem is readily proven once the following equivalences are established:

$$Q \circ C_{2^k} \cong D \circ C_{2^k}, \quad k > 1, \tag{9}$$

$$Q \circ Q \cong D \circ D, \tag{10}$$

$$Q \circ C_4 \cong D \circ C_4 \cong C_4 \times C_2 \times C_2. \tag{11}$$

Bringing these results together, the group G we are interested in belongs to one of the following five classes:

$$\text{if } n = 2m, \quad G = E_{m+} \quad \text{or } E_{m-}, \tag{12a}$$

if $n = 2m + 1$,

$$G = E_m \circ C_4 \text{ or } G = E_{m\pm} \circ (C_2 \times C_2) = E_{m\pm} \times C_2. \tag{12b}$$

Here the two classes of (12a) are those given by Theorem 3. The odd case uses (8)–(10) for the first possibility; the second two possibilities are shown by direct calculation. We note that when $n = 2m + 1$, the class of extraspecial factor does not matter if the center is C_4.

Thus far we have identified the groups G associated with a Clifford algebra as being one of five classes. We must now determine which class is associated with a given signature and algebra $C^{r,s}$. This is done in the following section.

III. CLASSIFICATION OF THE GROUPS

In the last section we associated a finite two-group with a Clifford algebra $C^{r,s}$. This group was one of five canonical forms (12) and the purpose of this section is to provide a simple way of determining which type of group we have, solely from the signature of the metric (2) used in the algebra's construction. The method of classification is based upon a quadratic form defined by the group. This method simplifies previous enumerations,[10] where the group was constructed from a knowledge of the orders of the elements. Another advantage of this method is that the periodicities associated with a Clifford algebra may be clearly seen.

The steps involved in this classification are as follows. First we shall use the group to define vector space and a quadratic and bilinear form, which act upon this. Then we shall decompose this vector space into a pairwise direct sum, noting how the quadratic form so changes. Associated with each pair we can attach a group, and overall we have their central product. We conclude the section with an example.

We now associate a quadratic form with our group. Because $\Phi(G) = G'$, the commutator quotient G/G' is elementary abelian and this may be naturally regarded as a vector space V^n over the field F_2 of two elements.[21] If $x = (\alpha_1^{i_1}, \alpha_2^{i_2}, \ldots, \alpha_n^{i_n}) \in G/G'$, then associated with this is the vector $(i_1, i_2, \ldots, i_n)$ where $i_j = 0,1$. Here, V^n is equipped with a quadratic form q and a bilinear form f given by (with $x,y,z \in G$)

$$q(x) = a, \quad \text{where } x^2 = C^a, \quad \langle C \rangle = G', \tag{13}$$

$$f(x,y) = b, \quad \text{where } [x,y] = C^b. \tag{14}$$

It is verified that q is well defined, and using

$$(xy)^p = X^p Y^p [x,y^{-1}]^{(1/2)p(p-1)}, \tag{15}$$

we get

$$q(xy) = q(x) + q(y) + f(x,y), \tag{16}$$

$$q(xyz) = q(x) + q(y) + q(z) + f(x,y) + f(x,z) + f(y,z). \tag{17}$$

These hold because $G' \leq Z(G)$ and so the following commutator relations are true.

Lemma 4: For $G' \leq Z(G)$

$$[x,xz] = [x,z], \tag{18a}$$

$$[x,yz] = [x,y][x,z], \tag{18b}$$

$$[x,xyz] = [x,y][x,z], \tag{18c}$$

$$[xy,xz] = [x,z][x,y][y,z]. \tag{18d}$$

From the defining relations (4) of G, we have

$$f(\Gamma_i,\Gamma_j) = 1, \quad i \neq j, \tag{19a}$$

$$q(\Gamma_i) = 1, \quad i > r. \tag{19b}$$

We now proceed to decompose the vector space pairwise, such that $V^n = V^2 \otimes V^{n-2}$, and see the effect on the quadratic form q and bilinear form f. Let $\Gamma_1, \Gamma_2, \ldots, \Gamma_n$ be the natural basis of V associated with the Γ_i. Then if we take as our basis of $V^{n-2}, \bar\Gamma_1 \bar\Gamma_2 \bar\Gamma_j, j = 3,\ldots,n$, the lemma gives

$$f(\bar\Gamma_1, \bar\Gamma_1\bar\Gamma_2\bar\Gamma_j) = f(\bar\Gamma_2, \bar\Gamma_1\bar\Gamma_2\bar\Gamma_j) = 0, \quad 3 \leq j \leq n, \tag{20a}$$

$$f(\bar\Gamma_1\bar\Gamma_2\bar\Gamma_j, \bar\Gamma_1\bar\Gamma_2\bar\Gamma_k) = f(\bar\Gamma_j, \bar\Gamma_k). \tag{20b}$$

Therefore the bilinear form f is left unchanged, while the quadratic form upon using (19) and (20a) becomes

$$q(\bar\Gamma_1\bar\Gamma_2\bar\Gamma_j) = q(\bar\Gamma_1) + q(\bar\Gamma_2) + q(\bar\Gamma_j) + f(\bar\Gamma_1,\bar\Gamma_2) + f(\bar\Gamma_2,\bar\Gamma_j) + f(\bar\Gamma_1,\bar\Gamma_j) \tag{21a}$$

$$= \begin{cases} q(\bar\Gamma_j) + 1, & \text{if } q(\bar\Gamma_1) = q(\bar\Gamma_2), \\ q(\Gamma_j), & \text{if } q(\bar\Gamma_1) \neq q(\bar\Gamma_2). \end{cases} \tag{21b}$$

Thus setting $(l,m) = (q(\bar\Gamma_1), q(\bar\Gamma_2))$, an initial (0,1) leaves the quadratic form on V^{n-2} unchanged, while (1,1) or (0,0) changes the remaining q's.

Continuing the procedure, we decompose V^n pairwise until we are left with no elements (n even) or a single element (n odd) remaining. According to the quadratic form on these orthogonal subspaces we may associate a group:

(i) (0,0) and (0,1) yield D,

(ii) (1,1) yield Q,

(iii) (0) gives $C_2 \times C_2 = V$,

(iv) (1) gives C_4.

These identifications come from our choice of signs for G, and by applying q and f in the manner described. Here, G is the central product of the groups that appear in this decomposition.

Some comments on this procedure are in order. First, the independence of the order of the basis of V (and so the order of the 0's and 1's in our precription) is reflected in the isomorphisms (10) and (11). Second, when n is even, the resulting groups are the two extraspecial two-groups. These correspond to the two possible nondegenerate quadratic forms[22] over F_2. We have[20]

(a) if $E_{m+} = D_1 \circ D_2 \circ \ldots \circ D_m$

and

$$D_i = \langle A_i, B_i \,|\, A_i^2 = B_i^4 = E, A_i B_i = B_i^{-1} A_i \rangle,$$

then

$$q(\bar A_1^{r_1} \bar B_1^{s_1} \bar A_2^{r_2} \bar B_2^{s_2} \ldots \bar A_m^{r_m} \bar B_m^{s_m}) = r_1 s_1 + r_2 s_2 + \cdots + r_m s_m.$$

(b) if $E_{m-} = D_1 \circ D_2 \circ \ldots \circ D_{m-1} \circ Q$

and

$$Q_m = \langle A_m, B_m \rangle,$$

then

$$q(\bar{A}_1^{r_1} \bar{B}_1^{s_1} \cdots \bar{A}_m^{r_m} \bar{B}_m^{s_m})$$
$$= r_1 s_1 + \cdots + r_{m-1} s_{m-1} + r_m^2 + r_m s_m + s_m^2.$$

The space V constructed from E_{m+}, E_{m-}, with f as its fundamental form, is a symplectic space. When n is odd we have a degenerate quadratic form. The last entry in this case corresponds to the Δ constructed earlier (6) which squares either to $+1$ or -1.

Third, we have from $D \circ D = Q \circ Q$, the periodicity

$$G^{p,q} = G^{p+4,q-4}, \tag{22}$$

where $G^{p,q}$ is the group corresponding to $C^{p,q}$. Finally, because combinations of four 0's or four 1's may be separated out without effecting the quadratic form on the remaining space, we never need consider combinations of more than three 0's or three 1's. All the possible groups are readily enumerated. We have

$$G^{p+4,q} = G^{p,q+4} = D \circ Q \circ G^{p,q}. \tag{23}$$

This modulo 4 periodicity changes the extraspecial group structure. There is the modulo 8 (or Bott periodicity) which preserves the extraspecial group structure:

$$G^{p+8,q} = G^{p,q+8} = D \circ D \circ D \circ D \circ G^{p,q}, \tag{24}$$

where $G^{p+8,q}$ and $G^{p,q}$ are of the same $(+,-)$ extraspecial group isomorphism class. Furthermore, we have the periodicity

$$G^{p,q} = G^{q+1,p-1}. \tag{25}$$

The periodicities [(22)–(25)] have long been known in the study of Clifford algebras (see, for example, Porteous[23] and references therein); in the context of these finite groups, Salingaros[9] also noted them. It is perhaps useful to note that these periodicities are by no means limited to Clifford algebras or these extraspecial groups. In the context we are working, they evidence a more general structure, the Brauer–Wall group.[24] For their application to Clifford algebras, we mention the work of Lounesto.[25] Collecting the results of this section, we have the following classification.

Lemma 5: The group $G^{r,n-r}(4)$ associated with the algebra $C^{r,n-r}$ is

(a) if $n = 2m$, it is E_{m+}, when $r - m \equiv 0,1 \bmod 4$,
E_{m-}, when $r - m \equiv 2,3 \bmod 4$;

(b) if $n = 2m+1$, it is $E_{m+} \times C_2$, when $r - m \equiv 1 \bmod 4$,
$E_{m-} \times C_2$, when $r - m \equiv 3 \bmod 4$,
$E_m \circ C_4$, when $r - m \equiv 0,2 \bmod 4$.

Table I gives the group structure associated with a given metric along with information about their representations, which is the subject of the next section. We conclude this section with an example.

Example: We determine the group structure of $G^{4,1}$ associated with a metric (2) $g_{ij} = \mathrm{diag.}\,(+\ +\ +\ +\ -)$. The quadratic form (13) is (0,0,0,0,1). Upon decomposition this gives

$$(0,0,0,0,1) = (0,0) \otimes (1,1,0) = (0,0) \otimes (1,1) \otimes (1),$$

which gives the group structure

$$G^{4,1} = D \circ Q \circ C = E_2 \circ C.$$

IV. REPRESENTATIONS

In this section we shall study the representations[26] of $G^{p,q}$. The representations of degree greater than one are representations of the algebra $C^{p,q}$; this enables the possible representations of the algebra—the gamma matrices—to be quickly classified. We firstly classify the irreducible representations of G and then describe their inductive construction, making contact with the work of the previous section. After this we derive the conditions for a representation to be either pure real or imaginary. The latter is of physical importance and will be used in the next section; it also will enable us to describe generally the group G in terms of the orders of its elements. We conclude this section by showing the group $G^{p,q}$ is simply reducible, reflecting the underlying spin group.

First we note that a representation of G can always be taken to be unitary, because G is finite. Thus the representation matrix of g is either Hermitian or anti-Hermitian according to whether $g^2 = 1$ or $g^2 = -1$. The number of linear or one-dimensional representations of G is given by

TABLE I. The group $G^{r,n-r} = \langle \Gamma_1,\ldots,\Gamma_n, -1 | \Gamma_i^2 = (-1)^2 = 1\ i \leq r; \Gamma_j^2 = -1\ j > r; [\Gamma_i,\Gamma_j] = -1, i \neq j, [\Gamma_i, -1] = 1 \rangle$ in terms of its extraspecial structure, and its representations. $r,i,-$ means there exists a pure real, imaginary, or only mixed representation. $E_\pm$ are the two distinct extraspecial groups of the appropriate order. $C \equiv C_4$. $V \cong C_2 \times C_2$. EC means the central product of E and C, $E \circ C$. $E_\pm \circ V \cong E_\pm \times C_2$.

$n-r$ \ n	1	2	3	4	5	6	7	8	9	10	11
0	$V\,r$	$E_+ r$	$EC-$	$E_-\,-$	$E_- V-$	$E_- i$	$EC\,i$	$E_+ r$	$E_+ V r$	$E_+ r$	$EC-$
1	$C\,i$	$E_+ r,i$	$E_+ V r$	$E_+ r$	$EC-$	$E_-\,-$	$E_- V-$	$E_- i$	$EC\,i$	$E_+ r$	$E_+ V r$
2		$E_-\,i$	$EC\,i$	$E_+ r,i$	$E_+ V r$	$E_+ r$	$EC-$	$E_-\,-$	$E_- V-$	$E_- i$	$EC\,i$
3			$E_- V-$	$E_- i$	$EC\,i$	$E_+ r,i$	$E_+ Vr$	$E_+ r$	$EC-$	$E_-\,-$	$E_- V-$
4				$E_-\,-$	$E_- V-$	$E_- i$	$EC\,i$	$E_+ r,i$	$E_+ Vr$	$E_+ r$	$EC-$
5					$EC-$	$E_-\,-$	$E_- V-$	$E_- i$	$EC\,i$	$E_+ r,i$	$E_+ Vr$
6						$E_+ r$	$EC-$	$E_-\,-$	$E_- V-$	$E_- i$	$EC\,i$
7							$E_+ Vr$	$E_+ r$	$EC-$	$E_-\,-$	$E_- V-$
8								$E_+ r$	$E_+ Vr$	$E_+ r$	$EC-$
9									$EC\,i$	$E_+ r$	$E_+ Vr$
10										$E_- i$	$EC\,i$
11											$E_- V-$

$[G{:}G'] = 2^{n-1}$. This, together with the class equation (26) determines completely the dimensions n_ν of the νth irreducible representations:

$$|G| = \sum_{\substack{\text{conjugacy}\\ \text{classes } \nu}} n_\nu^2. \tag{26}$$

Using the results of Sec. II, we have (a) for $n = 2m$ (i.e., G extraspecial) there is one nonlinear irreducible representation of degree 2^m; and (b) for $n = 2m+1$ [i.e., $G = E \circ Z(G)$] there are two nonlinear irreducible representations of degree 2^m.

Thus for even space-time dimensions there is only one type of spinor, while for odd dimensions there are two distinct types. Because, for $g \in Z(G)$, g and $-g$ are in the same conjugacy class, the characters of g vanish for these nonlinear representations: that is, they are represented by traceless matrices. These results are expressed in the following theorem (Dornhoff[19]).

Theorem 6: Let $G = E \circ C_p k$, where E is extraspecial of order p^{2m+1}. Then G has exactly the following irreducible complex characters: (i) $p^{2m+k-1} = [G{:}G']$ linear characters; and (ii) $p^k - p^{k-1}$ faithful irreducible linear characters χ_i of degree p^m, which vanish outside $C_p k$ and satisfy $\chi_i|_{C_p k} = p^m \lambda_i$, where λ_i is a faithful linear character of $Z(G)$.

Such a group, whose characters vanish outside of $Z(G)$, is sometimes known as central. For even dimensions all the gamma matrices are traceless, apart from $\chi_1(1) = -\chi_1(-1) = 2^m$. In odd space-time dimensions, $\chi_i(\Delta)$ is nonzero and is either $\pm 2^m$ or $\pm i2^m$, according to whether $Z(G)$ is either $C_2 \times C_2$ or C_4. [Note: the representations of $G = E \circ (C_2 \times C_2) = E \times C_2$ are just the direct product of the representations of E and C_2.] This structure of the representations reflects the isoclinism of the groups of fixed dimension. Before describing the reality properties of the irreducible representations of G, we comment on their construction and relation to the previous section.

Given a group $G^{r,n-r}$ we may embed this in either $G^{r,n+1-r}$ or $G^{r+1,r}$, corresponding to whether we add an extra generating element that squares to -1 or 1, respectively: that is, adding a 1 or 0 to the quadratic form. Schematically this is shown in Fig. 1. Two cases must now be distinguished between, according to whether n is even or odd.

Case 1: n even. Here we have only one representation D_1 of G. In this we have $D(\Delta) = D(\Gamma_1) \cdots D(\Gamma_n)$ with $D(\Delta)^2 = (-1)^{(n(n+1)/2)-r}$. We may now choose $D_\pm(\Gamma_{n+1}) = \pm D(\Delta)$ or $\pm iD(\Delta)$ according to whether we wish $\Gamma_{n+1}^2 = +1$ or (-1). The choice of $\pm$ corresponds to the two inequivalent representations of G_{odd}. The periodicity modulo 4 is reflected in the choice of $\pm D(\Delta)$ or $\pm iD(\Delta)$. If we choose $\pm D(\Delta)$ to go from $G^{r,2k-r}$ to $G^{r+1,2k-r}$, we must then choose $\pm iD(\Delta)$ to go from $G^{r+2,n-r}$ to $G^{r+3,n-r}$. This is shown in Fig. 2.

Case 2: n odd. In going from G_{2m+1} to G_{2m+2}, we are able to use the induced representation of G_{2m+1} in G_{2m+2}.

$$\begin{tikzcd} G^{r,n-r} \arrow[r] \arrow[dr] & G^{r+1,n-r} \\ & G^{r,n+1-r} \end{tikzcd}$$

FIG. 1. Extensions of $G^{r,n-r}$.

$$\begin{tikzcd} & G^{r,n-r} \arrow[r, "\pm iD(\Delta)"'] & G^{r+1,n-r} \arrow[r] & G^{r+2,n-r} \arrow[r, "\pm D(\Delta)"'] & G^{r+s,n-r} \\ n-r=0 & D \arrow[r] & D \circ C \arrow[r] & D \circ Q \arrow[r] & D \circ Q \circ V \end{tikzcd}$$

FIG. 2. Periodicities in adjoining elements to the representation $G^{r,n-r}$, for n even.

We have, if $H \lhd G$ and $G = \bigcup_{x_i} Hx_i$ for some $\{x_i\}$ a cross section of G/H, that the matrix

$$A(g) = \begin{bmatrix} D(x_E g x_E^{-1}) & \cdots & D(x_E g x_n^{-1}) \\ D(x_n g x_E^{-1}) & \cdots & D(x_n g x_n^{-1}) \end{bmatrix}$$

forms a representation of G, where D is a representation of H and $D(y) = 0$ for $y \notin H$. In our case we have

$$A(g) = \begin{bmatrix} D(g) & 0 \\ 0 & -D(g) \end{bmatrix}, \quad g \in G_{2m+1}, \quad g \notin Z(G_{2m+2}),$$

and

$$A(\Gamma_{2m+2}) = \begin{bmatrix} 0 & 1 \\ \pm 1 & 0 \end{bmatrix},$$

according to whether $\Gamma_{2m+2}^2 = \pm 1$. This representation is seen to be irreducible by Schur's lemma, as the only matrices which commute with all the elements of the group are multiples of the identity. It is readily seen that the other possible choices of $A(\Gamma_{2m+2})$ are equivalent to this. Denoting equivalence by $\sim$, then

$$\begin{bmatrix} 0 & 1 \\ \pm 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 0 & -1 \\ \mp 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 0 & i \\ \mp i & 0 \end{bmatrix} \sim \begin{bmatrix} 0 & -i \\ \pm i & 0 \end{bmatrix}.$$

Within this representation

$$A(\Delta_{2m+2}) = \begin{bmatrix} 0 & D(\Delta_{2m+1})\eta \\ \pm D(\Delta_{2m+1})\eta & 0 \end{bmatrix},$$

with $\eta \in \mathbb{C}^*$ suitably chosen. This procedure again has an alternating choice of the $\pm$ sign reflecting the modulo 4 periodicity.

The connection between the possible representations formed in these two cases is contained in the following lemma, which may be proven straightforwardly.

Lemma 7: We have equivalent representations by (i) first adjoining $D^2(\Delta_{2n+1}) = 1$ and then $A^2(\Gamma_{2n+2}) = -1$, and (ii) first adjoining $D^2(\Delta_{2n+1}) = -1$ and then $A^2(\Gamma_{2n+2}) = 1$. Thus the diagram in Fig. 3 is commutative, giving equivalent representations for $G^{r+1,2n-r+1}$.

The constructions just given interpolate between the usual representations of the gamma matrices inductively defined for even space-time dimensions.[7] The procedure for forming the representations clearly shows the modulo 4 periodicity observed from the study of the associated quadratic form.

We now discuss the reality properties of the representations of $G^{r,n-r}$, which enables us to give a general classifica-

$$\begin{tikzcd} G^{r,2n-r} \arrow[r, "D^2(\Delta)=1"] \arrow[dr, "D^2(\Delta)=-1"'] & G^{r+1,2n-r} \arrow[dr, "A^2(\Gamma_{2n+2})=-1"] & \\ & G^{r,2n-r+1} \arrow[r, "A^2(\Gamma_{2n+2})=1"'] & G^{r+1,2n-r+1} \end{tikzcd}$$

FIG. 3. Independence of the path, a representation is induced.

tion of the groups $G^{r,n-r}$ in terms of the orders of their elements. For the complex irreducible character χ_i of a finite group G, we have that

$$\nu(\chi_i) \equiv \frac{1}{|G|} \sum_{g\in G} \chi_i(g^2) = \begin{cases} 1, & \text{if } \chi_i = \bar{\chi}_i \text{ and } S_{\mathbb{R}}(\chi_i) = 1, \\ -1, & \text{if } \chi_i = \bar{\chi}_i \text{ and } S_{\mathbb{R}}(\chi_i) = 2, \\ 0, & \text{if } \chi_i \neq \bar{\chi}_i \end{cases} \quad (27)$$

Here $S_{\mathbb{R}}(\chi_i)$ denotes the Schur index[19] of χ_i over $\mathbb{R}$. The representation D_i corresponding to the character χ_i, is real if and only if $S_{\mathbb{R}}(\chi_i) = 1$. [When $S_{\mathbb{R}}(\chi_i) = -1$, we have a symplectic representation.] This leads to the Frobenius–Schur theorem.[28]

Theorem 8: Let G be a finite group and $1_G = \chi_1, \chi_2, ..., \chi_h$ the irreducible complex characters. With $\nu(\chi_i)$ defined in (27) we have (i) for any $g\in G$, let $t(g) = |\{x\in G \,|\, x^2 = g\}|$, then

$$t(g) = \sum_{i=1}^{h} \nu(\chi_i)\chi_i(g);$$

and (ii) the number of elements of order 2 in G is

$$\sum_{i=1}^{h} \nu(\chi_i)\chi_i(1) - 1 \leq \sum_{i=1}^{h} \chi_i(1) - 1.$$

Equality holds if and only if $\mathbb{R}$ is a splitting field for G.

Dornhoff[19] uses this theorem to characterize extraspecial two-groups in terms of their involutions. We generalize this slightly to cover the possible G arising here: this extends the explicit calculations of Salingaros[9] giving a general characterization of G.

Corollary 9: (i) E_{m+} contains exactly $2^{2m} + 2^m - 1$ nontrivial involutions, and $\nu(\chi) = 1$ for the 2^m-dimensional representation.

(ii) E_{m-} contains exactly $2^{2m} - 2^m - 1$ nontrivial involutions and $\nu(\chi) = -1$ for the 2^m-dimensional representation.

(iii) $G = E_m \circ C_{2^k}$, $k \geq 2$, contains exactly $2^{2m+k-1} - 1$ nontrivial involutions and $\nu(\chi_i) = 0$ for the 2^m-dimensional representations, i.e., they are complex.

Parts (i) and (ii) are an application of the Frobenius–Schur theorem (see Ref. 18). Part (iii) comes from Theorem 6, which says $\lambda_i \neq \bar{\lambda}_i$, as λ_i is faithful, and so the only real representations are the one-dimensional ones. This is then used in Theorem 7.

Corollary 9 enables us to classify the groups in terms of the orders of their elements, as an element is either an involution or of order 4. The group $G = E_m \times C_2$ has characters which are the products of the characters of E_m and C_2, and so their reality depends solely on that of E_m. Table II summarizes these properties of the group G: the reality of its representations, and the orders of its elements.

We conclude this section by showing the group G is simply reducible[29] for $n = 2m$. This reflects the underlying spin group and enables one to calculate the $3j$ and $6j$ symbols for these groups in terms of our finite group G (Braden[30] and de Vries[31]). A group is simply reducible if (i) every element of G is equivalent to its inverse and (ii) the Kronecker product of two irreducible representations of G contains each irreducible representation no more than once. We have shown condition (i) true for the case $n = 2m$ in Sec. II; this means all the characters are real and the representation is either integral or half-integral. Condition (ii) is also called "multiplicity-free," and we now show this is the case for both n even and odd.

It is useful to label the one-dimensional representations of G. A convenient choice is the following: if χ_I is the character of (1_I), then it is given by

$$\chi_I(\Gamma^B)1 = \Gamma^B \Gamma^A (\Gamma^B)^{-1} (\Gamma^A)^{-1}. \quad (28)$$

Being in the commutator subgroup, the right-hand side is ± 1. Here Γ^A is some element of a basis of gamma matrices, say $\Gamma^{\mu_1 \cdots \mu_s}$. The right-hand side of (28) shows that if $\Gamma^{\mu_j} \in \Gamma^B$, it anticommutes with all those $\mu_r \in A$ different from μ_j. Thus

$$\chi_A(\Gamma^B) = \prod_{\mu_j \in B} (-1)^{\text{No. } \mu\text{'s in } A \text{ different from } \mu_j} = \chi_B(\Gamma^A). \quad (29)$$

We see that (28) labels distinct representations as follows. Suppose $\chi_A(g) = \chi_{A'}(g)$, $g\in G$. Then $\Gamma_A^{-1}\Gamma_{A'} \in Z(G)$. Now for a one-dimensional representation, the center is trivial and so $\Gamma_A = \Gamma_{A'}$. With definition (28) we have

$$\chi_A(g)\chi_B(g) = \chi_{A^{-1}B}(g). \quad (30)$$

Consider now the tensor product of two irreducible representations $D^{(\mu)}, D^{(\nu)}$; call it $D^{(\mu\times\nu)}$. Then we have

$$D^{(\mu\times\nu)}(g) = \sum_{\sigma} a_\sigma D^{(\sigma)}(g), \quad (31a)$$

$$a_\sigma = \frac{1}{|G|} \sum_{g\in G} \chi^{(\mu)}(g)\chi^{(\nu)}(g)\bar{\chi}^{(\sigma)}(g). \quad (31b)$$

When $D^{(\mu)}$ is one of the 2^m-dimensional representations of G, (31b) is readily solved. Theorem 6 tells us the sum vanishes outside of G'. The case when we have two one-dimensional representations is aided by (30). Upon solving the $2^d + 1$

TABLE II. $G = (1,a,b)$ has a nontrivial involutions and b elements of order 4. $m > 0$.

G	$(1,a,b)$	Type of the 2^m dimensional of representations
E_{m+}	$(1,2^{2m}+2^m-1,2^{2m}-2^m)$	Real
E_{m-}	$(1,2^{2m}-2^m-1,2^{2m}+2^m)$	Symplectic
$E_m \circ C_4$	$(1,2^{2m+1}-1,2^{2m+1})$	Complex conjugate pair
$E_{m+} \times C_2$	$(1,2^{2m+1}+2^{m+1}-1,2^{2m+1}-2^{m+1})$	Two real
$E_{m-} \times C_2$	$(1,2^{2m+1}-2^{m+1}-1,2^{2m+1}+2^{m+1})$	Two symplectic

(+ 2, for d odd) equations (31b) we find, using a straightforward notation,

$$(1_A)\otimes(1_B)=(1_{A^{-1}B}), \tag{32}$$

$$(2^m)\otimes(1_A)=(2^m), \tag{33}$$

$$(2^m)\otimes(2^m)=\sum_A(1_A). \tag{34}$$

The sum in (34) is over all 2^{2m} A for $n=2m$; when $n=2m+1$ the 2^{2m+1} possible A are restricted by

$$\chi_A(1)=-\eta^2\chi_A(\Delta), \quad \text{with } \chi_{2^m}(\Delta)=\eta 2^m. \tag{35}$$

The 2^m-dimensional representation on the right-hand side of (33) is the same as that on the left-hand side. Lastly (32) reflects the orthogonality of the characters. In conclusion, (32)–(34) show that the group G is multiplicity-free. Thus G is simply reducible when n is even, or in the odd case, when G has a direct product structure $C_2\times E_{m\pm}$.

V. APPLICATIONS

One of the immediate uses of the classification scheme developed is to describe the possible representations of the Dirac matrices for a given metric. Using Tables I and II, we see that the metric $(-\ +\ +\ +)$ admits a real representation; thus the symplectic representation of $(+\ -\ -\ -)$ may be taken to be pure imaginary. With the former metric the Dirac equation is

$$(\partial\!\!\!/ + m)\psi=0. \tag{36}$$

A real representation of the Dirac matrices enables the Majorana condition to be implemented. Had we considered the metric $(-\ +\ +\ +\ +\ +)$ however, our tables show rather that a pure imaginary representation exists. This means that the Majorana condition can be implemented only for massless particles for such a metric.

We have also, that for the even-dimensional space-time with equal numbers of spacelike and timelike directions, both pure real and imaginary representations exist. This follows from our analysis of $G^{n,n}$, which is obviously $D_1\circ D_2\circ\cdots\circ D_n$. From this we see that the group $G^{2n,0}$ has a representation in which all the gamma matrices are Hermitian,[32] half being real (symmetric) and the other half being imaginary (antisymmetric).

The case of odd dimensions is a little different. Our analysis tells us immediately the type of representation associated with a given metric, but now there are two nontrivial irreducible representations. Given a representation $D(\Gamma_i)$ of Γ_i, the representation resulting from $D(\Gamma_i)$ is inequivalent to this. Similarly, the representation given by $D^*(\Gamma_i)$ is sometimes equivalent to $D(\Gamma_i)$ and other times not: this reflects whether the group is $E_\pm\times C_2$ or $E\circ C_4$, the latter having a pair of complex conjugate representations. Other representations may be obtained from $D(\Gamma_i)$ of Γ_i, by transposition, $\tilde{D}(\Gamma_i)$, and Hermitian conjugation, $D^\dagger(\Gamma_i)$. Table III, showing conditions of equivalence of these representations, is readily obtained by examining the properties of $D(\Delta)$, which by Schur's lemma is a scalar multiple of the identity. To illustrate the odd-dimensional case, consider the metric $(+\ -\ -)$. Table I shows this is associated with the group $E\circ C$ and so has a pair of complex representations. The metric $(-\ +\ +)$ has a pure real representation and so $(+\ -\ -)$ has a pure imaginary one; the other inequivalent representation is here the complex conjugate of this. The Majorana condition can be implemented here.

The Weyl condition reflects itself in the symplectic structure associated with the group; when the dimension is even we can decompose the attendant vector space into two equal parts. This is clearly seen in the representation theory, where $A(\Gamma_i)$ for $1\leq i\leq n-1$ acts on two orthogonal subspaces. We may now ask about applying both Weyl and Majorana restrictions together when this is possible, i.e., those even-dimensional spaces which have either a pure, real, or imaginary representation. Two cases arise. First, that where the restrictions are equivalent. For instance $(-\ +\ +\ +)$ has a real representation, as does $(-\ +\ +)$. The second possibility arises when the reduced space has no pure representation properties to be implemented. This case then yields independent restrictions. For example, the ten-dimensional space whose metric has one minus has a real representation, while the restricted nine-dimensional space with one minus has an imaginary representation. These different representation properties lead to separate and distinct restrictions. For a Lorentz signature, this gives the usual result of a Weyl–Majorana restriction holding for $n=2$ (modulo 8).[33] Such results are of use in supersymmetry today.[34]

One final application comes from Table II, which characterizes the orders of the group elements. The unitarity of any representations means $D^\dagger(\Gamma_i)=\pm D(\Gamma_i)$ according to whether $\Gamma_i^2=\pm 1$. Table II tells us how many $D^\dagger(g)=D(g)$ [and similarly $D^\dagger(g)=-D(g)$]; the number is just the number of elements of order 2. If we have a pure real (imaginary) representation, then this is just the number of symmetric (antisymmetric) matrices. The number of symmetric matrices of degree 2 is obviously $2^\nu(2^\nu-1)/2+2^\nu=\frac{1}{2}[2^{2\nu}+2^\nu]$ which is half that given in the table. The factor of two comes because $D(\Gamma_i)$ and $-D(\Gamma_i)$ count as separate elements in the group. Similar applications of Table II tell us the number of symmetric and antisymmetric matrices in the chosen representation. This connects the periodic group structure with the periodicities observed in these symmetry properties of the representation.[35]

VI. CONCLUSIONS

A new and direct method of classification of the finite groups associated with a Clifford algebra has been present-

TABLE III. Equivalence among representations of $G^{r,n-r}$ for n odd. $D(\Gamma_i)$ and $-D(\Gamma_i)$ are always inequivalent.

	r odd	r even
$n=4k+1$	$D(\Gamma_i), D^*(\Gamma_i), \tilde{D}(\Gamma_i), D^\dagger(\Gamma_i)$	$D(\Gamma_i), -D^*(\Gamma_i), \tilde{D}(\Gamma_i), -D^\dagger(\Gamma_i)$
$n=4k+3$	$D(\Gamma_i), -D^*(\Gamma_i), -\tilde{D}(\Gamma_i), D^\dagger(\Gamma_i)$	$D(\Gamma_i), D^*(\Gamma_i), -\tilde{D}(\Gamma_i), -D^\dagger(\Gamma_i)$

ed, and some of their general properties described. This group is useful in examining the representation properties of gamma matrices, and so the behavior of spinors in spaces of arbitrary dimension and signature. The classification highlights the periodic properties observed and provides new characterizations of the representations and group involved.

[1]W. K. Clifford, Am. J. Math. **1**, 350 (1878).

[2]M. Riesz, *Clifford Numbers and Spinors* (Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, 1958).

[3]R. H. Good, Rev. Mod. Phys. **97**, 810 (1955).

[4]This group has no universal name. Some call it the Dirac group, others the Clifford group. This latter definition is poor, as there exists another (and different) group with this name arising in the study of spin groups.

[5]A. S. Eddington, *Relativity Theory of Protons and Electrons* (Cambridge U. P., Cambridge, 1936).

[6]H. Boerner, *Representation of Groups* (North-Holland, Amsterdam, 1963).

[7]R. Brauer and H. Weyl, Am. J. Math. **57**, 447 (1935).

[8]A. Ramakrishnan, J. Math. Anal. Appl. **31**, 1 (1970).

[9]N. Salingaros, J. Math. Phys. **22**, 226 (1981).

[10]N. Salingaros, J. Math. Phys. **23**, 1 (1982).

[11]A. O. Morris, Proc. London. Math. Soc. **32**, 403 (1976).

[12]I. Schur, J. Reine Angew. Math. **139**, 155 (1911).

[13]J. E. Mandula and E. Shipiz, Nucl. Phys. B **232**, 180 (1984).

[14]P. Goddard and D. Olive, DAMTP preprint 83/22, University of Cambridge, 1984.

[15]S. D. Smith, "The Classification of finite groups with large extra-special 2-subgroups" in *Proceedings of the Symposium on Pure Mathematics* (A. M. S., Providence, RI, 1980), Vol. 37.

[16]R. Coqueraux, Phys. Lett. B **115**, 389 (1982).

[17]J. S. Rose, *A Course on Group Theory* (Cambridge U. P., Cambridge, 1978).

[18]D. Gorenstein, *Finite Groups* (Harper and Row, New York, 1968).

[19]L. Dornhoff, *Group Representation Theory* (Dekker, New York, 1971).

[20]B. Huppert, *Endliche Gruppen* I (Springer-Verlag, Berlin, 1967).

[21]Indeed we may associate with G a Lie algebra over the field of two elements. (See, for example, a general description for p-groups given in Gorenstein.[18]) We do not use this extra structure in the following account.

[22]J. Dieudonne, *La Geometrie des Groupes Classiques* (Springer-Verlag, Berlin, 1963).

[23]I. Porteous, *Topological Geometry* (Van Nostrand, New York, 1969).

[24]C. T. C. Wall, J. Reine Angew. Math. **213**, 187 (1964).

[25]P. Lounesto, Found. Phys. **11**, 721 (1981).

[26]By representation, we mean ordinary representation. For work pertaining to projective representations N. Blackburn and L. Evans[27] calculate the Schur multiplicator of an extra-special group.

[27]N. Blackburn and L. Evans, J. Reine Angew. Math. **309**, 100 (1979).

[28]G. Frobenius and I. Schur, Sitzungsber. Preuss. Akad. Wiss. 186, Berlin (1906).

[29]L. Biedenharn and J. D. Louck, "The Racah–Wigner Algebra in Quantum Theory," in *Encyclopedia of Mathematics and its Applications* (Addison-Wesley, Reading, MA, 1981), Vol. 9.

[30]H. W. Braden, J. Phys. A **17**, 2927 (1984).

[31]E. de Vries and A. J. van Zanten, Commun. Math. Phys. **17**, 332 (1970).

[32]A. Pais, J. Math. Phys. **3**, 1135 (1962).

[33]J. Scherk, "Extended supersymmetry," in *Recent Developments in Gravitation*, Cargese, 1978, edited by M. Levy and S. Deser (Plenum, New York, 1979).

[34]F. Gliozzi, J. Scherk, and D. Olive, Nucl. Phys. B **122**, 1253 (1977).

[35]K. M. Case, Phys. Rev. **97**, 810 (1955).

Perturbation of Schrödinger Hamiltonians by measures—Self-adjointness and lower semiboundedness

Johannes Brasche
Fakultät für Mathematik, Universität Bielefeld, Universitätsstrasse 1, 4800 Bielefeld 1, Federal Republic of Germany

(Received 24 May 1984; accepted for publication 14 September 1984)

We study the Hamiltonians for nonrelativistic quantum mechanics in one dimension, in terms of energy forms $\int |df/dx|^2\, dx + \int |f|^2\, d(\mu - \nu)$, where μ and ν are positive, not necessarily finite measures on the real line. We cover, besides regular potentials, cases of very singular interactions (e.g., a particle interacting with an infinite number of fixed particles by "delta function potentials" of arbitrary strengths). We give conditions for lower semiboundedness and closability of the above energy forms, which are sufficient and, for certain classes of potentials (e.g., $\mu - \nu$ a signed measure), also necessary. In contrast to the results in other approaches, no regularity conditions and no restrictions on the growth of the measures μ and ν at infinity are needed.

I. INTRODUCTION

In this paper we discuss self-adjointness and lower semiboundedness of Hamiltonians, which are formally given by $-(d^2/dx^2)+\mu$, where μ is a (neither necessarily finite nor necessarily positive) Radon measure on $\mathbb{R}$. Such Hamiltonians have been introduced in the physical literature for the description of singular interactions, e.g., point interactions (the case μ a sum of delta functions), and in problems of solid state physics, nuclear physics, and electromagnetism (see, e.g., the reference given in Ref. 1; see also, e.g., Refs. 2–8). There are various mathematical definitions of $-(d^2/dx^2)+\mu$. (See Refs. 2, 3, 6, and 9–11.)

If the energy form[12]

$$\int_{\mathbb{R}} |f'(x)|^2\, dx + \int_{\mathbb{R}} |f(x)|^2\, d\mu(x)$$

is lower semibounded (l.s.b.) and closable, the Hamiltonian $-(d^2/dx^2)+\mu$ can be defined as the self-adjoint operator in $L^2(\mathbb{R},dx)$ uniquely associated with this form, for this approach to self-adjointness brings into question previous works (see, e.g., Refs. 11,13). Thus the question arises, when is the above energy form l.s.b. and closable. Both properties are also important because lower semiboundedness is an expression of stability for the physical system, and closability is necessary for any determination of the dynamics.

First, we suppose that μ is a positive (Radon) measure on $\mathbb{R}$. Then the energy form

$$\int_{\mathbb{R}} |f'(x)|^2\, dx + \int_{\mathbb{R}} |f(x)|^2\, d\mu(x)$$

is closed if μ is regular [i.e., $d\mu = V\,dx$, for some $V \in L^1_{\mathrm{loc}}(\mathbb{R},dx)$] of if μ is finite, since in the former case the form

$$\int_{\mathbb{R}} |f(x)|^2\, d\mu(x) = \int_{\mathbb{R}} |f(x)|^2 V(x) dx$$

is closed, and in the latter case $\int_{\mathbb{R}} |f(x)|^2\, d\mu(x)$ is infinitesimally form-bounded with respect to $\int_{\mathbb{R}} |f'(x)|^2\, dx$ (see Ref. 11). But in general, if μ is neither a finite nor regular Radon measure on $\mathbb{R}$, the form $\int_{\mathbb{R}} |f(x)|^2\, d\mu(x)$ is neither closed nor bounded with respect to $\int_{\mathbb{R}} |f'(x)|^2\, dx$. We show that even in this case the form

$$\int_{\mathbb{R}} |f'(x)|^2\, dx + \int_{\mathbb{R}} |f(x)|^2\, d\mu(x)$$

is closed. (See Theorem 1(a).) Thus, for arbitrary positive Radon measures μ on $\mathbb{R}$, we can define a self-adjoint operator $-(d^2/dx^2)+\mu$. Moreover, we show that monotone convergence from below of positive measures μ_n implies convergence of the Hamiltonians $-(d^2/dx^2)+\mu_n$ in the strong resolvent sense. (See Theorem 1(b).)

Next, we consider perturbations of the free Hamiltonian by arbitrary, not necessarily positive, Radon measures. If $\mu-\nu$ is a signed Radon measure, or if μ, ν are positive Radon measures and μ is finite on an ϵ-neighborhood of the support of ν, we give a necessary and sufficient condition in order that the energy form

$$\int_{\mathbb{R}} |f'(x)|^2\, dx + \int_{\mathbb{R}} |f(x)|^2\, d(\mu-\nu)(x)$$

is l.s.b. and closed. (See Theorem 3(b).)

For arbitrary positive Radon measures μ,ν on $\mathbb{R}$, we show that the energy form

$$\int_{\mathbb{R}} |f'(x)|^2\, dx + \int_{\mathbb{R}} |f(x)|^2\, d(\mu-\nu)(x)$$

is l.s.b. and closed if $\mathbb{R}$ can be written as the disjoint union of intervals I_n so that $\inf_{n\in\mathbb{N}} |I_n| > 0$ and $\sup_{n\in\mathbb{N}} \nu(I_n) < \infty$, where $|\ |$ means Lebesgue measure, (See Theorem 3(a).) Thus, to show that

$$\int_{\mathbb{R}} |f'(x)|^2 + \int_{\mathbb{R}} |f(x)|^2 d(\mu-\nu)(x)$$

is l.s.b., only a *local* estimate of ν is needed. Under the stronger assumption of regular measures μ,ν a similar result (in the multidimensional case) has been given in Ref. 14.

Let μ,ν be positive Radon measures so that support μ $\cap$ support $\nu = \varnothing$. Then the Hamiltonian $-(d^2/dx^2)-\nu$ is not l.s.b. if, for some compact set K, $\nu(\{y+x \mid y\in K\})\to\infty$ as $x\to\pm\infty$, and one might expect that $-(d^2/dx^2)+\mu-\nu$ is not l.s.b., too. But we shall see that even in this case the operator $-(d^2/dx^2)+\mu-\nu$ can be l.s.b. and self-adjoint, i.e., the potentials μ and ν can "cancel out" each other though their supports are disjoint. (See Theorem 4 and example 2.)

 0022-2488/85/040621-06$02.50

In Theorem 2 we give an explicit description of the Hamiltonian $-(d^2/dx^2)+\mu$, associated with the energy form

$$\int_{\mathbb{R}}|f'(x)|^2\,dx+\int_{\mathbb{R}}|f(x)|^2\,d\mu(x).$$

The results of this paper extend some of the results in Ref. 3. We refer to the latter work for some further examples and discussions. All results in this paper are formulated for Hamiltonians defined in $L^2(\mathbb{R},dx)$; however, they extend easily to the case where $\mathbb{R}$ is replaced by any open subset D of $\mathbb{R}$ with the cone property (i.e., such that there is an $\epsilon>0$, so that for any $x\in D$, $[x-\epsilon,x]\subset D$ or $[x,x+\epsilon]\subset D$), and Dirichlet boundary conditions are chosen.

II. THE SCOPE OF THE HAMILTONIAN $-(d^2/dx^2)+\mu-\nu$

Let μ be a positive Radon measure on $\mathbb{R}$. We define a quadratic form P_μ with domain $D(P_\mu)$ on $L^2(\mathbb{R},dx)$ by

$$D(P_\mu)=H^{2,1}(\mathbb{R})\cap L^2(\mathbb{R},d\mu),$$

with $H^{2,1}(\mathbb{R})$ the space of L^2 functions with (generalized) L^2 derivatives, and[15]

$$P_\mu(f,g)=\int_{\mathbb{R}}f^*(x)g(x)d\mu(x),$$

for all $f,g\in D(P_\mu)$.

The energy form E_μ is defined by $E_\mu=E_0+P_\mu$, where

$$D(E_0)=H^{2,1}(\mathbb{R}),$$

$$E_0(f,g)=\int_{\mathbb{R}}f'^*(x)g'(x)\,dx,\quad f,g\in D(E_0).$$

It is well known that the free (kinetic) energy form E_0 is closed.

If the Radon measure μ is finite, the quadratic form P_μ is infinitesimally form-bounded with respect to E_0 (in the terminology of Ref. 6) since, by Sobolev's inequality, for any $a>0$, there is a number b so that

$$\|f\|_\infty^2\leqslant aE_0(f,f)+b\int_{\mathbb{R}}|f(x)|^2\,dx,\tag{1}$$

for any $f\in H^{2,1}(\mathbb{R})$. Thus, by the Kato–Lax–Milgram–Nelson (KLMN) theorem (see Ref. 6), E_μ is closed.

If μ is an arbitrary positive Radon measure on $\mathbb{R}$, we can choose an increasing sequence $\{\mu_n\}$ of finite positive Radon measures on $\mathbb{R}$, so that a function f is μ-square integrable if and only if it is μ_n-square integrable for any $n\in\mathbb{N}$, and

$$\sup_{n\in\mathbb{N}}\int_{\mathbb{R}}|f(x)|^2\,d\mu_n(x)<\infty.$$

For example, set $\mu_n=\chi K_n\cdot\mu$, where χ means characteristic function and $\{K_n\}$ is a sequence of compact sets increasing to $\mathbb{R}$). Then E_μ is the monotone limit form of the forms E_{μ_n}, i.e.,

$$D(E_\mu)=\{f\in\bigcap_{n\in\mathbb{N}}D(E_{\mu_n})|\sup_{n\in\mathbb{N}}E_{\mu_n}(f,f)<\infty\},$$

$$E_\mu(f,g)=\lim_{n\to\infty}E_{\mu_n}(f,g),\quad f,g\in D(E_\mu).$$

(The limits exist by polarization.)

Thus, by a theorem of Kato[13] and Simon[16] on monotone sequences of positive closed quadratic forms, E_μ is closed. It is easily seen that the space $C_0^\infty(\mathbb{R})$ of infinitely differentiable functions of compact support is a core of E_μ. Thus we have proven part (a) of the following theorem. Part (b) of Theorem 1 shows that monotone convergence from below of positive measure-valued potentials μ_n implies convergence (in the strong resolvent sense) of the Hamiltonians $-(d^2/dx^2)+\mu_n$.

Theorem 1:

(a) Let μ be a positive Radon measure on $\mathbb{R}$. Then the energy from E_μ, defined by

$$D(E_\mu)=H^{2,1}(\mathbb{R})\cap L^2(\mathbb{R},d\mu),$$

$$E_\mu(f,g)=\int_{\mathbb{R}}f'^*(x)g'(x)dx+\int_{\mathbb{R}}f^*(x)g(x)d\mu(x),\quad f,g\in D(E_\mu),$$

is closed on the Hilbert space $L^2(\mathbb{R},dx)$, and $C_0^\infty(\mathbb{R})$ is a core of this form.

(b) Let $\{\mu_n\}$ be an increasing sequence of positive Radon measures on $\mathbb{R}$, so that $\sup_{n\in\mathbb{N}}\mu_n(B)=\mu(B)$ for any Borel set B. Let H_{μ_n},H_μ be the positive self-adjoint operators in $L^2(\mathbb{R},dx)$ uniquely associated with the energy forms E_{μ_n},E_μ. Then $H_{\mu_n}\to H_\mu$, $n\to\infty$ in the strong resolvent sense.

Proof of (b): By hypothesis, E_μ is the monotone limit form of the E_{μ_n}. Thus (b) follows from Theorem 3.1 in Ref. 16. ■

Let μ,ν be positive Radon measures on $\mathbb{R}$. If the quadratic form $E_\mu-P_\nu$ is l.s.b. and closable on $L^2(\mathbb{R},dx)$ the energy form $E_{\mu-\nu}$ is defined as the closure of $E_\mu-P_\nu$. By the definition of $E_\mu-P_\nu$ and $E_{\mu-\nu}$, we have

$$E_{\mu-\nu}(f,g)=\int_{\mathbb{R}}f'^*(x)g'(x)dx+\int_{\mathbb{R}}f^*(x)g(x)d(\mu-\nu)(x),\tag{2}$$

for any $f,g\in D(E_\mu-P_\nu)$. However, in general there are functions f in the domain of $D(E_{\mu-\nu})$ which are neither μ-square integrable nor ν-square integrable. (See example 2 below.) Thus, in general there are functions $f,g\in D(E_{\mu-\nu})$ for which the right-hand side of (2) is not defined. But we shall see that (2) holds whenever f or g has compact support. (See the following, Lemma 1.) As an immediate consequence of this lemma, we shall prove that the l.s.b. self-adjoint operator $H_{\mu-\nu}$, uniquely associated with the energy form $E_{\mu-\nu}$, is given by $H_{\mu-\nu}=-(d^2/dx^2)+\mu-\nu$ [in the sense that, acting on a function f, we have $H_{\mu-\nu}f=-(d^2/dx^2)f+(\mu-\nu)f$, with the derivative, and $(\mu-\nu)f$ to be understood in the distributional sense].

Lemma 1: Let μ,ν be positive Radon measures on $\mathbb{R}$ so that the quadratic form $E_\mu-P_\nu$ is l.s.b. and closable on $L^2(\mathbb{R},dx)$ and let $E_{\mu-\nu}$ be the closure of $E_\mu-P_\nu$. Let $f\in D(E_{\mu-\nu})$ and $g\in C_0^\infty(\mathbb{R})$. Then

$$E_{\mu-\nu}(f,g)=\int_{\mathbb{R}}f'^*(x)g'(x)dx+\int_{\mathbb{R}}f^*(x)g(x)d(\mu-\nu)(x).$$

Proof: By the definition of $E_{\mu-\nu}$, we can choose a sequence $\{f_n\}$ in the domain of $E_\mu - P_\nu$, so that $E_{\mu-\nu}(f_n - f, f_n - f) \to 0$, $n\to\infty$, and $f_n \to f$, $n\to\infty$ in $L^2(\mathbb{R},dx)$. Subtracting a subsequence if necessary, we may assume that $\{f_n\}$ converges pointwise Lebesgue, a.e. on $\mathbb{R}$. Let D be a nonempty open bounded interval on $\mathbb{R}$ with support $g\subset D$, so that $\{f_n\}$ converges pointwise on the boundary of D.

Suppose that $\sup_{n\in\mathbb{N}} \int_D |f_n'(x)|^2\, dx = \infty$. Then, by Sobolev's inequality, and since the Radon measures μ,ν are finite on the bounded interval D and the sequence $\{\int_D |f_n(x)|^2\, dx\}$ is bounded, we have

$$\sup_{n\in\mathbb{N}} \left(\int_D |f_n'(x)|^2\, dx + \int_D |f_n(x)|^2\, d(\mu-\nu)(x) \right) = \infty. \quad (3)$$

Let $\tilde f_n$ be the continuous function, which equals f_n on $\mathbb{R}\setminus D$ and is affine linear on D. If any of the sequences $\{f_n(\inf D)\}$, $\{f_n(\sup D)\}$, $\{\int_{\mathbb{R}} |f_n(x)|^2\, dx\}$, and $\{E_{\mu-\nu}(f_n, f_n)\}$ are bounded and (3) holds, it is straightforward to show that $\sup_{n\in\mathbb{N}} \int_{\mathbb{R}} |\tilde f_n(x)|^2\, dx < \infty$ and $\inf_{n\in\mathbb{N}} E_{\mu-\nu}(\tilde f_n, \tilde f_n) = -\infty$, which contradicts the fact that $E_{\mu-\nu}$ is l.s.b.

Thus we have proven that $\sup_{n\in\mathbb{N}} \int_D |f_n'(x)|^2 dx < \infty$. Thus the sequence $\{f_n \restriction D\}$ of the restrictions of the f_n on D is bounded in the Hilbert space $(H^{2,1}(D), (\ ,\ )_s)$ with inner product

$$(h,\tilde h)_s := \int_D h'*(x)\tilde h'(x)dx + \int_D h*(x)\tilde h(x)dx.$$

Thus, subtracting a subsequence if necessary, we may assume that $\{f_n \restriction D\}$ converges weakly in $(H^{2,1}(D), (\ ,\ )_s)$ (see Ref. 17). This implies that there is a subsequence $\{f_{n_j}\}$ of $\{f_n\}$ so that the sequence $\{h_m \restriction D\}$, with $h_m := (1/m)\Sigma_{j=1}^m f_{n_j}$ for any $m\in\mathbb{N}$, converges strongly in $(H^{2,1}(D), (\ ,\ )_s)$ (see Ref. 17). Since $h_m \restriction D \to f \restriction D$, $m\to\infty$ strongly in $L^2(D,dx)$, we get $h_m \restriction D \to f\restriction D$, $m\to\infty$ strongly in $(H^{2,1}(D), (\ ,\ )_s)$. By Sobolev's inequality, this implies $\sup_{x\in D} |h_m(x) - f(x)| \to 0$, $m\to\infty$. Thus

$$\begin{aligned}
E_{\mu-\nu}(f,g) &= \lim_{m\to\infty} E_{\mu-\nu}(h_m, g)\\
&= \lim_{m\to\infty} \left(\int_D h_m'*(x) g'(x)dx \right.\\
&\quad \left. + \int_D h_m*(x) g(x) d(\mu-\nu)(x) \right)\\
&= \int_D f'*(x) g'(x)dx\\
&\quad + \int_D f*(x) g(x) d(\mu-\nu)(x).
\end{aligned}$$

The last but one step follows from $g\in D(E_\mu - P_\nu)$, $h_m \in D(E_\mu - P_\nu)$ for any $m\in\mathbb{N}$ and support $g\subset D$. ■

Theorem 2: Let μ,ν be positive Radon measures on $\mathbb{R}$ so that the quadratic form $E_\mu - P_\nu$ is l.s.b. and closable on $L^2(\mathbb{R},dx)$ and let $E_{\mu-\nu}$ be the closure of $E_\mu - P_\nu$. Let $H_{\mu-\nu}$ be the l.s.b. self-adjoint operator in $L^2(\mathbb{R},dx)$ uniquely associated with the energy form $E_{\mu-\nu}$. Then

$$D(H_{\mu-\nu}) = \{ f\in D(E_{\mu-\nu}) | (-f'' + (\mu-\nu)f)_{\text{dist}} \in L^2(\mathbb{R},dx)\},$$

where $(-f'' + (\mu-\nu)f)_{\text{dist}}$ means the distributions

$$g \to \int_{\mathbb{R}} f*(x)(-g''(x))dx + \int_{\mathbb{R}} f*(x)g(x)d(\mu-\nu)(x),\ g\in C_0^\infty(\mathbb{R}),$$

and

$$H_{\mu-\nu} f = (-f'' + (\mu-\nu)f)_{\text{dist}},$$

for any $f\in D(H_{\mu-\nu})$.

Proof: Let $f\in D(E_{\mu-\nu})$. Lemma 1 yields

$$E_{\mu-\nu}(f,g) = \int_{\mathbb{R}} (-f'' + (\mu-\nu)f)*_{\text{dist}}(x)\, g(x)\, dx, \quad (4)$$

for any $g\in C_0^\infty(\mathbb{R})$. Suppose that in addition $(-f'' + (\mu-\nu)f)_{\text{dist}} \in L^2(\mathbb{R},dx)$. Then (4) holds for any $g\in D(E_{\mu-\nu})$, since $C_0^\infty(\mathbb{R})$ is a core of $E_{\mu-\nu}$ (as is easily seen by the Theorem 1(a)). This implies $f\in D(H_{\mu-\nu})$ and $H_{\mu-\nu} f = (-f'' + (\mu-\nu)f)_{\text{dist}}$ (see Ref. 13). Conversely, let $f\in D(H_{\mu-\nu})$. Then there is a function $\tilde f \in L^2(\mathbb{R},dx)$ so that

$$E_{\mu-\nu}(f,g) = \int_{\mathbb{R}} \tilde f*(x) g(x) dx,$$

for any $g\in D(E_{\mu-\nu})$ (see Ref. 13). By (4) and since $C_0^\infty(\mathbb{R})$ is dense in $L^2(\mathbb{R},dx)$, this implies $\tilde f = (-f'' + (\mu-\nu)f)_{\text{dist}}$.

Remark: Hamilton operators, formally given by a perturbation of the free Hamiltonian H_0 by certain distribution-valued potentials, have also been discussed by methods of Dirichlet forms (in the multidimensional case) as follows. (See, e.g., Refs. 1, 3, and 18). Let $H^{(\varphi)}$ be the self-adjoint operator in $L^2(\mathbb{R}^n, \varphi^2\, dx)$ uniquely associated with the Dirichlet form

$$E^{(\varphi)}(f,f) = \int_{\mathbb{R}^n} |\operatorname{grad} f(x)|^2 \varphi^2(x) dx$$

on $L^2(\mathbb{R}^n, \varphi^2\, dx)$. Then we have $\varphi H^{(\varphi)} \varphi^{-1} = H_0 + V$, with

$$V = \frac{\Delta\varphi(x)}{\varphi(x)} \left(\Delta := \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} \right).$$

In particular, results on closability of these Dirichlet forms $E^{(\varphi)}$ have been obtained. For example, in Ref. 3 we proved that the Dirichlet form $E^{(\varphi)}$ is closable on $L^2(\mathbb{R}^n, \varphi^2\, dx)$ if there is a closed set N of Lebesgue measure zero, so that for any compact set $K\subset \mathbb{R}^n \setminus N$, there is a strict positive number $r(K)$ with $\varphi(x) > r(K)$ on K. This result has recently been extended in Ref. 19 (to which we refer also for a survey of known criteria for closability).

Although in the one-dimensional case a necessary and sufficient condition is known in order that a Dirichlet form $E^{(\varphi)}$ is closable on $L^2(\mathbb{R}, \varphi^2\, dx)$ (see Refs. 20 and 21) and the correspondence between Dirichlet forms $E^{(\varphi)}$ on $L^2(\mathbb{R}, \varphi^2\, dx)$ and energy forms E_μ on $L^2(\mathbb{R},dx)$ is well known, too [namely, $\mu(x) = \varphi''(x)/\varphi(x)$], in general it is difficult to decide by methods of Dirichlet forms whether, given measures μ and ν, the quadratic form $E_\mu - P_\nu$ is closable on $L^2(\mathbb{R},dx)$, since one has to find a solution φ of the differential equation $(\mu - \nu)(x) = \varphi''(x)/\varphi(x)$. In general, this equation cannot be solved exactly. Thus, it seems to be desirable to have criteria in order that the quadratic form $E_\mu - P_\nu$ is l.s.b. and closable on $L^2(\mathbb{R},dx)$, which only involve the measures μ and ν. Such criteria will be given in the following section.

III. CLOSABILITY AND LOWER SEMIBOUNDEDNESS OF ENERGY FORMS

Suppose that the following condition is satisfied.

(A) There is a sequence $\{I_n\}$ of pairwise disjoint intervals, so that $\mathbb{R} = \cup_{n\in\mathbb{N}} I_n$, $\sup_{n\in\mathbb{N}} \nu(I_n) < \infty$, and $\inf_{n\in\mathbb{N}} |I_n| > 0$ (where $|\ |$ means Lebesgue measure). Then, by Sobolev's inequality, for any $a > 0$ and $n \in \mathbb{N}$

$$\sup_{x\in I_n} |f(x)|^2 \leqslant a \int_{I_n} |f'(x)|^2\,dx + b \int_{I_n} |f(x)|^2\,dx, \tag{5}$$

for any $f \in H^{2,1}(\mathbb{R})$ where the number b can be chosen independent of n, since $\inf_{n\in\mathbb{N}} |I_n| > 0$. By (5),

$$\int_{\mathbb{R}} |f(x)|^2\,d\nu(x) \leqslant \sup_{n\in\mathbb{N}} \nu(I_n)(aE_0(f,f) + b\int_{\mathbb{R}} |f(x)|^2\,dx).$$

Thus, the quadratic form P_ν is infinitesimally form-bounded with respect to the free energy form E_0. Thus, the KLMN theorem and Theorem 1(a) yield part (a) of the following theorem. Part (b) follows from a simple computation.

Theorem 3: Let μ,ν be positive Radon measures on $\mathbb{R}$.

(a) Suppose that condition (A) holds. Then the quadratic form $E_\mu - P_\nu$ is l.s.b. and closed on $L^2(\mathbb{R},dx)$.

(b) Suppose that $\mu - \nu$ is a signed Radon measure or that $\mu(\{x \mid d(x,\text{support } \nu) < \epsilon\}) < \infty$, for some $\epsilon > 0$ (where d means distance). Let $\{I_n\}$ be a sequence of pairwise disjoint intervals so that $\mathbb{R} = \cup_{n\in\mathbb{N}} I_n$, $\inf_{n\in\mathbb{N}} |I_n| > 0$, and $\sup_{n\in\mathbb{N}} |I_n| < \infty$. Then the quadratic form $E_\mu - P_\nu$ on $L^2(\mathbb{R},dx)$ is l.s.b. if and only if $\sup_{n\in\mathbb{N}} \nu(I_n) < \infty$.

Remark: As mentioned in the Introduction, this result implies that $-(d^2/dx^2) + \mu - \nu$, defined by the method of quadratic forms as the operator uniquely associated with the closed form $E_\mu - P_\nu$, is l.s.b. and self-adjoint, whenever the assumption (A), which only involves a *local* estimate for ν, is satisfied. This extends a result of Ref. 14 to the case of measure valued potentials.

As an application of Theorems 2 and 3, we give the following example, which extends a result of Ref. 5 on perturbations of the free Hamiltonian by infinite sums of delta functions to the case of strengths of both signs and not restricted to be bounded.

Example 1: Let $\{x_n\}$, $\{m_n\}$ be sequences in $\mathbb{R}$, so that for some $\epsilon > 0$, $|x_n - x_m| \geqslant \epsilon$, if $n \neq m$. Let

$$\mu(x) = \sum_{m_n > 0} m_n\,\delta(x - x_n),$$

$$\nu(x) = \sum_{m_n < 0} m_n\,\delta(x - x_n),$$

where δ means delta function.

By Theorem 3, the quadratic form $E_\mu - P_\nu$ is l.s.b. if and only if $\inf_{n\in\mathbb{N}} m_n > -\infty$. In this case $E_\mu - P_\nu$ is closed on $L^2(\mathbb{R},dx)$ (Theorem 3) and the energy form $E_{\mu-\nu} = E_\mu - P_\nu$ is given by

$$D(E_{\mu-\nu}) = \left\{ f \in H^{2,1}(\mathbb{R}) \,\middle|\, \sum_{n=1}^{\infty} m_n |f(x_n)|^2 < \infty \right\},$$

$$E_{\mu-\nu}(f,g) = \int_{\mathbb{R}} f'^*(x) g'(x)\,dx + \sum_{n=1}^{\infty} m_n f^*(x_n) g(x_n),$$
$$f,g \in D(E_{\mu-\nu}).$$

Let $f \in D(E_{\mu-\nu})$. By Theorem 2, $f \in D(H_{\mu-\nu})$ if and only if the distribution

$$g \to \int_{\mathbb{R}} f^*(x)(-g''(x))\,dx + \sum_{n=1}^{\infty} m_n f^*(x_n) g(x_n), \quad g \in C_0^\infty(\mathbb{R}),$$

is in $L^2(\mathbb{R},dx)$. It is straightforward to see that this is equivalent to

$$f \upharpoonright \mathbb{R} \setminus \{x_n \mid n \in \mathbb{N}\} \in H^{2,2}(\mathbb{R} \setminus \{x_n \mid n \in \mathbb{N}\}),$$

and $f'(x_n +) - f'(x_n -) = m_n f(x_n)$ for any $n \in \mathbb{N}$. Thus, we have an explicit description of the functions in the domain of the Hamiltonian

$$-\frac{d^2}{dx^2} + \sum_{n=1}^{\infty} m_n \delta(\cdot - x_n),$$

by boundary conditions at the singular points x_n. Moreover, by Theorem 2,

$$H_{\mu-\nu} f = \Big(-f'' + \sum_{n=1}^{\infty} m_n f(x_n) \delta(\cdot - x_n)\Big)_{\text{dist}}.$$

Thus, $H_{\mu-\nu} f = -(f \upharpoonright \mathbb{R} \setminus \{x_n \mid n \in \mathbb{R}\})''$.

We shall now prove a corresponding result under a condition different from (A), which involves both μ and ν, but allows, in contrast to assumption (A) of the preceding theorem, for arbitrary growth of ν at infinity.

Let I be a bounded interval on $\mathbb{R}$. Let $f \in H^{2,1}(\mathbb{R})$. Suppose that for some number $r > 0$ we have $\mu(I) - \nu(I) \geqslant -r|I|$. Then we have the elementary estimates

$$\begin{aligned}
\int_I |f(x)|^2\,d(\mu - \nu)(x) &\geqslant (\mu(I) - \nu(I)) \inf_{x\in I} |f(x)|^2 \\
&\quad + \nu(I)((\inf_{x\in I} |f(x)|)^2 - (\sup_{x\in I} |f(x)|)^2) \\
&\geqslant -r \int_I |f(x)|^2\,dx - 2\nu(I) \int_I |f'(x)|\,dx \|f\|_\infty \\
&\geqslant -r \int_I |f(x)|^2\,dx - 2\nu(I)|I|^{1/2} \\
&\quad \times \Big(\int_I |f'(x)|^2\,dx\Big)^{1/2} \|f\|_\infty.
\end{aligned} \tag{6}$$

The last step follows from Hölder's inequality. Sobolev's inequality and (6) yield that for any $a>0$ there is a number b independent of f and I, so that

$$\left|\int_I |f(x)|^2\, d(\mu-\nu)(x)\right|$$
$$\leqslant r\int_I |f(x)|^2\,dx + \left(aE_0(f,f) + b\int_{\mathbb{R}} |f(x)|^2\,dx\right)|I|^{1/2}\sup(\nu(I),\mu(I)),$$
$$\text{if } |\mu(I)-\nu(I)|\leqslant r|I|. \tag{7}$$

Suppose now that there is a sequence $\{I_n\}$ of pairwise disjoint bounded intervals and a number r, so that

$$\mathbb{R}=\bigcup_{n\in\mathbb{N}} I_n,\quad |\mu(I_n)-\nu(I_n)|\leqslant r|I_n|,$$

for any $n\in\mathbb{N}$ and

$$\sum_{n=1}^{\infty} |I_n|^{1/2}\sup(\mu(I_n),\nu(I_n))<\infty.$$

We define a quadratic form $P_{\mu-\nu}$ on $L^2(\mathbb{R},dx)$ by

$$D(P_{\mu-\nu})=H^{2,1}(\mathbb{R}),$$
$$P_{\mu-\nu}(f,g)=\sum_{n=1}^{\infty}\int_{I_n} f^*(x)g(x)\times d(\mu-\nu)(x),\quad f,g\in D(P_{\mu-\nu}).$$

Let $\tilde{\mu}$ be a positive Radon measure on $\mathbb{R}$. By (7), $P_{\mu-\nu}$ is infinitesimally form-bounded with respect to $E_{\tilde{\mu}}$. Thus, by the KLMN theorem and Theorem 1(a), $E_{\tilde{\mu}}+P_{\mu-\nu}$ is l.s.b. and closed on $L^2(\mathbb{R},dx)$ and $C_0^\infty(\mathbb{R})$ is a core of $E_{\tilde{\mu}}+P_{\mu-\nu}$. This implies that $E_{\tilde{\mu}}+P_{\mu-\nu}$ is the closure of the quadratic form $E_{\tilde{\mu}+\mu}-P_\nu$ (i.e., $E_{\tilde{\mu}}+P_{\mu-\nu}=E_{\tilde{\mu}+\mu-\nu}$) since $E_{\tilde{\mu}+\mu}-P_\nu\subset E_{\tilde{\mu}}+P_{\mu-\nu}$ and $C_0^\infty(\mathbb{R})\subset D(E_{\tilde{\mu}+\mu}-P_\nu)$. Hence we are lead to the following theorem.

Theorem 4: Let μ,ν be positive Radon measures on $\mathbb{R}$. Suppose the following condition holds.

(B) There is a sequence $\{I_n\}$ of pairwise disjoint intervals and a number r, so that $\mathbb{R}=\bigcup_{n\in\mathbb{N}} I_n$, $\Sigma_{n=1}^{\infty}|I_n|^{1/2}\nu(I_n)<\infty$, and $\mu(I_n)-\nu(I_n)\geqslant -r(I_n)$, for any $n\in\mathbb{N}$. Then the quadratic form $E_\mu-P_\nu$ is l.s.b. and closable on $L^2(\mathbb{R},dx)$ and its closure $E_{\mu-\nu}$ is given by

$$D(E_{\mu-\nu})=\left\{f\in H^{2,1}(\mathbb{R})\,\middle|\,\sum_{n=1}^{\infty}\left|\int_{I_n}\right| |f(x)|^2\,d(\mu-\nu)(x)<\infty\right\},$$
$$E_{\mu-\nu}(f,g)=E_0(f,g)+\sum_{n=1}^{\infty}\int_{I_n} f^*(x)g(x)d(\mu-\nu)(x),\quad f,g\in D(E_{\mu-\nu}).$$

Moreover, if μ_1,μ_2 are positive Radon measures on $\mathbb{R}$, so that

$$\mu=\mu_1+\mu_2,\ |\mu_1(I_n)-\nu(I_n)|\leqslant r|I_n|,\quad \text{for any } n\in\mathbb{N}, \tag{8}$$

and

$$\sum_{n=1}^{\infty}\mu_1(I_n)|I_n|^{1/2}<\infty,$$

then $D(E_{\mu-\nu})=D(E_{\mu_2})$.

Proof: By the considerations preceding the statement of the theorem, we have only to show that there are positive Radon measures μ_1,μ_2 on $\mathbb{R}$ which satisfy (8). (Take $\tilde{\mu}=\mu_2$ and $\mu=\mu_1$ in the above arguments.) For this it is sufficient to define, for $x\in I_n$,

$$\mu_1(x)=\begin{cases}[\nu(I_n)/\mu(I_n)]\mu(x), & \text{if } \nu(I_n)<\mu(I_n),\\ \mu(x), & \text{if } \nu(I_n)\geqslant\mu(I_n),\end{cases}$$

and $\mu_2=\mu-\mu_1$. ■

Remark: It is possible that the quadratic form $E_\mu-P_\nu$ is l.s.b. and closable but not closed, and the domain of the closure contains functions which are neither in $L^2(\mathbb{R},d\mu)$ nor in $L^2(\mathbb{R},d\nu)$. Thus the definition of $E_{\mu-\nu}$, through the infinite sum in Theorem 4, is actually necessary, since for such functions $\int_{\mathbb{R}}|f(x)|^2\,d(\mu-\nu)(x)$ is not defined. Moreover, we remark that $E_\mu-P_\nu$ can be l.s.b. and closable on $L^2(\mathbb{R},dx)$ even in the case where E_0-P_ν is not l.s.b. and support $\mu\cup$ support $\nu=\varnothing$, i.e., the measures μ and ν can "cancel out" each other even if their supports are disjoint. The following example will illustrate the facts mentioned in this remark.

Example 2: Let $\{\epsilon_n\}$ be a sequence in the open interval $(0,1)$ so that $\Sigma_{n=1}^{\infty}\epsilon_n^{1/2}a^n<\infty$, where $a>1$ is an arbitrary real number. Let

$$\mu(x)=\sum_{n=1}^{\infty}a^n\,\delta(x-n),$$
$$\nu(x)=\sum_{n=1}^{\infty}a^n\,\delta(x-n-\epsilon_n).$$

The assumption (B) of Theorem 4 and formula (8) are obviously satisfied if we choose $I_1=(-\infty,1)$, $I_{2n}=[n,n+\epsilon_n]$, and $I_{2n+1}=(n+\epsilon_n,n+1)$ for any $n\in\mathbb{N}$, $\mu_1=\mu$, and $\mu_2=0$. Thus, by Theorem 4, the quadratic form $E_\mu-P_\nu$ is l.s.b. and closable on $L^2(\mathbb{R},dx)$ and its closure $E_{\mu-\nu}$ is given by

$$D(E_{\mu-\nu})=H^{2,1}(\mathbb{R}),$$
$$E_{\mu-\nu}(f,g)=\int_{\mathbb{R}} f'(x)^*g'(x)dx+\sum_{n=1}^{\infty}a^n[f^*(n)g(n)-f^*(n+\epsilon_n)g(n+\epsilon_n)],\quad f,g\in D(E_{\mu-\nu}).$$

The function $f(x)=a^{-|x|/2}$ is in $D(E_{\mu-\nu})=H^{2,1}(\mathbb{R})$, but

$$\int_{\mathbb{R}}|f(x)|^2\,d\mu(x)=\int_{\mathbb{R}}|f(x)|^2\,d\nu(x)=\infty,$$

and thus $f\notin D(E_\mu-P_\nu)$ so that, in particular, $E_\mu-P_\nu$ is not closed. In this example, E_0-P_ν is not l.s.b., and by definition the supports of μ and ν are disjoint. Hence all points made above are verified. The Hamiltonian

$$-\frac{d^2}{dx^2}+\sum_{n=1}^{\infty}a^n(\delta(\cdot-n)-\delta(\cdot-n-\epsilon_n))$$

is of the same class as the Hamiltonians discussed in more detail in example 1.

ACKNOWLEDGMENT

I am grateful to Professor Dr. S. Albeverio, who induced me to write this paper. I thank him for his interest and helpful discussions.

[1]S. Albeverio, F. Gesztesy, R. Høegh-Krohn, and H. Holden, "Some exactly solvable models in quantum mechanics and the low energy expansion," in *Proceedings of the Conference on Operator Algebras*, edited by G. Lassner (Teubner, Stuttgart, to be published).
[2]S. Albeverio, F. Gesztesy, R. Høegh-Krohn, and S. Kirsch, "On point interactions in one dimension" to appear in J. Operator Theory.
[3]J. Brasche, *Störungen von Schrödingeroperatoren durch Masse* (Staatsexamensarbeit, Bielefeld, 1983).
[4]C. N. Friedman, J. Funct. Anal. **10**, 346 (1975).
[5]S. Kirsch and F.Martinelli, Commun.Math. Phys. **85**, 329 (1982).
[6]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II.
[7]L. Streit, Phys. Rep. **77** (1981).
[8]J. Zorbas, J. Math. Phys. **21**, 840 (1980).
[9]S. Albeverio, J. E.Fenstad, and R. Høegh-Krohn, Trans. Am. Math. Soc. **252**, 275 (1979).
[10]J. D. Dollard and C. N. Friedman, "Product integration," in *Encyclopedia of Mathematics and Its Applications* (Addison–Wesley, Reading, MA, 1979).
[11]W. G. Faris, "Self-adjoint operators," in *Lecture Notes in Mathematics*, Vol. 433 (Springer, Berlin, 1975).
[12]Prime means derivative.
[13]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1976), 2nd ed.
[14]J. D. Morgan, III, J. Operator Theory **1**, 109 (1979).
[15]Asterisk means conjugate complex.
[16]B. Simon, J. Funct. Anal. **28**, 377 (1978), Theorem 3.1.
[17]J. Wiedmann, *Lineare Operatoren in Hilberträumen* (Teubner, Stuttgart, 1976).
[18]S. Albeverio, R. Høegh-Krohn, and L. Streit, J. Math. Phys. **18**, 907 (1977).
[19]M. Röckner and N. Wielens, "Dirichlet forms—Closability and change of speed measure," Bielefeld preprint, 1983.
[20]J. M. Hamza, Thèse de le cycle, Orsay, 1975).
[21]K. Rullkötter and U. Spönemann, *Dirichletformen und Diffusionsprozesse* (Diplomarbeit, Bielefeld, 1983).

Sums of products of ultraspherical functions

Mizan Rahman and Mihr J. Shah
Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada K1S 5B6

(Received 1 June 1984; accepted for publication 28 September 1984)

Analytical expressions for the sum $\Sigma_{n=0}^{\infty}(n+\lambda)[\Gamma(n+1)/\Gamma(n+2\lambda)]^2 \times C_n^\lambda(x)C_n^\lambda(y)C_n^\lambda(z)D_n^\lambda(u)$ are given, where C_n^λ and D_n^λ are ultraspherical polynomials and functions of the second kind, respectively, on the sets $\{|x|,|y|,|z|<1, u>1\}$ and $\{|x|,|y|,|z|,|u|<1\}$.

I. INTRODUCTION

Integrals and sums of products of classical orthogonal polynomials have been a matter of curiosity for a long time (see the Bateman Manuscript Project[1] for an impressive list of integrals and Hansen[2] for a similar list of sums). Some of these formulas have proved very useful in physical applications. For example, the Clebsch–Gordan coefficients with zero magnetic numbers have the following representation[3,4]:

$$(C_{k0m0}^{n0})^2 = \frac{2n+1}{2}\int_{-1}^{1} P_k(x)P_m(x)P_n(x)\,dx, \tag{1.1}$$

where $P_k(x)$ is the Legendre polynomial of degree k. By symmetry of the integral (1.1) and the orthogonality properties of these polynomials, it is clear that the integral vanishes if $k+m+n$ is odd or the triangle inequality

$$|n-m| \leqslant k \leqslant n+m \tag{1.2}$$

is not satisfied.

In trying to classify the various fluctuation modes in an investigation of the stability properties of some special solutions of the $O(n)$ nonlinear σ model, Din and Zakrzewski[5] found that the classification depended crucially on the vanishing of the sum

$$\sum_{k=|n-m|}^{n+m} \frac{2k+1}{k(k+1)-m(m+1)}(C_{k0m0}^{n0})^2.$$

Din[4] showed that this vanishing is equivalent to that of

$$\int_{-1}^{1} Q_k(x)P_m(x)P_n(x)\,dx, \tag{1.3}$$

when either $k+m+n$ is even or $k+m+n$ is odd and $|n-m|<k<n+m$, $Q_k(x)$ being the Legendre function of the second kind[6] on the branch cut $-1<x<1$. It is interesting to note that the integrals in (1.1) and (1.3) vanish in complementary sets. The same feature holds for the well-known generalization[7] of (1.1),

$$\int_{-1}^{1} C_k^\lambda(x)C_m^\lambda(x)C_n^\lambda(x)(1-x^2)^{\lambda-1/2}\,dx = \frac{\Gamma(\frac12)\Gamma(\lambda+\frac12)}{\Gamma(\lambda+1)}\,\frac{(\lambda)_{s-k}\lambda_{s-m}(\lambda)_{s-n}(2\lambda)_s}{(s-k)!(s-m)!(s-n)!(\lambda+1)_s}, \tag{1.4}$$

where $k+m+n=2s$ is even, and $|n-m|\leqslant k\leqslant n+m$, and zero otherwise. The corresponding generalization of (1.3) has recently been found by Askey, Koornwinder, and Rahman,[8] which states that the integral

$$\int_{-1}^{1} D_k^\lambda(x)C_m^\lambda(x)C_n^\lambda(x)(1-x^2)^{2\lambda-1}\,dx \tag{1.5}$$

vanishes when (i) $k+m+n$ is even, and (ii) $k+m+n$ is odd and $|n-m|<k<n+m$. Here $C_k^\lambda(x)$ and $D_k^\lambda(x)$ are the ultraspherical polynomials and ultraspherical functions of the second kind defined, respectively, by[9]

$$\begin{aligned} C_k^\lambda(\cos\theta) &= \sum_{j=0}^{k}\frac{(\lambda)_j(\lambda)_{k-j}}{j!(k-j)!}\cos(k-2j)\theta \\ &= \sin^{1-2\lambda}\theta\,\frac{2\Gamma(\lambda+\frac12)(2\lambda)_k}{\Gamma(\lambda+1)\Gamma(\frac12)k!} \\ &\quad\times\sum_{j=0}^{\infty}\frac{(1-\lambda)_j(1)_{k+j}}{j!(\lambda+1)_{k+j}} \\ &\quad\times\sin(k+2j+1)\theta,\quad 0<\theta<\pi,\quad \operatorname{Re}\lambda>0, \end{aligned} \tag{1.6}$$

and

$$\begin{aligned} D_k^\lambda(\cos\theta) &= \sin^{1-2\lambda}\theta\,\frac{2\Gamma(\lambda+\frac12)(2\lambda)_k}{\Gamma(\lambda+1)\Gamma(\frac12)k!} \\ &\quad\times\sum_{j=0}^{\infty}\frac{(1-\lambda)_j(1)_{k+j}}{j!(\lambda+1)_{k+j}} \\ &\quad\times\cos(k+2j+1)\theta,\quad 0<\theta<\pi,\quad \operatorname{Re}\lambda>0. \end{aligned} \tag{1.7}$$

The shifted factorial $(a)_k$ is defined by $(a)_k=1$, $k=0$, and $(a)_k=a(a+1)\cdots a(a+k-1)$, $k=1,2,\ldots$.

The duals of these results are equally interesting. Dougall[7] showed that

$$\begin{aligned} \sum_{n=0}^{\infty}(n+\lambda)\left[\frac{\Gamma(n+1)}{\Gamma(n+2\lambda)}\right]^2 & C_n^\lambda(\cos\alpha) \\ &\times C_n^\lambda(\cos\beta)C_n^\lambda(\cos\gamma) \\ &= \frac{\pi(\sin\alpha\sin\beta\sin\gamma)^{1-2\lambda}(16D)^{\lambda-1}}{2^{2\lambda}\Gamma^4(\lambda)}, \end{aligned} \tag{1.8}$$

where

$$16D = \sin\frac{\alpha+\beta+\gamma}{2}\sin\frac{\beta+\gamma-\alpha}{2}\times\sin\frac{\gamma+\alpha-\beta}{2}\sin\frac{\alpha+\beta-\gamma}{2}, \tag{1.9}$$

provided $0<\alpha$, β, $\gamma<\pi$, $0<\operatorname{Re}\lambda$, and a triangle can be drawn with sides α,β,γ, assuming that the sum of any two of them is less than or equal to π. The value of the infinite sum in (1.8) is zero if this triangle condition is not satisfied. On the other hand, the dual of (1.5) states that[10]

 0022-2488/85/040627-06$02.50

$$\sum_{n=0}^{\infty}(n+\lambda)\frac{\Gamma(n+1)}{\Gamma(n+2\lambda)}D_n^\lambda(\cos\alpha)C_n^\lambda(\cos\beta)C_n^\lambda(\cos\gamma)=\frac{2^{1-4\lambda}\sin\pi\lambda}{\Gamma^2(\lambda)}\begin{cases}0, & \text{if} |\beta-\gamma|<\alpha<\beta+\gamma\leqslant\pi,\\ (-16D)^{-\lambda}, & \text{if}\ \alpha<|\beta-\gamma| \text{ or } \pi<\beta+\gamma<2\pi,\\ & \text{and}\ \alpha+\beta+\gamma<2\pi\\ -(-16D)^{-\lambda}, & \text{if}\ \alpha>\beta+\gamma,\end{cases} \tag{1.10}$$

where $0<\alpha, \beta, \gamma<\pi$. Convergence of the infinite series on the left hand side of (1.10) requires that $0<\operatorname{Re}\lambda<1$.

Generalizations of these results for Jacobi polynomials $P_n^{(\alpha,\beta)}(x)$ and Jacobi functions of the second kind $Q_n^{(\alpha,\beta)}(x)$ (see Ref. 11) would seem to be a natural thing to try. A generalization of (1.4) was given by Rahman[12] and of (1.8) by Gasper.[13] Recently van Haeringen[14,15] considered a family of infinite sums of the form

$$\sum_{n=0}^{\infty}(2n+1)P_n^\alpha(x)P_n^\beta(y)P_n^\gamma(z)Q_n^\mu(u), \tag{1.11}$$

where P_n^α and Q_n^μ are Legendre functions of the first and second kinds, respectively, defined by[16]

$$P_\nu^\mu(z)=\frac{1}{\Gamma(1-\mu)}\left(\frac{z+1}{z-1}\right)^{\mu/2}\times{}_2F_1(-\nu,\nu+1;1-\mu;\tfrac{1}{2}-\tfrac{1}{2}z),\quad |(1-z)/2|<1, \tag{1.12}$$

$$Q_\nu^\mu(z)=e^{\pi i\mu}\,2^{-\nu-1}\,\pi^{1/2}\frac{\Gamma(\nu+\mu+1)}{\Gamma(\nu+\frac{3}{2})}\times z^{-\nu-\mu-1}(z^2-1)^{\mu/2}\times{}_2F_1\left(\frac{\nu+\mu+1}{2},\frac{\nu+\mu+2}{2};\nu+\tfrac{3}{2};z^{-2}\right),|z|>1. \tag{1.13}$$

For example, van Haeringen[15] showed that

$$\sum_{n=0}^{\infty}(2n+1)P_n(x)P_n(y)P_n(z)Q_n(u)=W^{-1/2}{}_2F_1(\tfrac{1}{4},\tfrac{3}{4};1;TW^{-2}), \tag{1.14}$$

where

$$W=x^2+y^2+z^2+u^2-2xyzu-2,$$
$$T=4(1-x^2)(1-y^2)(1-z^2)(1-u^2),$$

provided $-1\leqslant x, y, z\leqslant 1$, and u is outside an ellipse with arbitrary axes and foci at ± 1. He restricted all his calculations in Ref. 15 to the region where $\operatorname{Re} u$ is positive and sufficiently large in order to avoid difficulties arising from the branch cuts. Even though van Haeringen's work seems to be motivated by the unitarity relation for the Coulomb T matrix, it should be clear from (1.10) that very interesting things can happen when $|u|<1$. Of course, (1.13) is not valid when $|z|<1$, but there are known prescriptions[16] for computing $Q_n^\mu(x\pm i0)$, $-1<x<1$; that is, the values of $Q_n^\mu(z)$ as z approaches the cut from above or below.

Our principal objective in this paper is to compute the sum

$$\sum_{n=0}^{\infty}(n+\lambda)\left[\frac{\Gamma(n+1)}{\Gamma(n+2\lambda)}\right]^2 D_n^\lambda(\cos\alpha)C_n^\lambda(\cos\beta)\times C_n^\lambda(\cos\gamma)C_n^\lambda(\cos\delta)\equiv J_\lambda(\alpha,\beta,\gamma,\delta),\quad\text{say.} \tag{1.15}$$

This will give an evaluation of the sum in (1.14) for $\lambda=\frac{1}{2}$ and for the case $-1<u<1$. However, in Sec. II we shall also compute the sum

$$\sum_{n=0}^{\infty}(n+\lambda)\left[\frac{\Gamma(n+1)}{\Gamma(n+2\lambda)}\right]^2 D_n^\lambda(u)C_n^\lambda(x)C_n^\lambda(y)C_n^\lambda(z)\equiv K_\lambda(x,y,z,u),\quad\text{say,} \tag{1.16}$$

for $-1\leqslant x,y,z\leqslant 1$ and $1<|u|$. This will give a direct generalization of van Haeringen's sum (1.14).

The ultraspherical functions $C_n^\lambda(z)$ and $D_n^\lambda(z)$ are defined in different ways in different regions. One of their representations on the branch cut $-1<x<1$ is given in (1.6) and (1.7). However, for complex z with $\operatorname{Im} z\neq 0$, Durand[17] defines them as follows:

$$C_n^\lambda(z)=\frac{\Gamma(n+2\lambda)}{\Gamma(2\lambda)\Gamma(n+1)}\times{}_2F_1(-n,n+2\lambda;\lambda+\tfrac{1}{2};(1-z)/2)=2^{1-2\lambda}\sqrt{\pi}\frac{\Gamma(n+2\lambda)}{\Gamma(\lambda)\Gamma(n+\lambda+\frac{1}{2})}P_n^{(\lambda-1/2,\lambda-1/2)}(z), \tag{1.17}$$

$$D_n^\lambda(z)=e^{i\pi\lambda}\frac{\Gamma(n+2\lambda)}{\Gamma(\lambda)\Gamma n+\lambda+1)}(2z)^{-n-2\lambda}\times{}_2F_1((n/2)+\lambda,(n+1/2)+\lambda;n+\lambda+1;z^{-2})=e^{i\pi\lambda}\frac{2^{1-2\lambda}}{\sqrt{\pi}}\frac{\Gamma(n+2\lambda)}{\Gamma(\lambda)\Gamma(n+\lambda+\frac{1}{2})}Q_n^{(\lambda-1/2,\lambda-1/2)}(z), \tag{1.18}$$

where $P_n^{(\alpha,\beta)}(z)$ and $Q_n^{(\alpha,\beta)}(z)$ are Jacobi polynomials and Jacobi functions of the second kind defined, respectively, by[11]

$$P_n^{(\alpha,\beta)}(z)=\frac{\Gamma(n+\alpha+1)}{\Gamma(n+1)\Gamma(\alpha+1)}\times{}_2F_1(-n,n+\alpha+\beta+1;\alpha+1;(1-z)/2), \tag{1.19}$$

$$Q_n^{(\alpha,\beta)}(z)=2^{n+\alpha+\beta}\frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(2n+\alpha+\beta+2)}\times(z-1)^{-n-\alpha-1}(z+1)^{-\beta}\times{}_2F_1(n+1,n+\alpha+1;2n+\alpha+\beta+2;2/(1-z)). \tag{1.20}$$

Durand introduces the phase factor $e^{i\pi\lambda}$ in the definition of $D_n^\lambda(z)$ so that D_n^λ and C_n^λ satisfy the same recurrence relation in λ (see Ref. 18). The ultraspherical functions for real argument x on the cut $-1<x<1$ are then defined by

$$C_n^\lambda(x)=D_n^\lambda(x+i0)+e^{-2\pi i\lambda}D_n^\lambda(x-i0)=C_n^\lambda(x\pm i0),\quad -1<x\leqslant 1, \tag{1.21}$$

$$D_n^\lambda(x)=-iD_n^\lambda(x+i0)+ie^{-2\pi i\lambda}D_n^\lambda(x-i0),\quad -1<x<1. \tag{1.22}$$

It requires a bit of manipulation to reduce (1.21) and (1.22) to (1.6) and (1.7), respectively, but the work is fairly straightforward, based on some quadratic transformation properties of the Gaussian hypergeometric function ${}_2F_1$ (see Ref. 16).

II. EVALUATION OF $K_\lambda(x,y,z,u)$

By a formula due to Feldheim,[19,20] we have

$$\sum_{n=0}^{\infty}\frac{(2n+\alpha+\beta+1)n!}{\Gamma(n+\alpha+1)}\left[\frac{\Gamma(n+\alpha+\beta+1)}{\Gamma(n+\beta+1)}\right]^2 \times P_n^{(\alpha,\beta)}(x)P_n^{(\alpha,\beta)}(y)Q_n^{(\alpha,\beta)}(u)$$
$$=\frac{2^{\alpha+\beta}\Gamma(\alpha+\beta+1)}{\Gamma(\beta+1)}(u+1)^{-\alpha-\beta-1} \times F_4\left(\alpha+\beta+1,\alpha+1;\beta+1;\alpha+1;\ \frac{(1+x)(1+y)}{2(1+u)},\frac{(1-x)(1-y)}{2(1+u)}\right), \tag{2.1}$$

where

$$F_4(a,b;c,c';x,y)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(a)_{m+n}(b)_{m+n}}{m!n!c_m(c')_n}x^my^n \tag{2.2}$$

is an Appell function.[16] The double series on the right of (2.1) converges if

$$\left|\frac{(1+x)(1+y)}{2(1+u)}\right|^{1/2}+\left|\frac{(1-x)(1-y)}{2(1+u)}\right|^{1/2}<1, \tag{2.3}$$

which we shall assume to hold. For the sake of simplicity let us assume $u>1$; this will guarantee the convergence of the infinite series on either side of (2.1) for $-1\leqslant x,y\leqslant 1$.

The parameters of the F_4 function in (2.1) are so related that it can be transformed into a ${}_2F_1$ by a formula due to Bailey[21]

$$F_4(\alpha+\beta+1,\alpha+1;\beta+1,\alpha+1;\ -s/(1-s)(1-t),-t/(1-s)(1-t))$$
$$=(1-t)^{\alpha+\beta+1}{}_2F_1\left(\alpha+\beta+1,\alpha+1;\beta+1;\frac{s(t-1)}{1-s}\right). \tag{2.4}$$

Furthermore, denoting $v=s(t-1)/(1-s)$ and using the quadratic transformation formula[16]

$${}_2F_1(\alpha_1,\alpha_2;1-\alpha_2+\alpha_1;v)=(1+v)^{-\alpha_1}{}_2F_1\left(\frac{\alpha_1}{2},\frac{\alpha_1+1}{2};1-\alpha_2+\alpha_1;\frac{4v}{(1+v)^2}\right), \tag{2.5}$$

one can show after some tedious but straightforward calculations that

$$\sum_{n=0}^{\infty}\frac{(2n+\alpha+\beta+1)n!}{\Gamma(n+\alpha+1)}\left[\frac{\Gamma(n+\alpha+\beta+1)}{\Gamma(n+\beta+1)}\right]^2 P_n^{(\alpha,\beta)}(x)P_n^{(\alpha,\beta)}(y)Q_n^{(\alpha,\beta)}(u)$$
$$=\frac{2^{\alpha+\beta}\Gamma(\alpha+\beta+1)}{\Gamma(\beta+1)}(1+x+y+u)^{-\alpha-\beta-1}{}_2F_1\left(\frac{\alpha+\beta+1}{2},\frac{\alpha+\beta+2}{2};\beta+1;\frac{2(1+x)(1+y)(1+u)}{(1+x+y+u)^2}\right). \tag{2.6}$$

This assumes a particularly simple form in the ultraspherical case $\alpha=\beta$. Replacing α and β by $\lambda-\frac{1}{2}$ and using (1.17) and (1.18) we then obtain

$$\sum_{n=0}^{\infty}(n+\lambda)\frac{\Gamma(n+1)}{\Gamma(n+2\lambda)}C_n^\lambda(x)C_n^\lambda(y)D_n^\lambda(u)=[e^{i\pi\lambda}/2^{2\lambda}\Gamma^2(\lambda)](x^2+y^2+u^2-2xyu-1)^{-\lambda}, \tag{2.7}$$

which is valid for $\operatorname{Re}\lambda>0$ and $-1\leqslant x,y\leqslant 1$, $u>1$.

To compute $K_\lambda(x,y,z,u)$ we need Gegenbauer's product formula[22]

$$C_n^\lambda(x)C_n^\lambda(y)=\Gamma(n+2\lambda)/2^{2\lambda-1}\Gamma^2(\lambda)\Gamma(n+1)\int_0^\pi C_n^\lambda(xy+\sqrt{1-x^2}\sqrt{1-y^2}\cos\phi)(\sin\phi)^{2\lambda-1}\,d\phi. \tag{2.8}$$

Thus, from (1.16), (2.7), and (2.8) we get

$$K_\lambda(x,y,z,u)=[e^{i\pi\lambda}/2^{4\lambda-1}\Gamma^4(\lambda)]\times\int_0^\pi(f+2g\cos\phi+h\cos^2\phi)^{-\lambda}\times(\sin\phi)^{2\lambda-1}\,d\phi, \tag{2.9}$$

where

$$f=z^2+u^2+x^2y^2-2xyzu-1,\quad g=\sqrt{(1-x^2)(1-y^2)}\,(xy-zu),\quad h=(1-x^2)(1-y^2), \tag{2.10}$$

with $-1\leqslant x,y,z\leqslant 1$, $u>1$. The integral in (2.9) is evaluated in the Appendix. Using (A18) we then have

$$K_\lambda(x,y,z,u)=[e^{i\pi\lambda}/2^{2\lambda}\Gamma^2(\lambda)\Gamma(2\lambda)]\times W^{-\lambda}{}_2F_1((\lambda/2),(\lambda+1)/2;\lambda+\tfrac{1}{2};TW^{-2}), \tag{2.11}$$

where T and W were defined in the previous section. Setting $\lambda=\frac{1}{2}$ and noting that from the definitions (1.17) and (1.18), $C_n^{1/2}(z)=P_n(z)$ and $D_n^{1/2}(z)=(e^{\pi i/2}/\pi)Q_n(z)$, one can see that (2.11) agrees with (1.14).

III. EVALUATION OF $J_\lambda(\alpha,\beta,\gamma,\delta)$

As we shall see now, the evaluation of $J_\lambda(\alpha,\beta,\gamma,\delta)$ which is essentially the same as $K_\lambda(x,y,z,u)$ when x, y, z, u are all restricted to $(-1,1)$ is much more difficult. We can no longer use (2.7) since it is valid when $u>1$. Rather, we must use (1.10) whose derivation in Ref. 10 needed very careful use of the jump conditions (1.21) and (1.22).

One can deduce from (1.10) that, for $0<\operatorname{Re}\lambda<1$,

$$\sum_{n=0}^{\infty}(n+\lambda)\frac{\Gamma(n+1)}{\Gamma(n+2\lambda)}C_n^\lambda(y)C_n^\lambda(z)D_n^\lambda(u)=\begin{cases}0, & \text{if } yz-\sqrt{(1-y^2)(1-z^2)}<u<yz+\sqrt{(1-y^2)(1-z^2)},\\ \pm\frac{2^{2\lambda-1}}{\pi}\left[\frac{\Gamma(\lambda+\frac{1}{2})}{\Gamma(2\lambda)}\right]^2\sin\pi\lambda\,\{(u-yz)^2-(1-y^2)(1-z^2)\}^{-\lambda}, & \\ & \text{if } u<yz-\sqrt{(1-y^2)(1-z^2)} \text{ or } u>yz+\sqrt{(1-y^2)(1-z^2)}.\end{cases}\tag{3.1}$$

We now replace y by $xy+\sqrt{(1-x^2)(1-y^2)}\cos\phi$, $(x,y\neq\pm 1)$, multiply both sides by $(\sin\phi)^{2\lambda-1}$, and integrate with respect to ϕ from 0 to π. By (3.1), the integrand on the right-hand side is zero if

$$|xyz+\sqrt{(1-x^2)(1-y^2)}\,z\cos\phi-u|<\{1-(xy+\sqrt{(1-x^2)(1-y^2)}\cos\phi)^2\}^{1/2}(1-z^2),$$

which is equivalent to the inequality $x_1<\cos\phi<x_2$, where

$$x_1=[zu-xy-\sqrt{(1-z^2)(1-u^2)}]/\sqrt{(1-x^2)(1-y^2)},\quad x_2=[zu-xy+\sqrt{(1-z^2)(1-u^2)}]/\sqrt{(1-x^2)(1-y^2)}.\tag{3.2}$$

Also, we take the positive sign in (3.1) when $\cos\phi<x_1$ and the negative sign when $\cos\phi>x_2$.

Let us set

$$u=\cos\alpha,\quad x=\cos\beta,\quad y=\cos\gamma,\quad z=\cos\delta,\tag{3.3}$$

$0<\alpha,\beta,\gamma,\delta<\pi$. Then (3.2) gives

$$x_1=[\cos(\alpha+\delta)-\cos\beta\cos\gamma]/\sin\beta\sin\gamma,\quad x_2=[\cos(\alpha-\delta)-\cos\beta\cos\gamma]/\sin\beta\sin\gamma.\tag{3.4}$$

There are six possibilities.

(i) $x_1,x_2<-1$, which gives $\cos(\alpha\pm\delta)<\cos(\beta+\gamma)$.

(ii) $x_1<-1$, $-1<x_2<1$, which implies $\cos(\alpha+\delta)<\cos(\beta+\gamma)<\cos(\alpha-\delta)<\cos(\beta-\gamma)$.

(iii) $-1<x_1,x_2<1$, which means $\cos(\beta+\gamma)<\cos(\alpha\pm\delta)<\cos(\beta-\gamma)$.

(iv) $x_1<-1, x_2>1\Rightarrow\cos(\alpha+\delta)<\cos(\beta+\gamma)$, $\cos(\alpha-\delta)>\cos(\beta-\gamma)$.

(v) $-1<x_1<1, x_2>1\Rightarrow\cos(\beta+\gamma)<\cos(\alpha+\delta)<\cos(\beta-\gamma)<\cos(\alpha-\delta)$.

(vi) $x_1,x_2>1\Rightarrow\cos(\beta-\gamma)<\cos(\alpha\pm\delta)$.

Clearly, the contributions to $J_\lambda(\alpha,\beta,\gamma,\delta)$ in these six regions arise from three integrals

$$\int_{-1}^{1}(1-t^2)^{\lambda-1}[(x_1-t)(x_2-t)]^{-\lambda}\,dt,\quad \text{if } |x_1|,|x_2|>1,$$

$$\int_{-1}^{x_1}(1-t^2)^{\lambda-1}[(x_1-t)(x_2-t)]^{-\lambda}\,dt,\quad \text{if } \cos\phi<x_1<1,$$

and

$$\int_{x_2}^{1}(1-t^2)^{\lambda-1}[(x_1-t)(x_2-t)]^{-\lambda}\,dt,\quad \text{if } \cos\phi>x_2,$$

where $f+2gt+ht^2=h(x_1-t)(x_2-t)$. These integrals are evaluated in the Appendix. From (2.10) and (3.3)

$$f-h=x^2+y^2+z^2+u^2-2xyzu-2=\cos^2\alpha+\cos^2\beta+\cos^2\gamma+\cos^2\delta-2\cos\alpha\cos\beta\cos\gamma\cos\delta-2=W,\tag{3.5}$$

$$4(g^2-fh)=(1-x^2)(1-y^2)(1-z^2)(1-u^2)=(\sin\alpha\sin\beta\sin\gamma\sin\delta)^2=T,\tag{3.6}$$

where W,T are the same as those introduced in (1.14) except that x,y,z,u are to be replaced by $\cos\beta$, $\cos\gamma$, $\cos\delta$, and $\cos\alpha$, respectively. If we denote

$$\frac{2^{1-2\lambda}\sin\pi\lambda}{\Gamma^2(\lambda)\Gamma(2\lambda)}W^{-\lambda}{}_2F_1\Big(\frac{\lambda}{2},\frac{\lambda+1}{2};\lambda+\frac{1}{2};TW^{-2}\Big)\equiv G_\lambda(\alpha,\beta,\gamma,\delta),\tag{3.7}$$

and

$$\frac{2^{1-3\lambda}\pi}{\Gamma^4(\lambda)}T^{-\lambda/2}{}_2F_1\Big(\frac{1-\lambda}{2},\frac{\lambda}{2};1;1-\frac{W^2}{T}\Big)\equiv H_\lambda(\alpha,\beta,\gamma,\delta),\tag{3.8}$$

then we get the following evaluation of $J_\lambda(\alpha,\beta,\gamma,\delta)$:

$$J_\lambda(\alpha,\beta,\gamma,\delta)=\begin{cases}-G_\lambda(\alpha,\beta,\gamma,\delta), & \text{if } \cos(\alpha\pm\delta)<\cos(\beta+\gamma),\\ -H_\lambda(\alpha,\beta,\gamma,\delta), & \text{if } \cos(\alpha+\delta)<\cos(\beta+\gamma)<\cos(\alpha-\delta)<\cos(\beta-\gamma),\\ 0, & \text{if } \cos(\beta+\gamma)<\cos(\alpha\pm\delta)<\cos(\beta-\gamma),\\ 0, & \text{if } \cos(\alpha+\delta)<\cos(\beta+\delta),\ \cos(\alpha-\delta)<\cos(\beta-\gamma),\\ H_\lambda(\alpha,\beta,\gamma,\delta), & \text{if } \cos(\beta+\gamma)<\cos(\alpha+\delta)<\cos(\beta-\gamma)<\cos(\alpha-\delta),\\ G_\lambda(\alpha,\beta,\gamma,\delta), & \text{if } \cos(\beta-\gamma)<\cos(\alpha\pm\delta),\end{cases}\tag{3.9}$$

with $0<\operatorname{Re}\lambda<1$ and $0<\alpha,\beta,\gamma,\delta<\pi$.

ACKNOWLEDGMENT

The work of M. Rahman was supported by National Science and Engineering Research Council Grant No. A6197.

APPENDIX: EVALUATION OF INTEGRALS

We shall evaluate the integral

$$I=\int_R (f+2gt+ht^2)^{-\lambda}(1-t^2)^{\lambda-1}\,dt, \qquad \text{(A1)}$$

where R is either $[-1,1]$ or an appropriate subset thereof and f, g, h are as defined in (2.10). Clearly, $f+2gt+ht^2=h(t-x_1)(t-x_2)$, where x_1 and x_2 are given by (3.2). For (i) $x,y,z\in(-1,1)$ and $|u|>1$, x_1 and x_2 are complex conjugates, while for (ii) $x,y,z\in(-1,1)$ and $|u|<1$, x_1 and x_2 are real, with $x_1<x_2$. The integrals that need to be evaluated are

$$I_1=h^{-\lambda}\int_{-1}^{1}(1-t^2)^{\lambda-1}(t-x_1)^{-\lambda}(t-x_2)^{-\lambda}\,dt, \qquad \text{(A2)}$$

for x_1,x_2 real but $\notin(-1,1)$, or x_1,x_2 complex with $x_2=\bar{x}_1$;

$$I_2=h^{-\lambda}\int_{-1}^{x_1}(1-t^2)^{\lambda-1}(x_1-t)^{-\lambda}(x_2-t)^{-\lambda}\,dt, \qquad \text{(A3)}$$

$-1<x_1<1$; and

$$I_3=h^{-\lambda}\int_{x_2}^{1}(1-t^2)^{\lambda-1}(t-x_1)^{-\lambda}(t-x_2)^{-\lambda}\,dt, \qquad \text{(A4)}$$

$-1<x_2<1$. By simple transformations of the integration variables one can show that

$$I_1=2^{2\lambda-1}[h(1-x_1)(1-x_2)]^{-\lambda}\int_0^1 t^{\lambda-1}(1-t)^{\lambda-1}\times\left(1-\frac{2t}{1-x_1}\right)^{-\lambda}\left(1-\frac{2t}{1-x_2}\right)^{-\lambda}dt,\quad \operatorname{Re}\lambda>0, \qquad \text{(A5)}$$

$$I_2=2^{\lambda-1}[h(x_2+1)]^{-\lambda}\int_0^1 t^{\lambda-1}(1-t)^{-\lambda}\times\left(1-\frac{1+x_1}{2}t\right)^{\lambda-1}\left(1-\frac{1+x_1}{1+x_2}t\right)^{-\lambda}dt,\quad 0<\operatorname{Re}\lambda<1, \qquad \text{(A6)}$$

$$I_3=2^{\lambda-1}[h(1-x_1)]^{-\lambda}\int_0^1 t^{\lambda-1}(1-t)^{-\lambda}\times\left(1-\frac{1-x_2}{2}t\right)^{\lambda-1}\left(1-\frac{1-x_2}{1-x_1}t\right)^{-\lambda}dt,\quad 0<\operatorname{Re}\lambda<1. \qquad \text{(A7)}$$

By Eq. (5), p. 231 of Ref. 16, each of the three integrals is an F_1 Appell function defined by

$$F_1(\alpha;\beta,\beta';\gamma;x,y)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(\alpha)_{m+n}(\beta)_m(\beta')_n}{m!n!(\gamma)_{m+n}}x^m y^n. \qquad \text{(A8)}$$

Also, by Eq. (1), p. 238 of Ref. 16,

$$F_1(\alpha;\beta,\beta';\beta+\beta';x,y)=(1-y)^{-\alpha}{}_2F_1(\alpha,\beta;\beta+\beta';(x-y)/(1-y)). \qquad \text{(A9)}$$

Use of (A8) and (A9) then gives

$$I_1=2^{2\lambda-1}[\Gamma^2(\lambda)/\Gamma(2\lambda)]\,[h(x_2+1)(x_1-1)]^{-\lambda}\times{}_2F_1\left(\lambda,\lambda;2\lambda;\frac{2(x_2-x_1)}{(x_2+1)(1-x_1)}\right), \qquad \text{(A10)}$$

$$I_2=\frac{2^{\lambda-1}\pi}{\sin\pi\lambda}[h(x_2-x_1)]^{-\lambda}\times{}_2F_1\left(\lambda,1-\lambda;1;\frac{(x_1+1)(x_2-1)}{2(x_2-x_1)}\right)=I_3, \qquad \text{(A11)}$$

$0<\operatorname{Re}\lambda<1$, since $\Gamma(\lambda)\Gamma(1-\lambda)=\pi/\sin\pi\lambda$.

These formulas are further simplified by means of the following quadratic transformation formulas[16]:

$$ {}_2F_1(a,b;2b;z)=(1-z/2)^{-a}{}_2F_1\left(\frac{a}{2},\frac{a+1}{2};b+\frac{1}{2};\left(\frac{z}{2-z}\right)^2\right), \qquad \text{(A12)}$$

and

$$ {}_2F_1(a,1-a;c;z)=(1-z)^{c-1}{}_2F_1\left(\frac{c-a}{2},\frac{c+a-1}{2};c;4z(1-z)\right). \qquad \text{(A13)}$$

So, by (A12)

$$I_1=2^{2\lambda-1}[\Gamma^2(\lambda)/\Gamma(2\lambda)]\,[h(x_1x_2-1)]^{-\lambda}\times{}_2F_1\left(\frac{\lambda}{2},\frac{\lambda+1}{2};\lambda+\frac{1}{2};\left(\frac{x_2-x_1}{x_1x_2-1}\right)^2\right), \qquad \text{(A14)}$$

$$I_2=I_3=(2^{\lambda-1}\pi/\sin\pi\lambda)\,[h(x_2-x_1)]^{-\lambda}\times{}_2F_1\left(\frac{1-\lambda}{2},\frac{\lambda}{2};1;-\frac{(x_1^2-1)(x_2^2-1)}{(x_2-x_1)^2}\right). \qquad \text{(A15)}$$

From the definition of x_1 and x_2, it follows that

$$h(x_1x_2-1)=f-h, \qquad \text{(A16)}$$

$$x_2-x_1=\begin{cases}(2/h)(g^2-fh)^{1/2}, & \text{if } \max(|x|,|y|,|z|,|u|)<1,\\ (2i/h)(fh-g^2)^{1/2}, & \text{if } \max(|x|,|y|,|z|)<1,\quad u>1.\end{cases} \qquad \text{(A17)}$$

Thus,

$$I_1=2^{2\lambda-1}[\Gamma^2(\lambda)/\Gamma(2\lambda)](f-h)^{-\lambda}\times{}_2F_1\left(\frac{\lambda}{2};\frac{\lambda+1}{2};\lambda+\frac{1}{2};\frac{4(g^2-fh)}{(f-h)^2}\right), \qquad \text{(A18)}$$

and

$$I_2=I_3=(\pi/2\sin\pi\lambda)(g^2-fh)^{-\lambda/2}\times{}_2F_1\left(\frac{1-\lambda}{2},\frac{\lambda}{2};1;\frac{(f+h)^2-4g^2}{4(fh-g^2)}\right),\quad 0<\operatorname{Re}\lambda<1. \qquad \text{(A19)}$$

[1]Bateman Manuscript Project, *Tables of Integral Transforms*, edited by A.Erdélyi (McGraw-Hill, New York, 1954), Vol. II.

[2]E. R. Hansen, *A Table of Series and Products* (Prentice-Hall, Englewood Cliffs, NJ, 1975).

[3]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1960), 2nd ed.

[4]A. M. Din, Lett. Math. Phys. **5**, 207 (1981).
[5]A. M. Din and W. J. Zakrzewski, Nucl. Phys. B **186**, 173 (1980).
[6]G. Szegö, *Orthogonal Polynomials*, Am. Math. Soc. Colloq. Publ. **23** (Am. Math. Soc., Providence, RI, 1975), 4th ed.
[7]J. Dougall, Proc. Edinburgh Math. Soc. **37**, 33 (1919).
[8]R. Askey, T. H. Koornwinder, and M. Rahman, Proc. London Math. Soc., to appear.
[9]G. Szegö, Math. Ann. **110**, 501 (1934); also *Collected Papers* (Birkhäuser, Boston, 1982), Vol. II, p. 545.
[10]M. Rahman and M. J. Shah, SIAM J. Math. Anal. (to appear).
[11]Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw–Hill, New York, 1953), Vol. II.
[12]M. Rahman, Can. J. Math. **33**, 915 (1981).
[13]G. Gasper, Ann. Math. **93**, 112 (1971).
[14]H. van Haeringen and L. P. Kok, J. Math. Phys. **22**, 2482 (1981).
[15]H. van Haeringen, J. Math. Phys. **23**, 964 (1982).
[16]Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw–Hill, New York, 1953), Vol. I.
[17]L. Durand, SIAM J. Math. Anal. **9**, 76 (1978), Eqs. (3.1) and (3.2) [note a misprint in Eq. (3.2)].
[18]L. Durand, Notices Am. Math. Soc. **20**, 392 (1973).
[19]E. Feldheim, Mat. Lapok **48**, 453 (1941) (in Hungarian, French summary).
[20]M. Rahman, SIAM J. Math. Anal. **8**, 1014 (1977).
[21]W. N.Bailey, *Generalized Hypergeometric Series* (Stechert–Hafner, New York, 1964), p. 102, problem 20(v).
[22]R. Askey, *Orthogonal Polynomials and Special Functions*, Regional Conference Series in Applied Mathematics, Vol. 21 (SIAM, Philadelphia, 1975), p. 30.

Integrals of three Bessel functions and Legendre functions. I

A. Gervois[a] and H. Navelet
Service de Physique Théorique, CEN-SACLAY, 91191 Gif sur Yvette Cedex, France

(Received 20 June 1984; accepted for publication 31 August 1984)

Integrals of products of three Bessel functions of the form $\int_0^\infty t^{\lambda-1} J_\mu(at) J_\nu(bt) H_\rho^{(1)}(ct)dt$ are calculated when some relations exist between the indices λ, μ, ν, ρ: in these cases, the Appell function F_4 factorizes into two hypergeometric functions of one variable, so that analytical continuation is possible. New results are given, mainly when a, b, and c are real and positive and $|a-b|<c<a+b$, which correspond to most physical situations.

I. INTRODUCTION

Few results exist for integrals of products of three Bessel functions. The most general case, given 50 years ago by Bailey[1] is the well-known formula

$$\int_0^\infty t^{\lambda-1} J_\mu(at) J_\nu(bt) K_\rho(ct)dt$$
$$= \frac{2^{\lambda-2}a^\mu b^\nu}{c^{\lambda+\mu+\nu}} \frac{\Gamma((\lambda+\mu+\nu+\rho)/2)\Gamma((\lambda+\mu+\nu-\rho)/2)}{\Gamma(\mu+1)\Gamma(\nu+1)}$$
$$\times F_4\left(\frac{\lambda+\mu+\nu+\rho}{2}, \frac{\lambda+\mu+\nu-\rho}{2}; \mu+1, \nu+1; -\frac{a^2}{c^2}, -\frac{b^2}{c^2}\right), \tag{1.1}$$

for complex a,b,c (parameters) and λ, μ, ν, ρ (indices) provided that

$$\mathrm{Re}(\lambda+\mu+\nu\pm\rho)>0, \tag{1.2}$$
$$\mathrm{Re}(c\pm ia\pm ib)>0. \tag{1.3}$$

F_4 is the Appell function[2] which is defined as a double series inside the domain $|a|+|b|<|c|$ [which is more or less condition (1.3)]. In this work, we consider these integrals for $\lambda, \mu, \nu, \rho, a, b$, and c real. Using the formula

$$i\pi J_\rho(z) = e^{-i\pi\rho/2} K_\rho(e^{-i\pi/2}z) - e^{i\pi\rho/2} K_\rho(e^{i\pi/2}z), \tag{1.4}$$

we get the related integral for $c>a+b$,

$$\int_0^\infty t^{\lambda-1} J_\mu(at) J_\nu(bt) J_\rho(ct)dt$$
$$= 2^{\lambda-1} \frac{a^\mu b^\nu}{c^{\lambda+\mu+\nu}} \frac{\Gamma((\lambda+\mu+\nu+\rho)/2)\Gamma((\lambda+\mu+\nu-\rho)/2)}{\Gamma(\mu+1)\Gamma(\nu+1)}$$
$$\times \mathrm{Re}\left\{\frac{1}{i\pi} e^{-(i\pi/2)(\rho-\lambda-\mu-\nu)} F_4\left(\frac{\lambda+\mu+\nu+\rho}{2}, \frac{\lambda+\mu+\nu-\rho}{2}; \mu+1, \nu+1; \frac{a^2}{c^2}, \frac{b^2}{c^2}\right)\right\}, \tag{1.5a}$$
$$= \frac{2^{\lambda-1}a^\mu b^\nu}{c^{\lambda+\mu+\nu}} \frac{\Gamma((\lambda+\mu+\nu+\rho)/2)}{\Gamma(1-(\lambda+\mu+\nu-\rho)/2)} \frac{1}{\Gamma(\mu+1)\Gamma(\nu+1)}$$
$$\times F_4\left(\frac{\lambda+\mu+\nu+\rho}{2}, \frac{\lambda+\mu+\nu-\rho}{2}; \mu+1, \nu+1; \frac{a^2}{c^2}, \frac{b^2}{c^2}\right), \quad \lambda<\frac{5}{2}, \quad \lambda+\mu+\nu+\rho>0, \tag{1.5b}$$

as F_4 is real in this region.

Formulas (1.1)–(1.5) are actually not very useful for two reasons. First, function F_4 being a double series is not very tractable (especially in numerical tests). Second, integral (1.5b) is given for $c>a+b$ only, which means that it presumably does not hold when real positive parameters a,b,c can be the sides of a triangle (i.e., $|a-b|<c<a+b$), which is actually the case of interest in most physical situations.

The reason is that the behavior of function F_4 outside the convergence region $|a|+|b|<|c|$ is not well-known despite some analytical continuation properties.[3] For example, we can continue F_4 into the region $b>a+c$,[4] from

$$F_4\left(\alpha, \beta; \gamma, \gamma'; \frac{a^2}{c^2}, \frac{b^2}{c^2}\right)$$
$$= \frac{\Gamma(\gamma')\Gamma(\beta-\alpha)}{\Gamma(\gamma'-\alpha)\Gamma(\beta)} \left(e^{i\pi}\frac{b^2}{c^2}\right)^{-\alpha}$$
$$\times F_4\left(\alpha, \alpha+1-\gamma'; \gamma, \alpha+1-\beta; \frac{a^2}{b^2}, \frac{c^2}{b^2}\right)$$
$$+ \frac{\Gamma(\gamma')\Gamma(\alpha-\beta)}{\Gamma(\gamma'-\beta)\Gamma(\alpha)} \left(e^{i\pi}\frac{b^2}{c^2}\right)^{\beta}$$
$$\times F_4\left(\beta+1-\gamma', \beta; \gamma, \beta+1-\alpha; \frac{a^2}{b^2}, \frac{c^2}{b^2}\right),$$

[a] Chercheur C.N.R.S.

 0022-2488/85/040633-12$02.50

which exchanges b and c and ρ and $\pm\nu$. A similar expression holds for $a > b + c$ but not for $|a-b| > c > a+b$.

Bailey already pointed out that more could be said when the function F_4 factorizes into functions of one variable only with simpler behavior and listed them.[1] All known factorizations of F_4 imply the hypergeometric function ${}_2F_1$ which can be analytically continued in all the plane, except at most a cut along the real axis. Then, in those cases, formula (1.5a) can be continued for $|a-b| < c < a+b$. This property has been verified by quite different methods for special values of the indices such as $\lambda = 2$, $\rho = \nu - \mu$ (see Ref. 5) or $\lambda = 2 - \mu$, $\nu = \rho$ where factorization occurs.[6]

We first list all known factorizations of F_4. Some supplementary relations may be obtained by using contiguity relations but we shall not deal with them. Setting

$$\alpha = \frac{\lambda+\mu+\nu+\rho}{2}, \quad \beta = \frac{\lambda+\mu+\nu-\rho}{2},$$

$$\gamma = \mu + 1, \quad \gamma' = \nu + 1,$$

we have the following identities[1,7]:

(i) if $\alpha+\beta+1 = \gamma+\gamma'$ (i.e., $\lambda = 1$ any μ,ν,ρ), (1.6a)

$$F_4(\alpha,\beta;\gamma,\gamma';X(1-Y), Y(1-X))$$
$$= {}_2F_1(\alpha,\beta;\gamma;X)\,{}_2F_1(\alpha,\beta;\gamma';Y). \tag{1.6b}$$

(ii) $$F_4\left(\alpha,\beta;\beta,\beta;-\frac{x}{(1-x)(1-y)},-\frac{y}{(1-x)(1-y)}\right)$$
$$= (1-x)^\alpha(1-y)^\alpha\,{}_2F_1(\alpha,1+\alpha-\beta;\beta;xy), \tag{1.7a}$$

which holds provided

$$\mu = \nu, \quad \lambda = 2 \pm \rho. \tag{1.7b}$$

(iii) $$F_4\left(\alpha,\beta;1+\alpha-\beta;\beta;\right.$$
$$\left.-\frac{x}{(1-x)(1-y)},-\frac{y}{(1-x)(1-y)}\right) \tag{1.8a}$$
$$= (1-y)^\alpha\,{}_2F_1\left(\alpha,\beta;1+\alpha-\beta;-\frac{x(1-y)}{(1-x)}\right),$$

which implies

$$\pm\rho = \mu, \quad \lambda = \nu + 2$$

or (1.8b)

$$\pm\rho = \nu, \quad \lambda = \mu + 2.$$

(iv) $$F_4\left(\alpha,\beta;\alpha,\beta;-\frac{x}{(1-x)(1-y)},-\frac{y}{(1-x)(1-y)}\right)$$
$$= \frac{(1-x)^\beta(1-y)^\alpha}{(1-xy)} \tag{1.9a}$$

for

$$\lambda = 2, \quad \pm\rho = \mu - \nu, \tag{1.9b}$$

where we have explicitly used the symmetry (α,β) and (μ,ν). Cases (ii) and (iii) are nearly the same and case (iv) has already been studied.[5] They are special cases of the three-index formula:

(v) $$F_4\left(\alpha,\beta;\gamma;\beta;-\frac{x}{(1-x)(1-y)},-\frac{y}{(1-x)(1-y)}\right)$$
$$= (1-x)^\alpha(1-y)^\alpha F_1(\alpha;\gamma-\beta,1+\alpha-\gamma;\gamma;x,xy), \tag{1.10}$$

where F_1 is again an Appell's function[2] with convergence domain $|x|<1$, $|xy|<1$. Finally, we have a last formula[8]

(vi) $$F_4(\alpha,\alpha+\tfrac{1}{2};\gamma,\tfrac{1}{2};z,z')$$
$$= \tfrac{1}{2}(1+\sqrt{z'})^{-2\alpha}\,{}_2F_1(\alpha,\alpha+\tfrac{1}{2};\gamma;z/(1+\sqrt{z'})^2)$$
$$+ \tfrac{1}{2}(1-\sqrt{z'})^{-2\alpha}\,{}_2F_1(\alpha,\alpha+\tfrac{1}{2};\gamma;z/(1-\sqrt{z'})^2). \tag{1.11}$$

This latter case is merely the Fourier transform of $t^{\lambda-2}J_\mu(at)$ and can be found in any table.[9] In this paper, we deal mainly with cases (ii)–(v). Case (i) is investigated in a companion paper.[10] Actually, they are not fundamentally different as will be seen by inspection of the two families of variable transformations we study first in Sec. II:

$$(z,z')\mapsto(x,y), \quad (z,z')\mapsto(X,Y).$$

In the following other sections, we explicitly calculate the integral

$$\int_0^\infty t^{\lambda-1}J_\mu(at)J_\nu(bt)H_\rho^{(1)}(ct)\,dt, \tag{1.12}$$

$$H_\rho^{(1)}(ct) = J_\rho(ct) + iY_\rho(ct), \tag{1.13}$$

when a,b,c are real and, respectively,

$\lambda = 2 \pm \rho$, $\mu = \nu$

[Sec. III, corresponding to case (ii)],

$\lambda = \nu + 2$, $\mu = \pm\rho$

[Sec. IV, corresponding to case (iii)],

$\lambda = 2$, $\rho = \mu - \nu$

[Sec. V, corresponding to case (iv)].

Some further remarks about cases (v) and (vi) are given in the last sections. In the conclusion, we indicate some possible generalizations.

II. STUDY OF THE TWO FAMILIES OF VARIABLE TRANSFORMATIONS

We have to consider at length the changes of variables

$$\frac{-x}{(1-x)(1-y)} = \frac{a^2}{c^2}, \quad \frac{-y}{(1-x)(1-y)} = \frac{b^2}{c^2} \tag{2.1}$$

and

$$X(1-Y) = \frac{a^2}{c^2}, \quad Y(1-X) = \frac{b^2}{c^2}, \tag{2.2}$$

where a and b are two real variables and c has a small positive imaginary part. Actually by the involutive transformation[7] $(x,y)\mapsto(-X/(1-X), -Y/(1-Y))$, the family (x,y) transforms into (X,Y).

A. Family $-c^2x = a^2(1-x)(1-y)$, $-c^2y = b^2(1-x)(1-y)$

Here, x (resp. y) is a solution of

$$x^2b^2 - (a^2+b^2-c^2)x + a^2 = 0 \tag{2.3}$$
$$[\text{resp. } y^2a^2 - (a^2+b^2-c^2)y + b^2 = 0].$$

The discriminant of this second-order equation reads

$$\delta = [(a+b)^2-c^2][(a-b)^2-c^2]$$
$$= a^4+b^4+c^4-2a^2b^2-2b^2c^2-2c^2a^2,$$

where δ is positive for $c>a+b$ or $c<|a-b|$ and δ is negative for $a+b>c>|a-b|$ (triangle inequality). To remove the ambiguity when solving (2.3), we have to take into account that the F_4 reduction is valid where x and y lie in certain regions surrounding $x=0$ and $y=0$ (see Ref. 1). For $c>a+b$ and large, Eq. (2.3) has one solution which behaves as c^2/b^2, which is large and the other as a^2/c^2, which is small. As a consequence the F_4 reduction is valid if $2b^2x=a^2+b^2-c^2+\sqrt{\delta}$ for $c>a+b$, where $\sqrt{\delta}$ is positive. It is easy now to follow the determination of $\delta^{1/2}$ for $0<c^2<\infty$. Let $\delta=[Z-(a+b)^2][Z-(a-b)^2]$. In the Z plane, $\delta^{1/2}$ has a cut on the real axis for $|a-b|^2<\operatorname{Re} Z<(a+b)^2$.

We are interested in the case $\operatorname{Im} Z=2ic\eta$, for η small and positive. In this configuration

$$
\begin{aligned}
&c>a+b,\quad \delta^{1/2}=\sqrt{\delta}=4\tilde{\Delta},\\
&\qquad \operatorname{Im}\delta^{1/2}=(2\eta c/\sqrt{\delta})(c^2-a^2-b^2)>0,\\
&|a-b|<c<a+b,\quad \delta^{1/2}=e^{i\pi/2}\sqrt{-\delta}=e^{i\pi/2}\cdot 4\Delta,\\
&c<|a-b|,\quad \delta^{1/2}=e^{i\pi}\sqrt{\delta}=-\sqrt{\delta},\\
&\qquad \operatorname{Im}\delta^{1/2}=(2\eta c/\sqrt{\delta})(a^2+b^2-c^2)>0.
\end{aligned}
\tag{2.4}
$$

Note that Δ is the area of the triangle whose sides are a, b, and c. To derive the x and y variables, it is convenient to introduce angles[5]

$$
\begin{aligned}
&0<c<|a-b|,\\
&\qquad x=(a/b)e^{-u_c},\quad 2ab\cosh u_c=a^2+b^2-c^2,\\
&\qquad y=(b/a)e^{-u_c},\quad 2ab\sinh u_c=\sqrt{\delta},\\
&c>a+b,\\
&\qquad x=-(a/b)e^{-u_c},\quad 2ab\cosh u_c=c^2-a^2-b^2,\\
&\qquad y=-(b/a)e^{-u_c},\quad 2ab\sinh u_c=\sqrt{\delta},\\
&|a-b|<c<a+b,\\
&\qquad x=(a/b)e^{i\varphi_c},\quad 2ab\cos\varphi_c=a^2+b^2-c^2,\\
&\qquad y=(b/a)e^{-i\varphi_c},\quad 2ab\sin\varphi_c=\sqrt{-\delta}.
\end{aligned}
\tag{2.5}
$$

In case $|a-b|<c<a+b$,

$$
\operatorname{Im}\cos\varphi_c=-(c\eta/ab)<0. \tag{2.6}
$$

For $c<|a-b|$ or $c>a+b$, $2b^2\operatorname{Im}x=-2c\eta+\operatorname{Im}\delta^{1/2}=(2c\eta/\sqrt{\delta})2ab\,[\cosh u_c-\sinh u_c]$, which yields $\operatorname{Im}x=2c\eta/\sqrt{\delta}|x|$ positive. The same is true for $\operatorname{Im}y=(2c/\sqrt{\delta})|y|$. As a consequence,

$$
c>a+b,\quad x=e^{i\pi}(a/b)e^{-u_c},\quad y=e^{i\pi}(b/a)e^{-u_c}. \tag{2.7}
$$

To complete this study, we have also to define the variable $1-x$(resp. $1-y$), which is a solution of the equation

$$
\begin{aligned}
&b^2(1-x)^2-(b^2+c^2-a^2)(1-x)+c^2=0\\
&\qquad [\text{resp. } a^2(1-y)^2-(a^2+c^2-b^2)(1-y)+c^2=0];
\end{aligned}
\tag{2.8}
$$

$$
\begin{aligned}
&c>a+b,\quad 1-x=(c/b)e^{-u_a},\quad (1-y)=(c/a)e^{-u_b},\\
&\qquad 2bc\cosh u_a=c^2+b^2-a^2,\quad 2bc\sinh u_a=\sqrt{\delta},\\
&\qquad 2ac\cosh u_b=c^2+a^2-b^2,\quad 2ac\sinh u_b=\sqrt{\delta};
\end{aligned}
\tag{2.9a}
$$

$$
\begin{aligned}
&c<|a-b|,\quad 1-x=\operatorname{sgn}(b-a)(c/a)e^{u_a\operatorname{sgn}(b-a)},\\
&(1-y)=\operatorname{sgn}(a-b)(c/a)e^{u_b\operatorname{sgn}(a-b)},\\
&\qquad 2bc\cosh u_a=(b^2+c^2-a^2)\operatorname{sgn}(b-a),\\
&\qquad 2bc\sinh u_a=\sqrt{\delta},\\
&\qquad 2ac\cosh u_b=(a^2+c^2-b^2)\operatorname{sgn}(a-b),\\
&\qquad 2ac\sinh u_b=\sqrt{\delta},
\end{aligned}
\tag{2.9b}
$$

where

$$
\operatorname{sgn}(b-a)=-\operatorname{sgn}(a-b)=\begin{cases}+1, & \text{if } b>a,\\ -1, & \text{if } b<a;\end{cases}
$$

$$
\begin{aligned}
&|a-b|<c<a+b,\quad 1-x=\frac{c}{b}e^{-i\varphi_a},\\
&1-y=\frac{c}{a}e^{-i\varphi_b},\\
&\qquad 2bc\cos\varphi_a=b^2+c^2-a^2,\quad 2bc\sin\varphi_a=\sqrt{\delta},\\
&\qquad 2ac\cos\varphi_b=a^2+c^2-b^2,\quad 2ac\sin\varphi_b=\sqrt{\delta}.
\end{aligned}
\tag{2.9c}
$$

Note that since $x/(1-x)(1-y)=-a^2/c^2$, we get at once $e^{i\pi(\varphi_a+\varphi_b+\varphi_c)}=-1$ or

$$
\varphi_a+\varphi_b+\varphi_c=\pi. \tag{2.10}
$$

The geometrical interpretation of the φ_i's is thus straightforward.

For the same reason

$$
e^{-u_c+u_a+u_b}=1\quad\text{or}\quad u_c=u_a+u_b \tag{2.11}
$$

and

$$
\exp(-u_c+\operatorname{sgn}(a-b)(u_a-u_b))=1
$$

or

$$
u_c=\operatorname{sgn}(a-b)(u_a-u_b). \tag{2.12}
$$

From the previous study, both $\operatorname{Im}(1-x)$ and $\operatorname{Im}(1-y)$ are negative.

As a consequence

$$
\begin{aligned}
&(1-x)=e^{-(i\pi/2)[1-\operatorname{sgn}(b-a)]}(c/b)e^{u_a\operatorname{sgn}(b-a)},\\
&(1-y)=e^{-(i\pi/2)[1-\operatorname{sgn}(a-b)]}(c/a)e^{u_b\operatorname{sgn}(a-b)},\\
&\qquad c<|a-b|.
\end{aligned}
\tag{2.13}
$$

B. Family $c^2X(1-Y)=a^2$, $c^2Y(1-X)=b^2$

Taking advantage of the involutive transformation

$$
(x,y)\mapsto(-X/(1-X),-Y/(1-Y))
$$

we get at once the unique solution for the (X,Y) family; for instance,

$$
c>a+b,\quad \frac{-x}{1-x}=X=\frac{-(a/b)e^{-u_c}}{+(c/b)e^{-u_a}}=\frac{a}{c}e^{-u_c+u_a},
$$

or

$$
X=(a/c)e^{-u_b}.
$$

We summarize the results in formulas (2.14):

$$
\begin{aligned}
&c>a+b,\quad X=(a/c)e^{-u_b},\\
&\qquad\qquad Y=(b/c)e^{-u_a};\\
&c<|a-b|,\quad X=\operatorname{sgn}(a-b)(a/c)e^{\operatorname{sgn}(a-b)u_b},\\
&\qquad\qquad Y=\operatorname{sgn}(b-a)(b/c)e^{\operatorname{sgn}(b-a)u_a};\\
&|a-b|<c<a+b,\quad X=(a/c)e^{-i\varphi_b}
\end{aligned}
$$

and

$$
Y=(b/c)e^{-i\varphi_a}. \tag{2.14}
$$

For the sake of completeness, we study the family obtained by the change $c \to ic$, namely $X(1-Y) = -a^2/c^2$ and $X(1-Y) = -b^2/c^2$.

X (resp. Y) is the solution of the second-order equation

$$c^2X^2 - X(-a^2+b^2+c^2) - a^2 = 0$$

$$[\text{resp. } c^2Y^2 - Y(-b^2+a^2+c^2) - b^2 = 0] .$$

The discriminant δ' is now always positive for a, b, and c real:

$$\delta' = (a^2-b^2)^2 + c^2(c^2+2(a^2+b^2)) . \tag{2.15}$$

The unique solution reads

$$X = \frac{c^2+b^2-a^2-\sqrt{\delta'}}{2c^2}$$

$$\left(\text{resp. } Y = \frac{c^2+a^2-b^2-\sqrt{\delta'}}{2c^2}\right), \tag{2.16}$$

because for large c, X and Y have to be small.

Here again we may introduce the hyperbolic parametrization

$$\begin{aligned} &X = -(a/c)e^{-v_b} \quad (\text{resp. } Y = (b/c)e^{-v_a}) ,\\ &2ac \sinh v_b = c^2+b^2-a^2 ,\\ &2bc \sinh v_a = c^2+a^2-b^2 ,\\ &2ac \cosh v_b = \sqrt{\delta'} , \quad 2bc \cosh v_a = \sqrt{\delta'} . \end{aligned} \tag{2.17}$$

III. CASE $\mu = \nu, \lambda = \pm\rho + 2$

We treat here a case for which the F_4 function takes a simple form, namely

$$F_4\left(\alpha,\beta,\beta,\beta; \frac{-x}{(1-x)(1-y)}, \frac{-y}{(1-x)(1-y)}\right)$$

$$= [(1-x)(1-y)]^{\alpha}\, {}_2F_1(\alpha, 1+\alpha-\beta; \beta; xy) . \tag{3.1}$$

Indeed

$$\alpha = \frac{\lambda+\mu+\nu\pm\rho}{2}, \quad \beta = \frac{\lambda+\mu+\nu\mp\rho}{2},$$

$$\gamma = \mu+1, \quad \gamma' = \nu+1 .$$

The conditions $\gamma = \gamma' = \beta$ imply $\mu = \nu$, $\lambda = \pm\rho+2$,

$$\alpha = \nu+1\mp\rho, \quad \beta = \nu+1, \quad 1+\alpha-\beta = 1\mp\rho .$$

Let us define

$$\begin{aligned} f^{\pm} &= \int_0^\infty dt\, t^{1\mp\rho} J_\nu(at) J_\nu(bt) H_\rho^{(1)}(ct)\\ &= \frac{2}{i\pi} e^{-i\pi/2} \lim_{\eta\to 0^+} \int_0^\infty t^{1\mp\rho}\, dt\, J_\nu(at) J_\nu(bt)\\ &\quad\times K_\rho((\eta - ic)t) , \end{aligned}$$

for

$$|\rho| < \frac{1}{2}, \quad \nu+1+\inf(0,\mp\rho) > 0 , \tag{3.2}$$

$$\begin{aligned} f^{\pm} &= \lim_{\eta\to 0}\left(\frac{2}{i\pi}\right) \frac{2^{\mp\rho}a^\nu b^\nu}{(\eta-ic)^{2+2\nu\mp\rho}} e^{-i\pi\rho/2} \frac{\Gamma(\nu+1\mp\rho)}{\Gamma(\nu+1)}\\ &\quad\times \frac{(\eta-ic)^{2(\nu+1\mp\rho)}}{a^{2(\nu+1\mp\rho)}}\, {}_2F_1(\nu+1\mp\rho, 1\mp\rho, \nu+1, xy)\\ &\quad\times x^{\nu+1\mp\rho} . \end{aligned} \tag{3.3}$$

This last equation is obtained by replacing $(1-x)(1-y)$ by $((\eta-ic)/a)^2x$:

$$\begin{aligned} f^{\pm} &= \left(\frac{2}{i\pi}\right) 2^{\mp\rho} \frac{a^\nu b^\nu}{c^\rho} e^{-(i\pi/2)\rho\{1\mp 1\}} \left(\frac{\Gamma(\nu+1\mp\rho)}{\Gamma(\nu+1)}\right)\\ &\quad\times\left(\frac{x}{a^2}\right)^{\nu+1\mp\rho} {}_2F_1(\nu+1\mp\rho, 1\mp\rho, \nu+1, xy) . \end{aligned} \tag{3.4}$$

At this stage we have to examine three different cases.

(i) $c < |a-b|$, $x/a^2 = (1/ab)e^{-u_c}$, $xy = e^{-2u_c}$ [see formula (2.5)]. Let $Z = \cosh u_c$ and $\zeta = e^{u_c}$. Thus, from

$$\begin{aligned} Q_{\nu'}^{\mu'}(Z) &= 2^{\mu'} e^{i\mu'\pi} \sqrt{\pi} \frac{\Gamma(\nu'+\mu'+1)}{\Gamma(\nu'+3/2)} \frac{(Z^2-1)^{\mu'/2}}{\zeta^{\nu'+\mu'+1}}\\ &\quad\times {}_2F_1(\tfrac{1}{2}+\mu', \nu'+\mu'+1, \nu'+\tfrac{3}{2}; 1/\zeta^2) , \end{aligned} \tag{3.5}$$

where $Q_{\nu'}^{\mu'}(Z)$ is the associated Legendre function, and setting $\nu' = \nu - \frac{1}{2}$ and $\mu' = \frac{1}{2}\mp\rho$, we get at once

$$\begin{aligned} &{}_2F_1(\nu+1\mp\rho, 1\mp\rho, \nu+1, e^{-2u_c})\\ &\quad= \frac{1}{\sqrt{\pi}} e^{u_c(\nu+1\mp\rho)} 2^{\pm\rho-1/2} e^{i(\pm\rho-1/2)\pi} \frac{\Gamma(\nu+1)}{\Gamma(\nu+1\mp\rho)}\\ &\quad\times(\sinh u_c)^{\mp\rho-1/2} Q_{\nu-1/2}^{\mp\rho+1/2}(\cosh u_c) \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} &\int_0^\infty dt\, t^{1\mp\rho} J_\nu(at) J_\nu(bt) [J_\rho(ct) + iY_\rho(ct)]\\ &\quad= \left(\frac{2}{i\pi}\right)\left[\frac{1}{\sqrt{2\pi}} \frac{(ab)^{\pm\rho-1}}{c^{\pm\rho}} (\sinh u_c)^{\pm\rho-1/2}\right.\\ &\quad\left.\times e^{i\pi(\pm\rho-1/2)} Q_{\nu-1/2}^{\mp\rho+1/2}(\cosh u_c)\right] e^{-(i\pi/2)\rho\{1\mp 1\}} . \end{aligned} \tag{3.7}$$

(ii) $c > a+b$. Now $x/a^2 = -1/abe^{-u_c}$. Thus a phase $e^{i\pi(\nu+1\mp\rho)}$ appears the remaining part does not change [see formula (2.5)]. Thus,

$$\begin{aligned} &\int_0^\infty dt\, t^{1\mp\rho} J_\nu(at) J_\nu(bt) [J_\rho(ct) + iY_\rho(ct)]\\ &\quad= \left(\frac{2}{i\pi}\right)\left[\frac{1}{\sqrt{2\pi}} \frac{(ab)^{\pm\rho-1}}{c^{\pm\rho}} (\sinh u_c)^{\pm\rho-1/2}\right.\\ &\quad\left.\times e^{i\pi(\pm\rho-1/2)} Q_{\nu-1/2}^{\mp\rho+1/2}(\cosh u_c)\right]\\ &\quad\times e^{-(i\pi/2)\rho\{1\mp 1\}} e^{i\pi(\nu+1\mp\rho)} . \end{aligned} \tag{3.8}$$

(iii) $|a-b| < c < a+b$, $x/a^2 = 1/abe^{2i\varphi_c}$, and $\operatorname{Im}\cos\varphi_c < 0$ [formula (2.5)].

Let $Z = \cos\varphi_c$. To apply a formula like (3.5), $(Z^2-1)^{1/2}$ has to be defined as $e^{\pm i\pi/2}(1-Z^2)^{1/2}$ depending on the sign of $\operatorname{Im} Z$. In our case $\operatorname{Im} Z$ is negative, thus $(Z^2-1)^{1/2} = e^{-i\pi/2}\sin\varphi_c(\sin\varphi_c > 0)$,

$$\begin{aligned} &Q_{\nu'}^{\mu'}(\cos\varphi_c - i0)\\ &\quad= 2^{\mu'} e^{\mu'\pi i} \frac{\Gamma(\nu'+\mu'+1)}{\Gamma(\nu'+\frac{3}{2})} e^{-(i\pi/2)\mu'}\\ &\quad\times(\sin\varphi_c)^{\mu'} e^{i\varphi_c(\nu'+\mu'+1)}\\ &\quad\times {}_2F_1(\tfrac{1}{2}+\mu', \nu'+\mu'+1, \nu'+\tfrac{3}{2}, e^{2i\varphi_c}) . \end{aligned} \tag{3.9}$$

Using the property[11]

$$e^{-(i\pi/2)\mu'}Q^{\mu'}_{\nu'}(x-i0)=Q^{\mu'}_{\nu'}(x)+(i\pi/2)P^{\mu'}_{\nu'}(x)\,,$$

where $P^{\mu'}_{\nu'}$ and $Q^{\mu'}_{\nu'}$ are the Legendre functions on the cut, we get

$$\int_0^\infty dt\,t^{1\mp\rho}J_\nu(at)J_\nu(bt)H^{(1)}_\rho(ct)$$
$$=\frac{2}{i\pi}\frac{1}{\sqrt{2\pi}}\frac{(ab)^{\pm\rho-1}}{c^{\pm\rho}}(\sin\varphi_c)^{\pm\rho-1/2}$$
$$\times\{Q^{\mp\rho+1/2}_{\nu-1/2}(\cos\varphi_c)+(i\pi/2)P^{\mp\rho+1/2}_{\nu-1/2}(\cos\varphi_c)\}$$
$$\times e^{-(i\pi/2)\rho\{1\mp1\}}\,. \tag{3.10}$$

Note that the results for $\lambda=2-\rho$ agree with those already found.[12] In the special case $\lambda=2+\rho$, $\mu=\nu=\rho$, one can obtain directly the result[13] and check our general formulas (3.7), (3.8), and (3.10):

as $\lambda=\rho+2$, $\mu=\nu=\rho$, $\frac{1}{2}(\lambda+\mu+\nu+\rho)=2\nu+1$,

$\frac{1}{2}(\lambda+\mu+\nu-\rho)=\nu+1$,

then

$$F_4\left(\nu+1,2\nu+1,\nu+1,\nu+1,\frac{-x}{(1-x)(1-y)},\frac{-y}{(1-x)(1-y)}\right)$$
$$=[(1-x)(1-y)]^{2\nu+1}{}_2F_1(2\nu+1,\nu+1,\nu+1,xy)$$
$$=\left[\frac{(1-x)(1-y)}{(1-xy)}\right]^{2\nu+1}$$

and

$$\int_0^\infty t^{1+\nu}J_\nu(at)J_\nu(bt)H^{(1)}_\nu(ct)dt$$
$$=\frac{2}{i\pi}e^{-i\pi\nu/2}\frac{(ab)^\nu2^\nu}{(\eta-ic)^{2+3\nu}}\left(\frac{\eta-ic}{a}\right)^{4\nu+2}$$
$$\times\left[\frac{x}{1-yx}\right]^{2\nu+1}\frac{\Gamma(2\nu+1)}{\Gamma(\nu+1)} \tag{3.11}$$
$$=\frac{2}{i\pi}e^{-i\pi\nu}(2abc)^\nu\left[\frac{x}{(1-xy)a^2}\right]^{2\nu+1}\frac{\Gamma(2\nu+1)}{\Gamma(\nu+1)}\,.$$

(i) $|a-b|>c$, $\dfrac{x}{a^2}=\dfrac{e^{-u_c}}{ab}$, $(1-xy)=1-e^{-2u_c}$,

$$\frac{x}{a^2(1-xy)}=\frac{1}{2ab\sinh u_c}=\frac{1}{4\tilde{\Delta}}\,.$$

Moreover,

$$\frac{\Gamma(2\nu+1)}{\Gamma(\nu+1)}=\frac{2^{2\nu}}{\sqrt{\pi}}\Gamma\left(\nu+\frac{1}{2}\right),$$

thus

$$\int_0^\infty dt\,t^{1+\nu}J_\nu(at)J_\nu(bt)H^{(1)}_\nu(ct)$$
$$=\left(\frac{1}{i\pi}\right)\frac{2^{-\nu-1}}{\sqrt{\pi}}(abc)^\nu\Gamma\left(\nu+\frac{1}{2}\right)\frac{1}{(\tilde{\Delta})^{2\nu+1}}e^{-i\pi\nu}\,. \tag{3.12}$$

(ii) $a+b<c$, $x/a^2=-e^{-u_c}/ab$. Thus a new phase $e^{i\pi(2\nu+1)}=-e^{2i\pi\nu}$ appears and

$$\int_0^\infty dt\,t^{1+\nu}J_\nu(at)J_\nu(bt)H^{(1)}_\nu(ct)$$
$$=-\left(\frac{1}{i\pi}\right)\frac{2^{-\nu-1}}{\sqrt{\pi}}(abc)^\nu\Gamma\left(\nu+\frac{1}{2}\right)\frac{1}{\tilde{\Delta}^{2\nu+1}}e^{i\pi\nu}\,. \tag{3.13}$$

(iii) $|a-b|<c<a+b$, $x=(a/b)e^{i\varphi_c}$,

$$\frac{x}{a^2(1-xy)}=\frac{e^{i\varphi_c}}{ab(1-e^{2i\varphi_c})}=\frac{e^{i\pi/2}}{2ab\sin\varphi_c}=\frac{e^{i\pi/2}}{4\Delta},$$

and

$$\int_0^\infty dt\,t^{1+\nu}J_\nu(at)J_\nu(bt)H^{(1)}_\nu(ct)$$
$$=\frac{2^{-\nu-1}}{i\pi\sqrt{\pi}}(abc)^\nu\Gamma\left(\nu+\frac{1}{2}\right)e^{-i\pi\nu}e^{i\pi(\nu+1/2)}$$
$$=\frac{2^{-\nu-1}}{\pi\sqrt{\pi}}(abc)^\nu\Gamma\left(\nu+\frac{1}{2}\right)\frac{1}{\Delta^{2\nu+1}}\,. \tag{3.14}$$

The general study would have given

$$f=\int_0^\infty dt\,t^{1+\nu}J_\nu(at)J_\nu(bt)H^{(1)}_\nu(ct)$$

as a function of $Q^{\nu+1/2}_{\nu-1/2}$ which is simple. Indeed

$$e^{-i\pi\mu'\pi}Q^{\mu'}_{\nu'}(Z)\Gamma(\nu'+\tfrac{3}{2})$$
$$=2^{-\nu'+1}\sqrt{\pi}\Gamma(\nu'+\mu'+1)Z^{-1-\nu'+\mu'}$$
$$\times(Z^2-1)^{-(1/2)\mu'}{}_2F_1\left(\frac{1+\nu'-\mu'}{2},\frac{2+\nu'-\mu'}{2},\nu'+\frac{3}{2},Z^{-2}\right). \tag{3.15}$$

Thus for $\mu'=\nu'+1$, ${}_2F_1=1$,

$$e^{-i\pi(\nu'+1)}Q^{\nu'+1}_{\nu'}(Z)$$
$$=2^{\nu'}\Gamma(\nu'+1)(Z^2-1)^{-(1/2)(\nu'+1)},\quad\text{for } Z>1$$
$$e^{-i\pi(\nu'+1)}Q^{\nu'+1}_{\nu'}(Z-i0)$$
$$=2^{\nu'}\Gamma(\nu'+1)(1-Z^2)^{-(1/2)(\nu'+1)}e^{+(i\pi/2)(\nu'+1)},$$
$$\text{for } |Z|<1\,. \tag{3.16}$$

Applying formulas (3.15) and (3.16) in the formulas (3.7)–(3.10) for $\rho=\nu$, $\nu'=\nu-\frac{1}{2}$, we get, for $c>b+a$ or $c<|a-b|$,

$$e^{-i\pi(\nu+1/2)}Q^{\nu+1/2}_{\nu-1/2}(\cosh u_c)$$
$$=(\sinh u_c)^{-(\nu+1/2)}2^{\nu-1/2}\Gamma(\nu+\tfrac{1}{2})$$
$$(\sinh u_c=2\tilde{\Delta}/ab)$$

and

$$e^{-i\pi(\nu+1/2)}Q^{\nu+1/2}_{\nu-1/2}(\cos\varphi_c-i0)$$
$$=(\sin\varphi_c)^{-(\nu+1/2)}2^{\nu-1/2}\Gamma(\nu+\tfrac{1}{2})e^{(i\pi/2)(\nu+1/2)},$$
$$\text{for } a+b>c>|a-b|,\quad \sin\varphi_c=2\Delta/ab\,.$$

Thus for $|a-b|>c$,

$$\int_0^\infty dt\, t^{1+\nu} J_\nu(at) J_\nu(bt) H_\nu^{(1)}(ct)$$

$$= \left(\frac{2}{i\pi}\right)\left[\frac{1}{\sqrt{2\pi}}(ab)^{-(\nu+1)} c^\nu (\sinh u_c)^{-(\nu+1/2)} e^{-i\pi\nu} \times \left\{(\sinh u_c)^{-(\nu+1/2)} 2^{\nu-1/2} \Gamma\left(\nu+\frac{1}{2}\right)\right\}\right]$$

$$= \frac{1}{i\pi}\frac{1}{\sqrt{\pi}}(abc)^\nu e^{-i\pi\nu}\Gamma\left(\nu+\frac{1}{2}\right) 2^{-\nu-1}\frac{1}{(\tilde{\Delta})^{2\nu+1}},$$

in agreement with formula (3.12) and similarly with formula (3.13) in the case $a+b<c$ by adding the phase $e^{i\pi(2\nu+1)} = -e^{+2i\pi\nu}$. For $a+b>c>|a-b|$, formula (3.10) reads

$$\int_0^\infty dt\, t^{1+\nu} J_\nu(at) J_\nu(bt) H_\nu^{(1)}(ct)$$

$$= \left(\frac{2}{i\pi}\right)\frac{1}{\sqrt{2\pi}}(ab)^{-\nu+1} c^\nu (\sin\varphi_c)^{-(\nu+1/2)} \times \{\sin\varphi_c\}^{-(\nu+1/2)} 2^{\nu-1/2}\Gamma\left(\nu+\frac{1}{2}\right) \times e^{+i\pi(\nu+1/2)} e^{-i\pi\nu}$$

$$= \frac{(abc)^\nu 2^{-\nu-1}\Gamma(\nu+\frac{1}{2})}{\pi\sqrt{\pi}\Delta^{2\nu+1}},$$

in agreement with formula 3.14.

All the results of this section are summarized in Table I.

TABLE I.

$$\int_0^\infty dt\, t^{1\mp\rho} J_\nu(at) J_\nu(bt) \begin{Bmatrix} J_\rho(ct) \\ Y_\rho(ct) \end{Bmatrix},$$

any real a,b,c. Case $\rho=\nu$ is given for completeness.

$$\int_0^\infty dt\, t^{1-\rho} J_\nu(at) J_\nu(bt) \begin{Bmatrix} J_\rho(ct) \\ Y_\rho(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ab)^{\rho-1}}{c^\rho}(\sinh u_c)^{\rho-1/2} e^{i\pi(\rho-1/2)} Q_{\nu-1/2}^{-\rho+1/2}(\cosh u_c)\begin{Bmatrix} 0 \\ 1 \end{Bmatrix},$$

$c<|a-b|$, $2ab\cosh u_c = a^2+b^2-c^2$, $\frac{1}{2}ab\sinh u_c = \tilde{\Delta}$; $4\tilde{\Delta} = \sqrt{[c^2-(a-b)^2][c^2-(a+b)^2]}$.

$$\int_0^\infty dt\, t^{1-\rho} J_\nu(at) J_\nu(bt) \begin{Bmatrix} J_\rho(ct) \\ Y_\rho(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ab)^{\rho-1}}{c^\rho}(\sinh u_c)^{\rho-1/2} e^{i\pi(\rho-1/2)} Q_{\nu-1/2}^{-\rho+1/2}(\cosh u_c)\begin{Bmatrix} -\sin(\nu-\rho)\pi \\ \cos(\nu-\rho)\pi \end{Bmatrix},$$

$c>a+b$, $2ab\cosh u_c = c^2-a^2-b^2$, $\frac{1}{2}ab\sinh u_c = \tilde{\Delta}$.

$$\int_0^\infty dt\, t^{1-\rho} J_\nu(at) J_\nu(bt) \begin{Bmatrix} J_\rho(ct) \\ Y_\rho(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ab)^{\rho-1}}{c^\rho}(\sin\varphi_c)^{\rho-1/2}\begin{Bmatrix} (\pi/2) P_{\nu-1/2}^{-\rho+1/2}(\cos\varphi_c) \\ -Q_{\nu-1/2}^{-\rho+1/2}(\cos\varphi_c) \end{Bmatrix},$$

$|a-b|<c<a+b$, $2ab\cos\varphi_c = a^2+b^2-c^2$, $\frac{1}{2}ab\sin\varphi_c = \Delta$,

$4\Delta = \sqrt{[c^2-(a-b)^2][(a+b)^2-c^2]}$, $\rho>-\frac{1}{2}$, $\begin{bmatrix} \nu+1>0 & \text{(for } J\rho) \\ \nu+1-(\rho+|\rho|)/2>0 & \text{(for } Y\rho) \end{bmatrix}$.

$$\int_0^\infty dt\, t^{1+\rho} J_\nu(at) J_\nu(bt) \begin{Bmatrix} J_\rho(ct) \\ Y_\rho(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ab)^{-\rho-1}}{c^{-\rho}}(\sinh u_c)^{-\rho-1/2} e^{-i\pi(\rho+1/2)} Q_{\nu-1/2}^{\rho+1/2}(\cosh u_c)\begin{Bmatrix} -\sin\pi\rho \\ -\cos\pi\rho \end{Bmatrix},$$

$c<|a-b|$, $2ab\cosh u_c = a^2+b^2-c^2$, $\frac{1}{2}ab\sinh u_c = \tilde{\Delta}$.

$$\int_0^\infty dt\, t^{1+\rho} J_\nu(at) J_\nu(bt) \begin{Bmatrix} J_\rho(ct) \\ Y_\rho(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ab)^{-\rho-1}}{c^{-\rho}}(\sinh u_c)^{-\rho-1/2} e^{-(i\pi/2)(\rho+1/2)} Q_{\nu-1/2}^{\rho+1/2}(\cosh u_c)\begin{Bmatrix} -\sin\pi\nu \\ \cos\pi\nu \end{Bmatrix},$$

$c>a+b$, $2ab\cosh u_c = c^2-a^2-b^2$, $\frac{1}{2}ab\sinh u_c = \tilde{\Delta}$.

$$\int_0^\infty dt\, t^{1+\rho} J_\nu(at) J_\nu(bt) \begin{Bmatrix} J_\rho(ct) \\ Y_\rho(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ab)^{-\rho-1}}{c^{-\rho}}(\sin\varphi_c)^{-\rho-1/2}\begin{Bmatrix} \cos\pi\rho P_{\nu-1/2}^{\rho+1/2}(\cos\varphi_c)(\pi/2) - \sin\pi\rho Q_{\nu-1/2}^{\rho+1/2}(\cos\varphi_c) \\ -\sin\pi\rho P_{\nu-1/2}^{\rho+1/2}(\cos\varphi_c)(\pi/2) - \cos\pi\rho Q_{\nu-1/2}^{\rho+1/2}(\cos\varphi_c) \end{Bmatrix},$$

$|a-b|<c<a+b$, $2ab\cos\varphi_c = a^2+b^2-c^2$, $\frac{1}{2}ab\sin\varphi_c = \Delta$, $\rho<\frac{1}{2}$, $\begin{bmatrix} \nu+1+\rho>0 & \text{(for } J\rho) \\ \nu+1+(\rho-|\rho|)/2>0 & \text{(for } Y\rho). \end{bmatrix}$

$$\int_0^\infty dt\, t^{1+\nu} J_\nu(at) J_\nu(bt) \begin{Bmatrix} J_\nu(ct) \\ Y_\nu(ct) \end{Bmatrix} = \frac{2^{-\nu-1}}{\pi\sqrt{\pi}}(abc)^\nu \Gamma\left(\nu+\frac{1}{2}\right)\frac{1}{(\tilde{\Delta})^{2\nu+1}}\begin{Bmatrix} -\sin\pi\nu \\ -\cos\pi\nu \end{Bmatrix},$$

$c<|a-b|$, $|\nu|<\frac{1}{2}$.

$$\int_0^\infty dt\, t^{1+\nu} J_\nu(at) J_\nu(bt) \begin{Bmatrix} J_\nu(ct) \\ Y_\nu(ct) \end{Bmatrix} = \frac{2^{-\nu-1}}{\pi\sqrt{\pi}}(abc)^\nu \Gamma\left(\nu+\frac{1}{2}\right)\frac{1}{(\tilde{\Delta})^{2\nu+1}}\begin{Bmatrix} -\sin\pi\nu \\ \cos\pi\nu \end{Bmatrix},$$

$c>a+b$, $|\nu|<\frac{1}{2}$.

$$\int_0^\infty dt\, t^{1+\nu} J_\nu(at) J_\nu(bt) \begin{Bmatrix} J_\nu(ct) \\ Y_\nu(ct) \end{Bmatrix} = \frac{2^{-\nu-1}}{\pi\sqrt{\pi}}(abc)^\nu \Gamma\left(\nu+\frac{1}{2}\right)\frac{1}{(\Delta)^{2\nu+1}}\begin{Bmatrix} 1 \\ 0 \end{Bmatrix},$$

$|a-b|<c<a+b$, $|\nu|<\frac{1}{2}$.

IV. CASE $\pm\rho=\mu, \lambda=\nu+2$

For these values of λ, μ, ν, and ρ, we use the factorization formula

$$F_4\left(\alpha,\beta,1+\alpha-\beta,\beta,\frac{-x}{(1-x)(1-y)}\right) = (1-y)^{\alpha}{}_2F_1\left(\alpha,\beta,1+\alpha-\beta,\frac{-x(1-y)}{1-x}\right), \quad (4.1)$$

where

$$\alpha=\nu+1+\mu\,,\quad \beta=\nu+1\,.$$

Let us calculate then

$$f=\int_0^\infty dt\, t^{\nu+1}J_\mu(at)J_\nu(bt)H^{(1)}_{\pm\mu}(ct)$$

$$=\frac{2}{i\pi}\lim_{\eta\to 0^+}\int_0^\infty dt\, t^{\nu+1}J_\mu(at)J_\nu(bt) \times e^{\mp(i\pi/2)\mu}K_\mu((\eta-ic)t)\,,$$

$$\nu<\tfrac{1}{2},\quad \nu+1+\inf(0,\mu)>0\,,$$

$$f=\left(\frac{2}{i\pi}\right)e^{\mp(i\pi/2)\mu}\frac{2^\nu a^\mu b^\nu}{(\eta-ic)^{2\nu+\mu+2}}\frac{\Gamma(\nu+1+\mu)}{\Gamma(1+\mu)} \times(1-y)^{\nu+\mu+1}{}_2F_1\left(\nu+1+\mu,\nu+1,\mu+1,\frac{-x(1-y)}{1-x}\right),$$

$$f=\left(\frac{2}{i\pi}\right)e^{i\pi(\nu+1)}e^{(i\pi/2)\mu[1\mp 1]} \times\frac{2^\nu a^\mu b^\nu}{(c)^{2\nu+2+\mu}}\frac{\Gamma(\nu+1+\mu)}{\Gamma(1+\mu)}(1-y)^{\nu+\mu+1} \times{}_2F_1\left(\nu+1+\mu,\nu+1,\mu+1,\frac{-x(1-y)}{1-x}\right); \quad (4.2)$$

$$-\frac{-x(1-y)}{1-x}=\frac{-x(1-y)^2}{(1-x)(1-y)}=\frac{a^2}{c^2}(1-y)^2\,.$$

(i) $c>a+b$, $-x(1-y)/(1-x)=e^{-2u_b}$, $1-y=(c/a)e^{-u_b}$. See formula (2.9a). From formula (3.5),

$${}_2F_1(\nu+1+\mu,\nu+1,\mu+1,e^{-2u_b}) =\frac{1}{\sqrt{\pi}}e^{u_b(\nu+1+\mu)}\frac{\Gamma(\mu+1)}{\Gamma(\nu+1+\mu)}e^{-i(\nu+1/2)\pi}2^{-\nu-1/2} \times(\sinh u_b)^{-\nu-1/2}Q^{\nu+1/2}_{\mu-1/2}(\cosh u_b)\,. \quad (4.3)$$

Thus

$$f=\left(\frac{2}{i\pi}\right)\frac{e^{i\pi(\nu+1)}}{\sqrt{2\pi}}e^{+(i\pi/2)\mu[1\mp 1]}\frac{b^\nu}{(ac)^{\nu+1}} \times(\sinh u_b)^{-\nu-1/2}\{e^{-i(\nu+1/2)\pi}Q^{\nu+1/2}_{\mu-1/2}(\cosh u_b)\}\,.$$

(ii) $c<|a-b|$, $-x(1-y)/(1-x)=e^{2u_b\,\mathrm{sgn}(a-b)}$,

$$(1-y)=e^{-(i\pi/2)[1-\mathrm{sgn}(a-b)]}(c/a)e^{u_b\mathrm{sgn}(a-b)}$$

[see formulas (2.9b) and (2.13)]. One configuration is simple; namely, for $a<b$, $(1-y)=e^{-i\pi}(c/a)e^{-u_b}$.

We still apply formula (3.5) and get the same result as before but for the phase $e^{-i\pi(\nu+\mu+1)}$. We get

$$f=\left(\frac{2}{i\pi}\right)\frac{e^{-(i\pi/2)\mu[1\pm 1]}}{\sqrt{2\pi}}\frac{b^\nu}{(ac)^{\nu+1}}(\sinh u_b)^{-\nu-1/2} \times\{e^{-i(\nu+1/2)\pi}Q^{\nu+1/2}_{\mu-1/2}(\cosh u_b)\}\,. \quad (4.4)$$

In the case $a>b$, formula (3.5) does not apply and we need the analytic continuation of ${}_2F_1(a,b,c,Z)$ as linear combination of ${}_2F_1(a',b',c',Z^{-1})$; namely (Ref. 4, p. 108),

$$\frac{{}_2F_1(a,b,c,Z)}{\Gamma(c)}=\frac{\Gamma(b-a)}{\Gamma(b)\Gamma(c-a)}(-Z)^{-a} \times{}_2F_1(a,1+a-c,1+a-b,Z^{-1}) +\frac{\Gamma(a-b)}{\Gamma(a)\Gamma(c-b)}(-Z)^{-b} \times{}_2F_1(b,1+b-c,1+b-a,Z^{-1})\,, \quad (4.5)$$

$$\begin{array}{lll} a=\nu+1+\mu\,, & a-c+1=\nu+1\,, & \mathrm{Re}\,Z=(a/c)^2(1-y)^2=e^{2u_b},\\ b=\nu+1\,, & a-b+1=\mu+1\,, & \\ c=1+\mu\,, & 1+b-c=1+\nu-\mu\,, & \mathrm{Im}\,Z<0\quad\text{[see formula (2.7)]}\,,\\ & 1-a+b=-\mu+1\,, & \text{thus } (-Z)=e^{i\pi}|Z|\,, \end{array}$$

$$\frac{{}_2F_1(\nu+1+\mu,\nu+1,1+\mu,e^{2u_b})}{\Gamma(1+\mu)} =\frac{\Gamma(-\mu)}{\Gamma(\nu+1)\Gamma(-\nu)}e^{-i\pi(\nu+1+\mu)}e^{-2u_b(\nu+1+\mu)} \times{}_2F_1(\nu+1+\mu,\nu+1,\mu+1,e^{-2u_b}) +\frac{\Gamma(\mu)}{\Gamma(\nu+1+\mu)\Gamma(-\nu+\mu)}e^{-i\pi(\nu+1)}e^{-2u_b(\nu+1)} \times{}_2F_1(\nu+1-\mu,\nu+1,-\mu+1,e^{-2u_b})\,.$$

Now we apply safely formula (4.5) for each ${}_2F_1$ in the right-hand side to get

$$\frac{{}_2F_1(\nu+1+\mu,\nu+1,1+\mu,e^{2u_b})}{\Gamma(1+\mu)} =\frac{1}{\sqrt{\pi}}\frac{e^{-i\pi(\nu+1)}2^{-\nu-1/2}(\sinh u_b)^{-\nu-1/2}e^{-u_b(\nu+1+\mu)}}{\Gamma(\nu+1+\mu)} \times e^{-i(\nu+1/2)\pi}\left[\frac{\Gamma(-\mu)}{\Gamma(\nu+1)\Gamma(-\nu)}e^{-i\pi\mu}\Gamma(\mu+1) \times Q^{\nu+1/2}_{\mu-1/2}(\cosh u_b)+\frac{\Gamma(\mu)\Gamma(1-\mu)}{\Gamma(-\nu+\mu)\Gamma(\nu+1-\mu)} \times Q^{\nu+1/2}_{-\mu-1/2}(\cosh u_b)\right],$$

$$f= +\left(\frac{2}{i\pi}\right)\frac{1}{\sqrt{2\pi}}e^{+(i\pi/2)\mu[1\mp 1]}(\sinh u_b)^{-\nu-1/2}$$
$$\times\frac{b^\nu}{(ac)^{\nu+1}}\frac{e^{-i(\nu+1/2)\pi/2}}{\sin\mu\pi}$$
$$\times[\sin\nu\pi e^{-i\mu\pi}Q_{\mu-1/2}^{\nu+1/2}(\cosh u_b)$$
$$+\sin(\mu-\nu)\pi Q_{-\mu-1/2}^{\nu+1/2}(\cosh u_b)]\,.$$

Now[14]

$$Q_{\nu'}^{\mu'}(Z)\sin(\nu'+\mu')\pi+Q_{-\nu'-1}^{\mu'}(Z)\sin(\mu'-\nu')\pi$$
$$=\pi e^{\mu'\pi i}\cos\nu'\pi P_{\nu'}^{\mu'}(Z)\,.$$

Thus

$$Q_{-\mu-1/2}^{\nu+1/2}(\cosh u_b)\sin[\nu-\mu+1]\pi$$
$$=\pi e^{i(\nu+1/2)\pi}\cos[(\mu-\tfrac{1}{2})\pi]P_{\mu-1/2}^{\nu+1/2}(\cosh u_b)$$
$$-Q_{\mu-1/2}^{\nu+1/2}(\cosh u_b)\sin(\nu+\mu)\pi$$

and

$$f= -\left(\frac{2}{i\pi}\right)\frac{1}{\sqrt{2\pi}}e^{+(i\pi/2)\mu(1\mp 1)}(\sinh u_b)^{-\nu-1/2}$$
$$\times\frac{b^\nu}{(ac)^{\nu+1}}[e^{+i\nu\pi}Q_{\mu-1/2}^{\nu+1/2}(\cosh u_b)e^{-i(\nu+1/2)\pi}$$
$$-\pi P_{\mu-1/2}^{\nu+1/2}(\cosh u_b)]\,. \tag{4.6}$$

(iii) $|a-b|<c<a+b$,

$$\frac{a^2}{c^2}(1-y)^2=e^{-2i\varphi_b},\quad 2a\,\mathrm{Im}\cos\varphi_b=\frac{c^2+b^2-a^2}{c^2}\mathrm{Im}\,c\,,$$

$$f=\left(\frac{2}{i\pi}\right)e^{i\pi(\nu+1)}e^{(i\pi/2)\mu[1\mp 1]}2^\nu a^\mu b^\nu c^{-\mu-2\nu-2}$$
$$\times\frac{\Gamma(\nu+1+\mu)}{\Gamma(1+\mu)}e^{-i\varphi_b(\nu+\mu+1)}$$
$$\times{}_2F_1(\nu+1+\mu,\nu+1,\mu+1,e^{-2i\varphi_b})\,,$$

and $\mathrm{Im}\cos\varphi_b\,\mathrm{sgn}(b-a)$ is positive.

We apply formula (4.5) in the case $a<b$. Indeed for

$$Z=\cos\varphi_b\,,(Z^2-1)^{1/2}=e^{i\,\mathrm{sgn}(b-a)\pi/2}\sin\varphi_b\,.$$

and

$$[Z-(Z^2-1)^{1/2}]/[Z+(Z^2-1)^{1/2}]=e^{-2i\,\mathrm{sgn}(b-a)\varphi_b}.$$

From formula 8.777.2 (Ref. 6, p. 1012), we get

$${}_2F_1(\nu+1+\mu,\nu+1,\mu+1,e^{-2i\varphi_b})$$
$$=\frac{1}{\sqrt{\pi}}e^{i\varphi_b(\nu+1+\mu)}(\sin\varphi_b)^{-\nu-1/2}e^{-(i\pi/2)(\nu+1/2)}$$
$$\times e^{-i(\nu+1/2)\pi}Q_{\mu-1/2}^{\nu+1/2}(\cos\varphi_b+i0)$$
$$\times\frac{\Gamma(\mu+1)}{\Gamma(\nu+1+\mu)}2^{-\nu-1/2}$$

and

$$f=\left(\frac{2}{i\pi}\right)\frac{e^{i\pi(\nu+1)}}{\sqrt{2\pi}}e^{+(i\pi/2)\mu[1\mp 1]}\frac{b^\nu}{a^{\nu+1}c^{\nu+1}}(\sin\varphi_b)^{-\nu-1/2}$$
$$\times[Q_{\mu-1/2}^{\nu+1/2}(\cos\varphi_b)-(i\pi/2)P_{\mu-1/2}^{\nu+1/2}(\cos\varphi_b)]\,. \tag{4.7}$$

In the case $a>b$, we need either the analytic continuation formula (4.5) or the relation

$${}_2F_1(\nu+1+\mu,\nu+1,\mu+1,e^{-2i\varphi_b})$$
$$={}_2F_1^*(\nu+1+\mu,\nu+1,\mu+1,e^{2i\varphi_b})\,,$$

where ${}_2F_1^*$ means the complex conjugate quantity. This yields

$${}_2F_1(\nu+1+\mu,\nu+1,\mu+1,e^{2i\varphi_b})$$
$$=\frac{1}{\sqrt{\pi}}e^{-i\varphi_b(\nu+1+\mu)}(\sin\varphi_b)^{-\nu-1/2}2^{-\nu-1/2}$$
$$\times\frac{\Gamma(\mu+1)}{\Gamma(\nu+1+\mu)}[Q_{\mu-1/2}^{\nu+1/2}(\cos\varphi_b)$$
$$+(i\pi/2)P_{\mu-1/2}^{\nu+1/2}(\cos\varphi_b)]\,,$$

since in this case $\mathrm{Im}\cos\varphi_b$ is negative. Thus

$${}_2F_1(\nu+1+\mu,\nu+1,\mu+1,e^{-2i\varphi_b})$$
$$=\frac{1}{\sqrt{\pi}}e^{i\varphi_b(\nu+1+\mu)}(\sin\varphi_b)^{-\nu-1/2}2^{-\nu-1/2}$$
$$\times\frac{\Gamma(\mu+1)}{\Gamma(\nu+1+\mu)}[Q_{\mu-1/2}^{\nu+1/2}(\cos\varphi_b)$$
$$-(i\pi/2)P_{\mu-1/2}^{\nu+1/2}(\cos\varphi_b)]\,,$$

and as expected the result is the same for $a>b$ or $b<a$ when $|a-b|<c<a+b$.

All the results of this section are summarized in Table II. Note that some integrals in Table II are identical to others in Table I by the change $\nu\leftrightarrow\rho$ and $b\leftrightarrow c$, which is a good check of our method.

V. CASE $\lambda=2, \rho=\mu-\nu$

We have to consider integrals of the form

$$\int_0^\infty t\,dtJ_\mu(at)J_\nu(bt)\{J_{\mu-\nu}(ct)+iY_{\mu-\nu}(ct)\}\,.$$

They were already calculated by using another method in a previous work.[5] Here we give another way to get the results which gives a consistency check.

For this configuration

$$(\mu+\nu+\lambda-\rho)/2=\nu+1\,,$$
$$(\mu+\nu+\lambda+\rho)/2=\mu+1\,,$$

and

$$F_4\left(\frac{1}{2}(\mu+\nu+\lambda+\rho),\frac{1}{2}(\lambda+\mu+\nu-\rho),\mu+1,\nu+1\,;\right.$$
$$\left.\frac{-x}{(1-x)(1-y)},\frac{-y}{(1-x)(1-y)}\right)$$
$$=\frac{(1-x)^{\nu+1}(1-y)^{\mu+1}}{1-xy}, \tag{5.1}$$

with

$$(1-x)(1-y)=\left(\frac{\eta-ic}{a}\right)^2x=\left(\frac{\eta-ic}{b}\right)^2y\,,$$

η small and positive. Thus

TABLE II.

$$\int_0^\infty dt\, t^{1+\nu} J_\mu(at) J_\nu(bt) \begin{Bmatrix} J_{\pm\mu}(ct) \\ Y_{\pm\mu}(ct) \end{Bmatrix}$$

and any real a,b,c.

$$\nu<\frac{1}{2}, \quad \begin{bmatrix} \nu+1+(\mu\pm\mu)/2>0 & (\text{for } J) \\ \nu+1+(\mu-|\mu|)/2>0 & (\text{for } Y) \end{bmatrix}$$

$$\int_0^\infty dt\, t^{1+\nu} J_\mu(at) J_\nu(bt) \begin{Bmatrix} J_\mu(ct) \\ Y_\mu(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ac)^{-\nu-1}}{b^{-\nu}}(\sinh u_b)^{-\nu-1/2} e^{-i\pi(\nu+1/2)} Q^{\nu+1/2}_{\mu-1/2}(\cosh u_b) \begin{Bmatrix} -\sin\mu\pi \\ -\cos\mu\pi \end{Bmatrix},$$

$$\int_0^\infty dt\, t^{1+\nu} J_\mu(at) J_\nu(bt) \begin{Bmatrix} J_{-\mu}(ct) \\ Y_{-\mu}(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ac)^{-\nu-1}}{b^{-\nu}}(\sinh u_b)^{-\nu-1/2} e^{-i\pi(\nu+1/2)} Q^{\nu+1/2}_{\mu-1/2}(\cosh u_b) \begin{Bmatrix} 0 \\ -1 \end{Bmatrix},$$

$b>a, \quad c<b-a, \quad 2ac\cosh u_b = b^2-a^2-c^2, \quad \frac{1}{2}ac\sinh u_b = \widetilde{\Delta}.$

$$\int_0^\infty dt\, t^{1+\nu} J_\mu(at) J_\nu(bt) \begin{Bmatrix} J_\mu(ct) \\ Y_\mu(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ac)^{-\nu-1}}{b^{-\nu}}(\sinh u_b)^{-\nu-1/2} \begin{Bmatrix} -\sin\nu\pi e^{-i\pi(\nu+1/2)} Q^{\nu+1/2}_{\mu-1/2}(\cosh u_b) \\ \cos\nu\pi e^{-i\pi(\nu+1/2)} Q^{\nu+1/2}_{\mu-1/2}(\cosh u_b) - \pi P^{\nu+1/2}_{\mu-1/2}(\cosh u_b) \end{Bmatrix},$$

$$\int_0^\infty dt\, t^{1+\nu} J_\mu(at) J_\nu(bt) \begin{Bmatrix} J_{-\mu}(ct) \\ Y_{-\mu}(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ac)^{-\nu-1}}{b^{-\nu}}(\sinh u_b)^{-\nu-1/2} \begin{Bmatrix} -\sin(\mu+\nu)\pi e^{-i\pi(\nu+1/2)} Q^{\nu+1/2}_{\mu-1/2}(\cosh u_b) - \pi\sin\mu\pi P^{\nu+1/2}_{\mu-1/2}(\cosh u_b) \\ \cos(\mu+\nu)\pi e^{-i\pi(\nu+1/2)} Q^{\nu+1/2}_{\mu-1/2}(\cosh u_b) - \pi\cos\mu\pi P^{\nu+1/2}_{\mu-1/2}(\cosh u_b) \end{Bmatrix},$$

$a>b, \quad c<a-b, \quad 2ac\cosh u_b = a^2-b^2+c^2, \quad \frac{1}{2}ac\sinh u_b = \widetilde{\Delta}.$

$$\int_0^\infty dt\, t^{1+\nu} J_\mu(at) J_\nu(bt) \begin{Bmatrix} J_\mu(ct) \\ Y_\mu(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ac)^{-\nu-1}}{b^{-\nu}}(\sinh u_b)^{-\nu-1/2} e^{-i(\nu+1/2)\pi} Q^{\nu+1/2}_{\mu-1/2}(\cosh u_b) \begin{Bmatrix} -\sin\nu\pi \\ \cos\nu\pi \end{Bmatrix},$$

$$\int_0^\infty dt\, t^{1+\nu} J_\mu(at) J_\nu(bt) \begin{Bmatrix} J_{-\mu}(ct) \\ Y_{-\mu}(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ac)^{-\nu-1}}{b^{-\nu}}(\sinh u_b)^{-\nu-1/2} e^{-i(\nu+1/2)\pi} Q^{\nu+1/2}_{\mu-1/2}(\cosh u_b) \begin{Bmatrix} -\sin(\nu+\mu)\pi \\ \cos(\nu+\mu)\pi \end{Bmatrix},$$

$c>a+b, \quad 2ac\cosh u_b = c^2+a^2-b^2, \quad \frac{1}{2}ac\sinh u_b = \widetilde{\Delta}.$

$$\int_0^\infty dt\, t^{1+\nu} J_\mu(at) J_\nu(bt) \begin{Bmatrix} J_\mu(ct) \\ Y_\mu(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ac)^{-\nu-1}}{b^{-\nu}}(\sin\varphi_b)^{-\nu-1/2} \begin{Bmatrix} \cos(\nu\pi)(\pi/2) P^{\nu+1/2}_{\mu-1/2}(\cos\varphi_b) - \sin\nu\pi Q^{\nu+1/2}_{\mu-1/2}(\cos\varphi_b) \\ -\sin(\nu\pi)(\pi/2) P^{\nu+1/2}_{\mu-1/2}(\cos\varphi_b) - \cos\nu\pi Q^{\nu+1/2}_{\mu-1/2}(\cos\varphi_b) \end{Bmatrix},$$

$$\int_0^\infty dt\, t^{1+\nu} J_\mu(at) J_\nu(bt) \begin{Bmatrix} J_{-\mu}(ct) \\ Y_{-\mu}(ct) \end{Bmatrix} = \frac{1}{\pi}\sqrt{\frac{2}{\pi}}\frac{(ac)^{-\nu-1}}{b^{-\nu}}(\sin\varphi_b)^{-\nu-1/2} \begin{Bmatrix} \cos[(\nu+\mu)\pi](\pi/2) P^{\nu+1/2}_{\mu-1/2}(\cos\varphi_b) - \sin[(\nu+\mu)\pi] Q^{\nu+1/2}_{\mu-1/2}(\cos\varphi_b) \\ -\sin[(\nu+\mu)\pi](\pi/2) P^{\nu+1/2}_{\mu-1/2}(\cos\varphi_b) - \cos[(\nu+\mu)\pi] Q^{\nu+1/2}_{\mu-1/2}(\cos\varphi_b) \end{Bmatrix},$$

$|a-b|<c<a+b, \quad 2ac\cos\varphi_b = a^2+c^2-b^2, \quad \frac{1}{2}ac\sin\varphi_b = \Delta.$

$$\int_0^\infty t\,dt\, J_\mu(at) J_\nu(bt) H^{(1)}_{\mu-\nu}(ct)$$
$$= \lim_{\eta\to 0^+}\left(\frac{2}{i\pi}\right)\frac{a^\mu b^\nu}{(\eta-ic)^{\mu+\nu+2}} e^{-(i\pi/2)(\mu-\nu)} \times \frac{(1-x)^{\nu+1}(1-y)^{\mu+1}}{1-xy},$$
$$= \left(\frac{2}{i\pi}\right) a^\mu b^\nu e^{-i\pi(\mu-\nu)} \left(\frac{1-x}{c}\right)^{\nu-\mu}\left(\frac{x}{a^2}\right)^{\mu+1}\frac{1}{1-xy}. \tag{5.2}$$

Here again, three cases have to be studied.

(i) $c<|a-b|$, $\quad \dfrac{x}{a^2} = \dfrac{1}{ab}e^{-u_c}, \quad xy = e^{-2u_c},$

and

$$\left(\frac{1-x}{c}\right) = e^{-(i\pi/2)\{1-\operatorname{sgn}(b-a)\}}\frac{1}{b}e^{u_a \operatorname{sgn}(b-a)},$$

where u_a and u_c are defined in formulas (2.5) and (2.8).

$$\int_0^\infty t\,dt\, J_\mu(at) J_\nu(bt) H^{(1)}_{\mu-\nu}(ct)$$
$$= \left(\frac{1}{i\pi}\right)\frac{e^{-\mu u_c} e^{+(\nu-\mu)u_a \operatorname{sgn}(b-a)} e^{(i\pi/2)(\nu-\mu)(1+\operatorname{sgn}(b-a))}}{2\widetilde{\Delta}}. \tag{5.3}$$

(ii) $c>a+b$. In this case

$$\frac{x}{a^2} = e^{+i\pi}\frac{1}{ab}e^{-u_c}, \quad \frac{1-x}{c} = \frac{1}{b}e^{-u_a},$$

and

$$\int_0^\infty t\,dt\, J_\mu(at) J_\nu(bt) H^{(1)}_{\mu-\nu}(ct)$$
$$= \left(\frac{1}{i\pi}\right)\frac{e^{-\mu u_c} e^{-(\nu-\mu)u_a} e^{i\pi\nu}}{2\widetilde{\Delta}}. \tag{5.4}$$

(iii) $|a-b|<c<a+b$,

$$\frac{x}{a^2} = \frac{1}{ab}e^{i\varphi_c}, \quad \frac{1-x}{c} = \frac{1}{b}e^{-i\varphi_a}, \quad xy = e^{2i\varphi_c},$$

where φ_c and φ_a are defined in formulas (2.5) and (2.8). We thus get

$$\int_0^\infty t\,dt\, J_\mu(at) J_\nu(bt) H^{(1)}_{\mu-\nu}(ct) = \frac{1}{\pi}\frac{e^{i[\nu\varphi_c-(\mu-\nu)\varphi_b]}}{2\Delta}, \tag{5.5}$$

where $\varphi_a+\varphi_b+\varphi_c=\pi$, and Δ is the area of the triangle (a,b,c):

$$\Delta = \tfrac{1}{4}\sqrt{(a+b+c)(a+b-c)(a-b+c)(b+c-a)}$$
$$= \tfrac{1}{2}ab\sin\varphi_c = \tfrac{1}{2}bc\sin\varphi_a = \tfrac{1}{2}ca\sin\varphi_b. \tag{5.6}$$

Note that to recover easily the results of Ref. 5, we have to make the following change of notation:

$$\mu \to \mu' + \nu', \quad a \to c', \quad \varphi_a \to \varphi_{c'}, \quad u_a \to u_{c'},$$
$$\mu - \nu \to \mu', \quad b \to b', \quad \varphi_b \to \varphi_{b'}, \quad u_b \to u_{b'},$$
$$\nu \to \nu', \quad c \to a', \quad \varphi_c \to \varphi_{a'}, \quad u_c \to u_{a'}.$$

For instance, (5,5) reads

$$\int_0^\infty t\,dt\,J_{\mu'+\nu'}(c't)J_{\nu'}(b't)H^{(1)}_{\mu'}(a't) = e^{i\{\nu'\varphi_{a'} - \mu'\varphi_{b'}\}}/2\pi\Delta. \tag{5.7}$$

This method does not give the result for the integral $\int_0^\infty tJ_\mu(at)J_\nu(bt)Y_{\mu+\nu}(ct)dt$. This integral is calculated in the following paper[10] by differentiating other known integrals.

VI. CASE $\lambda = \nu + \rho - \mu + 2$

All preceding examples are results of the more general formula (1.10)

$$F_4\left(\alpha,\beta;\gamma,\beta;\frac{-x}{(1-x)(1-y)},\frac{-y}{(1-x)(1-y)}\right)$$
$$= (1-x)^\alpha(1-y)^\alpha F_1(\alpha;\gamma-\beta,1+\alpha-\gamma;\gamma;x,xy), \tag{6.1}$$

which holds provided

$$\lambda + \mu = \nu + \rho + 2 \tag{6.2}$$

and corresponds to the calculation of

$$\int_0^\infty t^{\nu+\rho-\mu+1}J_\mu(at)J_\nu(bt)H^{(1)}_\rho(ct)dt$$
$$= \frac{2^{\nu+\rho-\mu+1}a^\mu b^\nu}{c^{2\nu+\rho+2}}\frac{\Gamma(\nu+\rho+1)}{\Gamma(\mu+1)}\frac{1}{i\pi}e^{i\pi(\nu+1)}$$
$$\times F_4\left(\nu+\rho+1,\nu+1;\mu+1,\nu+1;\frac{a^2}{c^2},\frac{b^2}{c^2}\right). \tag{6.3}$$

The function $F_1(\ldots;z,z')$ is again an Appell function and it has a series expansion in powers of the two complex variables z,z' in the domain $|z|<1$, $|z'|<1$, i.e., here

$$|x|<1 \quad \text{and} \quad |xy|<1, \tag{6.4}$$

but in the absence of supplementary conditions on the indices λ,μ,ν,ρ nothing more can be said. We just add some comments on its existence.

Because of the cyclic analytical conditions for F_4 already quoted in the Introduction we can extend these remarks to regions $b>a+c$ and $a>b+c$, i.e., for all configurations when a, b, and c cannot be the sides of a triangle.

We turn now to the case $|a-b|<c<a+b$. We recall that (Sec. II)

$$x = (a/b)e^{i\varphi_c}, \quad y = (b/a)e^{i\varphi_c},$$
$$1-x = (c/b)e^{-i\varphi_a}, \quad 1-y = (c/a)e^{-i\varphi_b},$$

where φ_a, φ_b, and φ_c are again the angles opposite to sides a, b, and c. Here,

$$F_1(\nu+\rho+1;\mu-\nu,\nu+\rho-\mu+1;\mu+1;(a/b)e^{i\varphi_c},e^{2i\varphi_c})$$

has a meaning as a double series when $a<b$, as the function $F_1(\alpha;\beta,\beta';\gamma;x,1)$ exists when $|x|<1$. It is no more true when $a>b$. Using the analytical continuation (Ref. 6 Page 1055, formula 9.183.1)

$$F_1(\alpha;\beta,\beta';\gamma;z,z') = (1-z)^{-\beta}(1-z')^{\gamma-\alpha-\beta'}$$
$$\times F_1\left(\gamma-\alpha;\beta,\gamma-\beta-\beta';\gamma;\frac{z-z'}{z-1},z'\right),$$

we get here $(z=x, z'=xy)$

$$F_1 = \left(e^{-i\varphi_a}\frac{c}{b}\right)^{\nu-\mu}(1-e^{2i\varphi_c})^{2\mu-2\nu-2\rho-1}$$
$$\times F_1(\mu-\nu-\rho;\mu-\nu,\mu-\rho;\mu+1;e^{-2i\varphi_b},e^{2i\varphi_c}),$$

which is a convergent double series as $F_1(\alpha;\beta,\beta';\gamma;1,1)$ exists provided $\alpha+\beta+\beta'<\gamma$, i.e., here $\lambda+\nu+\rho>\mu+1$ or $2(\nu+\rho-\mu)+1>0$. Collecting all these results, we get

$$\int_0^\infty t^{\nu+\rho-\mu+1}J_\mu(at)J_\nu(bt)H^{(1)}_\rho(ct)dt$$
$$= \frac{2^{\mu-\rho-\nu}}{\pi c}\left(\frac{ab}{c}\right)^{\mu-\nu-\rho-1}$$
$$\times\frac{\Gamma(\nu+\rho+1)}{\Gamma(\mu+1)}(\sin\varphi_c)^{2\mu-2\rho-2\nu-1}$$
$$\times e^{i\pi(\mu-\rho)}e^{-i\varphi_a(\mu-\rho)}e^{-i\varphi_b(2\mu-\rho-\nu)}$$
$$\times F_1(\mu-\nu-\rho;\mu-\nu,\mu-\rho;\mu+1;$$
$$e^{-2i\varphi_b},e^{2i\varphi_c}), \tag{6.5}$$
$$\text{for } |a-b|<c<a+b. \tag{6.6}$$

VII. CASE $\nu = \pm\frac{1}{2}, \rho = \pm\frac{1}{2}$

In order to use formula 1.11, namely,

$$F_4(\alpha,\alpha+\tfrac{1}{2},\gamma,\tfrac{1}{2},z,z')$$
$$= \tfrac{1}{2}(1+\sqrt{z'})^{-2\alpha}\,{}_2F_1(\alpha,\alpha+\tfrac{1}{2},\gamma,z/(\sqrt{z'}+1)^2)$$
$$+\tfrac{1}{2}(1-\sqrt{z'})^{-2\alpha}\,{}_2F_1(\alpha,\alpha+\tfrac{1}{2},\gamma,z/(\sqrt{z'}-1)^2), \tag{7.1}$$

we have to impose $\nu=-\frac{1}{2}$ and $\rho=\pm\frac{1}{2}$. We deal with Fourier transforms. We thus consider the integral

$$f_\pm = \int_0^\infty dt\,t^{\lambda-1}J_\mu(at)J_{-1/2}(bt)J_{\pm 1/2}(ct).$$

For $c>b+a$ we get the result by using the general formula (3.3) and the factorization, namely,

$$\left(\frac{2}{i\pi}\right)\int_0^\infty dt\,t^{\lambda-1}J_\mu(at)J_{-1/2}(bt)H^{(1)}_{\pm1/2}(ct)e^{-i\pi/4}$$
$$= \left(\frac{1}{i\pi}\right)2^{\lambda-2}\frac{a^\mu}{\sqrt{\pi bc}}e^{(i\pi/2)(\lambda+\mu)}e^{-(i\pi/4)[1\pm1]}$$
$$\times\frac{\Gamma((\mu+\lambda)/2-\frac{1}{2})\Gamma((\mu+\lambda)/2)}{\Gamma(\mu+1)}\left[\frac{1}{(c+b)^{\lambda+\mu-1}}\right.$$
$$\times{}_2F_1\left(\frac{\mu+\lambda}{2},\frac{\mu+\lambda}{2}-\frac{1}{2},\mu+1,\frac{a^2}{(b+c)^2}\right)$$
$$+\frac{1}{(c-b)^{\lambda+\mu-1}}{}_2F_1\left(\frac{\mu+\lambda}{2},\frac{\mu+\lambda}{2}\right.$$
$$\left.\left.-\frac{1}{2},\nu+1,\frac{a^2}{(b-c)^2}\right)\right]. \tag{7.2}$$

Actually it is simpler to note that

$$J_{1/2}(z) = \sin z(\sqrt{2/\pi z})$$

and

$$J_{-1/2}(z) = \cos z(\sqrt{2/\pi z}),$$

so that

TABLE III.
$\int_0^\infty dt\, t^{\lambda-1} J_\mu(at) J_{-1/2}(bt) J_{\pm 1/2}(ct)$, $\lambda+\mu > \frac{1\mp 1}{2}$,
and any real a,b,c.

$$\lambda < \tfrac{5}{2}$$

$$\int_0^\infty dt\, t^{\lambda-1} J_\mu(at) J_{-1/2}(bt) J_{-1/2}(ct) = -\Gamma(\mu+\lambda-1)\Gamma(\lambda-\mu-1)\frac{a^{1-\lambda}}{\sqrt{2\pi bc}}\sin\Big[(\lambda-\mu)\frac{\pi}{2}\Big]\Big\{\big[P_{\mu-1/2}^{3/2-\lambda}(\cos\varphi_+) + P_{\mu-1/2}^{3/2-\lambda}(-\cos\varphi_+)\big](\sin\varphi_+)^{3/2-\lambda}$$
$$+ \big[P_{\mu-1/2}^{3/2-\lambda}(\cos\varphi_-) + P_{\mu-1/2}^{3/2-\lambda}(-\cos\varphi_-)\big](\sin\varphi_-)^{3/2-\lambda}\Big\}.$$

$$\lambda+\mu>1$$

$$\int_0^\infty dt\, t^{\lambda-1} J_\mu(at) J_{-1/2}(bt) J_{+1/2}(ct) = -\Gamma(\mu+\lambda-1)\Gamma(\lambda-\mu-1)\frac{a^{1-\lambda}}{\sqrt{2\pi bc}}\cos\Big[(\lambda-\mu)\frac{\pi}{2}\Big]\Big\{\big[P_{\mu-1/2}^{3/2-\lambda}(\cos\varphi_+) - P_{\mu-1/2}^{3/2-\lambda}(-\cos\varphi_+)\big](\sin\varphi_+)^{3/2-\lambda}$$
$$+ \operatorname{sgn}(c-b)\big[P_{\mu-1/2}^{3/2-\lambda}(\cos\varphi_-) - P_{\mu-1/2}^{3/2-\lambda}(\cos\varphi_-)\big](\sin\varphi_-)^{3/2-\lambda}\Big\},$$

$$a>b+c,\quad \cos\varphi_\pm = |b\pm c|/a,\quad \sin\varphi_\pm = \sqrt{a^2-|b\pm c|^2}/a.$$

$$\lambda+\mu>0$$

$$\int_0^\infty dt\, t^{\lambda-1} J_\mu(at) J_{-1/2}(bt) J_{-1/2}(ct) = \frac{a^{1-\lambda}}{\pi\sqrt{bc}}\Gamma(\lambda+\mu-1)\sin\Big[(\lambda+\mu)\frac{\pi}{2}\Big]\big\{(\sinh v_+)^{1-\lambda}P_{\lambda-2}^{-\mu}(\coth v_+) + (\sinh v_-)^{1-\lambda}P_{\lambda-2}^{-\mu}(\coth v_-)\big\}.$$

$$\lambda+\mu>1$$

$$\int_0^\infty dt\, t^{\lambda-1} J_\mu(at) J_{-1/2}(bt) J_{1/2}(ct) = -\frac{a^{1-\lambda}}{\pi\sqrt{bc}}\Gamma(\lambda+\mu-1)\cos\Big[(\lambda+\mu)\frac{\pi}{2}\Big]\big\{(\sinh v_+)^{1-\lambda}P_{\lambda-2}^{-\mu}(\coth v_+) + \operatorname{sgn}(c-b)(\sinh v_-)^{1-\lambda}P_{\lambda-2}^{-\mu}(\coth v_-)\big\},$$

$$a<|c-b|,\quad \cosh v_\pm = |b\pm c|/a,\quad \sinh v_\pm = \sqrt{|b\pm c|^2-a^2}/a.$$

$$\lambda+\mu>0$$

$$\int_0^\infty t^{\lambda-1}\, dt\, J_\mu(at) J_{-1/2}(bt) J_{-1/2}(ct)$$
$$= \frac{-a^{\lambda-1}}{\pi\sqrt{bc}}\Gamma(\mu+\lambda-1)\Big[-\sin\Big[(\mu+\lambda)\frac{\pi}{2}\Big](\sinh v_+)^{1-\lambda}P_{\lambda-2}^{-\mu}(\coth v_+) + \Gamma(\lambda-\mu-1)\sqrt{2\pi}\sin\Big[(\lambda-\mu)\frac{\pi}{2}\Big](\sin\varphi_-)^{3/2-\lambda}$$
$$\times\big\{P_{\mu-1/2}^{3/2-\lambda}(\cos\varphi_-) + P_{\mu-1/2}^{3/2-\lambda}(-\cos\varphi_-)\big\}\Big].$$

$$\lambda+\mu>1$$

$$\int_0^\infty t^{\lambda-1}\, dt\, J_\mu(at) J_{-1/2}(bt) J_{1/2}(ct)$$
$$= \frac{-a^{\lambda-1}}{\pi\sqrt{bc}}\Gamma(\lambda+\mu-1)\Big[\cos\Big[(\mu+\lambda)\frac{\pi}{2}\Big](\sinh v_+)^{1-\lambda}P_{\lambda-2}^{-\mu}(\coth v_+) + \operatorname{sgn}(c-b)\Gamma(\lambda-\mu-1)\sqrt{2\pi}\sin\Big[(\lambda-\mu)\frac{\pi}{2}\Big](\sin\varphi_-)^{3/2-\lambda}$$
$$\times\big\{P_{\mu-1/2}^{3/2-\lambda}(\cos\varphi_-) - P_{\mu-1/2}^{3/2-\lambda}(-\cos\varphi_-)\big\}\Big],$$

$$b+c>a>|b-c|$$

$$\cosh v_+ = \frac{b+c}{a},\quad \sinh v_+ = \frac{\sqrt{(b+c)^2-a^2}}{a};\quad \cos\varphi_- = \frac{|b-c|}{a},\quad \sin\varphi_- = \frac{\sqrt{a^2-|b-c|^2}}{a}.$$

$$\lambda+\mu>0$$

$$J_{-1/2}(cz)J_{-1/2}(bz) = (1/\sqrt{2\pi bcz})\big[J_{-1/2}((c+b)z)\sqrt{c+b} + J_{-1/2}(|c-b|z)\sqrt{|c-b|}\big] \tag{7.3}$$

and

$$J_{1/2}(cz)J_{-1/2}(bz) = (1/\sqrt{2\pi bcz})\big[J_{1/2}((c+b)z)\sqrt{c+b} + \operatorname{sgn}(c-b)J_{+1/2}(|c-b|z)\sqrt{|c-b|}\big].$$

Function $f_\pm$ of Eq. (3.2) is now simpler since

$$\int_0^\infty t^{\lambda-3/2}\, dt\, J_\mu(at)J_{\pm 1/2}(|b\pm c|t)dt$$

is well defined[15]

$$\int_0^\infty dt\, t^{\lambda-3/2}\, dt\, J_\mu(at)J_{\pm 1/2}(|b+\epsilon c|t)$$
$$= \frac{2^{\lambda-3/2}a^\mu}{\pi|b+\epsilon c|^{\mu+\lambda-1/2}}$$
$$\times\frac{\Gamma((\mu+\lambda)/2)\Gamma((\mu+\lambda-1)/2)}{\Gamma(\mu+1)}$$
$$\times\cos\Big[\Big\{\mu+\lambda-\frac{3}{2}\mp\frac{1}{2}\Big\}\frac{\pi}{2}\Big]$$
$$\times{}_2F_1\Big(\frac{\mu+\lambda}{2},\frac{\mu+\lambda}{2}-\frac{1}{2},\mu+1,\frac{a^2}{|b+\epsilon c|^2}\Big), \tag{7.4}$$

for $a<|b+\epsilon c|$, $\epsilon=\pm 1$, and

$$\int_0^\infty dt\, t^{\lambda-3/2} J_\mu(at) J_{\pm 1/2}(|b+\epsilon c|t)$$

$$= \frac{2^{\lambda-3/2}|b+\epsilon c|^{\pm 1/2}}{a^{\lambda-1/2\pm 1/2}}$$

$$\times \frac{\Gamma((\lambda+\mu-1\pm 1)/2)\Gamma((\lambda-\mu-1\pm 1)/2)}{\Gamma(1\pm 1/2)}$$

$$\times \cos\left[\left(\lambda-\frac{3}{2}\pm\frac{1}{2}-\mu\right)\frac{\pi}{2}\right]$$

$$\times {}_2F_1\left(\frac{\lambda+\mu-1\pm 1}{2}, \frac{\lambda-\mu-1\pm 1}{2},\right.$$

$$\left. 1\pm\frac{\epsilon}{2}, \frac{|b+\epsilon c|^2}{a^2}\right), \tag{7.5}$$

$$\text{for } a>|b\pm c|, \quad \epsilon=\pm 1.$$

For the sake of completeness, we may replace the hypergeometric functions ${}_2F_1$ by the associated Legendre functions.

If $a<|c-b|$, we use the formula (Ref. 14, p. 53) valid for $0<Z<1$

$${}_2F_1(\alpha,\alpha+\tfrac{1}{2},\gamma,Z)$$

$$= 2^{\gamma-1}\Gamma(\gamma)Z^{(1-\gamma)/2}(1-Z)^{(\gamma-1-2\alpha)/2}$$

$$\times P^{1-\gamma}_{2\alpha-\gamma}(1/\sqrt{1-Z}), \tag{7.6}$$

for $a>|c-b|$ or $|c+b|$, we use the formulas (Ref. 14, pp. 53 and 54)

$${}_2F_1(\alpha,\beta,\tfrac{1}{2},Z)$$

$$= (1/\sqrt{\pi})\Gamma(\alpha+\tfrac{1}{2})\Gamma(\beta+\tfrac{1}{2})$$

$$\times 2^{\alpha+\beta-3/2}(1-Z)^{(1/2)(1/2-\alpha-\beta)}$$

$$\times [P^{1/2-\alpha-\beta}_{\alpha-\beta-1/2}(\sqrt{Z}) + P^{1/2-\alpha-\beta}_{\alpha-\beta-1/2}(-\sqrt{Z})] \tag{7.7a}$$

and

$${}_2F_1(\alpha,\beta,\tfrac{3}{2},Z)$$

$$= -(1/\sqrt{\pi Z})\Gamma(\alpha-\tfrac{1}{2})\Gamma(\beta-\tfrac{1}{2})$$

$$\times 2^{\alpha+\beta-7/2}(1-Z)^{(1/2)(3/2-\alpha-\beta)}$$

$$\times [P^{3/2-\alpha-\beta}_{\alpha-\beta-1/2}(\sqrt{Z}) - P^{3/2-\alpha-\beta}_{\alpha-\beta-1/2}(-\sqrt{Z})]. \tag{7.7b}$$

For $a<|c-b|$, we define $\cos v_\epsilon = |b+\epsilon c|/a$, $\sinh v_\epsilon = \sqrt{(b+\epsilon c)^2-a^2}/a$, and we get

$$\frac{|b+\epsilon c|^{1/2}}{\sqrt{2\pi bc}}\int_0^\infty dt\, t^{\lambda-3/2}J_\mu(at)J_{\pm 1/2}(|b+\epsilon c|t)$$

$$= \frac{2^{\lambda+\mu-2}}{\pi\sqrt{\pi}}\frac{a^{1-\lambda}}{\sqrt{bc}}\Gamma\left(\frac{\lambda+\mu}{2}\right)\Gamma\left(\frac{\lambda+\mu-1}{2}\right)$$

$$\times(\sinh v_\epsilon)^{1-\lambda}P^{-\mu}_{\lambda-2}(\cosh v_\epsilon)$$

$$\times \sin(\pi/2)[(1\pm 1)/2+\lambda+\mu], \tag{7.8}$$

$$\text{with } \Gamma\left(\frac{\lambda+\mu}{2}\right)\Gamma\left(\frac{\lambda+\mu-1}{2}\right)\frac{2^{\lambda+\mu-2}}{\sqrt{\pi}}$$

$$= \Gamma(\lambda+\mu-1).$$

For $a>b+c$, we define $\cos\varphi_\epsilon = |b+\epsilon c|/a$, $\sin\varphi_\epsilon = \sqrt{a^2-|b+\epsilon c|^2}/a$, and get

$$\frac{|b+\epsilon c|^{1/2}}{\sqrt{2\pi bc}}\int_0^\infty t^{\lambda-3/2}J_\mu(at)J_{\pm 1/2}(|b+\epsilon c|t)$$

$$= -\frac{a^{1-\lambda}}{\sqrt{2\pi}\sqrt{bc}}\Gamma(\mu+\lambda-1)\Gamma(\lambda-\mu-1)$$

$$\times(\sin\varphi_\epsilon)^{3/2-\lambda}[P^{3/2-\lambda}_{\mu-1/2}(\cos\varphi_\epsilon)$$

$$-\epsilon P^{3/2-\lambda}_{\mu-1/2}(-\cos\varphi_\epsilon)]$$

$$\times\cos([\lambda-\mu-(1\mp 1)/2]\pi/2). \tag{7.9}$$

All the results of this section are summarized in Table III.

VIII. CONCLUSION

Some generalizations may be obtained by differentiation with respect to parameters a, b, and c. For example,

$$\int_0^\infty t^\lambda J_{\mu+1}(at)J_\nu(bt)H^{(1)}_\rho(ct)dt$$

$$= \left(\frac{\mu}{a}-\frac{d}{da}\right)\int_0^\infty t^{\lambda-1}J_\mu(at)J_\nu(bt)H^{(1)}_\rho(ct)dt.$$

We can thus reach contiguous integrals, a consequence of contiguity relations both for F_4 and ${}_2F_1$ functions.

Another possible generalization is the calculation of the integral

$$\int_0^\infty t^{\lambda-1}K_\mu(at)K_\nu(bt)J_\rho(ct)dt,$$

$$\mathrm{Re}(\lambda-|\mu|-|\nu|+\rho)>0, \quad \mathrm{Re}(a+b)>|\mathrm{Im}\, c|$$

by using again analytical continuation for the same factorization cases. The F_4 function has then the same variables $-a^2/c^2$, $-b^2/c^2$ as in the initial integral (1.1). When a, b, and c are real, no restrictions are necessary in the parameter space. Some simple examples are given in the following paper.[10]

[1] W. N. Bailey, Proc. London Math. Soc. **40**, 37 (1936).

[2] P. Appell and J. Kampe de Feriet, *Fonctions hypergeometriques et hypersphériques* (Gauthier-Villars, Paris, 1926).

[3] The most recent works on analytical continuation of F_4 and its monodromy group can be found in J. Kaneko, Tokyo J. Math. **4**, 35 (1981); and K. Takano, Funkcial. Ekvac. **23**, 97 (1980).

[4] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Bateman Manuscript project (McGraw–Hill, New York, 1953), Vol. 1, p. 240.

[5] A. Gervois and H. Navelet, J. Math. Phys. **25**, 3350 (1984).

[6] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965), p. 694 and 695.

[7] W. N. Bailey, Q. J. Math. **4**, 305 (1933); W. N. Bailey, Q. J. Math. **5**, 291 (1934).

[8] Reference 6, p. 1055, Formula 9; and Ref. 2, p. 23.

[9] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Table of Integral Transforms*, Bateman Manuscript project (McGraw–Hill, New York, 1954), Vol. 1.

[10] A. Gervois and H. Navelet, J. Math. Phys. **26**, 645 (1985).

[11] Reference 4, p. 144.

[12] Reference 6, p. 695, Formula 6.578.8, but the sign of the result for the triangle case is incorrect.

[13] R. Askey, T. H. Koornwinder, and M. Rahman, "An Integral of Products of Ultraspherical Functions and a q-extension," to appear in J. London Math. Soc.

[14] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, Berlin, 1966), p. 164. In Ref. 6, p. 1006, Formula 8.736.7, you have to replace in the rhs $\cos\mu\pi$ by $\cos\nu\pi$.

[15] G. Petiau, *La Théorie des fonctions de Bessel* (CNRS, Paris, 1955), p. 204. See also Ref. 6.

Integrals of some three Bessel functions and Legendre functions. II

A. Gervois[a)] and H. Navelet
Service de Physique Théorique CEN Saclay, 91191 Gif-sur-Yvette Cedex, France

(Received 20 June 1984; accepted for publication 31 August 1984)

Integrals of three Bessel functions of the form $\int_0^\infty J_\mu(at)J_\nu(bt)H_\rho^{(1)}(ct)dt$ are calculated when $\mu,\nu,\rho,a,b,$ and c are arbitrary real numbers. For this, use is made of the factorization of the Appell function F_4 in two hypergeometric functions. Further simplifications occur if $\mu = \pm\nu$ or $\rho = \pm 1/2$. New results are given, mainly when real a, b, and c satisfy the inequalities $|a-b|<c<a+b$, which correspond to most physical situations.

I. INTRODUCTION

The integral $\int_0^\infty t^{\lambda-1}J_\mu(at)J_\nu(bt)K_\rho(ct)dt$ was given by Bailey[1] in terms of the Appell F_4 function[2] provided $\mathrm{Re}(\lambda+\mu+\nu\pm\rho)>0$, $\mathrm{Re}\,\lambda<5/2$, and $\mathrm{Re}(c\pm ia\pm ib)>0$. It can be extended to the calculation of $\int_0^\infty t^{\lambda-1}J_\mu(at)J_\nu(bt)J_\rho(ct)dt$, where a, b, and c are real parameters and $c>b+a$, which corresponds to the domain of convergence of the double series for F_4. This expression is not very tractable and the problem remains open when $|a-b|<c<a+b$, a case encountered in most physical situations.

In a companion paper denoted by I (see Ref. 3), we claim that, *for real* λ,μ,ν,ρ and *real* a,b, and c the expression

$$\int_0^\infty t^{\lambda-1}J_\mu(at)J_\nu(bt)H_\rho^{(1)}(ct)dt$$
$$=\lim_{\gamma\to0^+}2^{\lambda-1}\frac{a^\mu b^\nu}{(\gamma-ic)^{\lambda+\mu+\nu}}\times\frac{\Gamma((\lambda+\mu+\nu+\rho)/2)\,\Gamma((\lambda+\mu+\nu-\rho)/2)}{\Gamma(\mu+1)\Gamma(\nu+1)}$$
$$\times F_4\Big(\frac{\lambda+\mu+\nu+\rho}{2},\frac{\lambda+\mu+\nu-\rho}{2};\;\mu+1,\nu+1;\frac{a^2}{(c+i\gamma)^2},\frac{b^2}{(c+i\gamma)^2}\Big),\tag{1.1}$$

with

$$H_\rho^{(1)}(z)=J_\rho(z)+iY_\rho(z)\tag{1.2}$$

and

$$a+b<c,\quad \lambda<5/2,\quad \lambda+\mu+\nu-|\rho|>0,\tag{1.3}$$

has an analytical continuation when the function F_4 factorizes into products of hyper geometric functions of one variable only.

It is still probably true when λ,μ,ν,ρ are complex numbers but we did not try to look at it. We have listed all the known factorizations of F_4. All but one correspond to either one of the two changes of variables

$$a^2/c^2=-x/(1-x)(1-y),$$
$$b^2/c^2=-y/(1-y)(1-x)\tag{1.4}$$

or

$$a^2/c^2=X(1-Y),\quad b^2/c^2=Y(1-X)\tag{1.5}$$

with the involutive correspondence[4]

$$X=-x/(1-x),\quad Y=-y/(1-y).\tag{1.6}$$

All the factorization cases for the change of variable (1.4) were studied in I. In the present paper, we are interested by the transformation which appears in the factorization formula[1,4,5]

$$F_4(\alpha,\beta;\gamma,\gamma';a^2/c^2,b^2/c^2)={}_2F_1(\alpha,\beta;\gamma;X)\,{}_2F_1(\alpha,\beta;\gamma';Y)\tag{1.7}$$

when $\alpha+\beta+1=\gamma+\gamma'$, i.e.,

$$\lambda=1.\tag{1.8}$$

In spite of condition (1.8), factorization (1.7) is not too restrictive as μ,ν,ρ remain arbitrary. By derivation, for example, we may—at least formally—calculate integrals like

$$\int_0^\infty t^{l-1}J_\mu(at)J_\nu(bt)H_\rho^{(1)}(ct)dt,$$

where l is a positive integer; it is simply a consequence of contiguity relations for the functions F_4 and ${}_2F_1$.

This paper is organized in the following way. In Sec. II, we recall results for X and Y (Eq. 1.5) and discuss the general expression for $\lambda=1$ and any μ,ν,ρ (and $1+\mu+\nu-|\rho|>0$). Function ${}_2F_1$ reduces to Legendre functions when μ,ν,ρ satisfy an additional condition. We study the case $\mu=\pm\nu$ (Sec. III) and the case $\rho=\pm\frac12$ (Sec. IV) both for arbitrary real a,b,c and when $a=b$. Condition $\rho=\pm\frac12$ corresponds to the Fourier transform of $t^{-1/2}\times J_\mu(at)J_\nu(bt)$ which is not tabulated in the usual textbooks. Lastly, in the Sec. V, we indicate some possible generalizations (i) by analytical continuation [integral $\int_0^\infty K_\mu(at)K_\nu(bt)H_\rho^{(1)}(ct)dt$] and (ii) by derivation.

The most general result is Eqs. (2.9) and (2.10). Specific cases are given in Tables I and II and formulas (5.2), (5.4), (5.5), and (5.7).

II. GENERAL FORMULAS

A. Pair $X(1-Y)=a^2/c^2$, $Y(1-X)=b^2/c^2$

We recall briefly the results stated in I; X is a solution of

$$c^2X^2+X(b^2-c^2-a^2)+a^2=0,\tag{2.1}$$

where the discriminant

$$\delta=[(a+b)^2-c^2][(a-b)^2-c^2]$$
$$=a^4+b^4+c^4-2a^2b^2-2b^2c^2-2c^2a^2\tag{2.2}$$

is positive when $c>a+b$ or $c<|a-b|$ and negative when $|a-b|<c<a+b$ (triangle inequalities). We get Y by exchanging a and b:

[a)] Chercheur CNRS.

 0022-2488/85/040645-11$02.50

$$X=(c^2+a^2-b^2\pm\sqrt{\delta})/2c^2,$$
$$Y=(c^2+b^2-a^2\pm\sqrt{\delta})/2c^2, \tag{2.3}$$

where X and Y correspond to the *same* determination of $\sqrt{\delta}$. This determination was found[3] to be

$$\begin{array}{ll} -\sqrt{\delta}, & \text{if } c>a+b,\\ -i\sqrt{(-\delta)}, & \text{if } |a-b|<c<a+b,\\ \sqrt{\delta}, & \text{if } 0<c<|a-b|, \end{array}$$

where $\sqrt{\ }$ denotes the positive square root. Introducing hyperbolic or trigonometric angles,[6] we rewrite X, Y for the three cases in the following ways.

(i) $c>a+b$, $X=(a/c)e^{-u_b}$, $Y=(b/c)e^{-u_a}$,

$$\text{and } 0<X,Y<1,\ a^2=b^2+c^2-2bc\cosh u_a,\quad b^2=a^2+c^2-2ac\cosh u_b. \tag{2.4}$$

(ii) $|a-b|<c<a+b$,

$$X=(a/c)e^{-i\varphi_b},\quad Y=(b/c)e^{-i\varphi_a},$$
$$a^2=b^2+c^2-2bc\cos\varphi_a,$$
$$b^2=c^2+a^2-2ac\cos\varphi_b, \tag{2.5}$$

and $c^2=b^2+a^2-2ab\cos\varphi_c$. Note that $1-Y=X^*$, where X^* denotes the complex conjugate of X.

(iii) $0<c<|a-b|$,

$$X=\operatorname{sgn}(a-b)(a/c)e^{\operatorname{sgn}(a-b)u_b},$$
$$Y=\operatorname{sgn}(b-a)(b/c)e^{\operatorname{sgn}(b-a)u_a}, \tag{2.6}$$

where $\operatorname{sgn} t=+1$ (resp. $-1,0$) when $t>0$ (resp. $t<0$, $t=0$). When $a>b$, $u_a=u_b+u_c, a^2=b^2+c^2+2bc\cosh u_a$, and $X>1$, $Y<0$; similar formulas hold when $b>a$ leading to $X<0$ and $Y>1$.

We shall also need $X'(1-Y')=-a^2/c^2$, $Y'(1-X')=-b^2/c^2$, already studied by Bailey.[1] No problem of determination occurs then and for every a,b,c the unique solution reads ($c^2\to-c^2$),

$$X'=(c^2+b^2-a^2-\sqrt{\delta'})/2c^2,$$
$$Y'=(c^2+a^2-b^2-\sqrt{\delta'})/2c^2, \tag{2.7}$$

where $\delta'=[c^2+(a+b)^2][c^2+(a-b)^2]$ is always positive. With a hyperbolic parametrization

$$\sqrt{\delta'}=2ac\cosh v_b=2bc\cosh v_a,$$
$$c^2+b^2-a^2=2ac\sinh v_b,$$
$$c^2+a^2-b^2=2bc\sinh v_a$$

we rewrite X',Y' as

$$X'=-(a/c)e^{-v_b},\quad Y'=-(b/c)e^{-v_a}.$$

N.B. In the case $a=b$, we get

$$X=Y=\tfrac{1}{2}(1-\sqrt{1-4a^2/c^2}),\quad \text{when } c>2a,$$
$$X=Y=\tfrac{1}{2}(1-i\sqrt{4a^2/c^2-1}),\quad c<2a,$$
$$X'=Y'=\tfrac{1}{2}(1-\sqrt{1+4a^2/c^2}),\quad \forall\, a,c.$$

B. General formula

From (1.1) and (1.7) we get the general formula

$$\int_0^\infty J_\mu(at)J_\nu(bt)H_\rho^{(1)}(ct)dt=\frac{a^\mu b^\nu}{\pi c^{1+\mu+\nu}}\frac{\Gamma((1+\mu+\nu+\rho)/2)\,\Gamma((1+\mu+\nu-\rho)/2)}{\Gamma(\mu+1)\Gamma(\nu+1)}e^{-i(\pi/2)(\rho-\mu-\nu)}$$
$$\times{}_2F_1\Big(\frac{1+\mu+\nu+\rho}{2},\frac{1+\mu+\nu-\rho}{2};\mu+1;X\Big){}_2F_1\Big(\frac{1+\mu+\nu+\rho}{2},\frac{1+\mu+\nu-\rho}{2};\nu+1;Y\Big), \tag{2.8}$$

where X, Y are given by (2.3)–(2.6).

We may note the following.

(i) Both ${}_2F_1$ functions have imaginary parts when $|a-b|<c<a+b$. The symmetry in the parameters a,b,c on one hand and indices μ,ν,ρ on the other hand is by no means trivial though it surely derives from the analytical continuation for F_4 (see Ref. 2): $F_4(\cdots;z,z')\to F_4(\cdots;z/z',1/z')$, i.e., $F_4(\ldots;a^2/c^2,b^2/c^2)\to F_4(\ldots;a^2/b^2,c^2/b^2)$ or equivalently, ${}_2F_1(\ldots;X)\to{}_2F_1(\ldots,1-X)$ or ${}_2F_1(\ldots;1-1/X)$ (see Ref. 7).

(ii) When $c>a+b$, $0<X<1$, $0<Y<1$; both functions ${}_2F_1$ are defined through a real series, hence are real; in that case, the separation between real and imaginary parts in (2.8) is obvious.

(iii) When $c<|a-b|$, and for example, $a>b$, we get $X>1$ and $Y<0$ (Sec. II A). The function ${}_2F_1(\ldots;Y)$ is real when $b>c$ as $-1<Y<0$; in all other cases the analytical continuation of ${}_2F_1(\ldots;Y)$ and ${}_2F_1(\ldots;X)$ has an imaginary part. Rewriting (2.8) and using the "angles" u and φ of Sec. II A, we get, provided $\mu+\nu\pm\rho>-1$,

$$\int_0^\infty J_\mu(at)J_\nu(bt)\begin{Bmatrix}J_\rho(ct)\\ Y_\rho(ct)\end{Bmatrix}dt=\frac{a^\mu b^\nu}{\pi\times c^{1+\mu+\nu}}\frac{\Gamma((1+\mu+\nu+\rho)/2)\,\Gamma((1+\mu+\nu-\rho)/2)}{\Gamma(\mu+1)\ \ \Gamma(\nu+1)}\times\mathrm{II}, \tag{2.9}$$

$$\mathrm{II}=\begin{Bmatrix}\cos(\pi/2)(\rho-\mu-\nu)\\ \sin(\pi/2)(\rho-\mu-\nu)\end{Bmatrix}\times{}_2F_1\Big(\frac{1+\mu+\nu+\rho}{2},\frac{1+\mu+\nu-\rho}{2};\mu+1;\frac{a}{c}e^{-u_b}\Big)$$
$$\times{}_2F_1\Big(\frac{1+\mu+\nu+\rho}{2},\frac{1+\mu+\nu-\rho}{2};\nu+1;\frac{b}{c}e^{-u_a}\Big),\quad c>a+b,$$

$$= \begin{Bmatrix}\mathrm{Re}\\ \mathrm{Im}\end{Bmatrix} \times e^{-i(\pi/2)(\rho-\mu-\nu)}\, {}_2F_1\Big(\frac{1+\mu+\nu+\rho}{2}, \frac{1+\mu+\nu-\rho}{2}; \mu+1; \frac{a}{c} e^{-i\varphi_b}\Big)$$

$$\times {}_2F_1\Big(\frac{1+\mu+\nu+\rho}{2}, \frac{1+\mu+\nu-\rho}{2}; \nu+1; \frac{b}{c} e^{-i\varphi_a}\Big), \quad |a-b| < c < a+b,$$

$$= \begin{Bmatrix}\mathrm{Re}\\ \mathrm{Im}\end{Bmatrix} \times e^{-i(\pi/2)(\rho-\mu-\nu)}\, {}_2F_1\Big(\frac{1+\mu+\nu+\rho}{2}, \frac{1+\mu+\nu-\rho}{2}; \mu+1; \frac{a}{c} e^{-u_b}\Big) \tag{2.10}$$

$$\times {}_2F_1\Big(\frac{1+\mu+\nu+\rho}{2}, \frac{1+\mu+\nu-\rho}{2}; \nu+1; -\frac{b}{c} e^{-u_a}\Big), \quad a > b \text{ and } c < a-b,$$

and a similar formula holds when $b > a$ and $c < b - a$ (exchange b and a and μ and ν). No real simplification occurs in the case $a = b$.

Remark: The same work can be done with Bailey's result and X', Y' pair as the function F_4 again factorizes. To our knowledge, the expression in terms of the ${}_2F_1$ function does not appear in tables of Ref. 8. We have

$$\int_0^\infty J_\mu(at) J_\nu(bt) K_\rho(ct)\, dt$$

$$= \frac{a^\mu b^\nu}{c^{1+\mu+\nu}} \frac{\Gamma((1+\mu+\nu+\rho)/2)\,\Gamma((1+\mu+\nu-\rho)/2)}{\Gamma(\mu+1)\;\;\Gamma(\nu+1)}\, {}_2F_1\Big(\frac{1+\mu+\nu+\rho}{2}, \frac{1+\mu+\nu-\rho}{2}; \mu+1; -\frac{a}{c} e^{-v_b}\Big)$$

$$\times {}_2F_1\Big(\frac{1+\mu+\nu+\rho}{2}, \frac{1+\mu+\nu-\rho}{2}; \nu+1, -\frac{b}{c} e^{-v_a}\Big),$$

if $\mu + \nu - |\rho| > 1$.

III. CASE $\mu = \pm \nu$

A. General formulas for $a \neq b$

Formula (2.8) simplifies when the indices μ, ν, ρ satisfy an additional condition. The simplest case corresponds to $\mu = \pm \nu$. We first notice that

$${}_2F_1\Big(\frac{1+2\mu+\rho}{2}, \frac{1+2\mu-\rho}{2}; 1+\mu; X\Big)$$

$$= (1-X)^{-\mu}\, {}_2F_1\Big(\frac{1-\rho}{2}, \frac{1+\rho}{2}; 1+\mu; X\Big),$$

where $(1-X)^{-\mu}$ may introduce a phase factor and the determination of the power function corresponds to $|\arg(1-X)| < \pi$.

Now, we investigate the following three cases.

(i) $c > a + b$. As $0 < X, Y < 1$,

$${}_2F_1\Big(\frac{1-\rho}{2}, \frac{1+\rho}{2}; 1+\mu; X\Big)$$

$$= (1-X)^{\mu/2} X^{-\mu/2} \Gamma(1+\mu) P^{-\mu}_{(\rho-1)/2}(1-2X),$$

whence formulas

$${}_2F_1\Big(\frac{1+2\mu+\rho}{2}, \frac{1+2\mu-\rho}{2}; \mu+1; X\Big)$$

$$\times {}_2F_1\Big(\frac{1+2\mu+\rho}{2}, \frac{1+2\mu-\rho}{2}; \mu+1; Y\Big)$$

$$= \Gamma^2(1+\mu)\Big(\frac{ab}{c^2}\Big)^{-\mu} P^{-\mu}_{(\rho-1)/2}(1-2X)$$

$$\times P^{-\mu}_{(\rho-1)/2}(1-2Y),$$

and (3.1)

$${}_2F_1\Big(\frac{1+\rho}{2}, \frac{1-\rho}{2}; 1+\mu; X\Big)\, {}_2F_1\Big(\frac{1+\rho}{2}, \frac{1-\rho}{2}; 1-\mu; Y\Big)$$

$$= \Gamma(1+\mu)\Gamma(1-\mu)\Big(\frac{b}{a}\Big)^{\mu} P^{-\mu}_{(\rho-1)/2}(1-2X)$$

$$\times P^{\mu}_{(\rho-1)/2}(1-2Y),$$

where the P^τ_σ are the Legendre functions *on the cut* and are real because $-1 < (1-2X), (1-2Y) < 1$. We get then

$$\int_0^\infty J_\mu(at) J_\mu(bt) H^{(1)}_\rho(ct)\, dt$$

$$= \frac{\Gamma((1+2\mu+\rho)/2)\,\Gamma((1+2\mu-\rho)/2)}{\pi c}$$

$$\times e^{-(i\pi/2)(\rho-2\mu)} P^{-\mu}_{(\rho-1)/2}(1-2X)$$

$$\times P^{-\mu}_{(\rho-1)/2}(1-2Y), \tag{3.2}$$

$$\int_0^\infty J_\mu(at) J_{-\mu}(bt) H^{(1)}_\rho(ct)\, dt$$

$$= \frac{\Gamma((1+\rho)/2)\,\Gamma((1-\rho)/2)}{\pi c}$$

$$\times e^{-i\pi\rho/2} P^{-\mu}_{(\rho-1)/2}(1-2X) P^{\mu}_{(\rho-1)/2}(1-2Y),$$

where, except for the phases $e^{-(i\pi/2)(\rho-2\mu)}$, $e^{-i\pi\rho/2}$, all other factors are real.

(ii) $0 < c < |a-b|$ and, for example, $a > b$. Replacing c by $c + i\eta$ where η is small and positive, it is easy to see that X changes to $X - i0$ (small negative imaginary part). As $X > 1$ we rewrite the power function as $(1-X)^{-\mu} = e^{-i\pi\mu}(X-1)^{-\mu}$ and $(-X)^{-\mu} = e^{-i\pi\mu}(X)^{-\mu}$,

$${}_2F_1\Big(\frac{1-\rho}{2}, \frac{1+\rho}{2}; 1+\mu; X\Big)$$

$$= \Gamma(1+\mu)(X-1)^{\mu/2} X^{-\mu/2} P^{-\mu}_{(\rho-1)/2}(1-2X),$$

$${}_2F_1\Big(\frac{1-\rho}{2}, \frac{1+\rho}{2}; 1+\mu; Y\Big)$$

$$= \Gamma(1+\mu)(1-Y)^{\mu/2}(-Y)^{-\mu/2} P^{-\mu}_{(\rho-1)/2}(1-2Y),$$

where the P_σ^τ are now the Legendre functions defined in Ref. 9 and the arguments $\tilde{X}=1-2X$, $\tilde{Y}=1-2Y$ are such that $\tilde{Y}>1$ [whence $P_{(\rho-1)/2}^{-\mu}(\tilde{Y})$ is real] and $\tilde{X}<-1$ [whence $P_{(\rho-1)/2}^{-\mu}(\tilde{X})$ has an imaginary part]. By taking care of the phase factors

$$
\begin{aligned}
&{}_2F_1\Big(\frac{1+2\mu+\rho}{2},\frac{1+2\mu-\rho}{2};1+\mu;X\Big)\\
&\quad\times{}_2F_1\Big(\frac{1+2\mu+\rho}{2},\frac{1+2\mu-\rho}{2};\mu+1;Y\Big)\\
&\quad=e^{-i\pi\mu}\Gamma^2(1+\mu)(ab/c^2)^{-\mu}\\
&\quad\times P_{(\rho-1)/2}^{-\mu}(1-2X)P_{(\rho-1)/2}^{-\mu}(1-2Y),\\
&{}_2F_1\Big(\frac{1+\rho}{2},\frac{1-\rho}{2};1+\mu;X\Big)\\
&\quad\times{}_2F_1\Big(\frac{1+\rho}{2},\frac{1-\rho}{2};1-\mu;Y\Big)\\
&\quad=\Gamma(1+\mu)\Gamma(1-\mu)(b/a)^{\mu}\\
&\quad\times P_{(\rho-1)/2}^{-\mu}(1-2X)P_{(\rho-1)/2}^{-\mu}(1-2Y),
\end{aligned}
\tag{3.3}
$$

whence the final result

$$
\begin{aligned}
&\int_0^\infty J_\mu(at)J_\mu(bt)H_\rho^{(1)}(ct)dt\\
&\quad=\frac{\Gamma((1+2\mu+\rho)/2)\,\Gamma((1+2\mu-\rho)/2)}{\pi c}\\
&\quad\times e^{-i\pi\rho/2}P_{(\rho-1)/2}^{-\mu}(1-2X)P_{(\rho-1)/2}^{-\mu}(1-2Y),\\
&\int_0^\infty J_\mu(at)J_{-\mu}(bt)H_\rho^{(1)}(ct)dt\\
&\quad=\frac{\Gamma((1+\rho)/2)\,\Gamma((1-\rho)/2)}{\pi c}\\
&\quad\times e^{-i\pi\rho/2}P_{(\rho-1)/2}^{-\mu}(1-2X)P_{(\rho-1)/2}^{\mu}(1-2Y),
\end{aligned}
\tag{3.4}
$$

which holds both for $a>b$ and $a<b$. We recall that the P_σ^τ with a negative real argument is *not* real and may be calculated using, for example, the formula[9]

$$P_\nu^\mu(-z)=e^{\mp i\pi\nu}P_\nu^\mu(z)-(2/\pi)e^{-i\pi\mu}\sin[\pi(\nu+\mu)]Q_\nu^\mu(z),$$

where now the P, Q in the rhs are *real*.

(iii) $|a-b|<c<a+b$. In that case $\operatorname{Im}X$, $\operatorname{Im}Y<0$; thus $\operatorname{Im}(1-X)>0$, $\operatorname{Im}(1-Y)>0$ and with the correct determination of the power function

$$
\begin{aligned}
&{}_2F_1\Big(\frac{1-\rho}{2},\frac{1+\rho}{2};1+\mu;X\Big)\\
&\quad=\Gamma(1+\mu)(1-X)^{\mu/2}(-X)^{-\mu/2}P_{(\rho-1)/2}^{-\mu}(1-2X),\\
&{}_2F_1\Big(\frac{1-\rho}{2},\frac{1+\rho}{2};1+\mu;Y\Big)\\
&\quad=\Gamma(1+\mu)(1-Y)^{\mu/2}(-Y)^{-\mu/2}P_{(\rho-1)/2}^{-\mu}(1-2Y),
\end{aligned}
$$

where the P_σ^τ are again the Legendre functions.

We get again formulas (3.3) with the same phase factor $e^{-i\pi\mu}$ which occurs because the real positive variables are $X(1-Y)$ and $Y(1-X)$ [and not $-X(1-Y)$ and $-Y(1-X)$]. Formulas (3.4) then hold also when $|a-b|<c<a+b$ though, of course, none of the functions P_σ^τ are now real.

Results are reported in Table I (lines 1 to 6) for all cases. The variables $\tilde{X},\tilde{Y}$ have no very interesting meaning in terms of the angles and hyperbolic angles φ and u except when $a=b$ (see below). In lines 5 and 6, we give the integrals $\int_0^\infty J_\mu J_{\pm\mu}K_\rho$ which are already known[1] in terms of the F_4 function but not in this simple factorized form except when $a=b$ (see Ref. 8). For the sake of completeness, we give in line 7 $\int_0^\infty K_\mu(at)K_\mu(bt)J_\rho(ct)$, which will be calculated in Sec. V (Eq. 5.2).

B. Case $a=b$

The results simplify when $a=b$ and some were already known in case $c>2a$ (for examples, see Ref. 8, p. 675 and following). They are reported in Table II, lines 1 to 4. The difficulty comes in when $c<2a$, as now the argument $\tilde{X}=\tilde{Y}=i\sqrt{4a^2/c^2-1}$ is pure imaginary and we have to go back to real arguments for practical purposes. This can be done by using Whipple's formula,[10] which gives in our case, setting $\sin\varphi=c/2a$, $0<\varphi<\pi/2$ $(2\varphi=\varphi_c)$,

$$
\begin{aligned}
&P_{(\rho-1)/2}^{-\mu}(i\cot\varphi)\\
&\quad=\sqrt{\frac{2}{\pi}\sin\varphi}\,\frac{e^{+i\pi\rho/2-i\pi/4}}{\Gamma((1+2\mu-\rho)/2)}Q_{\mu-1/2}^{-\rho/2}(\cos\varphi-i0),
\end{aligned}
$$

where the argument lies under the real axis.

Now, we express Q in terms of the functions P,Q on the cut[11]

$$
\begin{aligned}
&e^{i\pi\rho/4}Q_\sigma^{-\rho/2}(\cos\varphi-i0)\\
&\quad=Q_\sigma^{-\rho/2}(\cos\varphi)+i(\pi/2)P_\sigma^{-\rho/2}(\cos\varphi).
\end{aligned}
$$

Recalling that $\varphi=\varphi_c/2$, some easy manipulations give

$$
\begin{aligned}
&\int_0^\infty J_\mu^2(at)H_\rho^{(1)}(ct)dt\\
&\quad=-\frac{i}{\pi^2a}\frac{\Gamma((1+2\mu+\rho)/2)}{\Gamma((1+2\mu-\rho)/2)}\\
&\quad\times\Big[Q_{\mu-1/2}^{-\rho/2}\Big(\cos\frac{\varphi_c}{2}\Big)+i\frac{\pi}{2}P_{\mu-1/2}^{-\rho/2}\Big(\cos\frac{\varphi_c}{2}\Big)\Big]^2
\end{aligned}
\tag{3.5a}
$$

and

$$
\begin{aligned}
&\int_0^\infty J_\mu(at)J_{-\mu}(at)H_\rho^{(1)}(ct)dt\\
&\quad=-\frac{i}{\pi^2a}\frac{\Gamma((1+\rho)/2)\,\Gamma((1-\rho)/2)}{\Gamma((1+2\mu-\rho)/2)\,\Gamma((1-2\mu-\rho)/2)}\\
&\quad\times\Big[Q_{\mu-1/2}^{-\rho/2}\Big(\cos\frac{\varphi_c}{2}\Big)+i\frac{\pi}{2}P_{\mu-1/2}^{-\rho/2}\Big(\cos\frac{\varphi_c}{2}\Big)\Big]\\
&\quad\times\Big[Q_{-\mu-1/2}^{-\rho/2}\Big(\cos\frac{\varphi_c}{2}\Big)+i\frac{\pi}{2}P_{-\mu-1/2}^{-\rho/2}\Big(\cos\frac{\varphi_c}{2}\Big)\Big],\\
&\quad|\rho|<1.
\end{aligned}
\tag{3.5b}
$$

The final expression is given in Table II, lines 5 and 6. For $\rho=1$, the integral $\int_0^\infty J_\mu(at)J_{-\mu}(at)J_1(ct)dt$ exists and is calculated by taking the limit $\rho\to1$ of the real part of (3.5b). We get

$$
\begin{aligned}
&\int_0^\infty J_\mu(at)J_{-\mu}(at)J_1(ct)dt\\
&\quad=(1/\pi\mu c)[\sin\mu\pi-\sin\mu(\pi-\varphi_c)],
\end{aligned}
\tag{3.6}
$$

which can be checked with Sec. IV of I.

TABLE I. Integrals

$\int J_\mu(at)J_{\pm\mu}(bt)J_\rho(ct)dt$ and $\int J_\mu(at)J_{\pm\mu}(bt)Y_\rho(ct)dt$.

Integral

$\int J_\mu(at)J_{\pm\mu}(bt)K_\rho(ct)dt$

calculated by Bailey[1] is given for completeness. The P and Q functions are the Legendre functions.

$$\int_0^\infty J_\mu(at)J_\mu(bt)\begin{Bmatrix} J_\rho(ct) \\ Y_\rho(ct)\end{Bmatrix}dt = \frac{\Gamma((1+2\mu+\rho)/2)\Gamma((1+2\mu-\rho)/2)}{\pi c}\begin{Bmatrix}\cos(\pi/2)(2\mu-\rho) \\ \sin(\pi/2)(2\mu-\rho)\end{Bmatrix} P^{-\mu}_{(\rho-1)/2}\left(1-2\frac{a}{c}e^{-u_b}\right)P^{-\mu}_{(\rho-1)/2}\left(1-2\frac{b}{c}e^{-u_a}\right), \quad \text{(i)}$$

$$= \frac{\Gamma((1+2\mu+\rho)/2)\Gamma((1+2\mu-\rho)/2)}{\pi c}\begin{pmatrix}\mathrm{Re} \\ \mathrm{Im}\end{pmatrix}\left\{e^{-i\pi\rho/2}P^{-\mu}_{(\rho-1)/2}\left(1-2\frac{a}{c}e^{-u_b}\right)P^{-\mu}_{(\rho-1)/2}\left(1+2\frac{b}{c}e^{-u_a}\right), 0<c<|a-b|\right. \quad \text{(ii) with } \begin{matrix}1+2\mu+p>0, & \text{for } J_p, \\ 1+2\mu-|p|>0, & \text{for } Y_p.\end{matrix}$$

$$= \frac{\Gamma((1+2\mu+\rho)/2)\Gamma((1+2\mu-\rho)/2)}{\pi c}\begin{pmatrix}\mathrm{Re} \\ \mathrm{Im}\end{pmatrix}\left\{e^{-i\pi\rho/2}P^{-\mu}_{(\rho-1)/2}\left(1-2\frac{a}{c}e^{-i\varphi_b}\right)P^{-u}_{(\rho-1)/2}\left(1-2\frac{b}{c}e^{-i\varphi_a}\right)\right\}, \quad \text{(iii)}$$

$$\int_0^\infty J_\mu(at)J_{-\mu}(bt)\begin{Bmatrix} J_\rho(ct) \\ Y_\rho(ct)\end{Bmatrix}dt = \frac{\Gamma((1+\rho)/2)\Gamma((1-\rho)/2)}{\pi c}\begin{Bmatrix}\cos\pi\rho/2 \\ -\sin\pi\rho/2\end{Bmatrix} P^{-\mu}_{(\rho-1)/2}\left(1-2\frac{a}{c}e^{-u_b}\right)P^{\mu}_{(\rho-1)/2}\left(1-2\frac{b}{c}e^{-u_a}\right) \quad \text{(i)}$$

$$= \frac{\Gamma((1+\rho)/2)\Gamma((1-\rho)/2)}{\pi c}\begin{pmatrix}\mathrm{Re} \\ \mathrm{Im}\end{pmatrix}\left\{e^{-i\pi\rho/2}P^{-\mu}_{(\rho-1)/2}\left(1-2\frac{a}{c}e^{-u_b}\right)P^{\mu}_{(\rho-1)/2}\left(1+2\frac{b}{c}e^{-u_a}\right)\right\} \quad \text{(ii) with } \begin{matrix}1+p>0, & \text{for } J_p, \\ 1-|p|>0, & \text{for } Y_p.\end{matrix}$$

$$= \frac{\Gamma((1+\rho)/2)\Gamma((1-\rho)/2)}{\pi c}\begin{pmatrix}\mathrm{Re} \\ \mathrm{Im}\end{pmatrix}\left\{e^{-i\pi\rho/2}P^{-\mu}_{(\rho-1)/2}\left(1-2\frac{a}{c}e^{-i\varphi_b}\right)P^{\mu}_{(\rho-1)/2}\left(1-2\frac{b}{c}e^{-i\varphi_a}\right)\right\} \quad \text{(iii)}$$

(i) $c>a+b$ $\quad a^2=b^2+c^2-2bc\cosh u_a, \quad b^2=a^2+c^2-2ac\cosh u_b, \quad c^2=a^2+b^2+2ab\cosh u_c$

(ii) $0<c<|a-b|$ $\quad$ and $a>b$ (exchange of a and b and exchange $\mu \leftrightarrow -\mu$ in the case $J_\mu J_{-\mu}$ if $b>a$)
$a^2=b^2+c^2+2bc\cosh u_a, \quad b^2=a^2+c^2-2ac\cosh u_b, \quad c^2=a^2+b^2-2ab\cosh u_c$

(iii) $|a-b|<c<a+b$ $\quad a^2=b^2+c^2-2bc\cos\varphi_a, \quad b^2=a^2+c^2-2ac\cos\varphi_b, \quad c^2=a^2+b^2-2ab\cos\varphi_c$

$$\int_0^\infty J_\mu(at)J_\mu(bt)K_\rho(ct)dt = \frac{\Gamma((1+2\mu+\rho)/2)\Gamma((1+2\mu-\rho)/2)}{2c}P^{-\mu}_{(\rho-1)/2}\left(1+2\frac{a}{c}e^{-v_b}\right)P^{-\mu}_{(\rho-1)/2}\left(1+2\frac{b}{c}e^{-v_a}\right) \quad \text{with } 1+2\mu-|p|>0$$

$$\int_0^\infty J_\mu(at)J_{-\mu}(bt)K_\rho(ct)dt = \frac{\Gamma((1+\rho)/2)\Gamma((1-\rho)/2)}{2c}P^{-\mu}_{(\rho-1)/2}\left(1+2\frac{a}{c}e^{-v_b}\right)P^{\mu}_{(\rho-1)/2}\left(1+2\frac{b}{c}e^{-v_a}\right) \quad \text{with } 1-|p|>0$$

$$\int_0^\infty K_\mu(at)K_\mu(bt)J_\rho(ct)dt = \frac{1}{c}\frac{\Gamma((1+2\mu+\rho)/2)}{\Gamma((1+2\mu-\rho)/2)}e^{2i\pi\mu}Q^{-\mu}_{(\rho-1)/2}\left(1+2\frac{a}{c}e^{-v_b}\right)Q^{-\mu}_{(\rho-1)/2}\left(1+2\frac{b}{c}e^{-v_a}\right), \quad c^2+b^2-a^2=2ac\sinh v_b, \quad c^2+a^2-b^2=2bc\sinh v_a \quad \text{with } 1+p-2|\mu|>0$$

TABLE II. $\mu = \pm\nu$ and $a = b$. Integrals $\int J_\mu(at)J_{\pm\mu}(at)J_\rho(ct)dt$ and $\int J_\mu(at)J_{\pm\mu}(at)Y_\rho(ct)dt$. Formulas were already known for $c > 2a$ (See Ref. 8). The others are new. The P and Q are again the Legendre functions.

$$\left.\begin{aligned}
&\int_0^\infty J_\mu^2(at)\begin{pmatrix} J_\rho(ct) \\ Y_\rho(ct)\end{pmatrix}dt = \frac{1}{\pi c}\Gamma\left(\frac{1+2\mu+\rho}{2}\right)\Gamma\left(\frac{1+2\mu-\rho}{2}\right)\begin{pmatrix}\cos(\pi/2)(2\mu-\rho) \\ \sin(\pi/2)(2\mu-\rho)\end{pmatrix}\left[P^{-\mu}_{(\rho-1)/2}\left(\sqrt{1-4a^2/c^2}\right)\right]^2 \\
&\int_0^\infty J_\mu(at)J_{-\mu}(at)\begin{pmatrix} J_\rho(ct) \\ Y_\rho(ct)\end{pmatrix}dt = \frac{\Gamma((1+\rho)/2)\Gamma((1-\rho)/2)}{\pi c}\begin{pmatrix}\cos(\pi\rho/2) \\ -\sin(\pi\rho/2)\end{pmatrix}P^{-\mu}_{(\rho-1)/2}\left(\sqrt{1-4a^2/c^2}\right)
\end{aligned}\right\}\quad c > 2a$$

$$\left.\begin{aligned}
&\int_0^\infty J_\mu^2(at)\begin{pmatrix} J_\rho(ct) \\ Y_\rho(ct)\end{pmatrix}dt = \frac{\Gamma((1+2\mu+\rho)/2)}{\pi^2 a\Gamma((1+2\mu-\rho)/2)}\begin{pmatrix}\mathrm{Re} \\ \mathrm{Im}\end{pmatrix}\left\{-i\left[Q^{-\rho/2}_{\mu-1/2}(\sqrt{1-4a^2/c^2}+i\frac{\pi}{2}P^{-\rho/2}_{\mu-1/2}(\sqrt{1-c^2/4a^2})\right]^2\right\} \\
&\int_0^\infty J_\mu(at)J_{-\mu}(at)\begin{pmatrix} J_\rho(ct) \\ Y_\rho(ct)\end{pmatrix}dt = \frac{1}{\pi^2 a}\frac{\Gamma((1+\rho)/2)\Gamma((1-\rho)/2)}{\Gamma((1+2\mu-\rho)/2)\Gamma((1-2\mu-\rho)/2)}\begin{pmatrix}\mathrm{Re} \\ \mathrm{Im}\end{pmatrix}\left\{-i\left[Q^{-\rho/2}_{\mu-1/2}(u)+i\frac{\pi}{2}P^{-\rho/2}_{\mu-1/2}(u)\right]\left[Q^{-\rho/2}_{-\mu-1/2}(u)+i\frac{\pi}{2}P^{-\rho/2}_{-\mu-1/2}(u)\right]\right\}
\end{aligned}\right\}\quad c < 2a \quad \mu = (\sqrt{1-c^2/4a^2})$$

$$\int^\infty J_\mu^2(at)K_\rho(ct)dt = \frac{\Gamma((1+2\mu+\rho)/2)\Gamma((1+2\mu-\rho)/2)}{2c}\left[P^{-\mu}_{(\rho-1)/2}\left(\sqrt{1+4a^2/c^2}\right)\right]^2$$

$$\int_0^\infty J_\mu(at)J_{-\mu}(at)K_\rho(ct)dt = \frac{\Gamma((1+\rho)/2)\Gamma((1-\rho)/2)}{2c}P^{-\mu}_{\rho-1/2}\left(\sqrt{1+4a^2/c^2}\right)P^{\mu}_{\rho-1/2}\left(\sqrt{1+4a^2/c^2}\right)$$

$$\int_0^\infty K_\mu^2(at)J_\rho(ct)dt = \frac{1}{c}\frac{\Gamma((1+2\mu+\rho)/2)}{\Gamma((1-2\mu+\rho)/2)}e^{2i\pi\mu}\left[Q^{-\mu}_{\rho-1/2}\left(\sqrt{1+4a^2/c^2}\right)\right]^2$$

IV. Case $\rho = \pm\frac{1}{2}$

We can restrict ourselves to $\rho = +\frac{1}{2}$ as $H^{(1)}_{-1/2} = -iH^{(1)}_{1/2}$. We get the sine (and cosine) Fourier transforms of $J_\mu(at)J_\nu(bt)/\sqrt{t}$ which are not listed in the usual books.[8,12]

We use the definitions[13]

$$\begin{aligned}
&{}_2F_1(\alpha,\alpha+\tfrac{1}{2};\gamma;z) \\
&\quad = 2^{\gamma-1}\Gamma(\gamma)z^{(1-\gamma)/2}(1-z)^{(\gamma-1-2\alpha)/2} \\
&\qquad \times P^{1-\gamma}_{2\alpha-\gamma}(1/\sqrt{1-z}),
\end{aligned}\tag{4.1a}$$

which is valid everywhere except when z is real and $z < 0$ or $z > 1$.

For $X < 0$, it becomes

$$\begin{aligned}
&{}_2F_1(\alpha,\alpha+\tfrac{1}{2};\gamma;X) \\
&\quad = 2^{\gamma-1}\Gamma(\gamma)(-X)^{(1-\gamma)/2}(1-X)^{(\gamma-1-2\alpha)/2} \\
&\qquad \times P^{1-\gamma}_{2\alpha-\gamma}(1/\sqrt{1-X}).
\end{aligned}\tag{4.1b}$$

We shall need a third possibility $X > 1$ (actually X stands for $X - i0$ with $X > 1$). Using (3.1a) we have

$$\begin{aligned}
&{}_2F_1(\alpha,\alpha+\tfrac{1}{2};\gamma;X-i0) \\
&\quad = 2^{\gamma-1}\Gamma(\gamma)X^{(1-\gamma)/2}(X-1)^{(\gamma-1-2\alpha)/2} \\
&\qquad \times e^{i(\pi/2)(\gamma-1-2\alpha)}P^{1-\gamma}_{2\alpha-\gamma}(-i(1/\sqrt{X-1}+i0)),
\end{aligned}$$

where the variable is on the imaginary axis. Using again the Whipple formula,[10] we get a Legendre function of $1/\sqrt{X}+i0$,

$$\begin{aligned}
&P^{1-\gamma}_{2\alpha-\gamma}[-i(1/\sqrt{X-1}+i0)] \\
&\quad = \frac{ie^{i\pi(2\alpha-\gamma)}}{\Gamma(2\gamma-2\alpha-1)}\sqrt{\frac{2}{\pi}}e^{i\pi/4}\left(\frac{X-1}{X}\right)^{1/4} \\
&\qquad \times Q^{\gamma-2\alpha-1/2}_{\gamma-3/2}\left(\frac{1}{\sqrt{X}}+i0\right)
\end{aligned}$$

and Q near the cut is replaced by functions on the cut

$$\begin{aligned}
&Q^{\gamma-2\alpha-1/2}_{\gamma-3/2}\left(\frac{1}{\sqrt{X}}+i0\right) \\
&\quad = e^{(3i\pi/2)(\gamma-2\alpha-1/2)}\left[Q^{\gamma-2\alpha-1/2}_{\gamma-3/2}\left(\frac{1}{\sqrt{X}}\right)\right. \\
&\qquad \left. -i\frac{\pi}{2}P^{\gamma-2\alpha-1/2}_{\gamma-3/2}\left(\frac{1}{\sqrt{X}}\right)\right],
\end{aligned}$$

whence

$$\begin{aligned}
&{}_2F_1(\alpha,\alpha+\tfrac{1}{2};\gamma;X-i0) \\
&\quad = \frac{2^{\gamma-1}\Gamma(\gamma)}{\Gamma(2\gamma-2\alpha-1)}\sqrt{\frac{2}{\pi}}X^{(1-\gamma)/2} \\
&\qquad \times(X-1)^{(\gamma-1-2\alpha)/2}\left(\frac{X-1}{X}\right)^{1/4} \\
&\qquad \times e^{i(\pi/2)(2\gamma-4\alpha-1)}\times\left[Q^{\gamma-2\alpha-1/2}_{\gamma-3/2}\left(\frac{1}{\sqrt{X}}\right)\right. \\
&\qquad \left. -i\frac{\pi}{2}P^{\gamma-2\alpha-1/2}_{\gamma-3/2}\left(\frac{1}{\sqrt{X}}\right)\right], \quad X > 1.
\end{aligned}\tag{4.1c}$$

Going back to our original problem, we rewrite the general formula (2.8) when $\rho = \frac{1}{2}$

$$\int_0^\infty J_\mu(at)J_\nu(bt)H^{(1)}_{1/2}(ct)dt$$

$$=\frac{a^\mu b^\nu}{\pi c^{1+\mu+\nu}}\frac{\Gamma((\mu+\nu+1/2)/2)\Gamma((\mu+\nu+3/2)/2)}{\Gamma(\mu+1)\Gamma(\nu+1)}$$
$$\times e^{-i(\pi/2)(1/2-\mu-\nu)}$$
$$\times{}_2F_1\Big(\frac{\mu+\nu+1/2}{2},\frac{\mu+\nu+3/2}{2};\mu+1;X\Big)$$
$$\times{}_2F_1\Big(\frac{\mu+\nu+1/2}{2},\frac{\mu+\nu+3/2}{2};\nu+1;Y\Big),\quad(4.2)$$

which we express in terms of the Legendre functions in the three cases.

(*i*) *When* $c>a+b$: $0<X,Y<1$. Then, from (4.1a)

$${}_2F_1(...;X){}_2F_1(...;Y)$$
$$=2^{\mu+\nu}\Gamma(\mu+1)\Gamma(\nu+1)X^{-\mu/2}$$
$$\times(1-Y)^{-\mu/2}Y^{-\nu/2}(1-X)^{-\nu/2}$$
$$\times\frac{1}{\sqrt[4]{(1-X)(1-Y)}}P^{-\mu}_{\nu-1/2}\Big(\frac{1}{\sqrt{1-X}}\Big)$$
$$\times P^{-\nu}_{\mu-1/2}\Big(\frac{1}{\sqrt{1-Y}}\Big),$$

$$\int_0^\infty J_\mu(at)J_\nu(bt)H^{(1)}_{1/2}(ct)dt$$
$$=\frac{2^{\mu+\nu}}{\pi c}\frac{\Gamma((\mu+\nu+1/2)/2)\,\Gamma((\mu+\nu+3/2)/2)}{\sqrt[4]{(1-X)(1-Y)}}$$
$$\times P^{-\mu}_{\nu-1/2}\Big(\frac{1}{\sqrt{1-X}}\Big)P^{-\nu}_{\mu-1/2}\Big(\frac{1}{\sqrt{1-Y}}\Big),$$

as $1-X=(b/c)e^{u_a}$, $1-Y=(a/c)e^{u_b}$, and $u_c=u_a+u_b$, we get

$$\int_0^\infty J_\mu(at)J_\nu(bt)\begin{Bmatrix}J_{1/2}(ct)\\Y_{1/2}(ct)\end{Bmatrix}dt$$
$$=\frac{2^{\mu+\nu}}{\pi}\frac{e^{-(u_a+u_b)/4}}{\sqrt{c}\,\sqrt[4]{ab}}$$
$$\times\Gamma\Big(\frac{\mu+\nu+1/2}{2}\Big)\Gamma\Big(\frac{\mu+\nu+3/2}{2}\Big)$$
$$\times P^{-\mu}_{\nu-1/2}\Big(\sqrt{\frac{c}{b}}\,e^{-u_a/2}\Big)P^{-\nu}_{\mu-1/2}\Big(\sqrt{\frac{c}{a}}\,e^{-u_b/2}\Big),$$
$$\times\begin{pmatrix}\cos(\pi/2)(\mu+\nu-1/2)\\\sin(\pi/2)(\mu+\nu-1/2)\end{pmatrix},\quad(u_a+u_b=u_c),\quad(4.3a)$$

where the arguments in functions P are real and larger than 1 so that the P^τ_σ are real.

(*ii*) $|a-b|<c<a+b$: Taking care of all the phases and using (4.7a) again, as Im $X<0$, Im $Y<0$, we get similarly

$$\int_0^\infty J_\mu(at)J_\nu(bt)\begin{Bmatrix}J_{1/2}(ct)\\Y_{1/2}(ct)\end{Bmatrix}dt$$
$$=\frac{2^{\mu+\nu}}{\pi}\frac{1}{\sqrt{c}\,\sqrt[4]{ab}}$$
$$\times\Gamma\Big(\frac{\mu+\nu+1/2}{2}\Big)\Gamma\Big(\frac{\mu+\nu+3/2}{2}\Big)$$
$$\times\begin{pmatrix}\mathrm{Re}\\\mathrm{Im}\end{pmatrix}\Big\{e^{-(i/4)(\varphi_a+\varphi_b)}e^{-i(\pi/2)(1/2-\mu-\nu)}$$
$$\times P^{-\mu}_{\nu-1/2}\Big(\sqrt{\frac{c}{b}}e^{-i\varphi_a/2}\Big)P^{-\nu}_{\mu-1/2}\Big(\sqrt{\frac{c}{a}}e^{-i\varphi_b/2}\Big)\Big\}.\quad(4.3b)$$

(*iii*) $0<c<|a-b|$ *and, for example,* $a>b$:

$X=(a/c)e^{u_b}$, $X>1$ (and actually X stands for $X-i0$),

$Y=-(b/c)e^{-u_a}$, $Y<0$.

The second ${}_2F_1$ function is written explicitly with (4.1b) and the first ${}_2F_1$ with (4.1c). We get finally

$${}_2F_1(...;X){}_2F_1(...;Y)$$
$$=\frac{2^{\mu+\nu}\Gamma(\mu+1)\Gamma(\nu+1)}{\Gamma(\mu-\nu+1/2)}\sqrt{\frac{2}{\pi}}(-Y)^{-\nu/2}$$
$$\times(1-Y)^{-(\mu+1/2)/2}P^{-\nu}_{\mu-1/2}\Big(\frac{1}{\sqrt{1-Y}}\Big)$$
$$\times X^{-\mu/2}(X-1)^{-(\nu+1/2)/2}\Big(\frac{X-1}{X}\Big)^{1/4}$$
$$\times e^{-i\pi\nu}\Big[Q^{-\nu}_{\mu-1/2}\Big(\frac{1}{\sqrt{X}}\Big)-i\frac{\pi}{2}P^{-\nu}_{\mu-1/2}\Big(\frac{1}{\sqrt{X}}\Big)\Big],$$

whence

$$\int_0^\infty J_\mu(at)J_\nu(bt)\begin{Bmatrix}J_{1/2}(ct)\\Y_{1/2}(ct)\end{Bmatrix}dt$$
$$=\frac{2^{\mu+\nu}}{\pi\sqrt{ac}}\frac{\Gamma((\mu+\nu+1/2)/2)\,\Gamma((\mu+\nu+3/2)/2)}{\Gamma(\mu-\nu+1/2)}$$
$$\times\sqrt{\frac{2}{\pi}}\begin{pmatrix}\mathrm{Re}\\\mathrm{Im}\end{pmatrix}\{e^{-i(\pi/2)(1/2-\mu+\nu)}$$
$$\times P^{-\nu}_{\mu-1/2}\Big(\sqrt{\frac{c}{a}}e^{u_b/2}\Big)$$
$$\times\Big[Q^{-\nu}_{\mu-1/2}\Big(\sqrt{\frac{c}{a}}e^{-u_b/2}\Big)-i\frac{\pi}{2}$$
$$\times P^{-\nu}_{\mu-1/2}\Big(\sqrt{\frac{c}{a}}e^{-u_b/2}\Big)\Big]$$
$$(u_a-u_b=u_c)\quad a>b,\quad(4.3c)$$

and a similar formula when $b>a$, by exchanging a and b, μ and ν.

In the derivation, many phases appear and we have preferred to verify these formulas for $\nu=1/2$ as the integrals

$$\int_0^\infty\frac{J_\mu(at)}{t}\begin{Bmatrix}\sin(tu)\\\cos(tu)\end{Bmatrix}dt$$

are known[12]. Part of the proof is sketched in Appendix A.

V. POSSIBLE GENERALIZATIONS

We investigate two kinds of generalizations.

A. Analytical continuation

From the identity

$$(2\sin\pi\mu)/\pi K_\mu(at)$$
$$=e^{-i\pi\mu/2}J_{-\mu}(e^{-i\pi/2}at)-e^{i\pi\mu/2}J_\mu(e^{-i\pi\mu/2}at),\quad(5.1)$$

we can calculate the integral

$$\int_0^\infty t^{\lambda-1} K_\mu(at) K_\nu(bt) H_\rho^{(1)}(ct)\,dt$$

in terms of the Appell function

$$F_4\left(\frac{\lambda+\mu+\nu+\rho}{2}, \frac{\lambda+\mu+\nu-\rho}{2}; \mu+1, \nu+1; -\frac{a^2}{c^2}, -\frac{b^2}{c^2}\right),$$

at least when

$$\mathrm{Re}(a+b) > |\mathrm{Im}\, c|.$$

For $\lambda = 1$, F_4 factorizes into two ${}_2F_1$ functions; the same work as above may be done with the pair $X'(1-Y') = -a^2/c^2$, $Y'(1-X') = -b^2/c^2$, which was previously introduced by Bailey[1] for the integral $\int_0^\infty J_\mu(at) J_\nu(bt) K_\rho(ct)\,dt$, the main results being indicated in Sec. II A. In that case, the analytical continuation for real a,b,c, is straightforward; we have

$$X' = (c^2 + b^2 - a^2 - \sqrt{\delta'})/2c^2,$$

$$Y' = (c^2 + a^2 - b^2 - \sqrt{\delta'})/2c^2,$$

$$\delta' = (c^2 + b^2 - a^2)^2 + 4a^2c^2 = [c^2 + (a+b)^2][c^2 + (a-b)^2],$$

and X', Y' are real negative for every real value of a,b,c.

The general formula is a lengthy sum which is of no use here. It reduces drastically when $\mu = \nu$ (which gives the same result as $\mu = -\nu$, as $K_\nu = K_{-\nu}$). After some tedious manipulations we get

$$\int_0^\infty K_\mu(at) K_\mu(bt) J_\rho(ct)\,dt = \frac{1}{c} \frac{\Gamma((1+2\mu+\rho)/2)}{\Gamma((1-2\mu+\rho)/2)} Q_{(\rho-1)/2}^{-\mu}(1-2X') \times Q_{(\rho-1)/2}^{-\mu}(1-2Y') e^{2i\pi\mu}, \tag{5.2}$$

where the $Q_\sigma^\tau e^{-i\pi\sigma}$ are real as $1-2X'$, $1-2Y' > 1$. This formula is a generalization of that of Ref. 8, page 668 (formula 6.513.5) for $a = b$. The expression for the other integral $\int_0^\infty K_\mu(at) K_\mu(bt) Y_\rho(ct)\,dt$ is more complicated and we do not give it here.

B. Derivation

As we pointed out in the Introduction, we may obtain the values of the integrals when λ is raised by one or more units, by derivation with respect to one of the parameters a,b,c. It is always theoretically possible, but it is interesting only in simple cases.

For example, one can verify easily that

$$\int_0^\infty t J_\mu(at) J_{\mu+1}(at) J_\rho(ct)\,dt = \left(\frac{\mu}{a} - \frac{1}{2}\frac{d}{da}\right) \int_0^\infty J_\mu^2(at) J_\rho(ct)\,dt,$$

$$\int_0^\infty t J_\mu^2(at) J_\rho(ct)\,dt = \left(\frac{\rho+1}{c} + \frac{d}{dc}\right) \int_0^\infty J_\mu^2(at) J_{\rho+1}(ct)\,dt,$$

$$\int_0^\infty t J_\mu(at) J_{\mu+1}(at) K_\rho(ct)\,dt = \left(\frac{\mu}{a} - \frac{1}{2}\frac{d}{da}\right) \int_0^\infty J_\mu^2(at) K_\rho(ct)\,dt,$$

$$\int_0^\infty t J_\mu^2(at) K_\rho(ct)\,dt = -\left(\frac{\rho+1}{c} + \frac{d}{dc}\right) \int_0^\infty J_\mu^2(at) K_{\rho+1}(ct)\,dt, \tag{5.3}$$

$$\int_0^\infty t K_\mu^2(at) J_\rho(ct)\,dt = \left(\frac{\rho+1}{c} + \frac{d}{dc}\right) \int_0^\infty K_\mu^2(at) J_{\rho+1}(ct)\,dt,$$

and

$$\int_0^\infty t K_\mu(at) K_{\mu+1}(at) J_\rho(ct)\,dt = \left(\frac{\mu}{a} - \frac{1}{2}\frac{d}{da}\right) \int_0^\infty K_\mu^2(at) J_\rho(ct)\,dt.$$

The integrals on the rhs are known and tabulated in Table II. Those on the lhs are calculated using

$$(z^2-1)\frac{d}{dz} P_\nu^\mu(z) = (\nu+\mu)(\nu-\mu+1)\sqrt{z^2-1}\, P_\nu^{\mu-1} - \mu z P_\nu^\mu(z) = \nu z P_\nu^\mu(z) - (\mu+\nu) P_{\nu-1}^\mu(z),$$

$$(1-x^2)\frac{d}{dx} P_\nu^\mu(x) = -\nu x P_\nu^\mu(x) + (\nu+\mu) P_{\nu-1}^\mu(x) = (\mu+\nu)(\nu-\mu+1)\sqrt{1-x^2}\, P_\nu^{\mu-1}(x) + \mu x P_\nu^\mu(x),$$

and identical relations for $Q_\nu^\mu(z)$ and $Q_\nu^\mu(x)$ (see Ref. 9). The four last formulas of (5.3) give known results (see Ref. 8), pages 672–73, formulas 6.522.7–1.2 and 8). The first two formulas are not tabulated to our knowledge and are new. They give different results, according whether $c > 2a$ or $c < 2a$.

We get

$$\int_0^\infty t J_\mu(at) J_{\mu+1}(at) J_\rho(ct)\,dt = \frac{2}{c^2} \frac{\Gamma((3+2\mu+\rho)/2)}{\Gamma((-1-2\mu+\rho)/2)} \left(1 - \frac{4a^2}{c^2}\right)^{-1/2} \times P_{(\rho-1)/2}^{-\mu}\left(\sqrt{1-\frac{4a^2}{c^2}}\right) \times P_{(\rho-1)/2}^{-\mu-1}\left(\sqrt{1-\frac{4a^2}{c^2}}\right), \quad c > 2a,$$

$$= \frac{1}{2\pi a^2} \frac{\Gamma((3+2\mu+\rho)/2)}{\Gamma((-1-2\mu+\rho)/2)} \left(1 - \frac{c^2}{4a^2}\right)^{-1/2} \times \left\{ P_{\mu+1/2}^{-\rho/2}\left(\sqrt{1-\frac{c^2}{4a^2}}\right) \times Q_{\mu-1/2}^{-\rho/2}\left(\sqrt{1-\frac{c^2}{4a^2}}\right) \right.$$

$$+P_{\mu-1/2}^{-\rho/2}\left(\sqrt{1-\frac{c^2}{4a^2}}\right)$$
$$\times Q_{\mu+1/2}^{-\rho/2}\left(\sqrt{1-\frac{c^2}{4a^2}}\right)\Bigg\},$$
$$c<2a, \tag{5.4}$$

and

$$\int_0^\infty tJ_\mu^2(at)J_\rho(ct)dt$$
$$=\frac{2}{c^2}\frac{\Gamma((2+2\mu+\rho)/2)}{\Gamma((-2\mu+\rho)/2)}\left(1-\frac{4a^2}{c^2}\right)^{-1/2}$$
$$\times P_{\rho/2}^{-\mu}\left(\sqrt{1-\frac{4a^2}{c^2}}\right)$$
$$\times P_{\rho/2-1}^{-\mu}\left(\sqrt{1-\frac{4a^2}{c^2}}\right),$$
$$c>2a,$$
$$=\frac{1}{2\pi a^2}\frac{\Gamma((2+2\mu+\rho)/2)}{\Gamma((-2\mu+\rho)/2)}\left(1-\frac{c^2}{4a^2}\right)^{-1/2}$$
$$\times\{P_{\mu-1/2}^{-(\rho-1)/2}(u)P_{\mu-1/2}^{-(\rho+1)/2}(u)$$
$$+Q_{\mu-1/2}^{-(\rho+1)/2}(u)Q_{\mu-1/2}^{-(\rho-1)/2}(u)\},$$
$$u=\sqrt{1-c^2/4a^2},\quad c<2a. \tag{5.5}$$

Another application is the derivation of the integral

$$\textstyle\int_0^\infty tJ_\mu(at)\,J_\nu(bt)Y_{\mu+\nu}(ct)dt$$
$$=\left(\frac{\mu+\nu-1}{c}-\frac{d}{dc}\right)$$
$$\times\int_0^\infty J_\mu(at)J_\nu(bt)Y_{\mu+\nu-1}(ct)dt, \tag{5.6}$$

which could not be calculated with the methods of Refs. 3 and 6, while $\int tJ_\mu J_\nu J_{\mu+\nu}$, $\int tJ_\mu J_\nu J_{\nu-\mu}$, and $\int tJ_\mu J_\nu Y_{\nu-\mu}$ were determined. The details of the calculation are reported in Appendix B. We have

$$\frac{\int_0^\infty tJ_\mu(at)J_\nu(bt)Y_{\mu+\nu}(ct)dt}{(a^\mu b^\nu/2\pi c^{\mu+\nu})\,[\Gamma(\mu+\nu)/\Gamma(\mu+1)\Gamma(\nu+1)]}$$
$$=(1/\widetilde{\Delta})\{\mu_2F_1(\mu+\nu,1;\nu+1,Y)$$
$$+\nu_2F_1(\mu+\nu,1;\mu+1;X)\},$$
$$c>a+b,$$
$$=(1/\Delta)\mathrm{Im}\{\mu_2F_1(\mu+\nu,1;\nu+1,Y)$$
$$+\nu_2F_1(\mu+\nu,1;\mu+1;X)\},$$
$$|a-b|<c<a+b, \tag{5.7}$$
$$=-(\mu/\widetilde{\Delta})[{}_2F_1(\mu+\nu,1;\nu+1;Y)$$
$$-{}_2F_1(\mu+\nu,1;\nu+1-X)]-(\cos\pi\nu/\widetilde{\Delta})$$
$$\times e^{-\nu u_a-\mu u_b}\frac{c^{\mu+\nu}}{a^\mu b^\nu}\frac{\Gamma(\mu+1)\Gamma(\nu+1)}{\Gamma(\mu+\nu)},$$
$$a>b\ \text{and}\ c<a-b$$

and a similar result for $b>a$, $c<b-a$ (exchange of μ and ν and a and b). We have set

$$2\widetilde{\Delta}=ab\sinh u_c,\quad 2\Delta=ab\sin\varphi_c. \tag{5.8}$$

ACKNOWLEDGMENTS

The authors are grateful to Dr. M. L. Metha for many comments on the manuscript.

APPENDIX A: CHECK OF FORMULAS (4.3) FOR $\rho=\frac{1}{2}$

We give only verification for $c<|a-b|$, $a>b$, i.e., for formula (4.3c):

$$I=\int_0^\infty J_\mu(at)J_{1/2}(bt)H_{1/2}^{(1)}(ct)dt$$
$$=\frac{1}{\pi\sqrt{bc}}\int_0^\infty\frac{J_\mu(at)}{t}\{e^{-i(b-c)t}-e^{i(b+c)t}\}.$$

From Ref. 12, we get

$$I=(1/\pi\mu\sqrt{bc})[e^{-i\mu\arcsin[(b-c)/a]}-e^{i\mu\arcsin[(b+c)/a]}]$$

Now, from (4.3c), it reads

$$I=\frac{2^{\mu+1/2}}{\pi\sqrt{ac}}\sqrt{\frac{2}{\pi}}\frac{\Gamma(\mu/2+1/2)\Gamma(\mu/2+1)}{\Gamma(\mu)}$$
$$\times e^{-i(\pi/2)(1-\mu)}P_{\mu-1/2}^{-1/2}\left(\sqrt{\frac{c}{a}}e^{u_b/2}\right)$$
$$\times\Bigg[Q_{\mu-1/2}^{-1/2}\left(\sqrt{\frac{c}{a}}e^{-u_b/2}\right)$$
$$-i\frac{\pi}{2}P_{\mu-1/2}^{-1/2}\left(\sqrt{\frac{c}{a}}e^{-u_b/2}\right)\Bigg],$$

with[14]

$$P_{\mu-1/2}^{-1/2}(x)$$
$$=-\frac{i}{\sqrt{2\pi}\mu}\frac{1}{\sqrt[4]{1-x^2}}$$
$$\times[(x+i\sqrt{1-x^2})^\mu-(x-i\sqrt{1-x^2})^\mu],$$
$$Q_{\mu-1/2}^{-1/2}(x)$$
$$=\frac{1}{2}\sqrt{\frac{\pi}{2}}\frac{1}{\mu}\frac{1}{\sqrt[4]{1-x^2}}$$
$$\times[(x+i\sqrt{1-x^2})^{-\mu}+(x-i\sqrt{1-x^2})^{-\mu}],$$

and the duplication formula

$$\Gamma(\mu/2+\tfrac{1}{2})\Gamma(\mu/2+1)/\Gamma(\mu)=\mu\sqrt{\pi}/2^\mu.$$

Collecting all these results, we get for the integral

$$I=(1/\pi\mu\sqrt{bc})2\sin\mu\theta_1[\sin\mu\theta_2-i\cos\mu\theta_2]$$
$$=(1/\pi\mu\sqrt{bc})[e^{-i\mu(\theta_1-\theta_2)}-e^{i\mu(\theta_1+\theta_2)}],$$

where

$$\cos\theta_1=\sqrt{c/a}e^{u_b/2},\quad\sin\theta_2=\sqrt{c/a}e^{-u_b/2}.$$

The equality between the two expressions results from a straightforward application of the relation

$$\sin(\theta_1\pm\theta_2)=(b\pm c)/a.$$

For $c<|a-b|$ but $b>a$,

$$\int_0^\infty J_\mu(at)J_{1/2}(bt)H_{1/2}^{(1)}(ct)dt$$
$$=\frac{1}{\pi\mu\sqrt{bc}}\Bigg\{e^{-i\mu\pi/2}\left[\frac{a}{b-c+\sqrt{(b-c)^2-a^2}}\right]^\mu$$

$$- e^{i\mu\pi/2}\left[\frac{a}{b+c+\sqrt{(b+c)^2-a^2}}\right]^{\mu}\Bigg\}.$$

Formula (4.3c) exchanging a and b and μ and ν gives

$$\frac{2^{\mu+1/2}}{\pi\sqrt{bc}}\frac{\Gamma(\mu/2+\frac{1}{2})\Gamma(\mu/2+1)}{\Gamma(1-\mu)}$$

$$\times\sqrt{\frac{2}{\pi}}e^{-i\pi\mu/2}P_0^{-\mu}\left(\sqrt{\frac{c}{b}}e^{u_a/2}\right)$$

$$\times\left[Q_0^{-\mu}\left(\sqrt{\frac{c}{b}}e^{-u_a/2}\right)-i\frac{\pi}{2}P_0^{-\mu}\left(\sqrt{\frac{c}{b}}e^{-u_a/2}\right)\right]$$

$$=\frac{1}{\pi\mu\sqrt{bc}}e^{-i\pi\mu/2}\left(\tan\frac{\alpha_1}{2}\right)^{\mu}\left[\left(\tan\frac{\alpha_2}{2}\right)^{-\mu}\right.$$

$$\left.-\cos\mu\pi\left(\tan\frac{\alpha_2}{2}\right)^{\mu}-i\left(\tan\frac{\alpha_2}{2}\right)^{\mu}\sin\pi\mu\right]$$

$$=\frac{1}{\pi\mu\sqrt{bc}}\left[e^{-i\pi\mu/2}\left(\frac{\tan\alpha_1/2}{\tan\alpha_2/2}\right)^{\mu}\right.$$

$$\left.-e^{i\pi\mu/2}(\tan\alpha_1/2\tan\alpha_2/2)^{\mu}\right],$$

where

$$\cos\alpha_1=\sqrt{c/b}\,e^{u_a/2},\quad \cos\alpha_2=\sqrt{c/b}\;e^{-u_a/2}.$$

Using $\tan\alpha=\sqrt{(1-\cos\alpha)/(1+\cos\alpha)}$, it remains to verify that

$$\left(\frac{1-\sqrt{c/b}\,e^{u_a/2}}{1+\sqrt{c/b}\,e^{u_a/2}}\right)^{1/2}\left(\frac{1\pm\sqrt{c/b}\,e^{-u_a/2}}{1\mp\sqrt{c/b}\,e^{-u_a/2}}\right)^{1/2}$$

$$=\frac{a}{b\mp c+\sqrt{(b\mp c)^2-a^2}}.$$

This completes the proof.

APPENDIX B: CALCULATION OF $I=\int_0^\infty tJ_\mu(at)J_\nu(bt)H^{(1)}_{\mu+\nu}(ct)dt$

Starting from Eqs. (2.8)–(5.6), we rewrite I as

$$I=\frac{i\Gamma(\mu+\nu)}{\pi\Gamma(\mu+1)\Gamma(\nu+1)}\left(\frac{\mu+\nu-1}{c}-\frac{d}{dc}\right)$$

$$\times\left[\frac{a^\mu b^\nu}{c^{\mu+\nu+1}}{}_2F_1(\mu+\nu,1;\right.$$

$$\left.\mu+1;X){}_2F_1(\mu+\nu,1;\nu+1;Y)\right]$$

$$=\frac{i\Gamma(\mu+\nu)}{\pi\Gamma(\mu+1)\Gamma(\nu+1)}\left(\frac{\mu+\nu-1}{c}-\frac{d}{dc}\right)$$

$$\times[a^{-\mu}b^{-\nu}c^{\mu+\nu-1}X^{\mu}(1-X)^{\nu}$$

$$\times{}_2F_1(\mu+\nu,1;\mu+1;X)$$

$$\times Y^{\nu}(1-Y)^{\mu}\,{}_2F_1(\mu+\nu,1;\nu+1;Y)],$$

where X,Y are defined in Eqs. (2.3)–(2.6).

Using

$$\frac{d}{dx}X^{\mu}(1-X)^{\mu}\,{}_2F_1(\mu+\nu,1;\mu+1;X)$$

$$=\mu X^{\mu-1}(1-X)^{\nu-1}$$

we have now

$$I=-\frac{ia^\mu b^\nu}{c^{\mu+\nu+1}}\frac{\Gamma(\mu+\nu)}{\pi\Gamma(\mu+1)\Gamma(\nu+1)}$$

$$\times\left[\frac{\partial X}{\partial c}\frac{\mu}{X(1-X)}{}_2F_1(\mu+\nu,1;\nu+1;Y)\right.$$

$$\left.+\frac{\partial Y}{\partial c}\frac{\nu}{Y(1-Y)}{}_2F_1(\mu+\nu,1;\mu+1;X)\right].$$

We examine the three cases one by one.

(i) $c>a+b$. As $X=(a/c)e^{-u_b}$ and $2ac\cosh u_b=c^2+a^2-b^2$, we have $\partial X/\partial c=-cX(1-X)/2\tilde{\Delta}$ and similarly $\partial Y/\partial c=-cY(1-Y)/2\tilde{\Delta}$, where $\sqrt{\delta}=4\tilde{\Delta}=2ab\sinh u_c=2bc\sinh u_a=2ca\sinh u_b$,

$$I=\frac{ia^\mu b^\nu}{2\pi\tilde{\Delta}c^{\mu+\nu}}\frac{\Gamma(\mu+\nu)}{\Gamma(\mu+1)\Gamma(\nu+1)}$$

$$\times[\mu\,{}_2F_1(\mu+\nu,1;\nu+1;Y)+\nu\,{}_2F_1(\mu+\nu,1;\mu+1;X)],$$

as $0<X<1$, $0<Y<1$ both functions ${}_2F_1$ are real and I is pure imaginary. We recover the result of Ref. 6

$$\int_0^\infty tJ_\mu(at)J_\nu(bt)J_{\mu+\nu}(ct)dt=0.$$

(ii) $|a-b|<c<a+b$: From $X=(a/c)e^{-i\varphi_b}$ and $Y=(b/c)e^{-i\varphi_a}$ we get

$$\frac{\partial X}{\partial c}=\frac{icX(1-X)}{2\Delta},\quad\frac{\partial Y}{\partial c}=\frac{icY(1-Y)}{2\Delta},$$

where $\sqrt{-\delta}=4\Delta=2ab\sin\varphi_c=2bc\sin\varphi_a=2ca\sin\varphi_b$ and Δ is the area of the triangle with sides of length a,b,c. We have

$$I=\frac{a^\mu b^\nu}{2\pi\Delta c^{\mu+\nu}}\frac{\Gamma(\mu+\nu)}{\Gamma(\mu+1)\Gamma(\nu+1)}$$

$$\times[\mu\,{}_2F_1(\mu+\nu,1;\nu+1;Y)$$

$$+\nu\,{}_2F_1(\mu+\nu,1;\mu+1;X)].$$

As X and Y are complex numbers, the functions ${}_2F_1$ are complex numbers too. We can separate real and imaginary parts by using the analytical continuation[7]

$${}_2F_1(\mu+\nu,1;\nu+1;Y)$$

$$=-(\nu/\mu){}_2F_1(\mu+\nu,1;\mu+1;1-Y)$$

$$+\frac{\Gamma(\mu)\Gamma(\nu+1)}{\Gamma(\mu+\nu)}(1-Y)^{-\mu}$$

$$\times{}_2F_1(1-\mu,\nu;1-\mu;1-Y).$$

Remembering that $1-Y=X^*$, the above expression reads

$$-\frac{\nu}{\mu}{}_2F_1(\mu+\nu,1;\mu+1;X^*)$$

$$+\frac{\Gamma(\nu+1)\Gamma(\mu)}{\Gamma(\mu+\nu)}(1-Y)^{-\mu}Y^{-\nu},$$

whence as

$$(a^\mu b^\nu/c^{\mu+\nu})(1-Y)^{-\mu}Y^{-\nu}=e^{i(\nu\varphi_a-\mu\varphi_b)},$$

$$I=\frac{1}{2\pi\Delta}e^{i(\nu\varphi_a-\mu\varphi_b)}+i\frac{\nu}{\Delta\pi}\frac{a^\mu b^\nu}{c^{\mu+\nu}}\frac{\Gamma(\mu+\nu)}{\Gamma(\mu+1)\Gamma(\nu+1)}$$

$$\times\mathrm{Im}\,{}_2F_1(\mu+\nu,1;\mu+1;X)$$

or, in a more symmetrical way,

$$I=\frac{1}{2\pi\Delta}\cos(\nu\varphi_a-\mu\varphi_b)+\frac{i}{2\Delta\pi}\frac{\Gamma(\mu+\nu)}{\Gamma(\mu+1)\Gamma(\nu+1)}\frac{a^\mu b^\nu}{c^{\mu+\nu}}$$
$$\times\operatorname{Im}[\nu\,{}_2F_1(\mu+\nu,1;\mu+1;X)$$
$$+\mu\,{}_2F_1(\mu+\nu,1;\nu+1;Y)].$$

Again, we recover[6]

$$\int_0^\infty tJ_\mu(at)J_\nu(bt)J_{\mu+\nu}(ct)dt$$
$$=\frac{1}{2\Delta\pi}\cos(\nu\varphi_a-\mu\varphi_b).$$

(iii) $a>b$, $a-b>c$: If

$$X=\frac{a}{c}e^{u_b},\quad 1-X=-\frac{b}{c}e^{u_a},$$
$$Y=-\frac{b}{c}e^{-u_a},\quad 1-Y=\frac{a}{c}e^{-u_b},$$
$$\frac{\partial X}{\partial c}=\frac{cX(1-X)}{2\widetilde{\Delta}},\quad \frac{\partial Y}{\partial c}=\frac{cY(1-Y)}{2\widetilde{\Delta}},$$

where again $4\widetilde{\Delta}=\sqrt{\delta}$; then

$$I=-\frac{ia^\mu b^\nu}{2\pi\widetilde{\Delta}c^{\mu+\nu}}\frac{\Gamma(\mu+\nu)}{\Gamma(\mu+1)\Gamma(\nu+1)}$$
$$\times[\mu\,{}_2F_1(\mu+\nu,1;\nu+1;Y)$$
$$+\nu\,{}_2F_1(\mu+\nu,1;\mu+1;X)].$$

Though X, Y are real, at least one of them, X, has a modulus greater than 1. Then the term in bracket is not real. As[7]

$${}_2F_1(\mu+\nu,1;\mu+1,X)$$
$$=-\frac{\mu}{\nu}{}_2F_1(\mu+\nu,1;\nu+1,1-X)$$
$$+\frac{\Gamma(\mu+1)\Gamma(\nu)}{\Gamma(\mu+\nu)}e^{-i\pi\nu}(X-1)^{-\nu}$$
$$\times{}_2F_1(1-\nu,\mu;1-\nu;1-X)$$
$$=-\frac{\mu}{\nu}{}_2F_1(\mu+\nu,1;\nu+1,1-X)$$
$$+\frac{\Gamma(\mu+1)\Gamma(\nu)}{\Gamma(\mu+\nu)}e^{-i\pi\nu}(X-1)^{-\nu}X^{-\mu}$$
$$=-\frac{\mu}{\nu}{}_2F_1(\mu+\nu,1;\nu+1;1-X)$$
$$+e^{-i\pi\nu}\frac{\Gamma(\mu+1)\Gamma(\nu)}{\Gamma(\mu+\nu)}\frac{c^{\mu+\nu}}{a^\mu b^\nu}e^{-\nu u_a-\mu u_b},$$

and

$$I=-\frac{ie^{-i\pi\nu}}{2\pi\widetilde{\Delta}}e^{-\nu u_a-\mu u_b}$$
$$-\frac{ia^\mu b^\nu}{2\pi\widetilde{\Delta}c^{\mu+\nu}}\frac{\Gamma(\mu+\nu)}{\Gamma(\mu+1)\Gamma(\nu+1)}$$
$$\times[\mu_2F_1(\mu+\nu,1;\nu+1;Y)$$
$$-\mu_2F_1(\mu+\nu,1;\nu+1;1-X)].$$

As the real part is known[6] and reduces to

$$-(\sin\pi\nu/2\pi\widetilde{\Delta})e^{-\nu u_a-\mu u_b},$$

the term in brackets is real and the imaginary part of I reads

$$-\frac{\cos\pi\nu}{2\pi\widetilde{\Delta}}e^{-\nu u_a-\mu u_b}-\frac{a^\mu b^\nu}{2\pi\widetilde{\Delta}c^{\mu+\nu}}\frac{\Gamma(\mu+\nu)}{\Gamma(\mu+1)\Gamma(\nu+1)}$$
$$\times\mu[{}_2F_1(\mu+\nu,1;\nu+1;Y)$$
$$-{}_2F_1(\mu+\nu,1;\nu+1;1-X)],$$

which is, of course, not symmetrical in a and b.

[1]W. N. Bailey, Proc. London Math. Soc. **40**, 37 (1936).
[2]P. Appell and J. Kampé de Feriet, *Fonctions hypergéométriques et hypersphériques* (Gauthier–Villars, Paris, 1926).
[3]A. Gervois and H. Navelet, J. Math. Phys. **26**, 633 (1985).
[4]W. N. Bailey, Q. J. Math. **5**, 291 (1934).
[5]W. N. Bailey, Q. J. Math. **4**, 305 (1933).
[6]A. Gervois and H. Navelet, J. Math. Phys. **25**, 3350 (1984).
[7]A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Bateman Manuscript project (Mc Graw–Hill, New York, 1953), Vol. 1, pp. 108 and 109.
[8]I. S. Gradshteyn and I. M. Ryshik, *Tables of Integrals, Series and Products* (Academic, New York, 1965), pp. 667, 668, and 675.
[9]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, Berlin, 1966), p. 151 and on.
[10]Reference 9, p. 164 or Ref. 7, p. 141. Whipple's formula is incorrectly reported in Ref. 8, p. 1006, formulas 8.738.1 and 2. In 8.738.2, $\exp i\pi(\nu+\frac{1}{2})$ must be replaced by $\exp i\pi(\nu+\frac{1}{4})$. In 8.738.1, $\exp i\pi(\mu-(\nu+1)/2)$ must be replaced by $\exp i\pi(\mu-\frac{1}{4})$; the exchange of μ and ν in the rhs function P is quoted in the erratum.
[11]Reference 7, p. 144.
[12]A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Table of Integral Transforms*, Bateman manuscript project (Mc Graw–Hill, New York, 1954), Vol. 1, pp. 43 and 99.
[13]Reference 9, p. 53.
[14]In Reference 7, p. 150, formula (13) for $Q^{-1/2}$, an overall negative sign is missing.

Bessel function expansions of Coulomb wave functions

J. Humblet
W. K. Kellogg Radiation Laboratory, 106–038, California Institute of Technology, Pasadena, California 91125 and University of Liège, Institute of Physics, B.5, Sart Tilman, B-4000 Liège 1, Belgium

(Received 6 March 1984; accepted for publication 12 October 1984)

From the convergence properties of the expansion of the function $\Phi_l \propto r^{-l-1} F_l$ in powers of the energy, we successively obtain the expansions of F_l and G_l as single series of modified Bessel functions I_{2l+1+n} and K_{2l+1+n}, respectively, as well as corresponding asymptotic approximations of G_l for $|\eta| \to \infty$. Both repulsive and attractive fields are considered for real and complex energies as well. The expansion of F_l is not new, but its convergence is given a simpler and corrected proof. The simplest form of the asymptotic approximations obtained for G_l, in the case of a repulsive field and for low positive energies, is compared to an expansion obtained by Abramowitz.

I. INTRODUCTION AND NOTATION

The usual Coulomb wave functions F_l and G_l satisfy the differential equation

$$\left[\frac{d^2}{d\rho^2} + 1 - \frac{2\eta}{\rho} - \frac{l(l+1)}{\rho^2}\right] R = 0, \tag{1.1}$$

in which $\rho = kr$, while η is the Sommerfeld parameter, i.e.,

$$\eta = \alpha/k, \quad \alpha = Z_1 Z_2 e^2 M \hbar^{-2}. \tag{1.2}$$

Let us also define the Gamov factor $C_l(\eta)$, the polynomial in η^{-2}, $u_l(\eta)$, and the functions Φ_l and Θ_l by the equations[1–5]

$$C_l(\eta) = 2^l e^{-\pi\eta/2} \times [\Gamma(l+1-i\eta)\Gamma(l+1+i\eta)]^{1/2}/(2l+1)!, \tag{1.3a}$$

$$C_0(\eta) = \{2\pi\eta/[\exp(2\pi\eta) - 1]\}^{1/2}, \tag{1.3b}$$

$$u_l(\eta) = \frac{\Gamma(l+1+i\eta)}{(i\eta)^{2l+1}\Gamma(-l+i\eta)} = \frac{(2l+1)!^2 C_l^2}{(2\eta)^{2l} C_0^2} \tag{1.4a}$$

$$= (1 + l^2\eta^{-2}) \cdots (1 + 2^2\eta^{-2})(1 + 1^2\eta^{-2}), \tag{1.4b}$$

$$u_0(\eta) = 1, \tag{1.4c}$$

$$F_l = (2l+1)! C_l(\eta) \rho^{l+1} \Phi_l, \tag{1.5a}$$

$$G_l = (2\eta)^{2l+1} \rho^{l+1} \Theta_l / [(2l+l)! C_l(\eta)]. \tag{1.5b}$$

Rewriting Eq. (1.1) in terms of the variable

$$x = (8\eta\rho)^{1/2} = (8\alpha r)^{1/2}, \tag{1.6}$$

i.e., as

$$\left[\frac{d^2}{dx^2} + \frac{d}{x\,dx} - 1 - \frac{2l+1)^2}{x^2} + \frac{x^2}{16\eta^2}\right]\left(\frac{R}{x}\right) = 0, \tag{1.7}$$

the latter equation is then satisfied by the functions

$$(x/2)^{2l+1}\Phi_l = 2\eta C_0(\eta)^{-1} u_l(\eta)^{-1/2} (2/x) F_l, \tag{1.8a}$$

$$(x/2)^{2l+1}\Theta_l = C_0(\eta) u_l(\eta)^{1/2} (2/x) G_l. \tag{1.8b}$$

For $|\eta| = \infty$, Eq. (1.7) is satisfied by the linear combination of modified Bessel functions

$$aI_{2l+1}(x) + bK_{2l+1}(x), \tag{1.9}$$

and Yost, Wheeler, and Breit[2,3] have proved that

$$\lim_{\eta=\infty} \Phi_l = (x/2)^{-2l-1} I_{2l+1}(x), \tag{1.10a}$$

$$\lim_{\eta=\infty} \Theta_l = -2(-x/2)^{-2l-1} K_{2l+1}(x). \tag{1.10b}$$

Since the modified Bessel functions $I_n(x)$ and $(-)^n K_n(x)$ satisfy the same differential equation and the same recurrence relations, Eqs. (1.10) suggest the existence of an expansion of F_l and G_l as single series of functions $I_{2l+1+n}(x)$ and $K_{2l+1+n}(x)$, respectively, with $n = 0,1,2,\dots$. Such expansions would be particularly useful for large η, i.e., for large Z_1Z_2 and/or low energies, since

$$\eta^{-2} \propto E = \hbar^2 k^2/(2M). \tag{1.11}$$

Expansions of this type have been given by Abramowitz,[6] the one for G_l being asymptotic only. Since Abramowitz's derivation of these expansions is not entirely correct, we rederive the expansion of F_l in Sec. II and simultaneously obtain the explicit form of the polynomials in η^{-2} entering into the expansion. We obtain these results as a straightforward application of general properties we have recently established for the Kuhn–Ham expansion[1] of Φ_l. The same method is then applied in Sec. III to obtain the corresponding (exact) expansion of G_l as a simple series in $K_{2l+1+n}(x)$, but still as a double one in $I_{2l+1+i}(x)$ with $0 \leqslant i \leqslant n$. For large η and appropriate values of $\arg \eta$, the asymptotic approximation of G_l reduces to a single series in $K_{2l+1+n}(x)$ as it is proved in Sec. IV. This latter result is compared to that of Abramowitz,[6] which also appears in the *Handbook of Mathematical Functions*.[7] An appendix gives, up to $n = 9$, the polynomials in η^{-2}, $b_n^{(l)}(\eta)$, introduced in Sec. II.

II. EXPANSION OF THE REGULAR FUNCTION

With the notation of Sec. I, the Kuhn–Ham expansion of Φ_l reads[1]

$$\Phi_l = \sum_{\mu=0}^{\infty} (-)^\mu \eta^{-2\mu} \sum_{\lambda=0}^{\mu} \beta_{\mu\lambda}^{(l)} \left(\frac{x}{2}\right)^{-2l-1+2\mu+\lambda} \times I_{2l+1+2\mu+\lambda}(x), \tag{2.1}$$

where the $\beta_{\mu\lambda}^{(l)}$ are polynomials of degree $\mu - \lambda$ in l. They satisfy a recurrence relation also given in Ref. 1. The expansion (2.1) is a power series of E, since $\eta^{-2} \propto E$. It is absolutely and uniformly convergent, both in its dependence on η^{-2} and x considered as complex variables. This property has been proved in the Appendix of Ref. 1 by obtaining a double

 0022-2488/85/040656-04$02.50

series of convergent upper bounds for the modulus of each $(\mu\lambda)$ term under the conditions

$$|\eta^{-2}|\leqslant M<\infty, \quad |x^2/4|\leqslant X<\infty, \tag{2.2}$$

the only conditions where M and X are arbitrarily large but bounded.

Consequently, we can rewrite the expansion of Φ_l as a sum over $n=2\mu+\lambda$ and λ without modifying its convergence properties. It then reads

$$\Phi_l=\sum_{n=0}^{\infty}\left(\frac{x}{2}\right)^{-2l-1+n} I_{2l+1+n}(x) b_n^{(l)}(\eta), \tag{2.3}$$

with

$$b_n^{(l)}(\eta)=\sum_{\lambda}(-\eta^2)^{(n-\lambda)/2}\beta^{(l)}_{(n-\lambda)/2,\lambda}. \tag{2.4}$$

The latter sum extends to all integral values of λ satisfying the conditions

$$(n-\lambda)/2\geqslant\lambda\geqslant 0, \tag{2.5}$$

and having the same parity as n, so that $(n-\lambda)/2$ is an integer (by definition[1] $\beta^{(l)}_{\mu\lambda}=0$ when $\mu<0$ or $\lambda<0$ or $\lambda>\mu$).

Accordingly, we can also write

$$b_n^{(l)}(\eta)=\sum_{m=n'}^{n''}(-\eta^{-2})^m\beta^{(l)}_{m,n-2m}, \tag{2.6}$$

where[8] $n'=[(n+2)/3]$ and $n''=[n/2]$. We have, in particular,

$$b_0^{(l)}=1, \quad b_1^{(l)}=0, \tag{2.7}$$

since $\beta^{(l)}_{00}=1$, while the inequality (2.5) cannot be satisfied for $n=1$. Introducing the expansion (2.3) into Eq. (1.7), one obtains the recurrence relation

$$b_n^{(l)}(\eta)=-\frac{2l+n}{4\eta^2 n}b^{(l)}_{n-2}(\eta)-\frac{1}{4\eta^2 n}b^{(l)}_{n-3}(\eta), \tag{2.8}$$

for $n\geqslant 1$ and $b_m^{(l)}(\eta)=0$ when $m<0$.

Abramowitz's[6] proof of the absolute and uniform convergence of the expansion (2.3), given explicitly for $l=0$ only, is marred by the fact that he has erroneously deduced from the relation (2.8) that $|b_n^{(0)}|\sim|2\eta|^{-n}$. From Eq. (2.6), one rather has

$$b_n^{(l)}(\eta)=O(\eta^{-2n'}), \quad \text{with } n'=[(n+2)/3]. \tag{2.9}$$

From Eqs. (1.8a) and (2.3), the expansion of F_l as a single series of modified Bessel functions reads

$$F_l=(1/2\eta)C_0(\eta)u_l(\eta)^{1/2}\times\sum_{n=0}^{\infty}\left(\frac{x}{2}\right)^{n+1}I_{2l+n+1}(x)b_n^{(l)}(\eta). \tag{2.10a}$$

It directly applies to a repulsive field, i.e., when x is real and positive. For an attractive field, with the notation[1]

$$\zeta=e^{i\pi/2}x \quad (\zeta>0), \quad \hat\eta=e^{i\pi}\eta=|\alpha|/k, \tag{2.11}$$

we have $C_0(\eta)=C_0(\hat\eta)\exp(\pi\hat\eta)$, and hence

$$F_l=(-)^{l+1}(1/2\hat\eta)C_0(\hat\eta)e^{\pi\hat\eta}u_l(\hat\eta)^{1/2}\times\sum_{n=0}^{\infty}\left(\frac{-\zeta}{2}\right)^{n+1}J_{2l+n+1}(\zeta)b_n^{(l)}(\hat\eta), \tag{2.10b}$$

i.e., a series of Bessel functions of the first kind.

III. EXPANSION OF THE IRREGULAR FUNCTION

We have proved in Ref. 1 that the expansion of Θ_l, corresponding to that of Φ_l as given by Eq. (2.1), is

$$\begin{aligned}\Theta_l=u_l(\eta)\Bigg\{&h(\eta)\Phi_l-\sum_{\mu=0}^{\infty}(-)^\mu\eta^{-2\mu}\\&\times\Bigg[2\sum_{\lambda=0}^{\mu}\beta^{(l)}_{\mu\lambda}\left(\frac{-x}{2}\right)^{-2l-1+2\mu+\lambda}K_{2l+1+2\mu+\lambda}(x)\\&+\sum_{s=0}^{\mu}(-)^s b_s\sum_{\lambda=0}^{\mu-s}\beta^{(l)}_{\mu-s,\lambda}\left(\frac{x}{2}\right)^{-2l-1+2(\mu-s)+\lambda}\\&\times I_{2l+1+2(\mu-s)+\lambda}(x)\Bigg]\Bigg\},\end{aligned} \tag{3.1}$$

where $h(\eta)$ is defined in terms of psi functions $[\psi(z)=\Gamma'(z)/\Gamma(z)]$ by

$$h(\eta)=\tfrac{1}{2}[\psi(1-i\eta)+\psi(1+i\eta)]-\ln\eta, \tag{3.2}$$

while the b_s are defined in terms of the Bernoulli numbers[9] B_{2s} by

$$b_0=0, \quad b_s=|B_{2s}|/(2s) \quad (s\geqslant 1). \tag{3.3}$$

The function $h(\eta)$ has no convergent expansion in powers of η^{-2}, but only an asymptotic expansion for $|\eta|\to\infty$.

The series in Eq. (3.1) is absolutely and uniformly convergent under the conditions[10]

$$\begin{gathered}|\eta^{-2}|\leqslant M<\infty, \quad |\arg x|\leqslant\pi-\epsilon<\pi,\\ 0<\xi\leqslant|x^2/4|\leqslant X<\infty.\end{gathered} \tag{3.4}$$

Rearranging it by introducing $n=2\mu+\lambda$ as we have done for Φ_l, we readily obtain the desired expansions, namely

$$\begin{aligned}G_l=C_0(\eta)^{-1}u_l(\eta)^{1/2}\Bigg\{&h(\eta)\sum_{n=0}^{\infty}\left(\frac{x}{2}\right)^{n+1}I_{2l+n+1}(x)b_n^{(l)}(\eta)\\&-\sum_{n=0}^{\infty}\Bigg[2\left(\frac{-x}{2}\right)^{n+1}K_{2l+n+1}(x)b_n^{(l)}(\eta)\\&+\sum_{s=0}^{[n/2]}b_s\eta^{-2s}\left(\frac{x}{2}\right)^{n+1-2s}\\&\times I_{2l+n+1-2s}(x)\,b^{(l)}_{n-2s}(\eta)\Bigg]\Bigg\},\end{aligned} \tag{3.5a}$$

for a repulsive field $(\alpha>0)$, and

$$\begin{aligned}G_l=(-)^l C_0(\hat\eta)^{-1}e^{-\pi\hat\eta}u_l(\hat\eta)^{1/2}&\\\times\Bigg\{h(\hat\eta)&\sum_{n=0}^{\infty}\left(\frac{-\zeta}{2}\right)^{n+1}J_{2l+n+1}(\zeta)b_n^{(l)}(\hat\eta)\\&+\sum_{n=0}^{\infty}\Bigg[\pi\left(\frac{-\zeta}{2}\right)^{n+1}Y_{2l+n+1}(\zeta)b_n^{(l)}(\hat\eta)\\&-\sum_{s=0}^{[n/2]}b_s\eta^{-2s}\left(\frac{-\zeta}{2}\right)^{n+1-2s}\\&\times J_{2l+n+1-2s}(\zeta)\,b^{(l)}_{n-2s}(\hat\eta)\Bigg]\Bigg\},\end{aligned} \tag{3.5b}$$

for an attractive field $(\alpha<0)$.

IV. ASYMPTOTIC EXPANSION OF THE IRREGULAR FUNCTION FOR LARGE η

For real and positive η, Breit and Hull[11] have proved that one obtains an asymptotic expansion of G_l for $\eta\to\infty$ when an asymptotic expansion is substituted for $h(\eta)$. This remains valid for complex η, provided one takes into account the domain of validity of the expansion used for $h(\eta)$. From elementary properties of the ψ function,[12] $h(\eta)$ is easily given the forms

$$h(\eta)=h^{\pm}(\eta)\mp i\pi(e^{2\pi\eta}-1)^{-1}, \tag{4.1}$$

with[13]

$$h^{\pm}(\eta)=\psi(\pm i\eta)\pm(1/2i\eta)-\ln(e^{\pm i\pi/2}\eta) \tag{4.2}$$

$$\sim\sum_{s=1}^{\infty} b_s\eta^{-2s}. \tag{4.3a}$$

The latter asymptotic expansion holds when $|\eta|\to\infty$ for both h^+ and h^-, but, respectively, under the conditions

$$|\arg\eta\pm\pi/2|\leqslant\pi-\epsilon<\pi. \tag{4.3b}$$

Since $h(\eta)=[h^+(\eta)+h^-(\eta)]/2$, we also have

$$h(\eta)\sim\sum_{s=1}^{\infty} b_s\eta^{-2s}, \tag{4.4a}$$

for $|\eta|\to\infty$, but only when

$$|\arg\eta|\leqslant\pi/2-\epsilon<\pi/2. \tag{4.4b}$$

Rearranging the product $h(\eta)\Phi_l$ such as it is given by Eqs. (4.1), (4.3), and (2.3), we obtain, under condition (4.3b),

$$(x/2)^{2l+2}h(\eta)\Phi_l\sim\mp i\pi(e^{2\pi\eta}-1)^{-1}(x/2)^{2l+2}\Phi_l+\sum_{n=0}^{\infty}\sum_{s=0}^{[n/2]} b_s\eta^{-2s}\left(\frac{x}{2}\right)^{n+1-2s}\times I_{2l+n+1-2s}(x)b_{n-2s}^{(l)}(\eta). \tag{4.5}$$

Substituting this result for the first term in Eqs. (3.5), we then have, for a repulsive field ($\alpha>0$),

$$G_l\sim-C_0(\eta)^{-1}u_l(\eta)^{1/2}\Big[\pm i\pi(e^{2\pi\eta}-1)^{-1}\times\sum_{n=0}^{\infty}\left(\frac{x}{2}\right)^{n+1}I_{2l+n+1}(x)b_n^{(l)}(\eta)+2\sum_{n=0}^{\infty}\left(\frac{-x}{2}\right)^{n+1}K_{2l+n+1}(x)b_n^{(l)}(\eta)\Big], \tag{4.6a}$$

for $|\eta|\to\infty$ and under the conditions

$$\arg k\in[-\pi\pm\pi/2+\epsilon,\quad+\pi\pm\pi/2-\epsilon]. \tag{4.6b}$$

Similarly, for an attractive field ($\alpha<0$), we have

$$G_l\sim(-)^lC_0(\hat\eta)^{-1}e^{-\pi\hat\eta}u_l(\hat\eta)^{1/2}\Big[\mp i\pi(e^{2\pi\hat\eta}-1)^{-1}\times\sum_{n=0}^{\infty}\left(\frac{-\zeta}{2}\right)^{n+1}J_{2l+n+1}(\zeta)b_n^{(l)}(\hat\eta)+\pi\sum_{n=0}^{\infty}\left(\frac{-\zeta}{2}\right)^{n+1}Y_{2l+n+1}(\zeta)b_n^{(l)}(\hat\eta)\Big], \tag{4.6c}$$

under the same conditions (4.6b).

The results just obtained take a simpler form in the right half of the complex k plane, since $(e^{2\pi\eta}-1)\sim0$ for $|\eta|\to\infty$ in the domain (4.4b). Accordingly, we have

$$G_l\sim-2C_0(\eta)^{-1}u_l(\eta)^{1/2}\sum_{n=0}^{\infty}\left(\frac{-x}{2}\right)^{n+1}\times K_{2l+n+1}(x)b_n^{(l)}(\eta), \tag{4.7a}$$

for a repulsive field and

$$G_l\sim(-)^l\pi C_0(\hat\eta)^{-1}e^{-\pi\hat\eta}u_l(\hat\eta)^{1/2}\sum_{n=0}^{\infty}\left(\frac{-\zeta}{2}\right)^{n+1}\times Y_{2l+n+1}(\zeta)b_n^{(l)}(\hat\eta), \tag{4.7b}$$

for an attractive one, when k is real and positive, and more generally when

$$|\arg k|\leqslant\pi/2-\epsilon<\pi/2. \tag{4.7c}$$

Equations (4.7) also result directly from introducing the expansion (4.4) into Eqs. (3.5). The expansions (4.7) and the second series in the expansions (4.6) are divergent for $|\eta|^2<\infty$, as are the asymptotic expansions (4.3) and (4.4).

However, the expansions (4.6) and (4.7) are not asymptotic expansions proper, since any term is not always of an order of magnitude larger than the one next after it. From Eq. (2.6), it is indeed obvious that $b_{3N-2}^{(l)}$, $b_{3N-1}^{(l)}$, and $b_{3N}^{(l)}$ are all $O(\eta^{-2N})$, so that the three corresponding terms in the divergent series of Eqs. (4.6) and (4.7) should actually form a single term in an asymptotic approximation of G_l. Accordingly, instead of Eq. (4.7) e.g., we must rather write

$$G_l\sim2u_l(\eta)^{1/2}C_0(\eta)^{-1}\Big\{\left(\frac{x}{2}\right)K_{2l+1}(x)-\sum_{N=0}^{\infty}\Big[\sum_{i=1}^{3}\left(\frac{-x}{2}\right)^{1+3N+i}K_{2l+1+3N+i}(x)\times b_{3N+i}^{(l)}(\eta)\Big]\Big\} \tag{4.8a}$$

and

$$G_l\sim(-)^l\pi u_l(\hat\eta)^{1/2}C_0(\hat\eta)^{-1}e^{-\pi\hat\eta}\times\Big\{-\left(\frac{\zeta}{2}\right)Y_{2l+1}(\zeta)+\sum_{N=0}^{\infty}\Big[\sum_{i=1}^{3}\left(\frac{-\zeta}{2}\right)^{1+3N+i}\times Y_{2l+1+3N+i}(\zeta)b_{3N+i}^{(l)}(\hat\eta)\Big]\Big\}, \tag{4.8b}$$

for $|\eta|=|\hat\eta|\to\infty$ and $|\arg k|\leqslant\pi/2-\epsilon<\pi/2$.

In the case of an attractive field and for real positive energies only ($\eta>0$), the expansions (4.7a) and (4.8a) must be equivalent to the one of Breit and Hull[11] and Ham,[14] namely,[15] in the notation of Secs. I and II,

$$G_l\sim-2u_l(\eta)^{1/2}C_0(\eta)^{-1}\sum_{\mu=0}^{\infty}(-)^{\mu}\eta^{-2\mu}\times\sum_{\lambda=0}^{\mu}\beta_{\mu\lambda}^{(l)}\left(\frac{-x}{2}\right)^{2\mu+\lambda+1}K_{2l+1+2\mu+\lambda}(x), \tag{4.9}$$

for $\eta\to\infty$. Indeed, the very fact that the function

$$G_l/[u_l(\eta)^{1/2}C_0(\eta)^{-1}] \tag{4.10}$$

has one asymptotic power expansion in Poincaré's sense entails the uniqueness of that expansion.[16-18] The expansions (4.7a) and (4.9) only differ by a rearrangement of terms, the same that modified Eq. (2.1) into Eq. (2.3) and which we also used in Sec. III in absolutely and uniformly convergent double series.

In this context, it is of practical importance that we now turn to the asymptotic expansion of G_l as given by Abramowitz,[4,6] also for $\eta>0$ and $\eta\to\infty$. With the notation introduced in Sec. I, it reads

$$G_l\sim\lambda_l(\eta)(2l)!u_l(\eta)^{-1/2}C_0(\eta)^{-1}\times\sum_{n=0}^{\infty}\left(\frac{-x}{2}\right)^{n+1}K_{2l+n+1}(x)b_n^{(l)}(\eta), \tag{4.11}$$

where the overall factor $\lambda_l(\eta)$ remains to be fixed. Comparing Eqs. (4.7a) and (4.11), the uniqueness of the expansion of the ratio (4.10) leaves no other choice than defining $\lambda_l(\eta)$ as being the polynomial

$$-2u_l(\eta)/(2l)!, \tag{4.12}$$

or a function asymptotically equal to it.

Abramowitz[6] proceeds differently. He fixes $\lambda_l(\eta)$ by imposing the *divergent* expansion (4.11) to satisfy a condition[19] verified by the exact (unexpanded) G_l, namely

$$\{[(2l)!]^{-1}u_l^{1/2}C_0(\eta)(x/2)^{2l}G_l\}_{x=0} = 1. \tag{4.13}$$

This necessarily introduces in Eq. (4.11) the inverse of a divergent expansion and he actually obtains $\lambda_l(\eta)$ as

$$\lambda_l(\eta) = -2\left[\sum_{n=0}^{\infty}(-)^n(2l+n)!b_n^{(l)}(\eta)\right]^{-1}. \tag{4.14}$$

But, if we introduce the asymptotic expansion (4.7a) into Eq. (4.13), we then obtain[20]

$$u_l(\eta)\sum_{n=0}^{\infty}\frac{(-)^n b_n^{(l)}(\eta)(2l+n)!}{(2l)!}\sim 1, \tag{4.15}$$

so that the two factors (4.12) and (4.14) differ only by a factor asymptotically equal to 1, i.e., equal to $1 + O(\eta^{-2N})$ for any $N \gg 1$.

The former factor (4.12) is obviously much easier to use and normally more accurate in practice, since there is no need to expand or to approximate it. It partly reduces with the factor $u_l^{-1/2}$ inEq. (4.11).

On the other hand, it is worth noticing that, for $l>0$, when only a few terms are retained in the denominator of Abramowitz's factor $\lambda_l(\eta)$, one introduces poles in the corresponding approximate G_l. Although, in principle, such approximations are valid only for $\eta \gg 1$, this can, in practice, reduce the range of η's values in which the corresponding G_l remains a good approximation of the exact one.[21] For $l=0$, $\lambda_l(\eta)$ simply reduces to $\lambda_0 = -2[1+O(\eta^{-2N})]$ with N arbitrarily large.

To conclude, we observe that the practical advantage of the expansions (4.6)–(4.8) over those directly deduced from Ham's expansion[3,14] of G_l, such as the expansion (4.9), is that they come out as single series (rather that as a double sum over μ and λ). Moreover, the polynomials $b_n^{(l)}(\eta)$ can easily be obtained algebraically, by means of the single-index recurrence relation (2.8), up to any desired value of n. They are given in the Appendix up to $n=9$ [i.e., $N=2$ in Eq. (4.8)]. In contrast, obtaining algebraically the $\beta_{\mu\lambda}^{(l)}$ for any l and μ and for $0<\lambda<\mu-1$ proved to be very tedious.[22]

ACKNOWLEDGMENTS

Most of this paper has been prepared at the W. K. Kellogg Radiation Laboratory of the California Institute of Technology. The author is very grateful to C. A. Barnes, W. A. Fowler, and S. E. Koonin for their warm hospitality during his stay at Caltech and their interest in this work.

APPENDIX: THE FIRST POLYNOMIALS $b_n^{(l)}(\eta)$

The polynomials $b_n^{(l)}(\eta)$ given hereafter for $n\leqslant 9$, i.e., $N\leqslant 2$ in Eqs. (4.8), illustrate the increase with n of their lowest (n') and highest (n'') powers in η^{-2}. In agreement with Eqs. (2.4)–(2.9), we have

$$b_0^{(l)} = 1, \quad b_1^{(l)} = 0, \quad b_2^{(l)} = -\frac{l+1}{4\eta^2},$$

$$b_3^{(l)} = -\frac{1}{12\eta^2},$$

$$b_4^{(l)} = \frac{(l+1)(l+2)}{32\eta^4}, \quad b_5^{(l)} = \frac{5l+8}{240\eta^4},$$

$$b_6^{(l)} = \frac{1}{288\eta^4} - \frac{(l+1)(l+2)(l+3)}{384\eta^6},$$

$$b_7^{(l)} = -\frac{35l^2+147l+142}{13440\eta^6},$$

$$b_8^{(l)} = -\frac{5l+11}{5760\eta^6} + \frac{(l+1)(l+2)(l+3)(l+4)}{6144\eta^8},$$

$$b_9^{(l)} = -\frac{1}{10368\eta^6} + \frac{35l^3+273l^2+664l+496}{161280\eta^8}.$$

Then $b_{10}^{(l)}$ and $b_{11}^{(l)}$ have terms in η^{-8} and η^{-10}, and $b_{12}^{(l)}$ has terms in η^{-8}, η^{-10}, and η^{-12}. For any n and $\eta\to\infty$, the main term of $b_n^{(l)}$ has the sign of $(-)^{n'}$.

[1] J. Humblet, Ann. Phys. (N.Y.) **155**, 461 (1984).

[2] F. L. Yost, J. A. Wheeler, and G. Breit, Phys. Rev. **49**, 1 (1936).

[3] M. H. Hull and G. Breit, "Coulomb Wave Functions," in *Encyclopedia of Physics*, edited by S. Flügge (Springer, Berlin, 1959), Vol. 41/1, Chap. 2.

[4] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1968), 7th ed.

[5] The notation adopted in Eqs. (1.5) is the same as in Ref. 1. It slightly differs from the one of Refs. 2–4 for reasons given in Ref. 1.

[6] M. Abramowitz, J. Math. Phys. **33**, 111 (1954).

[7] See Ref. 4, Eqs. (14.4.1)–(14.4.4).

[8] $[\alpha]$ is the largest integer $\leqslant\alpha$.

[9] See Ref. 4, Eq. (23.1.3) and Table (23.2).

[10] See Ref. 1, end of Sec. 5.4 and end of the Appendix.

[11] G. Breit and M. H. Hull, Phys. Rev. **80**, 392 (1950); **80**, 561 (1950).

[12] See Ref. 4, Eqs. (6.3.5) and (6.3.7).

[13] See Ref. 4, Eq. (6.3.18).

[14] F. S. Ham, Q. Appl. Math. **15**, 31 (1957).

[15] Equation (4.9) is also given as Eq. (5.25) in Ref. 1, but it has been misprinted. A factor 2 is missing in its right-hand side.

[16] F. W. J. Olver, *Asymptotics and Special Functions* (Academic, New York, 1974).

[17] In principle, one could also include the factor $u_l(\eta)^{1/2}$ in the expansion of G_l, but that requires the use of the power expansion of $u_l^{1/2}$, so that G_l is more easily computed directly and exactly from the definition (1.4b) of u_l. The expansion of $u_l(\eta)$ is given in Ref. 18.

[18] F. S. Ham, Cruft Laboratory Technical Report 204, Harvard University, 1955, Appendix II.

[19] The condition (4.13) is easily obtained from Ref. 4, Eq. (14.1.14) or Ref. 1, Eq. (5.34b).

[20] When the *exact* expansion (3.5a) is introduced in Eq. (4.13), one still obtains the same relation (4.14) which contains a divergent asymptotic expansion. This is related to the fact that the second series in Eq. (3.5a) is convergent only under the conditions (3.4).

[21] For example, $\lambda_1(1)=\infty$ when terms of the order of η^{-4} are neglected in the denominator of $\lambda_1(\eta)$.

[22] See Eqs. (3.5) in Ref. 1; $\beta_{\mu\lambda}^{(l)}$ (here) $= (-)^{\mu+\lambda}a_{\mu,2\mu+\lambda}^{(l)}$ in Ref. 3.

Exponential convergence for nonlinear diffusion problems with positive lateral boundary conditions

Charles J. Holland
Code 411–Math Program, Office of Naval Research, Arlington, Virginia 22217
James G. Berryman
Lawrence Livermore National Laboratory, P. O. Box 808, L–201, Livermore, California 94550

(Received 25 May 1984; accepted for publication 31 August 1984)

It is established that the solution u of $u_t = \Delta(u^m) > 0$, with positive initial data, positive lateral boundary data, and positive exponent m, converges exponentially to the solution v of the corresponding stationary equation $\Delta(v^m) = 0$. The analysis also provides the form of the leading contribution to the difference $(u - v)$.

I. INTRODUCTION

Let B be a bounded domain in $R^n, n < 6$, and consider the solution $u = u(x,t)$ to the first boundary value problem for the nonlinear (concentration-dependent) diffusion equation

$$u_t = \Delta(u^m) \quad \text{in } B \times (0,\infty), \tag{1}$$

for $m > 0$ with positive initial data

$$u(x,0) = F(x) \ \text{ in } B, \tag{2}$$

and with fixed positive lateral boundary data

$$u(x,t) = G(x) > 0 \quad \text{on } dB \times (0,\infty). \tag{3}$$

The symbol Δ denotes the Laplacian in the variables x and dB denotes the boundary of the domain B. The objective of this paper is to establish the rate and the form of convergence of solutions $u(x,t)$ of (1)–(3) to solutions $v = v(x)$ of the stationary equation

$$\Delta(v^m) = 0, \tag{4}$$

with

$$v = G \quad \text{on } dB. \tag{5}$$

For $m = 1$, problem (1)–(3) represents the first boundary value problem for the linear heat equation. In this case, it is well known that under appropriate smoothness assumptions on the data and the domain, there exists a positive function H in B vanishing on the boundary, a positive constant k, and a constant c (which may be zero) such that

$$\exp kt\,[u(x,t) - v(x)] \to cH(x) \tag{6}$$

in $H_0^1(B)$ as $t \to \infty$. Here $(\Delta + k)H = 0$ in B, $H = 0$ on dB, and H is the eigenfunction corresponding to the principal eigenvalue k. Equation (6) thus establishes the exponential convergence of the solution u to v, where the limiting form is determined by $cH(x)\exp(-kt)$. In this paper we establish an exponential convergence result [corresponding to (6)] in the nonlinear diffusion case with $m > 0$. For the linear problem (1) with $m = 1$, higher-order expansions can also be established. The question of higher-order expansions for the nonlinear problem is not considered here.

For the linear case, the convergence result (6) holds even if $G = 0$ on dB. It is well known for the nonlinear case,[1,2] however, that if the positive lateral data is replaced by zero lateral boundary data, then one does not obtain an exponential rate of convergence to a limiting solution.

We now briefly discuss the physical origin of Eqs. (1)–(3). Notice that Eq. (1) can be written in the form

$$u_t = \nabla\circ(D(u)\nabla u), \tag{7}$$

where $D(u) = mu^{m-1}$ is called the diffusion coefficient. If $m > 1$, then $D(u) \to 0$ as $u \to 0$ and this case is called the slow diffusion problem. If $0 < m < 1$, then $D(u) \to \infty 0$ as $u \to 0$ and this case is called the fast diffusion problem. The slow diffusion problem arises in many contexts. Problem (1)–(3) is used to model the motion of a polytropic gas in a porous medium; then u represents the density of the gas. This problem occurs in the diffusion of biological populations whose rate of diffusion is population-dependent. This problem is also used to model radiation heat conduction for ionized gases; then u is interpreted as a temperature. It should be emphasized, as mentioned in Ref. 3, that the slow diffusion process may not be slow compared to other physical processes in the environment. For example, at temperatures greater than 10 000 K, where radiation heat conduction is important, the nonlinear heat conduction mechanism can transfer energy at a speed much faster than the speed of sound in the medium. Slow diffusion thus may be very fast, but the speed of propagation is still finite and therefore much slower than for *linear* diffusion in similar circumstances. The fast diffusion problem arises in plasma physics theory; then u represents plasma density. Okuda and Dawson[4] discussed a mechanism for explaining the experimental observation that crossfield diffusion of a plasma is faster than predicted by classical collision theory when the plasma is held in a strong magnetic field. Their modeling led to a one-dimensional problem with $m = 1/2$. See Refs. 2,4, and 5 for discussions.

Equation (1) and various modifications of it have also been the subject of much interest under boundary conditions other than (2) and (3). In Sec. V, some of this recent work is summarized.

II. PRELIMINARIES AND STATEMENT OF RESULTS

We now motivate the convergence result, which is stated below in Theorem 1. Throughout, we assume that B is a bounded domain with smooth boundary of class C^3. Write $w = u - v$ and assume that formally $w = z +$ higher-order terms. Then formal linearization shows that z satisfies the equation

 0022-2488/85/040660-04$02.50

$$\Delta\,(mv^{m-1}z) = z_t, \tag{8}$$

with zero lateral boundary conditions and appropriate initial conditions. The linear equation (8) has a slowest decaying "mode" corresponding to the separable solution $cS(x)\exp(-kt)$, where $S>0$ in B, and $S=0$ on dB is an $L^2(B)$ normalized eigenfunction corresponding to the principal eigenvalue for the non-self-adjoint eigenvalue problem

$$\Delta\,(mv^{m-1}S^*) = -K^*S^*,\quad S^*=0 \quad \text{on } dB. \tag{9}$$

In Theorem 1 we establish the result that the formal linearization process described above determines the precise rate and form of decay of $u(x,t)$ to the equilibrium solution $v(x)$.

Theorem 1: There exists a constant c, which depends upon the initial data and may be zero, such that

$$\exp kt\,[u(\cdot,t) - v(\cdot)] \rightarrow cS(\cdot) \tag{10}$$

in $H_0^1(B)$ as $t\rightarrow\infty$. Here S,k are determined by (9).

Theorem 1 is established by first deriving an analogous result for the difference $u^m - v^m$. Define $\bar{u}=u^m$, $\bar{v}=v^m$, $\bar{F}=F^m$, $\bar{G}=G^m$. Using these definitions, we see that $\bar{u}$ satisfies the equation

$$\Delta\bar{u} = (q+1)\bar{u}^q\,\bar{u}_t \quad \text{on } B\times(0,\infty), \tag{11}$$

with $q=(1-m)/m$ and $\bar{v}$ satisfies $\Delta\,(\bar{v})=0$. Let $\bar{w}=\bar{u}-\bar{v}$ denote the difference. Then $\bar{w}$ satisfies the equation

$$\Delta\bar{w} = (q+1)(\bar{v}+\bar{w})^q\bar{w}_t, \tag{12}$$

with $\bar{w}=0$ on the lateral boundary. Again using formal linearization and assuming that $\bar{w}=\bar{z}+$ higher-order terms, one obtains that $\bar{z}$ satisfies

$$\Delta\,(\bar{z}) = (q+1)\bar{v}^q\bar{z}_t \quad \text{in } B\times(0,\infty). \tag{13}$$

This equation has a slowest-decaying mode corresponding to the separable solution $c\bar{S}(x)\exp(-\bar{k}t)$, where $\bar{S}>0$ in B, $\bar{S}=0$ on dB is an L^2 normalized eigenfunction corresponding to the principal eigenvalue $\bar{k}$ for the eigenvalue problem

$$\Delta\,(\bar{\bar{S}}) = -\bar{\bar{k}}\,(q+1)\bar{v}^q\bar{\bar{S}} \quad \text{in } B, \tag{14}$$

with $\bar{\bar{S}}=0$ on dB. In Theorem 2 we establish that formal linearization yields the correct answer for the rate of convergence of $\bar{u}$ to $\bar{v}$.

Theorem 2: There exists a positive constant c, which depends upon $\bar{u}(x,0)$ and may be zero, such that

$$[\bar{u}(\cdot,t) - \bar{v}(\cdot)]\exp\bar{k}t \rightarrow c\bar{S}(\cdot) \tag{15}$$

in H_0^1 as $t\rightarrow\infty$. Here $\bar{S}$, $\bar{k}$ are as described above.

Remark: It should be noticed immediately that $k=\bar{k}$ and that $\bar{S}$ and $v^{m-1}S$ differ only by a multiplicative constant.

III. PROOFS

We need the following lemma for the proof of Theorem 2. Throughout the remainder of the paper integration with respect to x will be over the set B.

Lemma: Define $M(t) = \int\{|\nabla\bar{w}(t,x)|^2\}dx$. Then there exists a positive constant K such that for all $t>0$,

$$M(t) \leqslant K\exp(-2kt). \tag{16}$$

Proof: Define $f(x,\bar{w}) = (q+1)[\bar{v}(x)+\bar{w}]^q$. Using Eq. (12) for $\bar{w}$, one obtains that

$$\dot{M}(t) = -2\int\frac{\Delta\bar{w}(t,x)^2\,dx}{f(x,\bar{w}(t,x))}. \tag{17}$$

From (17), we conclude that $M(t)$ is decreasing. Next

$$M(t)^2 \leqslant \left(\int\frac{(\Delta\bar{w})^2}{f\,dx}\right)\left(\int\bar{w}^2f\,dx\right),$$

which can be rewritten, using the identity (17) for $\dot{M}(t)$, in the form

$$\frac{-M(t)}{\dot{M}(t)} \leqslant \frac{\int\bar{w}^2f\,dx}{2M(t)}. \tag{18}$$

Next we need to estimate the right-hand side of (18). There exists $a(x,t)$ in the interval $[\min(0,w(x,t)), \max(0,w(x,t))]$ such that

$$f(x,\bar{w}(x,t)) = f(x,0) + f_{\bar{w}}(x,a(x,t))\bar{w},$$

where the subscript $\bar{w}$ indicates partial differentiation with respect to the second ($\bar{w}$) argument.

Using appropriate maximum principles, it is easy to see, using (12), that both $\min(-\bar{w}(x,t))$ and $\max w(\bar{x},t)$ are decreasing functions of t. Hence (18) can be rewritten as

$$\frac{-M(t)}{\dot{M}(t)} < \left(\int\frac{\bar{w}^2f(x,0)dx}{2M(t)}\right) + \left(k'\int\frac{|\bar{w}|^3\,dx}{M(t)}\right), \tag{19}$$

for some constant k' independent of t. Since $\bar{w}(x,t)=0$ on the boundary of B and by using the Rayleigh–Ritz characterization of the minimum eigenvalue, one finds that the first term on the right side of the inequality of (19) is bounded above by $1/2k$. Sobolev estimates show that the second term is bounded above by $k''[M(t)]^{1/2}$ for some constant k'' since x is a member of R^n with $n<6$. Hence

$$(-M(t)/\dot{M}(t)) \leqslant (1/2k) + k''M(t)^{1/2}.$$

Since $\dot{M}<0$, then

$$(-2k) \geqslant (\dot{M}/M) + 2kk''M^{-1/2}\dot{M}.$$

Integrating from 0 to t, one obtains that

$$(-2kt) \geqslant \ln(M(t)/M(0)) + 4k\,k''(M^{1/2}(t) - M^{1/2}(0)).$$

Because $M(t)$ is bounded above for all $t>0$, the lemma follows from the previous inequality.

Proof of Theorem 2: Define $p=(\bar{w}\exp kt)$ and recall that $f(x,\bar{w}) = (q+1)(\bar{v}(x)+\bar{w})q$. Then $p(x,t)$ satisfies the equation

$$\Delta p + kp\,f(x,p\exp(-kt)) = f(x,p\exp(-kt))p_t. \tag{20}$$

Recall that the lemma has established bounds on the L^3 and H^1 norms of $p(\cdot,t)$ independent of t.

Now for functions $h(\cdot)$ in $H_0^1(B)$ define the time-dependent "energy functional"

$$E_t(h) = \int\left\{\frac{1}{2}|\nabla h|^2 - \int_0^{h(x)}kz \times f(x,z\exp(-kt))dz\right\}dx, \tag{21}$$

and $g(t) = E_t(p(\cdot,t))$. Since

$$g(t) \geqslant \frac{1}{2}\int\{|\nabla p(x,t)|^2 - k\max\{f(x,0), f(x,\bar{w}(x,t))\}p^2(x,t)\}dx,$$

the lemma guarantees that $g(\cdot)$ is bounded below for $t>0$.

Next we have that

$$\dot{g}(t) = \int [p_x p_{xt} + kpf(x,p\exp(-kt))p_t]\,dx$$

$$+ \int \int_0^p k^2 h \exp(-kt) f_w(x,h\exp(-kt))\,dh\,dx,$$

where $p = p(x,t)$. Integrating the first term by parts and using Eq. (20), one has that

$$\dot{g}(t) = \int \Big\{ -f(x,p\exp(-kt))p_t^2 + \int_0^p kh^2 \exp(-kt) f_w(x,h\exp(-kt))\,dh \Big\}\,dx. \quad (22)$$

Observing that $f_w(x,h\exp(-kt))$ is uniformly bounded for h between 0 and $p(x,t)$ and using Lemma 1 to bound the $L^3(B)$ norm of p, one can show that there exists a constant K such that the second term on the right side of (21) is bounded above by $K\exp(-kt)$. This estimate shows that the second term is absolutely integrable on $(0,\infty)$. Since $g(t)$ is bounded below and the first term on the right side of (21) is nonpositive, there exists an increasing sequence of times $t_n \to \infty$ such that $\dot{g}(t_n) \to 0$ and

$$\int f(x,p\exp(-kt))p_t^2\,dx \to 0 \quad \text{as } n\to\infty. \quad (23)$$

Moreover, $g(t)$ has a limit as $t\to\infty$.

Using (20), one obtains that $\int |\Delta p(x,t_n)|^2\,dx$ is uniformly bounded for the sequence. Hence there exists a subsequence of times, again denoted by t_n, and a function $R(\cdot)$ in $H_0^1(B)$ such that $p(\cdot,t_n) \to R(\cdot)$ as $t_n \to \infty$. We must show that R is a solution to (14). To accomplish this, multiply (20) by any function J in $H_0^1(B)$, integrate one term by parts, and take the limit as $t\to\infty$. One obtains, using the integral bounds on p and (22), that

$$\int (-\nabla J \nabla R + kJ(q+1)v^q R)\,dx = 0. \quad (24)$$

But this implies that $R(\cdot)$ is a weak solution to the linear equation (14) with zero boundary conditions. Since (14) is linear, then R is, in fact, a classical solution and therefore $R = c\bar{S}$ for some constant c. Note also that $E_t[R(\cdot)]$ has limit 0 as $t\to\infty$.

We need to establish that $p(\cdot,t) \to R(\cdot)$ for all t and not just for the above subsequence. Since $p(\cdot,t_n) \to R(\cdot)$, one can use estimates on $p(\cdot,t_n)$ to conclude from (21) that $g(t_n)$ has limit 0 as $n\to\infty$. Consequently, $g(t)$ has limit 0 as $t\to\infty$.

Suppose for the purpose of contradiction that $p(\cdot,t)$ does not converge to $R(\cdot)$ as $t\to\infty$. Then there exists a sequence of times t_n such that $p(\cdot,t_n)$ does not converge to $R(\cdot)$. Since $g(t)$ is bounded, the sequence $p(\cdot,t_n)$ is weakly bounded in $H_0^1(B)$. Hence there exists a subsequence again denoted by t_n, and a function $Z(\cdot)$ not equal to $R(\cdot)$, in $H_0^1(B)$ such that $p(\cdot,t_n) \to Z(\cdot)$ strongly in $L^3(B)$, and weakly in $H_0^1(B)$.

We next show that $Z(\cdot)$ is also a solution to (14). Since $\int |\nabla p|^2\,dx$ is lower semicontinuous with respect to weak convergence, one obtains that

$$\lim_{n\to\infty} g(t_n) = \frac{1}{2}\int [|\nabla Z|^2 + k(q+1)\bar{v}(x)^q Z^2]\,dx \leqslant 0.$$

However, by the Rayleigh–Ritz minimizing property of k, equality must hold in the above equation. Thus $Z(\cdot)$ also minimizes the Rayleigh–Ritz quotient and is also a solution to (14). Hence $Z = c''\bar{S}$ for some constant c''.

Thus Z, R are distinct solutions of (14). We need to show that this is not possible. We accomplish this by obtaining an estimate on the time derivative of $\int |\nabla p|^2\,dx$.

Differentiating, integrating by parts, and using (20), one obtains

$$\frac{1}{2}\frac{d}{dt}\int |\nabla p|^2\,dx = -\int (\Delta p)^2 f\,dx + \int k|\nabla p|^2\,dx. \quad (25)$$

Next we develop an estimate for the first term on the right side of (25). Integrating by parts and using the Cauchy–Schwarz inequality, one obtains

$$\Big(\int |\nabla p|^2\,dx\Big)^2 \leqslant \Big(\int \frac{(\Delta p)^2\,dx}{f}\Big)\Big(\int p^2 f\,dx\Big).$$

Since $\int |\nabla p|^2\,dx \geqslant k\int p^2 f(x,0)\,dx$, then

$$\frac{k\int |\nabla p|^2\,dx \int p^2 f(x,0)\,dx}{\int p^2 f(x,p\exp(-kt))\,dx} \leqslant \int \frac{|\Delta p|^2\,dx}{f(x,p\exp(-kt))}. \quad (26)$$

We then write $\int p^2 f(x,p\exp(-kt))\,dx = A(t) + B(t)\exp(-kt)$, where $A(t) = \int p^2 f(x,0)\,dx$ and $B(t)$ is a bounded term on $(0,\infty)$. Recall that the lemma bounds the L^3 norm of p. Substituting this expression into (26) and then using (26) in (25) one obtains

$$\frac{1}{2}\frac{d}{dt}\int |\nabla p|^2\,dx \leqslant k\,\frac{\int |\nabla p|^2\,dx\,(B(t)\exp(-kt))}{A(t) + B(t)\exp(-kt)}. \quad (27)$$

Recall that $\int |\nabla p|^2\,dx$ is also bounded for $t \geqslant 0$ by the lemma. We shall use the estimate (27) below.

It is easy to show that the set of limit points of $p(\cdot,t)$, denoted by Q, is a connected set in $H_0^1(B)$ of equilibrium solutions of (14). Let $a = \min(\int |\nabla h|^2\,dx : h \text{ in } Q)$, and let b denote the corresponding maximum. Define $4d = b - a$.

Since the set Q is connected, there exists a sequence of times $t_n \to \infty$ such that

$$\lim \int |\nabla p(x,t_n)|^2\,dx = a + 2d, \quad \lim A(t_n) = a + 2d.$$

For each t_n, let $s = s(t_n)$ be the first time after t_n such that either

$$(1)\int |\nabla p(x,s)|^2\,dx = a + d,$$

$$(2)\int |\nabla p(x,s)|^2\,dx = a + 3d,$$

or

$$(3)\ A(s) = a + d.$$

The differential inequality (27) shows that, for sufficiently large t_n, $s(t_n)$ occurs because either case (1) or case (3) occurs. Hence there cannot exist a sequence $\{t_n\}$ such that $p(\cdot,t_n) \to c^*S$ in $H_0^1(B)$ for constant c^* with $\int |\nabla c^*S(x)|^2\,dx > a + 3d$. This is a contradiction and Theorem 2 is proved.

Theorem 1 follows in a straightforward manner from Theorem 2. The details are omitted.

TABLE I. Zeroes (r_n) of $J_{(1/3)}$ and eigenvalues $k_n = 9/8\, r_n^2$ for the first five eigenfunctions of Eq. (32). The values of $n^2\pi^2$ are listed for comparison.

n	r_n	k_n	$n^2\pi^2$
1	2.902 59	9.478 13	9.869 60
2	6.032 75	40.943 29	39.478 42
3	9.170 51	94.610 47	88.826 44
4	12.310 19	170.483 48	157.913 67
5	15.450 65	268.562 87	246.740 11

IV. EXAMPLE

As a concrete example of the results presented in Sec. II, consider the one-dimensional problem

$$u_{xx} = 2uu_t \quad \text{on}(0,1)\times(0,\infty), \tag{28}$$

with

$$u(0,t) = 0, \quad u(1,t) = 1, \tag{29}$$

and

$$u(x,0) = F(x) > 0 \quad \text{on}(0,1). \tag{30}$$

Since $u(0,t) = 0$, Theorem 2 does not directly apply to this equation. However, since $u(1,t) = 1 > 0$, the solution to (28)–(30) does not vanish in finite time and it can be shown that the conclusion of Theorem 2 is also valid for this problem. Then

$$v(x) = x, \tag{31}$$

and $S(x)$ satisfies

$$S_{xx} + 2k\,xS = 0, \tag{32}$$

with $S(0) = S(1) = 0$.

Equation (32) can be transformed into Bessel's equation. Let

$$r = (2kx)^{1/3}/3k, \tag{33}$$

and

$$S(r) = (3kr)^{1/3}R(r). \tag{34}$$

Then substituting into (32), we find that R satisfies

$$r^2R_{rr} + rR_r + (r^2 - \tfrac{1}{9})R = 0. \tag{35}$$

Therefore,

$$R(r) = J_{1/3}(r), \tag{36}$$

where $J_a(r)$ is the Bessel function of the first kind of order a.

The various eigenvalues and eigenfunctions of (32) are found from (36) by locating the values r_n such that

$$J_{1/3}(r_n) = 0, \tag{37}$$

i.e., the zeroes of $J_{1/3}$. Writing $r = r_n x^{3/2}$ and using (33), we see that

$$k_n = \tfrac{9}{8} r_n^2. \tag{38}$$

The zeroes r_n can be found very accurately using a numerical scheme due to Temme.[6] The results for the first five zeroes and eigenvalues are listed in Table I, where the values of $n^2\pi^2$ are also listed for comparison.

Assuming that the conclusion of Theorem 2 is valid for this problem, we find that the solution of (28) satisfies

$$|u(x,t) - x|\exp k_1 t < \text{const}\, x^{1/2} J_{1/3}(r_1 x^{3/2}). \tag{39}$$

V. BRIEF REVIEW OF RELATED RESULTS

The asymptotic behavior of solutions to (1) in the nonlinear case has also been treated for the case when the positive lateral boundary data (3) is replaced with zero lateral boundary data. The slow diffusion case has been treated by Aronson and Peletier[1] and the fast diffusion case by Berryman and Holland.[2] Convergence of solutions of (1)–(3) to appropriate separable solutions $S(x)T(t)$ is established in these papers.

Problem (1)–(3) has also been treated in the slow case in Ref. 4, where u_t is replaced with $p(x)u_t$, with $p > 0$ in B and vanishing on dB. Under certain assumptions, convergence to a separable solution of (1)–(3) is obtained. In this case the equation mathematically models the thermal evolution of a heated plasma in which the density is stationary but inhomogeneous.

The one-dimensional Cauchy problem for (1) and (2) has also been treated in the slow diffusion case. Convergence of solutions of (1) and (2) to appropriately scaled self-similar solutions has been established under various conditions. See Refs. 3, 7, and 8.

The problem (1) has also been treated on the half-line ($x > 0$) when the concentration $u(0,t)$ is held at some constant value U for all $t > 0$. It has been established that the solution u converges to a similarity solution of (1). The fast diffusion case has been treated recently by Bertsch[5], while the slow diffusion case was treated much earlier by Peletier.[9]

Finally, the slow diffusion case for (1) and (2) with zero lateral boundary data has been treated in case Eq. (1) also contains an appropriate source of extinction terms $f(u)$. These problems have been studied in the one-dimensional case by Aronson, Crandall, and Peletier,[10] Rosenau,[11] and by Gurtin and MacCamy.[12]

ACKNOWLEDGMENTS

The work of J. G. B. was performed under the auspices of the U. S. Department of Energy by the Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48.

[1]D. Aronson and L. Peletier, J. Diff. Eqs. **39**, 378 (1981).
[2]J. Berryman and C. Holland, Arch. Ration. Mech. Anal. **74**, 379 (1980).
[3]J. Berryman, J. Math. Phys. **21**, 1326 (1980).
[4]H. Okuda and J. Dawson, Phys. Fluids **16**, 408 (1977).
[5]M. Bertsch, SIAM J. Appl. Math. **42**, 66 (1982).
[6]M. Temme, J. Comput. Phys. **32**, 270 (1979).
[7]S. Kamenomostskaya (S. Kamin), Israel J. Math. **14**, 76 (1973).
[8]S. Kamin, Arch. Ration. Mech. Anal. **60**, 171 (1976).
[9]L. Peletier, SIAM J. Appl. Math. **21**, 542 (1971).
[10]D. Aronson, M. Crandall, and L. Peletier, MRC Technical Summary Report 2220, May 1981.
[11]P. Rosenau, Physica D5, 136 (1982).
[12]M. Gurtin and R. MacCamy, Math. Biosci. **33**, 35 (1977).

A simple derivation of the addition theorems of the irregular solid harmonics, the Helmholtz harmonics, and the modified Helmholtz harmonics

E. Joachim Weniger and E. Otto Steinborn
Institut für Physikalische und Theoretische Chemie, Universität Regensburg, D-8400 Regensburg, Federal Republic of Germany

(Received 25 April 1984; accepted for publication 12 October 1984)

In this article a simple derivation of the addition theorems of the irregular solid harmonics, the Helmholtz harmonics, and the modified Helmholtz harmonics is presented. Our derivation is based upon properties of the differential operator $\mathscr{Y}_l^m(\nabla)$, which is obtained from the regular solid harmonic $\mathscr{Y}_l^m(\mathbf{r})$ by replacing the Cartesian components of $\mathbf{r}$ by the Cartesian components of ∇. With the help of this differential operator $\mathscr{Y}_l^m(\nabla)$, which is an irreducible spherical tensor of rank l, the addition theorems of the anisotropic functions are obtained by differentiating the addition theorems of the isotropic functions. The performance of the necessary differentiations is greatly facilitated by a systematic exploitation of the tensorial nature of the differential operator $\mathscr{Y}_l^m(\nabla)$.

I. INTRODUCTION

In molecular and solid state physics, systems with more than one electron and with more than one atomic nucleus are treated. Consequently, it frequently happens that the eigenfunctions or operators which occur there have arguments that are given as sums or differences of two vectors that represent the coordinates of electrons and nuclei. Since quantum mechanical computational procedures usually involve integrations, the dependence of eigenfunctions and operators on the sum or difference of two vectors may be very inconvenient and it is often imperative to obtain a separation of variables, which can be achieved with the help of addition theorems. The probably best-known example of such an addition theorem is the Laplace expansion of the Coulomb potential in spherical coordinates,

$$\frac{1}{|\mathbf{r}-\mathbf{R}|} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{4\pi}{2l+1} \frac{r_<^l}{r_>^{l+1}} Y_l^{m*}\left(\frac{\mathbf{r}}{r}\right) Y_l^m\left(\frac{\mathbf{R}}{R}\right),$$
$$r_< = \min(r,R), \quad r_> = \max(r,R). \tag{1.1}$$

There is an extensive literature on addition theorems. Particularly well-studied are the addition theorems of those solutions of the homogeneous Laplace, Helmholtz, and modified Helmholtz equations that are also eigenstates of the orbital angular momentum operators. The addition theorems of the regular and irregular solid harmonics which are solutions of the homogeneous Laplace equation were studied by Hobson,[1] Rose,[2] Chiu,[3] Sack,[4,5] Dahl and Barnett,[6] Steinborn,[7] Steinborn and Ruedenberg[8] and by Tough and Stone.[9] The addition theorems of the Helmholtz harmonics which are products of Bessel functions and spherical harmonics were studied by Friedman and Russek,[10] Stein,[11] Cruzan,[12] Sack,[5] Danos and Maximon,[13] Nozawa,[14] and by Steinborn and Filter.[15] The addition theorems of the modified Helmholtz harmonics which are products of modified Bessel functions and spherical harmonics were studied by Buttle and Goldfarb[16] and by Steinborn and Filter.[15]

In the articles cited a multitude of different methods was used for the derivation of these addition theorems. Most of these approaches, however, are relatively complicated and sometimes rather lengthy and are based upon some special properties of the functions under consideration. Therefore, it is the intention of this article to demonstrate that the addition theorems of the irregular solid harmonics, the Helmholtz harmonics, and the modified Helmholtz harmonics can be derived in a very simple and unified way. Our method has the additional advantage that it can also be applied in the case of other functions.

Our derivation is based upon some special differential operator, which we call the spherical tensor gradient $\mathscr{Y}_l^m(\nabla)$. It is obtained from the regular solid harmonic $\mathscr{Y}_l^m(\mathbf{r})$ by replacing the Cartesian components of $\mathbf{r}$—x, y, and z—by the differentials $\partial/\partial x$, $\partial/\partial y$, and $\partial/\partial z$. The properties of the spherical tensor gradient, which was in principle already used by Hobson,[1] were investigated by Santos,[17] Rowe,[18] Bayman,[19] Stuart,[20] and recently by Niukkanen[21,22] and ourselves.[23,24] We shall show that there exists an intimate relationship between the spherical tensor gradient and irregular solid harmonics of (modified) Helmholtz harmonics, respectively, which can be employed profitably for the derivation of addition theorems.

II. DEFINITIONS

For the commonly occurring special functions of mathematical physics we shall use the notations and conventions of Magnus, Oberhettinger, and Soni[25] unless explicitly stated. Hereafter, this reference will be denoted as MOS in the text.

For the spherical harmonics $Y_l^m(\theta,\phi)$ we use the phase convention of Condon and Shortley,[26] i.e., they are defined by the expression

$$Y_l^m(\theta,\phi) = i^{m+|m|}\left[\frac{(2l+1)}{4\pi}\frac{(l-|m|)!}{(l+|m|)!}\right]^{1/2} \times P_l^{|m|}(\cos\theta)e^{im\phi}. \tag{2.1}$$

Here, $P_l^{|m|}(\cos\theta)$ is an associated Legendre polynomial

$$P_l^m(x) = (1-x^2)^{m/2}\frac{d^{l+m}}{dx^{l+m}}\frac{(x^2-1)^l}{2^l l!}$$
$$= (1-x^2)^{m/2}\frac{d^m}{dx^m}P_l(x). \tag{2.2}$$

For the regular and irregular solid harmonics we use the notation

 0022-2488/85/040664-07$02.50

$$\mathcal{Y}_l^m(\mathbf{r}) = r^l Y_l^m(\theta,\phi), \tag{2.3}$$

$$\mathcal{Z}_l^m(\mathbf{r}) = r^{-l-1} Y_l^m(\theta,\phi). \tag{2.4}$$

For the integral of the product of three spherical harmonics over the surface of the unit sphere in $\mathbf{R}^3$ we write

$$\langle l_3 m_3 | l_2 m_2 | l_1 m_1 \rangle = \int Y_{l_3}^{m_3 *}(\Omega) Y_{l_2}^{m_2}(\Omega) Y_{l_1}^{m_1}(\Omega) d\Omega. \tag{2.5}$$

These Gaunt coefficients may be expressed in terms of Clebsch–Gordan coefficients[27] or $3jm$ symbols

$$\langle l_3 m_3 | l_2 m_2 | l_1 m_1 \rangle = (-1)^{m_3} \left[\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi} \right]^{1/2} \times \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & -m_3 \end{pmatrix}. \tag{2.6}$$

With the help of the Gaunt coefficients the product of two spherical harmonics can be linearized

$$Y_{l_1}^{m_1}(\theta,\phi) Y_{l_2}^{m_2}(\theta,\phi) = \sum_{l=l_{\min}}^{l_{\max}}{}^{(2)} \langle l m_1 + m_2 | l_1 m_1 | l_2 m_2 \rangle Y_l^{m_1+m_2}(\theta,\phi). \tag{2.7}$$

The symbol $\Sigma^{(2)}$ indicates that the summation is to be performed in steps of two. The summation limits in Eq. (2.7) are direct conseqences of the selection rules satisfied by the Gaunt coefficient and are given by[28]

$$l_{\max} = l_1 + l_2, \tag{2.8a}$$

$$l_{\min} = \begin{cases} \max(|l_1 - l_2|, |m_1 + m_2|), \\ \quad \text{if } l_{\max} + \max(|l_1 - l_2|, |m_1 + m_2|) \text{ is even,} \\ \text{and} \\ \max(|l_1 - l_2|, |m_1 + m_2|) + 1, \\ \quad \text{if } l_{\max} + \max(|l_1 - l_2|, |m_1 + m_2|) \text{ is odd.} \end{cases} \tag{2.8b}$$

III. SOME PROPERTIES OF THE SPHERICAL TENSOR GRADIENT

In this section we shall review only those properties of the spherical tensor gradient $\mathcal{Y}_l^m(\nabla)$ which are needed for our derivation of the addition theorems of the irregular solid harmonics and the (modified) Helmholtz harmonics. Further properties can be found elsewhere.[17–24]

The spherical tensor gradient is an irreducible spherical tensor of rank l.[29] Therefore, if the spherical tensor gradient is applied to a function $\phi(r)$ which only depends upon the distance r, i.e., to an irreducible spherical tensor of rank zero, we obtain in agreement with the angular momentum coupling rules an irreducible spherical tensor of rank l, which is given by

$$\mathcal{Y}_l^m(\nabla)\phi(r) = \left[\left(\frac{1}{r}\frac{d}{dr} \right)^l \phi(r) \right] \mathcal{Y}_l^m(\mathbf{r}). \tag{3.1}$$

As we showed recently[30] this relationship can be derived quite easily with the help of a theorem on differentiation which was published by Hobson[31] already in 1892. Equation (3.1) can also be obtained by considering special cases in more recent publications by Santos,[32] Bayman,[33] Stuart,[34] and Niukkanen[35] who, however, apparently were not aware of Hobson's theorem.[31] If the spherical tensor gradient is applied to another spherical tensor of nonvanishing rank, i.e., to a function that can be written as

$$F_{l_2}^{m_2}(\mathbf{r}) = f_{l_2}(r) Y_{l_2}^{m_2}(\theta,\phi), \tag{3.2}$$

the structure of the resulting expression can also be understood in terms of angular momentum coupling,[36]

$$\mathcal{Y}_{l_1}^{m_1}(\nabla) F_{l_2}^{m_2}(\mathbf{r}) = \sum_{l=l_{\min}}^{l_{\max}}{}^{(2)} \langle l m_1 + m_2 | l_1 m_1 | l_2 m_2 \rangle \times \gamma_{l_1 l_2}^l(r) Y_l^{m_1+m_2}(\theta,\phi). \tag{3.3}$$

For the functions $\gamma_{l_1 l_2}^l$ in Eq. (3.3) various representations could be derived, for instance[37]

$$\gamma_{l_1 l_2}^l(r) = \sum_{q=0}^{\Delta l} \frac{(-\Delta l)_q (-\sigma(l) - \frac{1}{2})_q}{q!} 2^q r^{l_1 + l_2 - 2q} \times \left(\frac{1}{r}\frac{d}{dr} \right)^{l_1 - q} \frac{f_{l_2}(r)}{r^{l_2}} \tag{3.4}$$

$$= \sum_{s=0}^{\Delta l_2} \frac{(-\Delta l_2)_s (\Delta l_1 + \frac{1}{2})_s}{s!} 2^s r^{l_1 - l_2 - 2s - 1} \times \left(\frac{1}{r}\frac{d}{dr} \right)^{l_1 - s} r^{l_2 + 1} f_{l_2}(r), \tag{3.5}$$

$$\Delta l = (l_1 + l_2 - l)/2, \quad \Delta l_1 = (l - l_1 + l_2)/2,$$
$$\Delta l_2 = (l + l_1 - l_2)/2, \quad \sigma(l) = (l_1 + l_2 + l)/2. \tag{3.6}$$

It is a direct consequence of the selection rules satisfied by the Gaunt coefficient in Eq. (3.3) that Δl, Δl_1, Δl_2, and $\sigma(l)$ are always either positive integers or zero.

Since the spherical tensor gradient is obtained from the regular solid harmonic by replacing the Cartesian components of $\mathbf{r}$ by the Cartesian components of ∇ we may conclude that the spherical tensor gradient and the regular solid harmonics must obey the same coupling law. Hence we obtain from Eq. (2.7) (see Refs. 38 and 39)

$$\mathcal{Y}_{l_1}^{m_1}(\nabla)\mathcal{Y}_{l_2}^{m_2}(\nabla) = \sum_{l=l_{\min}}^{l_{\max}}{}^{(2)} \langle l m_1 + m_2 | l_1 m_1 | l_2 m_2 \rangle \times \nabla^{l_1 + l_2 - l} \mathcal{Y}_l^{m_1+m_2}(\nabla). \tag{3.7}$$

Let us now assume that a spherical tensor $F_{l_2}^{m_2}(\mathbf{r})$ and a radially symmetric function $\phi(r)$ are known, which satisfy

$$F_{l_2}^{m_2}(\mathbf{r}) = \mathcal{Y}_{l_2}^{m_2}(\nabla)\phi(r). \tag{3.8}$$

If the spherical tensor gradient $\mathcal{Y}_{l_1}^{m_1}(\nabla)$ is applied to $F_{l_2}^{m_2}(\mathbf{r})$ we then can couple the two spherical tensor gradients according to Eq. (3.7) and finally obtain with the help of Eq. (3.1)

$$\mathcal{Y}_{l_1}^{m_1}(\nabla) F_{l_2}^{m_2}(\mathbf{r}) = \mathcal{Y}_{l_1}^{m_1}(\nabla)\mathcal{Y}_{l_2}^{m_2}(\nabla)\phi(r) = \sum_{l=l_{\min}}^{l_{\max}}{}^{(2)} \langle l m_1 + m_2 | l_1 m_1 | l_2 m_2 \rangle \nabla^{l_1 + l_2 - l} \times \left[\left(\frac{1}{r}\frac{d}{dr} \right)^l \phi(r) \right] \mathcal{Y}_l^{m_1+m_2}(\mathbf{r}). \tag{3.9}$$

This relationship is particularly well-suited for the functions which are treated in this article since in these cases the differential operators which occur in Eq. (3.9) can be applied quite easily. Under these circumstances Eq. (3.9) is in our opinion preferable to other, more general expressions which were, for instance, given by Santos,[17] Niukkanen,[21] and ourselves.[24] Relationships of the type of Eq. (3.9) were already used by Novosadov[40] and ourselves[41] in connection with functions related to modified Bessel functions.

IV. THE ADDITION THEOREM OF THE IRREGULAR SOLID HARMONICS

Our derivation of the addition theorem of the irregular solid harmonics will be based upon the fact that the addition theorem of the Coulomb potential, Eq. (1.1), is known and that the application of the spherical tensor gradient to the Coulomb potential yields the irregular solid harmonic

$$\mathscr{Z}_l^m(\mathbf{r}) = [(-1)^l/(2l-1)!!]\,\mathscr{Y}_l^m(\nabla)(1/r). \tag{4.1}$$

This relationship, which was already known to Hobson,[1,31] can be proved quite easily with the help of Eq. (3.1). In order to facilitate the application of the spherical tensor gradient we rewrite the Laplace expansion of the Coulomb potential, Eq. (1.1), in the following way, which is more convenient for our purposes:

$$\frac{1}{|\mathbf{r}_< + \mathbf{r}_>|} = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{(-1)^l}{2l+1}\,\mathscr{Y}_l^{m*}(\mathbf{r}_<)\mathscr{Z}_l^m(\mathbf{r}_>). \tag{4.2}$$

Here, $\mathbf{r}_<$ is the vector with the smaller magnitude and $\mathbf{r}_>$ is the vector with the greater magnitude.

The spherical tensor gradient is invariant with respect to translation. Consequently, Eq. (4.1) can be rewritten in the following ways:

$$\mathscr{Z}_l^m(\mathbf{r}_< + \mathbf{r}_>) = \frac{(-1)^l}{(2l-1)!!}\,\mathscr{Y}_l^m(\nabla_<)\frac{1}{|\mathbf{r}_< + \mathbf{r}_>|}, \tag{4.3}$$

$$= \frac{(-1)^l}{(2l-1)!!}\,\mathscr{Y}_l^m(\nabla_>)\frac{1}{|\mathbf{r}_< + \mathbf{r}_>|}. \tag{4.4}$$

Here, $\nabla_<$ implies a differentiation with respect to $\mathbf{r}_<$ and $\nabla_>$ implies a differentiation with respect to $\mathbf{r}_>$. If we combine Eqs. (4.2) and (4.4) we find

$$\mathscr{Z}_l^m(\mathbf{r}_< + \mathbf{r}_>) = \frac{(-1)^l 4\pi}{(2l-1)!!} \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} \frac{(-1)^{l_1}}{2l_1+1} \times \mathscr{Y}_{l_1}^{m_1*}(\mathbf{r}_<)\mathscr{Y}_l^m(\nabla_>)\mathscr{Z}_{l_1}^{m_1}(\mathbf{r}_>). \tag{4.5}$$

The remaining differentiation can be performed quite easily. The easiest way would be the use of Eqs. (3.8) and (3.9) in connection with Eq. (4.1). We then obtain

$$\mathscr{Y}_{\lambda_1}^{\mu_1}(\nabla)\mathscr{Z}_{\lambda_2}^{\mu_2}(\mathbf{r}) = (-1)^{\lambda_1} \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}}{}^{(2)} \frac{(2\lambda-1)!!}{(2\lambda_2-1)!!} \times \langle \lambda\mu_1+\mu_2|\lambda_1\mu_1|\lambda_2\mu_2\rangle \nabla^{\lambda_1+\lambda_2-\lambda}\mathscr{Z}_\lambda^{\mu_1+\mu_2}(\mathbf{r}). \tag{4.6}$$

If we take into account that the irregular solid harmonics are solutions of the homogeneous Laplace equation we see that in Eq. (4.6) only the term with $l = l_1 + l_2$ can be different from zero. This implies

$$\mathscr{Y}_{\lambda_1}^{\mu_1}(\nabla)\mathscr{Z}_{\lambda_2}^{\mu_2}(\mathbf{r}) = (-1)^{\lambda_1}(2\lambda_1+2\lambda_2-1)!!/(2\lambda_2-1)!! \times \langle \lambda_1+\lambda_2\mu_1+\mu_2|\lambda_1\mu_1|\lambda_2\mu_2\rangle \mathscr{Z}_{\lambda_1+\lambda_2}^{\mu_1+\mu_2}(\mathbf{r}). \tag{4.7}$$

Inserting this result into Eq. (4.5) yields the addition theorem for the irregular solid harmonics

$$\mathscr{Z}_l^m(\mathbf{r}_< + \mathbf{r}_>) = 4\pi \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} (-1)^{l_1} \frac{(2l+2l_1-1)!!}{(2l_1+1)!!(2l-1)!!} \times \langle l+l_1 m+m_1|lm|l_1m_1\rangle \mathscr{Y}_{l_1}^{m_1*}(\mathbf{r}_<)\mathscr{Z}_{l+l_1}^{m+m_1}(\mathbf{r}_>). \tag{4.8}$$

The Gaunt coefficient in Eq. (4.8) can be expressed in closed form. In that case one obtains the factorless form of the addition theorem which was given by Steinborn[7] and by Steinborn and Ruedenberg.[8]

If we now combine Eqs. (4.2) and (4.3) we find

$$\mathscr{Z}_l^m(\mathbf{r}_< + \mathbf{r}_>) = \frac{(-1)^l 4\pi}{(2l-1)!!} \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} \frac{(-1)^{l_1}}{2l_1+1} \times \mathscr{Y}_l^m(\nabla_<)\mathscr{Y}_{l_1}^{m_1*}(\mathbf{r}_<)\mathscr{Z}_{l_1}^{m_1}(\mathbf{r}_>). \tag{4.9}$$

The remaining differentiation again poses no problems. With the help of Eqs. (3.3) and (3.4) we obtain after some algebra

$$\mathscr{Y}_{\lambda_1}^{\mu_1}(\nabla)\mathscr{Y}_{\lambda_2}^{\mu_2}(\mathbf{r}) = \frac{(2\lambda_2+1)!!}{(2\lambda_2-2\lambda_1+1)!!}\langle \lambda_2-\lambda_1\,\mu_1+\mu_2|\lambda_1\mu_1|\lambda_2\mu_2\rangle \times \mathscr{Y}_{\lambda_2-\lambda_1}^{\mu_1+\mu_2}(\mathbf{r}). \tag{4.10}$$

If we insert this result into Eq. (4.9) we find another version of the addition theorem of the irregular solid harmonics

$$\mathscr{Z}_l^m(\mathbf{r}_< + \mathbf{r}_>) = 4\pi \sum_{l_1=l}^{\infty} \sum_{m_1=-l_1}^{l_1} \frac{(2l_1-1)!!}{(2l-1)!!(2l_1-2l+1)!!} \times \langle l_1m_1|lm|l_1-l\,m_1-m\rangle \times (-1)^{l_1+l}\mathscr{Y}_{l_1-l}^{m_1-m*}(\mathbf{r}_<)\mathscr{Z}_{l_1}^{m_1}(\mathbf{r}_>). \tag{4.11}$$

In order to prove the equivalence of Eqs. (4.8) and (4.11) we introduce new summation variables in Eq. (4.11)

$$l_2 = l_1 - l, \quad m_2 = m_1 - m. \tag{4.12}$$

With these definitions we find for Eq. (4.11)

$$\mathscr{Z}_l^m(\mathbf{r}_< + \mathbf{r}_>) = 4\pi \sum_{l_2=0}^{\infty} \sum_{m_2=-l_2}^{l_2} (-1)^{l_2} \frac{(2l+2l_2-1)!!}{(2l-1)!!(2l_2+1)!!} \times \langle l+l_2 m+m_2|lm|l_2m_2\rangle \mathscr{Y}_{l_2}^{m_2*}(\mathbf{r}_<)\mathscr{Z}_{l+l_2}^{m+m_2}(\mathbf{r}_>). \tag{4.13}$$

Obviously, Eqs. (4.8) and (4.13) are identical.

V. THE ADDITION THEOREMS OF THE HELMHOLTZ HARMONICS

In this section $C_\nu(z)$ stands for any of the Bessel functions $J_\nu(z)$ and $Y_\nu(z)$ or Hankel functions $H_\nu^{(1)}(z)$ and $H_\nu^{(2)}(z)$, which are defined by (MOS, pp. 65–66)

$$J_\nu(z) = \sum_{m=0}^{\infty} \frac{(-1)^m (z/2)^{\nu+2m}}{m!\Gamma(\nu+m+1)}$$
$$= \frac{(z/2)^\nu}{\Gamma(\nu+1)} {}_0F_1\left(\nu+1; -\frac{z^2}{4}\right), \tag{5.1}$$
$$Y_\nu(z) = [1/\sin(\pi\nu)][\cos(\pi\nu)J_\nu(z) - J_{-\nu}(z)], \tag{5.2}$$
$$H_\nu^{(1)}(z) = J_\nu(z) + iY_\nu(z), \tag{5.3}$$
$$H_\nu^{(2)}(z) = J_\nu(z) - iY_\nu(z). \tag{5.4}$$

This generalization is possible because for our derivation of the addition theorems we shall only need the following differential formulas and the recurrence relationship of these functions (MOS, p.67)

$$\left(\frac{1}{z}\frac{d}{dz}\right)^m z^\nu C_\nu(z) = z^{\nu-m} C_{\nu-m}(z), \tag{5.5}$$
$$\left(\frac{1}{z}\frac{d}{dz}\right)^m z^{-\nu} C_\nu(z) = (-1)^m z^{-\nu-m} C_{\nu+m}(z), \tag{5.6}$$
$$C_{\nu-1}(z) + C_{\nu+1}(z) = (2\nu/z) C_\nu(z). \tag{5.7}$$

With the help of these formulas the following relationships can be proved quite easily:

$$[1 + \alpha^{-2}\nabla^2](\alpha r)^{-l-1/2} C_{l+1/2}(\alpha r) \mathscr{Y}_l^m(\alpha \mathbf{r}) = 0, \tag{5.8}$$
$$[1 + \alpha^{-2}\nabla^2](\alpha r)^{-l-1/2} C_{-l-1/2}(\alpha r) \mathscr{Y}_l^m(\alpha \mathbf{r}) = 0. \tag{5.9}$$

The functions in Eqs. (5.8) and (5.9) are usually called Helmholtz harmonics. It seems that we have obtained two different classes of solutions of the homogeneous three-dimensional Helmholtz equation. However, in the case of half-integral orders $\nu = n + \frac{1}{2}$, $n \in \mathbb{Z}$, there exist symmetry relationships among Bessel functions, for instance (MOS, p. 72)

$$Y_{-n-1/2}(z) = (-1)^n J_{n+1/2}(z), \quad n \in \mathbb{N}. \tag{5.10}$$

Hence, if $C_{n+1/2}$ stands for one of the Bessel functions $J_{n+1/2}$, $Y_{n+1/2}$, $H_{n+1/2}^{(1)}$, and $H_{n+1/2}^{(2)}$, then $C_{-n-1/2}$ can also be expressed in terms of one of these functions. Consequently, it would in principle be sufficient to derive the addition theorems for either the functions in Eq. (5.8) or those in Eq. (5.9). However, since the derivation is in either case quite simple we shall derive the addition theorems for the functions in Eqs. (5.8) and (5.9) independently.

In Eqs. (5.5) and (5.6) the differential operator $z^{-1}\,d/dz$ acts as a kind of a shift operator for the order ν. Hence, if we combine Eq. (3.1) with either Eq. (5.5) or (5.6) we immediately find

$$(\alpha r)^\nu C_\nu(\alpha r) \mathscr{Y}_l^m(\alpha \mathbf{r}) = \alpha^{-l} \mathscr{Y}_l^m(\nabla)(\alpha r)^{\nu+l} C_{\nu+l}(\alpha r), \tag{5.11}$$
$$(\alpha r)^{-\nu} C_\nu(\alpha r) \mathscr{Y}_l^m(\alpha \mathbf{r}) = (-\alpha)^{-l} \mathscr{Y}_l^m(\nabla)(\alpha r)^{l-\nu} C_{\nu-l}(\alpha r). \tag{5.12}$$

Bessel and Hankel functions with orders $\nu = \pm\frac{1}{2}$ are essentially trigonometric functions, for instance (MOS, p. 73)

$$J_{1/2}(z) = [2/\pi z]^{1/2} \sin z. \tag{5.13}$$

Therefore, we see that the Helmholtz harmonics with higher angular momentum quantum numbers may be generated by applying the spherical tensor gradient to some trigonometric functions,

$$(\alpha r)^{-l-1/2} C_{-l-1/2}(\alpha r) \mathscr{Y}_l^m(\alpha \mathbf{r})$$
$$= \alpha^{-l} \mathscr{Y}_l^m(\nabla)(\alpha r)^{-1/2} C_{-1/2}(\alpha r), \tag{5.14}$$
$$(\alpha r)^{-l-1/2} C_{l+1/2}(\alpha r) \mathscr{Y}_l^m(\alpha \mathbf{r})$$
$$= (-\alpha)^{-l} \mathscr{Y}_l^m(\nabla)(\alpha r)^{-1/2} C_{1/2}(\alpha r). \tag{5.15}$$

These relationships suggest that the addition theorems of the Helmholtz harmonics can be derived in exactly the same way as we derived the addition theorem of the irregular solid harmonics in Sec. IV. We only have to apply the spherical tensor gradient to the addition theorems of the relatively simple functions $(\alpha r)^{-1/2} C_{\pm 1/2}(\alpha r)$, which are usually called Gegenbauer addition theorems (MOS, p. 107), and which can be compactly written as

$$(\alpha|\mathbf{r}_< + \mathbf{r}_>|)^{-1/2} C_{-1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|)$$
$$= (2\pi)^{3/2} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (\alpha r_<)^{-l-1/2} J_{l+1/2}(\alpha r_<) \mathscr{Y}_l^{m*}(\alpha \mathbf{r}_<)(\alpha r_>)^{-l-1/2} C_{-l-1/2}(\alpha r_>) \mathscr{Y}_l^m(\alpha \mathbf{r}_>), \tag{5.16}$$
$$(\alpha|\mathbf{r}_< + \mathbf{r}_>|)^{-1/2} C_{1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|)$$
$$= (2\pi)^{3/2} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (-1)^l (\alpha r_<)^{-l-1/2} J_{l+1/2}(\alpha r_<) \mathscr{Y}_l^{m*}(\alpha \mathbf{r}_<)(\alpha r_>)^{-l-1/2} C_{l+1/2}(\alpha r_>) \mathscr{Y}_l^m(\alpha \mathbf{r}_>). \tag{5.17}$$

Again, $\mathbf{r}_<$ is the vector with the smaller and $\mathbf{r}_>$ is the vector with the greater magnitude. Following our procedure in Sec. IV we differentiate Eq. (5.16) with respect to $\mathbf{r}_>$ and obtain with the help of Eq. (5.14)

$$(\alpha|\mathbf{r}_< + \mathbf{r}_>|)^{-l-1/2} C_{-l-1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|) \mathscr{Y}_l^m(\alpha[\mathbf{r}_< + \mathbf{r}_>])$$
$$= \alpha^{-l} \mathscr{Y}_l^m(\nabla_>)(\alpha|\mathbf{r}_< + \mathbf{r}_>|)^{-1/2} C_{-1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|)$$
$$= (2\pi)^{3/2} \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} (\alpha r_<)^{-l_1-1/2} J_{l_1+1/2}(\alpha r_<) \mathscr{Y}_{l_1}^{m_1*}(\alpha \mathbf{r}_<) \alpha^{-l} \mathscr{Y}_l^m(\nabla_>)(\alpha r_>)^{-l_1-1/2} C_{-l_1-1/2}(\alpha r_>) \mathscr{Y}_{l_1}^{m_1}(\alpha \mathbf{r}_>). \tag{5.18}$$

The remaining differentiation can be done quite easily. With the help of Eqs. (3.8), (3.9), (5.8), and (5.14) we obtain immediately

$$\alpha^{-\lambda_1} \mathscr{Y}_{\lambda_1}^{\mu_1}(\nabla)(\alpha r)^{-\lambda_2-1/2} C_{-\lambda_2-1/2}(\alpha r) \mathscr{Y}_{\lambda_2}^{\mu_2}(\alpha \mathbf{r})$$
$$= \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}}{}^{(2)} (-1)^{\Delta\lambda} \langle \lambda\mu_1 + \mu_2 | \lambda_1\mu_1 | \lambda_2\mu_2 \rangle (\alpha r)^{-\lambda-1/2} C_{-\lambda-1/2}(\alpha r) \mathscr{Y}_\lambda^{\mu_1+\mu_2}(\alpha \mathbf{r}), \quad \Delta\lambda = (\lambda_1 + \lambda_2 - l)/2. \tag{5.19}$$

If we insert this relationship into Eq. (5.18) we obtain the addition theorem

$$
\begin{aligned}
&(\alpha|\mathbf{r}_< + \mathbf{r}_>|)^{-l-1/2} C_{-l-1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|) \mathscr{Y}_l^m(\alpha[\mathbf{r}_< + \mathbf{r}_>]) \\
&\quad = (2\pi)^{3/2} \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} (\alpha r_<)^{-l-1/2} J_{l_1+1/2}(\alpha r_<) \mathscr{Y}_{l_1}^{m_1*}(\alpha \mathbf{r}_<) \sum_{l_2=l_2^{\min}}^{l_2^{\max}\,(2)} (-1)^{\Delta l_2} \\
&\qquad \times \langle l_2 m + m_1 | l m | l_1 m_1 \rangle (\alpha r_>)^{-l_2-1/2} C_{-l_2-1/2}(\alpha r_>) \mathscr{Y}_{l_2}^{m+m_1}(\alpha \mathbf{r}_>), \quad \Delta l_2 = (l + l_1 - l_2)/2.
\end{aligned} \tag{5.20}
$$

This addition theorem can also be derived by differentiating Eq. (5.16) with respect to $\mathbf{r}_<$. We only need

$$
\begin{aligned}
&(-\alpha)^{-\lambda_1} \mathscr{Y}_{\lambda_1}^{\mu_1}(\nabla) (\alpha r)^{-\lambda_2-1/2} C_{\lambda_2+1/2}(\alpha r) \mathscr{Y}_{\lambda_2}^{\mu_2}(\alpha \mathbf{r}) \\
&\quad = \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}\,(2)} (-1)^{\Delta\lambda} \langle \lambda \mu_1 + \mu_2 | \lambda_1 \mu_1 | \lambda_2 \mu_2 \rangle (\alpha r)^{-\lambda-1/2} C_{\lambda+1/2}(\alpha r) \mathscr{Y}_{\lambda}^{\mu_1+\mu_2}(\alpha \mathbf{r}), \quad \Delta\lambda = (\lambda_1 + \lambda_2 - \lambda)/2,
\end{aligned} \tag{5.21}
$$

which can be proved with the help of Eqs. (3.8), (3.9), (5.9), and (5.15) to obtain a somewhat different representation of the addition theorem,

$$
\begin{aligned}
&(\alpha|\mathbf{r}_< + \mathbf{r}_>|)^{-l-1/2} C_{-l-1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|) \mathscr{Y}_l^m(\alpha[\mathbf{r}_< + \mathbf{r}_>]) \\
&\quad = (2\pi)^{3/2} \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} (\alpha r_>)^{-l_1-1/2} C_{-l_1-1/2}(\alpha r_>) \mathscr{Y}_{l_1}^{m_1}(\alpha \mathbf{r}_>) \sum_{l_2=l_2^{\min}}^{l_2^{\max}\,(2)} (-1)^{\Delta l_1} \\
&\qquad \times \langle l_1 m_1 | l m | l_2 m_1 - m \rangle (\alpha r_<)^{-l_2-1/2} J_{l_2+1/2}(\alpha r_<) \mathscr{Y}_{l_2}^{m_1-m*}(\alpha \mathbf{r}_<), \quad \Delta l_1 = (l - l_1 + l_2)/2.
\end{aligned} \tag{5.22}
$$

To prove the equivalence of Eqs. (5.20) and (5.22) we only have to introduce in Eq. (5.22) the new summation variable $\mu_2 = m_1 - m$ and to change the order of the two l summations.

The addition theorem of the function $(\alpha r)^{-l-1/2} C_{l+1/2}(\alpha r) \mathscr{Y}_l^m(\alpha \mathbf{r})$ can be derived in exactly the same way. If we differentiate Eq. (5.17) with respect to $\mathbf{r}_>$ we find

$$
\begin{aligned}
&(\alpha|\mathbf{r}_< + \mathbf{r}_>|)^{-l-1/2} C_{l+1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|) \mathscr{Y}_l^m(\alpha[\mathbf{r}_< + \mathbf{r}_>]) \\
&\quad = (2\pi)^{3/2} \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} (-1)^{l_1} (\alpha r_<)^{-l_1-1/2} J_{l_1+1/2}(\alpha r_<) \mathscr{Y}_{l_1}^{m_1*}(\alpha \mathbf{r}_<) \\
&\qquad \times \sum_{l_2=l_2^{\max}}^{l_2^{\min}\,(2)} (-1)^{\Delta l_2} \langle l_2 m + m_1 | l m | l_1 m_1 \rangle (\alpha r_<)^{-l_2-1/2} C_{l_2+1/2}(\alpha r_>) \mathscr{Y}_{l_2}^{m+m_1}(\alpha \mathbf{r}_>), \quad \Delta l_2 = (l + l_1 - l_2)/2.
\end{aligned} \tag{5.23}
$$

If we differentiate Eq. (5.17) with respect to $\mathbf{r}_<$, we find

$$
\begin{aligned}
&(\alpha|\mathbf{r}_< + \mathbf{r}_>|)^{-l-1/2} C_{l+1/2}(\alpha|\mathbf{r}_< + \mathbf{r}_>|) \mathscr{Y}_l^m(\alpha[\mathbf{r}_< + \mathbf{r}_>]) \\
&\quad = (2\pi)^{3/2} \sum_{l_1=0}^{\infty} \sum_{m_1=-l_1}^{l_1} (-1)^{l_1} (\alpha r_>)^{-l_1-1/2} C_{l_1+1/2}(\alpha r_>) \mathscr{Y}_{l_1}^{m_1}(\alpha \mathbf{r}_>) \\
&\qquad \times \sum_{l_2=l_2^{\min}}^{l_2^{\max}\,(2)} (-1)^{\Delta l_2} \langle l_1 m_1 | l m | l_2 m - m_1 \rangle (\alpha r_<)^{-l_2-1/2} J_{l_2+1/2}(\alpha r_<) \mathscr{Y}_{l_2}^{m-m_1*}(\alpha \mathbf{r}_<).
\end{aligned} \tag{5.24}
$$

The equivalence of Eqs. (5.23) and (5.24) can be proved by introducing the new summation variable $\mu_2 = m - m_1$ into (5.24) and by changing the order of the two l summations.

VI. THE ADDITION THEOREM OF THE MODIFIED HELMHOLTZ HARMONICS

The differential operator of the modified Helmholtz equation, $1 - \alpha^{-2}\nabla^2$, can be obtained from the differential operator of the Helmholtz equation, $1 + \alpha^{-2}\nabla^2$, if the parameter α is replaced by $i\alpha$. Consequently, the solutions of the homogeneous modified Helmholtz equations can be expressed in terms of modified Bessel functions. This follows also from the following relationships, which can be proved quite easily using known differential and recursive properties of the modified Bessel functions,

$$[1 - \alpha^{-2}\nabla^2](\alpha r)^{-l-1/2} I_{-l-1/2}(\alpha r) \mathscr{Y}_l^m(\alpha \mathbf{r}) = 0, \tag{6.1}$$

$$[1 - \alpha^{-2}\nabla^2](\alpha r)^{-l-1/2} I_{l+1/2}(\alpha r) \mathscr{Y}_l^m(\alpha \mathbf{r}) = 0, \tag{6.2}$$

$$[1 - \alpha^{-2}\nabla^2](\alpha r)^{-l-1/2} K_{l+1/2}(\alpha r) \mathscr{Y}_l^m(\alpha \mathbf{r}) = 0. \tag{6.3}$$

Here, $I_\nu(z)$ is a modified Bessel function of the first kind (MOS, p. 66),

$$I_\nu(z) = \sum_{m=0}^{\infty} \frac{(z/2)^{\nu+2m}}{m!\Gamma(\nu+m+1)}$$
$$= \frac{(z/2)^\nu}{\Gamma(\nu+1)}\, {}_0F_1\left(\nu+1; \frac{z^2}{4}\right), \tag{6.4}$$

and $K_\nu(z)$ is a modified Bessel function of the second kind (MOS, p. 66),

$$K_\nu(z) = \pi/[2\sin(\pi\nu)][I_{-\nu}(z) - I_\nu(z)]. \tag{6.5}$$

The functions of the first kind, $I_\nu(z)$, increase exponentially for large arguments z whereas the functions of the second kind, $K_\nu(z)$, decline exponentially (MOS, p. 139). Consequently, it is not surprising that only the modified Helmholtz harmonics which occur in Eq. (6.3) have been of physical interest so far.

The modified Helmholtz harmonics in Eq. (6.3) may be considered to be some special B functions which are defined by[42]

$$B_{n,l}^m(\alpha,\mathbf{r}) = (2/\pi)^{1/2}/[2^{n+l}(n+l)!](\alpha r)^{n-1/2}$$
$$\times K_{n-1/2}(\alpha r)\mathscr{Y}_l^m(\alpha\mathbf{r}). \tag{6.6}$$

Because of the factorial in the denominator, B functions are only defined in the sense of classical analysis if the inequality $n + l \geqslant 0$ holds. However, it can be shown that the definition of the B functions, Eq. (6.6), remains meaningful even if n is a negative integer such that $n + l < 0$ holds. In those cases B functions are distributions which can be identified with derivatives of the delta function.[24]

If B functions are used Eq. (6.3) can be rewritten as

$$[1-\alpha^{-2}\nabla^2]B_{-l,l}^m(\alpha,\mathbf{r}) = 0. \tag{6.7}$$

If the spherical tensor gradient is applied to a scalar B function, one obtains[43]

$$B_{n,l}^m(\alpha,\mathbf{r}) = (4\pi)^{1/2}(-\alpha)^{-l}\mathscr{Y}_l^m(\nabla)B_{n+l,0}^0(\alpha,\mathbf{r}). \tag{6.8}$$

If we set in Eq. (6.8) $n = -l$ we find

$$B_{-l,l}^m(\alpha,\mathbf{r}) = (4\pi)^{1/2}(-\alpha)^{-l}\mathscr{Y}_l^m(\nabla)B_{0,0}^0(\alpha,\mathbf{r}). \tag{6.9}$$

However, the function $B_{0,0}^0$ is proportional to the Yukawa potential,

$$B_{0,0}^0(\alpha,\mathbf{r}) = (4\pi)^{-1/2}e^{-\alpha r}/(\alpha r), \tag{6.10}$$

for which an addition theorem is known (MOS, p. 107). We rewrite this addition theorem in the following way:

$$B_{0,0}^0(\alpha,\mathbf{r}_< + \mathbf{r}_>)$$
$$= (2\pi^2)^{1/2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-1)^l(\alpha r_<)^{-l-1/2}$$
$$\times I_{l+1/2}(\alpha r_<)\mathscr{Y}_l^{m*}(\alpha\mathbf{r}_<)B_{-l,l}^m(\alpha,\mathbf{r}_>). \tag{6.11}$$

Again, $\mathbf{r}_<$ is the vector with the smaller and $\mathbf{r}_>$ is the vector with the greater magnitude.

The derivation of the addition theorems of the modified Helmholtz harmonics can now be done in exactly the same way as the derivation of the addition theorems of the irregular solid harmonics and of the Helmholtz harmonics. If we differentiate Eq. (6.11) with respect to $\mathbf{r}$, we find

$$B_{-l,l}^m(\alpha,\mathbf{r}_< + \mathbf{r}_>)$$
$$= (4\pi)^{1/2}(-\alpha)^{-l}\mathscr{Y}_l^m(\nabla_>)B_{0,0}^0(\alpha,\mathbf{r}_< + \mathbf{r}_>)$$
$$= (2\pi)^{3/2}\sum_{l_1=0}^{\infty}\sum_{m_1=-l_1}^{l_1}(-1)^{l_1}(\alpha r_<)^{-l_1-1/2}$$
$$\times I_{l_1+1/2}(\alpha r_<)\mathscr{Y}_{l_1}^{m_1*}(\alpha\mathbf{r}_<)$$
$$\times(-\alpha)^{-l}\mathscr{Y}_l^m(\nabla_>)B_{-l_1,l_1}^{m_1}(\alpha,\mathbf{r}_>). \tag{6.12}$$

Now we only have to insert the relationship[44]

$$(-\alpha)^{\lambda_1}\mathscr{Y}_{\lambda_1}^{\mu_1}(\nabla)B_{-\lambda_2,\lambda_2}^{\mu_2}(\alpha,\mathbf{r})$$
$$= \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}}{}^{(2)}\langle\lambda\mu_1+\mu_2|\lambda_1\mu_1|\lambda_2\mu_2\rangle B_{-\lambda,\lambda}^{\mu_1+\mu_2}(\alpha,\mathbf{r}) \tag{6.13}$$

into Eq. (6.12) to obtain the addition theorem of the modified Helmholtz harmonics,

$$B_{-l,l}^m(\alpha,\mathbf{r}_< + \mathbf{r}_>)$$
$$= (2\pi)^{3/2}\sum_{l_1=0}^{\infty}\sum_{m_1=-l_1}^{l_1}(-1)^{l_1}(\alpha r_<)^{-l_1-1/2}$$
$$\times I_{l_1+1/2}(\alpha r_<)\mathscr{Y}_{l_1}^{m_1*}(\alpha\mathbf{r}_<)$$
$$\times\sum_{l_2=l_2^{\min}}^{l_2^{\max}}{}^{(2)}\langle l_2 m+m_1|lm|l_1m_1\rangle$$
$$\times B_{-l_2,l_2}^{m+m_1}(\alpha,\mathbf{r}_>). \tag{6.14}$$

This addition theorem can also be derived by differentiating Eq. (6.11) with respect to $\mathbf{r}_<$. We then obtain

$$B_{-l,l}^m(\alpha,\mathbf{r}_< + \mathbf{r}_>)$$
$$= (4\pi)^{1/2}(-\alpha)^{-l}\mathscr{Y}_l^m(\nabla_<)B_{0,0}^0(\alpha,\mathbf{r}_< + \mathbf{r}_>)$$
$$= (2\pi)^{3/2}\sum_{l_1=0}^{\infty}\sum_{m_1=-l_1}^{l_1}(-1)^{l_1}(-\alpha)^{-l}$$
$$\times\mathscr{Y}_l^m(\nabla_<)(\alpha r_<)^{-l_1-1/2}I_{l_1+1/2}(\alpha r_<)\mathscr{Y}_{l_1}^{m_1*}(\alpha\mathbf{r}_<)$$
$$\times B_{-l_1,l_1}^{m_1}(\alpha,\mathbf{r}_>). \tag{6.15}$$

To perform the remaining differentiation we use (MOS, p. 67)

$$\left(\frac{1}{z}\frac{d}{dz}\right)^m z^{-\nu}I_\nu(z) = z^{-\nu-m}I_{\nu+m}(z) \tag{6.16}$$

in connection with Eq. (3.1) to obtain

$$(\alpha r)^{-l-1/2}I_{l+1/2}(\alpha r)\mathscr{Y}_l^m(\alpha\mathbf{r})$$
$$= \alpha^{-l}\mathscr{Y}_l^m(\nabla)(\alpha r)^{-1/2}I_{1/2}(\alpha r). \tag{6.17}$$

If we now combine Eqs. (3.9), (6.2), and (6.17) we find

$$\alpha^{-\lambda_1}\mathscr{Y}_{\lambda_1}^{\mu_1}(\nabla)(\alpha r)^{-\lambda_2-1/2}I_{\lambda_2+1/2}(\alpha r)\mathscr{Y}_{\lambda_2}^{\mu_2}(\alpha\mathbf{r})$$
$$= \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}}{}^{(2)}\langle\lambda\mu_1+\mu_2|\lambda_1\mu_1|\lambda_2\mu_2\rangle(\alpha r)^{-\lambda-1/2}$$
$$\times I_{\lambda+1/2}(\alpha r)\mathscr{Y}_\lambda^{\mu_1+\mu_2}(\alpha\mathbf{r}). \tag{6.18}$$

If we insert this result into Eq. (6.15) we obtain a somewhat different representation of the addition theorem of the modified Helmholtz harmonics,

$$B_{-l,l}^m(\alpha,\mathbf{r}_< + \mathbf{r}_>)$$
$$= (2\pi)^{3/2}\sum_{l_1=0}^{\infty}\sum_{m_1=-l_1}^{l_1}B_{-l_1,l_1}^{m_1}(\alpha,\mathbf{r}_>)$$
$$\times\sum_{l_2=l_2^{\min}}^{l_2^{\max}}{}^{(2)Y}(-1)^{l_2}\langle l_1m_1|lm|l_2m_1-m\rangle$$
$$\times(\alpha r_<)^{-l_2-1/2}I_{l_2+1/2}(\alpha r_<)\mathscr{Y}_{l_2}^{m_1-m*}(\alpha\mathbf{r}_<). \tag{6.19}$$

To prove the equivalence of Eqs. (6.14) and (6.19) we only have to introduce the new summation variable $\mu_2 = m_1 - m$

into Eq. (6.19) and to change the order of the two l summations.

VII. SUMMARY AND CONCLUSIONS

In this article simple and unified derivations of the addition theorems of the irregular solid harmonics, the Helmholtz harmonics, and the modified Helmholtz harmonics are presented. Our derivations are based upon differential relationships of the following type:

$$F_l^m(\mathbf{r}) = \mathscr{Y}_l^m(\nabla)\phi(r). \tag{7.1}$$

Here, $F_l^m(\mathbf{r})$ is an irreducible spherical tensor, $\phi(r)$ is a function that only depends upon the distance r, i.e., a spherical tensor of rank zero, and $\mathscr{Y}_l^m(\nabla)$ is the spherical tensor gradient which is obtained from the regular solid harmonic $\mathscr{Y}_l^m(\mathbf{r})$ by replacing the Cartesian components of $\mathbf{r}$ by the Cartesian components of ∇.

The differential relationship (7.1) assumes a particularly simple form for the functions under consideration because in these cases the application of the spherical tensor gradient merely leads to a shift of angular momentum quantum numbers. If the spherical tensor gradient $\mathscr{Y}_l^m(\nabla)$ acts upon the Coulomb potential which is the irregular solid harmonic of rank zero we obtain $\mathscr{Z}_l^m(\mathbf{r})$. In the same way we obtain the (modified) Helmholtz harmonics of rank l by differentiating the (modified) Helmholtz harmonics of rank zero.

The remarkable differential properties of the irregular solid harmonics and the (modified) Helmholtz harmonics can be employed profitably for the derivation of addition theorems. We simply have to apply the spherical tensor gradient to the addition theorems of the Coulomb potential or the (modified) Helmholtz harmonics of rank zero and obtain the addition theorems of the anisotropic functions.

The idea of applying differentiation methods for the derivation of addition theorems is not at all new. Methods that are in some sense equivalent or closely related to our method, which is based upon the spherical tensor gradient and its tensor character, have already been employed by Hobson,[1] Rose,[2] Chiu,[3] Dahl and Barnett,[6] Steinborn and Ruedenberg,[8] Tough and Stone,[9] and Nozawa.[14] However, in the references cited the differential operators were applied in their Cartesian form and the tensorial nature of the differential operators was not exploited systematically. The direct application of differential operators, which involve differentiations with respect to x, y, and z to irreducible spherical tensors, leads to relatively complicated and sometimes rather messy expressions which cannot be manipulated easily. In our approach we utilize the fact that the application of the spherical tensor gradient to an irreducible spherical tensor leads to an angular momentum coupling. Therefore, only differentiations with respect to the radial variable r have to be done. It is the systematic exploitation of the tensor character of the differential operator $\mathscr{Y}_l^m(\nabla)$ which makes our derivation of the addition theorems almost trivial.

It should be noted that our method for the derivation of the addition theorem of an anisotropic function is not restricted to irregular solid harmonics and (modified) Helmholtz harmonics. If the addition theorem of an isotropic function $\phi(r)$ is known one only has to apply the spherical tensor gradient $\mathscr{Y}_l^m(\nabla)$ to it. According to Eq. (7.1) one then obtains the addition theorem of the anisotropic function $F_l^m(\mathbf{r})$.

ACKNOWLEDGMENT

E.O.S. thanks the Fonds der Chemischen Industrie for financial support.

[1] E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Chelsea, New York, 1965), Chap. IV.
[2] M. E. Rose, J. Math. Phys. (Cambridge, Mass.) **37**, 215 (1958).
[3] Y-N. Chiu, J. Math. Phys. **5**, 283 (1964).
[4] R. A. Sack, J. Math. Phys. **5**, 252 (1964).
[5] R. A. Sack, SIAM J. Math. Anal. **5**, 774 (1974).
[6] J. P. Dahl and M. P. Barnett, Mol. Phys. **9**, 175 (1965).
[7] E. O. Steinborn, Chem. Phys. Lett. **3**, 671 (1969).
[8] E. O. Steinborn and K. Ruedenberg, Adv. Quantum Chem. **7**, 1 (1973).
[9] R. J. A. Tough and A. J. Stone, J. Phys. A **10**, 1261 (1977).
[10] B. Friedman and J. Russek, Q. Appl. Math. **12**, 13 (1954).
[11] S. Stein, Q. Appl. Math. **19**, 15 (1961).
[12] O. R. Cruzan, Q. Appl. Math. **20**, 33 (1962).
[13] M. Danos and L. C. Maximon, J. Math. Phys. **6**, 766 (1965).
[14] R. Nozawa, J. Math. Phys. **7**, 1841 (1966).
[15] E. O. Steinborn and E. Filter, Int. J. Quantum Chem. Symp. **9**, 435 (1975).
[16] P. J. A. Buttle and L. J. B. Goldfarb, Nucl. Phys. **78**, 409 (1966).
[17] F. D. Santos, Nucl. Phys. A **212**, 341 (1973).
[18] E. G. P. Rowe, J. Math. Phys. **19**, 1962 (1978).
[19] B. F. Bayman, J. Math. Phys. **19**, 2558 (1978).
[20] S. N. Stuart, J. Austral. Math. Soc. Ser. B **22**, 368 (1981).
[21] A. W. Niukkanen, J. Math. Phys. **24**, 1989 (1983).
[22] A. W. Niukkanen, J. Math. Phys. **25**, 698 (1984).
[23] E. J. Weniger and E. O. Steinborn, J. Chem. Phys. **78**, 6121 (1983).
[24] E. J. Weniger and E. O. Steinborn, J. Math. Phys. **24**, 2553 (1983).
[25] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966). This reference will be denoted as MOS in the text.
[26] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge U.P., Cambridge, England, 1970), p. 48.
[27] L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics* (Addison–Wesley, Reading, MA, 1981), p. 86, Eq. (3.192)
[28] E. J. Weniger and E. O. Steinborn, Comput. Phys. Commun. **25**, 149 (1982), p. 151, Eq. (3.1).
[29] See Ref. 27, p. 312.
[30] See Ref. 23, p. 6126.
[31] E. W. Hobson, Proc. London Math. Soc. **24**, 55 (1892). See also Ref. 1, p. 127, Eq. (7).
[32] See Ref. 17, pp. 359–361, Appendix 2.
[33] See Ref. 19, p. 2559, Eq. (11).
[34] See Ref. 20, p. 373, Theorem 2.
[35] See Ref. 21, pp. 1989–1990, Eqs. (3)–(6), (12), and (13).
[36] See Ref. 24, p. 2555, Eq. (3.9).
[37] See Ref. 24, p. 2557, Eq. (3.29) and p. 2559, Eq. (4.24).
[38] See Ref. 21, p. 1990, Eq. (15) and Ref. 23, p. 6127, Eq. (4.24).
[39] B. K. Novosadov, Int. J. Quantum Chem. **24**, 1 (1983), p. 14, Eq. (65).
[40] See Ref. 39, p. 15, Eq. (66).
[41] See Ref. 23, p. 6127, Eqs. (4.21)–(4.28).
[42] E. Filter and E. O. Steinborn, Phys. Rev. A **18**, 1 (1978), p. 2, Eqs. (2.7) and (2.14).
[43] See Ref. 23, p. 6126, Eq. (4.12).
[44] See Ref. 24, p. 2562, Eq. (6.25).

On the body of supermanifolds

Roberto Catenacci[a)]
Dipartimento di Matematica, Via Strada Nuova 65, Pavia, Italy and Istituto Nazionale di Física Nucleare, Sezione di Pavia, Italy
Cesare Reina[a)]
Dipartimento di Fisica, Via Celoria 16, Milano, Italy
Paolo Teofilatto
Dipartimento di Fisica, Via Celoria 16, Milano, Italy

(Received 17 July 1984; accepted for publication 2 November 1984)

The problem of constructing the body of a G^∞ manifold is considered. It is shown that any such manifold is foliated, and the body is defined to be the space of the leaves of this foliation. Under certain regularity conditions on the foliation, the body is a smooth finite-dimensional real manifold.

I. INTRODUCTION

In supersymmetric field theories and supergravity one extends space-time to a "superspace," where four anticommuting coordinates appear, as well as the usual commuting ones. Superspace was introduced as a somewhat heuristic tool, which proved to be effective in handling very complex field theories, where one deals with commuting (bosonic) as well as anticommuting (fermionic) fields in supersymmetry.

Setting up such theories in a proper geometric framework was a bit of a problem, because one was forced to work either on a space (like superspace) where no proper differential calculus was established, or on a "supermanifold" (like those of Konstant and Batchelor) where all the fields are commuting (see, e.g., Ref. 1). The definition by Rogers[2] of G^∞ manifolds seems able to bypass both these shortcomings in physical application, because these are actually Banach manifolds, and the natural fields on them are Grassmann valued. So, anticommuting variables and fields can be treated on the same ground as the commuting ones.

After the introduction of G^∞ manifolds, some work has been devoted to the study of their relations with ordinary real differentiable manifolds. To understand these relations is crucial in view of possible applications to supersymmetric field theories and supergravity. Indeed the physical meaning of such theories can be understood only in terms of representation of the Poincaré group, that is, after the theory has been suitably reduced on ordinary space-time. It is therefore important to inquire to what extent G^∞ manifolds provide extension of space-time. Also from the purely mathematical point of view, it seems natural to inquire about the relations between the category of G^∞ manifolds and that of C^∞ manifolds.

This question was already considered by Rogers,[2] by introducing the notion of the "body" of a G^∞ manifold. This definition was stated in terms of local coordinates. After the work by Jadczyk and Pilch,[3] Percacci and Marchetti[4] and Hoyos *et al.*[5] came back to the problem, showing that the local definition by Rogers did not extend globally, unless the G^∞ structure was quite peculiar.

In this paper we came back to the problem of defining the body of a G^∞ manifold. Our approach is independent of charts, and it is based on the fact that any G^∞ manifold is foliated (as shown in Sec. II). Then the body arises as the quotient space of the G^∞ manifold by this foliation, which always exists as a topological space both in the finite- and the infinite-dimensional case. However, as is usual when taking the quotient by a foliation, the body does not admit a manifold structure, even at a topological level. A simple example of this phenomenon, relevant for the present case, is given in Sec. III. Finally we show that, under suitable regularity conditions on the foliation, a G^∞ manifold admits a smooth differentiable structure on its body. Examples of regular manifolds are the ρ manifolds of Ref. 5.

To avoid a long list of notation and definitions, we adopt the notation of Yadczyk and Pilch.[3] In particular, Q will usually denote a Banach–Grassmann algebra, and it is infinite dimensional over the reals. When we speak of finite-dimensional G^∞ manifolds, we mean a manifold which is finite dimensional over the reals; so in this cases Q will stand for a Grassmann algebra with L odd generators (i.e., we identify Q with B_L, according to the notations of Ref. 2).

II. FOLIATION AND EQUIVALENCE RELATIONS ON A G^∞ MANIFOLD

In this section we show that any G^∞ manifold $\tilde{X}$ is foliated. The basic fact is that one can define an involutive subbundle Σ of the tangent bundle $T\tilde{X}$, by considering tangent vectors whose components in any chart have vanishing real parts.

To be definite, let $(\tilde{U}_\alpha,\tilde{\varphi}_\alpha)$ be a G^∞ atlas for $\tilde{X}$, with coordinate maps $\tilde{\varphi}_\alpha: \tilde{U}_\alpha \to \tilde{A}_\alpha \subset Q^{m,n}$, and consider the map $\epsilon: Q^{m,n} \to R^m$ gotten by taking the real parts. The map $f_\alpha: \tilde{U}_\alpha \to \epsilon(Q_\alpha) \subset R^m$, given by $f_\alpha = \epsilon\cdot\tilde{\varphi}_\alpha$ is clearly a submersion. Its differential $df_\alpha: T\tilde{U}_\alpha \to R^m$ has a closed kernel. Then we set $\Sigma_\alpha = \ker df_\alpha$.

Clearly this local definition extends to a global one, because a tangent vector v at p belongs to $\Sigma_\alpha|_p$ if and only if its components in the chart $(\tilde{U}_\alpha, \tilde{\varphi}_\alpha)$ have vanishing real parts, a property which is clearly independent of charts. We call such a vector of type σ. More computatively if v^A are the

[a)] Also at Gruppo Nazionale di Fisica Matematica del CNR.

 0022-2488/85/040671-04$02.50

components of v in a chart $(\tilde U, \tilde\varphi)$ around p, its components $v^{A'}$ in another chart $(\tilde U', \tilde\varphi')$ around p are given by $v^{A'} = (d\tilde\psi)^{A}_{A}{}' v^{A}$, where $\tilde\psi = \tilde\varphi' \circ \tilde\varphi^{-1}$ is the coordinate transformation between the two charts. Since $\tilde\psi$ is G^∞, one has that $\epsilon(v^{A'}) = \epsilon(d\tilde\psi)^{A}_{A}{}'\epsilon(v^{A})$. Now $\epsilon(d\tilde\psi)^{A}_{A}{}'$ is invertible, and hence, being of type σ, is independent of charts. Accordingly two local distributions Σ_α and Σ_β agree at any $p \in U_\alpha \cap U_\beta$, i.e., $\Sigma_\alpha|_p = \Sigma_\beta|_p$.

From the construction above it is clear that Σ is integrable, its leaves being locally given by the equation $f_\alpha = \epsilon \cdot \tilde\varphi_\alpha = \text{const}$. Then we give the following definition.

Definition: Two points $p,q \in \tilde X$ are equivalent $(p \sim q)$ if they belong to the same connected maximal integral manifold (leaf) of Σ.

It is apparent that this is an equivalence relation independent of charts, which refines the definitions previously attempted in the literature.[2,4] To make contact with these, we notice that if p,q belong to the same chart $(\tilde U_\alpha, \tilde\varphi_\alpha)$, then $f_\alpha(p) = f_\alpha(q)$ (i.e., having the same real coordinates) implies that $p \sim q$. The converse is not true in general, because the intersection of a leaf of Σ with $\tilde U_\alpha$ may be not connected. Hence two points in $\tilde U_\alpha$, belonging to different connected components, may very well be equivalent, without having the same real coordinates.

One may argue that the present equivalence relation is in some sense unnatural, in that it seems better to start with the usual local relation[2,4,5] defined as follows. Whenever $p,q \in \tilde U_\alpha$, one sets $p \underset{\text{loc}}{\sim} q$ if and only if $f_\alpha(p) = f_\alpha(q)$. The trouble with this local relation is that (i) it is not independent of charts, and (ii) its global extension is not trivial. As to (i), it is sufficient to notice that if p,q belong to two disconnected components of the intersection of two charts $U_\alpha \cap U_\beta$, it may happen that $f_\alpha(p) = f_\alpha(q)$ but $f_\beta(p) \neq f_\beta(q)$. If $Q = B_L$ is finite dimensional, one can overcome this difficulty by taking a suitable refinement of the G^∞ atlas of $\tilde X$, as shown in Appendix A. Another possible way out is to assume that $\tilde X$ has a special G^∞ structure, i.e., that the images $\varphi_\alpha(U_\alpha \cap U_\beta) \subset Q^{m,n}$ are ϵ connected,[4] or that the manifold is as in Ref. 5.

Even when the relation $\underset{\text{loc}}{\sim}$ is suitably treated to yield independence of charts, one faces the fact that it is reflexive and symmetric, but fails, in general, to be transitive. To get, in any case, an equivalence relation, one follows the standard prescription of considering the transitive closure of the subset $R_{\text{loc}} = \{p,q/p \underset{\text{loc}}{\sim} q\}$, i.e., the minimal $R \subset \tilde X \times \tilde X$ which contains R_{loc} and is an equivalence relation $\underset{R}{\sim}$. In other words, one has that $p \underset{R}{\sim} q$ if and only if there exists a finite sequence $p_i, q_i (1 < i < N)$ of points such that $p_1 = p$, $q_N = q$, and $p_i \underset{\text{loc}}{\sim} q_i$. In this form this equivalence has been introduced in Ref. 5.

Notice that the existence of $\underset{R}{\sim}$ depends crucially on being independent of charts. In this case, we can show that the two equivalence $\underset{R}{\sim}$ and $\sim$ are actually the same.

Proposition: Two points $p,q \in \tilde X$ are R equivalent if and only if they belong to the same leaf of Σ.

Proof: Any curve $p(t) \subset \tilde X$ of R-equivalent points has tangent vectors of type σ. Hence $\forall t, p(t)$ belongs to the same leaf of Σ. Conversely if p,q belong to the same leaf of Σ, then there exists a compact curve $p(t)$ of type σ connecting them. Then we can choose points $p_i = p(t_i)$ and $q_i = p(t_i')$ such that $p_i \underset{\text{loc}}{\sim} q_i$ and $p_0 = p$, $q_N = q$.

Remark: As already mentioned, the proof above requires that $\underset{R}{\sim}$ be independent of charts. If not, we stress that $\underset{R}{\sim}$ cannot be consistently defined, while our equivalence relation $\sim$ exists in any case.

III. THE BODY OF A G^∞ MANIFOLD

Definition: The body X of a G^∞ manifold $\tilde X$ is the space of the leaves of Σ on $\tilde X$.

When R equivalence exists on $\tilde X$, one can as well say that $X = \tilde X / \underset{R}{\sim}$ is the set of R-equivalence classes of points in $\tilde X$.

We give $\tilde X/\sim$ the quotient topology, thus yielding that the canonical projection π: $\tilde X \to X$ is continuous and open. The question is now if X can be given a manifold structure. As is well known, the answer to this question for a generic foliated manifold is negative. In any case, to build a manifold structure on the space of leaves, one has at least to assume that the foliation was regular (see, e.g., Ref. 6).

Thanks to the properties of G^∞ manifolds, we can say a bit more in the present case. First notice that "concrete" G^∞ manifolds are built gluing together charts, and giving them the topology which makes the coordinate maps homeomorphisms. Now, around any $p \in \tilde X$ one can define a cubic and flat coordinate patch $(\tilde U_p, \tilde\varphi_p)$ centered at p as follows.

Let $(\tilde U, \tilde\varphi)$ be a chart containing p; we set $\tilde\varphi_p = \tilde\varphi - \tilde\varphi(p)$ so that $\tilde\varphi_p(p) = 0 \in Q^{m,n}$. If $(x^1,\dots,x^m) = \epsilon \cdot \tilde\varphi_p(q) \in R^m$ denote the real coordinates of $q \in \tilde U$, we consider a cube $c \subset R^m$, of width $2a$, given by $|x^i| < a$. Then let $\tilde U_p = \tilde\varphi_p^{-1}[\epsilon^{-1}(c) \cap \tilde\varphi_p(\tilde U)]$. From Sec. II it follows that the leaves of Σ in $\tilde U_p$ are parametrized by the real coordinates $(x^1,\dots,x^m) = \tilde\varphi_p(q) \in \tilde U_p$, that is, the coordinate patch $(\tilde U_p, \tilde\varphi_p)$ is "flat."

The trouble here is that the correspondence between leaves of Σ in $\tilde U_p$ and the real coordinates $(x^1,\dots,x^m)$ is not a bijection, that is in general one has no maps φ_p making the following diagram commutative:

$$\begin{array}{ccc} \tilde U_p & \xrightarrow{\ \tilde\varphi_p\ } & \tilde A \subset Q^{m,n} \\ \pi \downarrow & & \downarrow \epsilon \\ U_p & \xrightarrow[\ \varphi_p\ ]{} & A \subset R^m . \end{array} \qquad (3.1)$$

If, on the contrary, for any p one has a patch $(\tilde U_p, \tilde\varphi_p)$ and a map φ_p such that the diagram above commutes, we say that the Σ foliation is regular. A G^∞ manifold whose Σ foliation is regular will be called regular itself.

To see that regularity is missing in general consider the following example.

Example: We construct a torus over $Q_1^{1,1}$. If $x + \theta y \in Q_1$

with $\theta^2 = 0$, then we have coordinates $(x,y\theta) \in Q_1^{1,1}$. Consider the intersection of the two strips

$$\text{(I)}\ \alpha(x-1) \leqslant y \leqslant \alpha(x+1),\ (\alpha \in R^+)$$

$$\text{(II)}\ -\alpha(x+1) \leqslant y \leqslant -\alpha(x-1)$$

and identify the boundaries of (I) by $(a + b\theta) \mapsto (a+1,(b+\alpha)\theta)$ and of (II) by $(a,b\theta) \to (a-1,(b+\alpha)\theta)$. It is clear that this operation is G^∞, and that the resulting torus $\tilde{X}$ has a G^∞ structure. Now if α is not rational, the leaves of Σ are dense on $\tilde{X}$, and, therefore, the Σ foliation is not regular. From this example we see that regularity is by no means a local property. In other words the existence of a map φ in diagram (3.1) crucially depends on the global behavior of the leaves of Σ. Since φ_p is lacking, one has no coordinates on $\tilde{X}/\sim$, that is, the body of $\tilde{X}$ is not even a topological manifold.

Although regularity will be difficult to check in a generic case, one can give sufficient conditions. It is easy to show that the ρ supermanifolds of Ref. 5 have regular foliation. Conversely if the foliation was regular, than the flat coordinate charts are ϵ connected and, the diagram being commutative, it yields a ρ supermanifold structure on $\tilde{X}$. Examples of regular supermanifolds are the G^∞ extension of any ordinary C^∞ space-time constructed by Bonora, Pasti, and Tonin.[7]

Whenever $\tilde{X}$ is regular, its body is obviously a topological manifold. We can also prove the following theorem.

Theorem: Let $\tilde{X}$ be a regular G^∞ manifold. Then its body X is a C^∞ manifold.

Proof: Since $\tilde{X}$ is regular, one has an atlas $\{(\tilde{U}_p,\tilde{\varphi}_p)\}$ and bijections φ_p making the diagram (3.1) commute. Then one has bijections $\varphi_q\varphi_p^{-1}:A \to B$, $A,B \subset \mathbb{R}^n$ such that the diagram

$$\begin{tikzcd}
\tilde{U}_p \cap \tilde{U}_q \arrow[rr, "\tilde{\varphi}_p"] \arrow[dr, "\tilde{\varphi}_p"'] \arrow[dd, "\pi"'] & & \tilde{A} \arrow[dl, "\tilde{\varphi}_q \cdot \tilde{\varphi}_p^{-1}"] \arrow[dd, "\epsilon"] \\
& \tilde{B} \arrow[dd, "\epsilon"] & \\
U_p \cap U_q \arrow[rr, "\varphi_p"] \arrow[dr, "\varphi_q"'] & & A \arrow[dl, "\varphi_q \varphi_p^{-1}"] \\
& B &
\end{tikzcd}$$

commutes. Now the transition functions $\varphi_p \cdot \varphi_q^{-1}$ are clearly local homeomorphisms. They are also C^∞ diffeomorphisms. Indeed we can represent them by $\varphi_p \cdot \varphi_q^{-1} = \epsilon \cdot \tilde{\varphi}_p \cdot \tilde{\varphi}_q^{-1} \cdot \sigma$, where σ: $A \to \tilde{A}$ is a C^∞ section of $\tilde{A} \to \epsilon(\tilde{A}) = A$. Then $\varphi_p \cdot \varphi_q^{-1}$ arises as a composition of C^∞ maps, and hence it is C^∞. The same applies to the inverse $\varphi_q \cdot \varphi_p^{-1}$. Hence X is a C^∞ manifold.

Next, by similar arguments, one proves that if $\tilde{\psi}:\tilde{X} \to \tilde{X}$ is a G^∞ diffeomorphism, then there exists a unique $\psi:X \to X$ which is a C^∞ diffeomorphism and such that the diagram

$$\begin{tikzcd}
\tilde{X} \arrow[r, "\tilde{\psi}"] \arrow[d, "\pi"'] & \tilde{X} \arrow[d, "\pi"] \\
X \arrow[r, "\psi"] & X
\end{tikzcd}$$

is commutative. So the body $X = \pi(\tilde{X})$ is unique up to diffeomorphisms. More precisely one may say that π is a functor from the category of regular G^∞ manifolds with G^∞ diffeomorphisms to the category of C^∞ manifolds with C^∞ diffeomorphisms.

ACKNOWLEDGMENTS

G. Canuto, M. Cornalba, and R. Percacci are warmly thanked for several discussions and suggestions.

We thank also the referee of this paper for suggesting a comparison between our approach and Ref. 5.

We heard also from the referee that Boyer and Gitler[8] discussed the present problem very much on the same line as ours.

This work was partially supported by Progetto Nazionale "Geometria e Fisica" M. P. I.

APPENDIX A: THE EXISTENCE OF A "GOOD" SUPERATLAS

In this appendix we show the existence, in the finite-dimensional case, of a superatlas in which the $\underset{R}{\sim}$ relation is chart independent.

The existence of a "good" atlas is a consequence of a well-known result in the theory of ordinary differentiable manifold: Let X be a paracompact differentiable manifold of dim n. Then every open covering $\{V_\alpha\}$ of X has an open refinement $\{V_i\}$ such that (i) each V_i has compact closure; (ii) $\{V_i\}$ is locally finite; and (iii) any nonempty finite intersection of the V_i's is diffeomorphic to an open ball of R^n. Now, if we have a G^∞ supermanifold $\tilde{X}$ modeled on $B_L^{m,n}$ with a given superatlas $\{\tilde{V}_\alpha, \tilde{\varphi}_\alpha\}$, we can consider it as a real C^∞ manifold of dim $N = 2^{L-1}(m+n)$. In fact, every G^∞ manifold is a Banach manifold C^∞, and every G^∞ map between supermanifolds is also a C^∞ map, and hence there exists a forgetful functor

$$F{:}G^\infty \text{ supermanifolds} \to C^\infty \text{ manifolds.}$$

The identification of $B_L^{m,n}$ with $R^{2^{L-1}(m+n)}$ is as follows: We take a basis of B_L, $\{\beta_\mu\}$, and set $Z^A = Z^{A\mu}\beta_\mu$. We then define a map $f{:}\tilde{X} \to F\tilde{X}$, which, on the underlying topological spaces, is the identity and on suitable atlases $\{\tilde{V}_\alpha, \tilde{\varphi}_\alpha\}$ of $\tilde{X}$ and $\{\tilde{V}_\alpha, \psi_\alpha\}$ of $F\tilde{X}$ has the representation

$$\tilde{\varphi}_\alpha \cdot f \cdot \psi_\alpha^{-1}(Z^A) = Z^{A\mu}.$$

Now we can apply the proposition above to the manifold $F\tilde{X}$ getting the "good" covering $\{\tilde{V}_i\}$. We can then transfer the sets $\{\tilde{V}_i\}$ on $\tilde{X}$ and we have "good" superatlas $\{\tilde{V}_i, \tilde{\varphi}_\alpha|_{V_i} = \tilde{\varphi}_i\}$, where α corresponds to a $\tilde{U}_\alpha$ of the original superatlas containing $\tilde{V}_i$. In fact, as $\tilde{V}_i \cap \tilde{V}_j$ is connected and the $\tilde{\varphi}_i$'s are homeomorphisms, $\tilde{\varphi}_i(\tilde{V}_i \cap \tilde{V}_j)$ and $\tilde{\varphi}_j(\tilde{V}_i \cap \tilde{V}_j)$ are connected open sets in $B_L^{m,n}$.

We recall now the following proposition (see Rogers[2]).

Proposition: Let U be open and connected in $B_L^{m,n}$ and let $f \in G^\infty(U)$. Then there exists a unique $f \in G^\infty(\epsilon^{-1}(\epsilon(U)))$ such that $f'|_U = f$. It follows from the properties of Taylor series and the fact that ϵ is an algebra homomorphism that if $x,y \in U$ (U open and connected in $B_L^{m,n}$) with $\epsilon(x) = \epsilon(y)$, then

$\epsilon f(x) = \epsilon f(y)$. We have then proved that the transition function $\tilde{\psi}_{ij}$ of the "good" atlas are "body preserving," i.e., if $x, y \in \tilde{\varphi}_j(V_i \cap V_j)$ are such that $\epsilon(x) = \epsilon(y)$ then $\epsilon \tilde{\psi}_{ij}(x) = \epsilon \tilde{\psi}_{ij}(y)$. So the relation $\sim_R$ is chart independent.

[1]J. Drell and L. Smolin, Commun. Math. Phys. **66**, 197 (1979).
[2]A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[3]A. Jadczyk and K. Pilch, Commun. Math. Phys. **78**, 373 (1981).
[4]R. Percacci and M. Marchetti, "Categorial properties of supermanifolds," Preprint ISAS 3/81/E. P. Trieste, 1981.
[5]J. Hoyos, M. Quiros, J. Ramirez Mittelbrunn, and F. J. de Urries, J. Math. Phys. **25**, 847 (1984).
[6]R. Abraham, J. E. Marsden, and T. Ratiu, *Manifolds and Tensor Analysis and Applications* (Addison–Wesley, Reading, MA, 1983).
[7]L. Bonora, P. Pasti, and M. Tonin, J. Math. Phys. **23**, 839 (1982).
[8]C. Boyer and S. Gitler, Contemporary Math. **12**, 53 (1982).

Harmonic analysis on the Euclidean group in three-space

Jung Sik Rno
Departments of Mathematics and Physics, University of Cincinnati—RWC Campus, Cincinnati, Ohio 45236

(Received 13 September 1984; accepted for publication 1 November 1984)

We develop the extensive harmonic analysis on the universal covering group of the Euclidean group in three-space.

I. INTRODUCTION

This paper is a sequel to our previous work[1] (called hereafter Paper 1), which dealt with the explicit computation of Clebsch–Gordan (CG) coefficients of the (simply connected) twofold universal covering group of the Euclidean group in three-space $\widetilde{E(3)}$ and special functions associated with group representations and CG coefficients of $\widetilde{E(3)}$. Miller[2] initiated the harmonic analysis on $\widetilde{E(3)}$, although he did not identify it as such. The purpose of this paper is to complete his work by performing the formal harmonic analysis on $\widetilde{E(3)}$ through a new systematic, refined, and rigorous approach.

Physicists have recently been utilizing harmonic expansions to study dimensional reduction. Their interests lie, so far, in harmonic expansions[3–9] on coset spaces for the fields that occur in higher-(larger than four-) dimensional theories, Kaluza–Klein theories and supergravity theories. It is noteworthy that harmonic analysis on a group yields a harmonic expansion on a coset space. This is no surprise, since harmonic analysis requires reduction of group representations.

We will set up our groundwork in Secs. II and III. In Sec. II we give the necessary résumé of $\widetilde{E(3)}$. Section III provides a useful outline regarding how to construct the unitary irreducible representation (UIR) of $\widetilde{E(3)}$. The full scale of harmonic analysis on $\widetilde{E(3)}$ will be carried out in Sec. IV.

II. THE GROUP E(3)

In this paper we are concerned with the (simply connected) twofold covering group $\widetilde{E(3)}$ of the proper Euclidean group E(3) [$\widetilde{E(3)}$ is often named E(3) throughout this paper]. It is the semidirect product $R^3 \times_\eta \mathrm{SU}(2)$ relative to the homomorphism η of SU(2) into the group of automorphisms of R^3. The matrices $\pm A \in \mathrm{SU}(2)$ determine the same rotation $\eta(A)$ given by

$$A(r\cdot\sigma)A^{-1} = (\eta(A)r)\cdot\sigma, \tag{1}$$

where σ stands for the Pauli matrices

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2}$$

The multiplication law for $\widetilde{E(3)}$ is

$$\{r_1, A_1\}\{r_2, A_2\} = \{r_1 + \eta(A_1)r_2, A_1A_2\}. \tag{3}$$

In the following we shall usually write Ar instead of $\eta(A)r$. If

$$A = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}$$

with $a\bar{a} + b\bar{b} = 1$, then $\eta(A)$ has the expression[10]

$$\eta(A) = \begin{pmatrix} \frac{1}{2}(a^2 - b^2 + \bar{a}^2 - \bar{b}^2) & \frac{1}{2}[i(\bar{a}^2 + \bar{b}^2 - a^2 - b^2)] & \bar{a}\bar{b} + ab \\ \frac{1}{2}[i(a^2 - b^2 - \bar{a}^2 + \bar{b}^2)] & \frac{1}{2}(\bar{a}^2 + \bar{b}^2 + a^2 + b^2) & i(-\bar{a}\bar{b} + ab) \\ -(\bar{a}b + a\bar{b}) & i(-\bar{a}b + a\bar{b}) & a\bar{a} - b\bar{b} \end{pmatrix}. \tag{4}$$

III. THE UIR OF E(3)

The dual group $\hat{R}^3$ of R^3 consists of the unitary characters $\chi^p : a \to e^{ip\cdot a}$ for $a \in R^3$. We identify $\hat{R}^3$ with the momentum space P^3. Then the group SU(2) acts on P^3 as well as on R^3. The SU(2) orbit of a given $p \in P^3$ are spheres $\Omega_\rho = \{p \in P^3 : \|p\| = \rho \geqslant 0\}$. Thus we can characterize the partition of P^3 into orbits by choosing the following set K representing the standard momentum $\dot{p}$:

$$P^3 = \bigcup_{\dot{p} \in K} \Omega(\dot{p}) \equiv \bigcup_{\rho \geqslant 0} \Omega_\rho, \tag{5}$$

where

$$K = \{\dot{p} = (0,0,\rho) : \rho \geqslant 0\}.$$

Hence there are only two stability groups (little groups),

$$G_{\dot{p}} = \mathrm{SU}(2), \quad \text{for } \dot{p} \in \Omega_0,$$
$$G_{\dot{p}} = \widetilde{\mathrm{SO}(2)}, \quad \text{for } \dot{p} \in \Omega_\rho (\rho > 0), \tag{6}$$

where $\widetilde{\mathrm{SO}(2)}$ is the twofold covering group of SO(2), the group of rotations around the z axis, and it is isomorphic to the multiplicative group of the complex numbers $e^{i\psi/2}$, $0 \leqslant \psi < 4\pi$. Thus its UIR's are one-dimensional and of the form

$$\Gamma^s\left(\begin{bmatrix} e^{i\psi/2} & 0 \\ 0 & e^{-i\psi/2} \end{bmatrix}\right) = e^{is\psi}, \tag{7}$$

where $2s = 0, \pm 1, \pm 2, \ldots$.

The UIR's associated with the trivial orbit Ω_0 are of no interest in the present work. The UIR's (ρ, s) of $\widetilde{E(3)}$ associated with an orbit Ω_ρ $(\rho > 0)$ are given by

$$[U^{\rho,s}(a, A)f](p) = e^{ip\cdot a}(\Gamma^s \uparrow \mathrm{SU}(2))(p, A)f(A^{-1}p), \tag{8}$$

where $\uparrow$ denotes "induced."

The carrier space of (ρ, s) is $H(\rho, s)$, the Hilbert space of Lebesgue square integrable functions on the manifolds Ω_ρ with inner product

$$\langle f, g \rangle = \int_{\Omega_\rho} \overline{f(p)} \cdot g(p) dw(p), \quad f, g \in \mathrm{H}(\rho, s), \tag{9}$$

 0022-2488/85/040675-03$02.50

where $dw(p) = \sin\theta\, d\theta\, d\varphi$ for $p = (\rho\sin\theta\cos\varphi, \rho\sin\theta\sin\varphi, \rho\cos\theta)\in\Omega_\rho$. We recall that the set K in (5) meets each orbit just once, and it is certainly a Borel set in P^3. Thus $\widetilde{E(3)}$ is a regular semidirect product. Therefore one can conclude[11] that (i) every UIR of $\widetilde{E(3)}$ which acts nontrivially on the translation subgroup is unitarily equivalent to a representation of the form (8) for some choice of constants ρ and s and (ii) two such representations U_1 and U_2 are unitarily equivalent if and only if $\rho_1 = \rho_2$ and $s_1 = s_2$. Further explicit expression of (8) was given in Paper 1.

IV. HARMONIC ANALYSIS ON E(3)

We aim to complete the proof of the Plancherel theorem. Thus our task is the reduction of the regular representation of $\widetilde{E(3)}$. We first prove the following lemma.

Lemma:

$$U^{\rho,s}|_{SU(2)} \cong \sum_{l=|s|}^{\infty} D(l),$$

where $D(l)$ is a UIR of SU(2) and $\cong$ means an equivalence of representations.

Proof: Let

$$U^{\rho,s}|_{SU(2)} \cong \sum_{2l=0}^{\infty} n(l,s) D(l),$$

where $n(l,s)$ is a multiplicity of $D(l)$.

By (8),

$$U^{\rho,s}|_{SU(2)} \equiv \Gamma^{s}_{\widetilde{SO(2)}} \downarrow SU(2) \cong \sum_{2l=0}^{\infty} n(l,s) D(l).$$

Since $SU(2)\cap\widetilde{SO(2)} = \widetilde{SO(2)}$, the Frobenius reciprocity theorem[12] gives rise to

$$D(l)|_{\widetilde{SO(2)}} = \sum_{s=-l}^{l} n(l,s)\Gamma^{s}_{\widetilde{SO(2)}}.$$

Thus we have

$$n(l,s) = \begin{cases} 1, & \text{if} |s| \leqslant l, \\ 0, & \text{if } |s| > l. \end{cases} \qquad \blacksquare$$

The above lemma is known.[10,13] Our elegant proof is, however, new.

We consider the Hilbert space $H \equiv L^2(\widetilde{E(3)})$, the elements of which are complex functions, Lebesgue integrable with respect to the Haar measure $dA\, d^3r$. The inner product is given by

$$\langle f, g\rangle = \int_{E(3)} f(r,A) g(r,A) dA\, d^3r, \quad f,g\in H. \tag{10}$$

We define the unitary (left) regular representation

$$[U(r_0, A_0)f](r,A) = f(A_0^{-1}(r - r_0), A_0^{-1}A). \tag{11}$$

We shall explicitly decompose this representation into irreducible components.

We now define the Fourier transform $\mathscr{F}$ for $f\in H$

$$(\mathscr{F}\cdot f)(p,a) \equiv \hat{f}(p,A) = \frac{1}{(2\pi)^{3/2}}\int_{R^3} e^{ip\cdot r} f(r,A) d^3r, \tag{12}$$

where $p\cdot r$ is the Euclidean scalar product. By the classical Fourier theorem, we have

$$(\mathscr{F}^{-1}\cdot f)(r,A) \equiv f(r,A)$$
$$= \frac{1}{(2\pi)^{3/2}}\int_{P^3} e^{-ir\cdot p}\hat{f}(p,A) d^3p. \tag{13}$$

Then we obtain

$$\int_R\int_{SU(2)} |f(r,A)|^2 d^3r\, dA = \int_{P^3}\int_{SU(2)} |\hat{f}(p,A)|^2 d^3p\, dA. \tag{14}$$

Therefore (13) defines an isometric mapping into $\hat{H} = L^2(P^3\times_\eta SU(2))$, the Hilbert space of complex functions, Lebesgue square integrable with respect to the Haar measure $d^3p\, dA$.

The regular representation U in (11) induces in $\hat{H}$

$$[\hat{U}(r_0, A_0)\hat{f}](p,A) = e^{ir_0\cdot p}\hat{f}(A_0^{-1}p, A_0^{-1}A). \tag{15}$$

We recall, from Sec. III, that Ω_ρ denotes the SU(2) orbit of the standard momentum $\dot{p} = (0,0,\rho)$:

$$\Omega_\rho = \{ p\in P^3 \,||p|| = \rho\}.$$

We set

$$\hat{f}^\rho(p,A) \equiv \hat{f}(p,A), \quad \text{for } p\in\Omega_\rho. \tag{16}$$

By (15) we now have

$$[\hat{U}(r_0, A_0)\hat{f}^\rho](p,A) = e^{ir_0\cdot p}\hat{f}^\rho(A_0^{-1}p, A_0^{-1}A). \tag{17}$$

Notice that U and $\hat{U}$ are unitarily equivalent via $U = \mathscr{F}^{-1}\hat{U}\mathscr{F}$. Obviously, we have

$$\int_{p^3}\int_{SU(2)} |\hat{f}(p,A)|^2 d^3p\, dA$$
$$= \int_0^{\oplus\,\infty} \rho^2 d\rho \int_{\Omega_\rho} dw(p)\int_{SU(2)} |\hat{f}(p,A)|^2 dA, \tag{18}$$

where $dw(p)$ and dA are Haar measures for Ω_ρ and SU(2), respectively. This shows that the representation U defined by (15) is a direct integral of the representations defined by (17). Our problem will be solved if we further reduce these to simpler representations. We need to recall, from Paper 1, that the matrix element $A_{p\to\dot{p}}\in SU(2)$ has the property that satisfies $A_{p\to\dot{p}}\cdot\dot{p} = p$. Using this fact, we can write[14]

$$\hat{f}(p,A) = \hat{f}^\rho(A_{p\to\dot{p}}\dot{p}, A_{p\to\dot{p}}\cdot A_{p\to\dot{p}}^{-1}\cdot A)$$
$$= \sum_{2u=0}^{\infty}\sum_{s=-u}^{u}\sum_{m=-u}^{u} \hat{f}^{\rho,u}_{s,m}(\dot{p}, A_{p\to\dot{p}}^{-1}\cdot A) \times T^u_{s,m}(A_{p\to\dot{p}}), \tag{19}$$

where $T^u_{s,m}$ is the matrix element of SU(2), which is given in the Wigner D function. [See Eq. (34) and Appendix A in Paper 1.] Since $\hat{f}^{\rho,u}_{s,m}(\dot{p}, A_{p\to\dot{p}}^{-1}A)$ is also square integrable on SU(2), we can further write

$$\hat{f}^{\rho,u}_{s,m}(\dot{p}, A_{p\to\dot{p}}^{-1}A) = \sum_{2v=0}^{\infty}\sum_{t=-v}^{v}\sum_{n=-v}^{v} \hat{f}^{\rho,u,v}_{s,m;t,n}\, T^v_{t,n}(A_{p\to\dot{p}}^{-1}A), \tag{20}$$

where the $\hat{f}^{\rho,u,v}_{s,m;t,n}$ are functions of ρ. Choose $C\in SU(2)$ such that $C\dot{p} = \dot{p}$. We know $T^u_{s,m}(C) = \delta_{s,m}e^{-is\varphi}$ for some φ such that $0\leqslant\varphi<4\pi$. We then obtain

$$\hat{f}^{\rho,u}_{s,m}(\dot{p}, C^{-1}A) = e^{is\varphi}\hat{f}^{\rho,u}_{s,m}(\dot{p}, A). \tag{21}$$

Comparing (21) with (20) we can conclude[2] that

$$\hat{f}^{\rho,u,v}_{s,m;t,n} = 0, \quad \text{unless } t = s. \tag{22}$$

Furthermore, $u + v$ must be an integer. Substituting (20) and (22) into (19), we see that the functions $\hat{f}^\rho(p,A)$ have the form

$$\hat{f}^{\rho}(p,A) = \sum_{2u=0}^{\infty} \sum_{2v=0}^{\infty} \sum_{s=-u}^{u} \sum_{m=-u}^{u} \sum_{n=-v}^{v} \hat{f}^{\rho,u,v}_{s,m;s,n} \times T^{u}_{s,m}(A_{p\to\hat{p}}) T^{v}_{s,n}(A^{-1}_{p\to\hat{p}}A). \tag{23}$$

The quantity on the right side of (23) transforms properly under SU(2). In particular, for fixed s, the transformation under the little group $G = \widetilde{\mathrm{SO}(2)}$ shows that $\hat{U}(r_0, A)$ in (17) is unitarily equivalent to (ρ,s) when the carrier space is restricted to $H(\rho,s)$. Thus, by our lemma, we can write (23) as

$$\hat{f}^{\rho}(p,A) = \sum_{2s=-\infty}^{\infty} \sum_{u=|s|}^{\infty} \sum_{v=|s|} \sum_{m=-u}^{u} \sum_{n=-v}^{v} \hat{f}^{\rho,u,v}_{s,m;s,n} \times T^{u}_{s,m}(A_{p\to\hat{p}}) T^{v}_{s,n}(A^{-1}_{p\to\hat{p}}A). \tag{24}$$

We shall write $\hat{f}^{\rho,u,v}_{s,m;s,n}$ explicitly. Using (20), (19), and (12) successively we can obtain

$$\hat{f}^{\rho,u,v}_{s,m;s,n} = \int_{R^3} d^3r \int_{\mathrm{SU}(2)} dA\ \overline{\{u,m|\rho,s|v,n\}}(r,A) f(r,A), \tag{25}$$

where $\{u,m|\rho,s|v,n\}(r,A)$ is the matrix element of the operator $U^{\rho,s}(r,A)$ with respect to the orthonormal basis of $H(\rho,s)$. We refer its details to Paper 1. The bar on the matrix element signifies its complex conjugate. Thus $\hat{f}^{\rho,u,v}_{s,m;s,n}$ is exactly a matrix Fourier coefficient of f. Also making use of those appropriate inversion transforms, we can derive without any difficulty

$$f(r,A) = \sum_{2s=-\infty}^{\infty} \sum_{u=|s|}^{\infty} \sum_{v=|s|}^{\infty} \sum_{m=-u}^{u} \sum_{n=-v}^{v} \int_0^{\infty} \rho^2 d\rho \times \hat{f}^{\rho,u,v}_{s,m;s,n} \{u,m|\rho,s|v,n\}(r,A), \tag{26}$$

where we have used, from Paper 1, the orthogonality relation for the matrix element

$$\int_{R^3} d^3r \int_{\mathrm{SU}(2)} dA\ \overline{\{u_1,m_1,|\rho_1,s_1|v_1,n_1\}(r,A)} \times \{u_2,m_2|\rho_2,s_2|v_2,n_2\}(r,A) = \frac{4\pi}{\rho_1^2}\delta(\rho_1-\rho_2)\cdot\delta_{s_1,s_2}\cdot\delta_{m_1,m_2}\cdot\delta_{u_1,u_2}\cdot\delta_{n_1,n_2}\cdot\delta_{v_1,v_2}. \tag{27}$$

From (14), (18), and (23) we can finally obtain the Plancherel formula for $\widetilde{\mathrm{E}(3)}$

$$\int_{\widetilde{\mathrm{E}(3)}} |f(r,A)|^2 d^3r\, dA = \sum_{2s=-\infty}^{\infty} \sum_{u=|s|}^{\infty} \sum_{v=|s|}^{\infty} \sum_{m=-u}^{u} \sum_{n=-v}^{v} \int_0^{\oplus\infty} |\hat{f}^{\rho,u,v}_{s,m;s,n}|^2 \rho^2 d\rho. \tag{28}$$

Notice that the variable ρ and indices s, u, v, m, and n represent the dual space of $\widetilde{\mathrm{E}(3)}$, which is usually denoted as $\widehat{\widetilde{\mathrm{E}(3)}}$. Thus we have proved the following.

Plancherel theorem: There exists a Fourier–Plancherel transform, which is an isometric mapping between $L^2(\widetilde{\mathrm{E}(3)})$ and $L^2(\widehat{\widetilde{\mathrm{E}(3)}})$ with a natural measure as it appears in (28). This mapping is given by (25) and (26).

We remark that the Plancherel formula (28) can be expressed as

$$||f||_2^2 = \int_0^{\oplus\infty} \rho^2 d\rho \sum_{-2s=-\infty}^{\infty} \oplus\ |||\hat{f}(\rho,s)|||^2,$$

where $|||\hat{f}(\rho,s)|||^2 = \mathrm{trace}\{\hat{f}(\rho,s), \hat{f}^*(\rho,s)\}$ is the Hilbert–Schmidt norm of $\hat{f}(\rho,s)$ whose matrix elements are $\hat{f}^{\rho,u,v}_{s,m;s,n}$. The above type of theorem, in general, is valid for the locally compact, separable, unimodulator, and postliminar group.[15]

Finally, we wish to comment, from a physicist's viewpoint of dimensional reduction, that (26) is a harmonic expansion of a scalar field in six $(=3+2^2-1)$-dimensional space $\hat{R}^3 \times_\eta \mathrm{SU}(2)$ expressed on the coset space $\hat{R}^3 \times_\eta \mathrm{SU}(2)/\mathrm{SU}(2)$, which is three-dimensional.

ACKNOWLEDGMENTS

The author would like to thank Professor Willard Miller, Jr. for helpful discussion. He also wishes to thank Professor Louis Witten and Professor Freydoon Mansouri for discussion on dimensional reduction. It is the author's pleasure to acknowledge the superb secretarial help of Mrs. Sylvia Sprague.

This research has been supported in part by Raymond Walters College, University of Cincinnati.

[1] J. S. Rno, J. Math. Phys. **15**, 2042 (1974).
[2] W. Miller, Jr., Commun. Pure Appl. Math. **17**, 527 (1964).
[3] A. Salam and J. Strathdee, Ann. Phys. **141**, 316 (1982).
[4] W. Mecklenburg, "The Kaluza–Klein idea: Status and Prospects," I.C.T.P., Trieste, Italy preprint IC/83/32, 1983.
[5] K. S. Viswanathan "Dimensional Reduction and Harmonic Expansions on Coset Spaces," Simon Fraser University preprint, 1983.
[6] S. Tanaka, Prog. Theor. Phys. **70**, 563 (1983).
[7] D. Toms, "Introduction to Kaluza–Klein Theories," University of Wisconsin—Milwaukee preprint, 1983.
[8] S. P. de Alwis and I. G. Koh, "Harmonic Expansions and Quasi-Riemannian Geometry in Kaluza–Klein Theories," Imperial College, London preprint, 1984.
[9] G. Fogleman and K. S. Viswanathan, "Harmonic Expansions for $(4+K)$ Dimensional Rarita–Schwinger Fields on Coset Spaces and Effective Lagrangian in Four Dimensions," Simon Fraser University preprint, 1984.
[10] W. Miller, Jr., *Lie Theory and Special Functions* (Academic, New York, 1968).
[11] G. W. Mackey, in *Functional Analysis and Related Fields*, edited by F. E. Browder (Springer, Berlin, 1970).
[12] G. Warner, *Harmonic Analysis on Semi-Simple Lie Groups* I (Springer, Berlin, 1972).
[13] J. D. Talman, *Special Functions* (Benjamin, New York, 1968).
[14] M. A. Naimark, *Linear Representations of the Lorentz Group*, English translation (MacMillan, New York, 1964).
[15] J. Diximier, *Les C*-Algèbres et Leurs Représentations* (Gauthier–Villars, Paris, 1969).

Asymptotic eigenvalue degeneracy for a class of one-dimensional Fokker–Planck operators

Angelo Angeletti
Istituto Matematico, Universitá di Camerino, 62032 Camerino (MC), Italy
Cinzia Castagnari
Istituto Matematico, Universitá di Camerino, 62032 Camerino (MC), Italy
Francesco Zirilli[a)]
Department of Mathematical Sciences, Rice University, Houston, Texas 77001

(Received 1 June 1984; accepted for publication 4 October 1984)

Let $f(x)$, $x \in \mathbb{R}$, be a fourth-degree polynomial with $\lim_{|x|\to\infty} f(x) = +\infty$ with two minima, and let $L_\epsilon(\cdot) = \epsilon^2/2\, \partial^2(\cdot)/\partial x^2 + (\partial/\partial x)((\cdot)(\partial f/\partial x))$ be the corresponding Fokker–Planck operator. We study the spectrum of L_ϵ in the limit $\epsilon \to 0$. We show that in the limit $\epsilon \to 0$ the spectrum of L_ϵ degenerates in the spectrum of three decoupled harmonic oscillators.

I. INTRODUCTION

Asymptotic eigenvalue degeneracy due to singular perturbations is a common phenomenon to many different fields of applied mathematics such as quantum mechanics,[1–6] statistical mechanics, and quantum field theory.[7]

In this paper we study the behavior as the diffusion constant goes to zero of the spectrum of a class of one-dimensional Fokker–Planck operators. The problem considered here can be considered analogous for the Fokker–Planck equation of the anharmonic oscillator problem for the Schrödinger equation studied in Refs. 1, 2, and 4. In particular, we will follow the path of Isaacson in Ref. 2.

Let us consider the Smoluchowski approximation to Langevin's equation[8,9]

$$dx(t) = -\frac{\partial f}{\partial x}(x(t))dt + \epsilon\, dw(t), \tag{1.1}$$

where $f: \mathbb{R} \to \mathbb{R}$ is a smooth function called potential, $\mathbb{R}$ is the real line, ϵ is a real parameter, and $w(t)$ is a standard one-dimensional Wiener process. Equation (1.1) is an Ito stochastic differential equation widely used in mathematical physics and engineering, whose solution $x_\epsilon(t)$ is a stochastic process.

The transition probability density $p_\epsilon(x,x_0,t)$ of $x_\epsilon(t)$ is defined as

$$p_\epsilon(x,x_0,t)dx \equiv P_r\{x_\epsilon(t) \in (x, x+dx) | x_\epsilon(0) = x_0\}, \tag{1.2}$$

where $P_r\{\cdot\}$ = probability of $\{\cdot\}$ and $p_\epsilon(x,x_0,t)$ is the solution of the Fokker–Planck equation

$$\frac{\partial p}{\partial t} = L_\epsilon(p), \quad x \in \mathbb{R}, \quad t > 0, \tag{1.3}$$

where $L_\epsilon(\cdot)$, the Fokker–Planck operator, is given by

$$L_\epsilon(p) = \frac{\epsilon^2}{2}\frac{\partial^2 p}{\partial x^2} + \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x} p\right), \quad x \in \mathbb{R}, \tag{1.4}$$

subject to the condition

$$\lim_{t\to 0} p_\epsilon(x,x_0,t) = \delta(x - x_0), \tag{1.5}$$

where $\delta(\cdot)$ is the Dirac's delta.

The problem of deriving asymptotic formulas as $\epsilon \to 0$ for the first nonzero eigenvalue of the Fokker–Planck operator has been considered for a long time both on physical and mathematical grounds. We refer for reasons of brevity only to the recent paper by Matkowsky and Schuss,[10] where several Fokker–Planck operators, including some two-dimensional ones, are considered.

However, the problem of studying the spectrum of the Fokker–Planck operator as $\epsilon \to 0$ has received much less attention. In this paper we restrict our attention to the one-dimensional case when L_ϵ is given by (1.4) and f is a fourth-degree polynomial with two minimizers.

Even in this particular case the resulting problem is an interesting singular perturbation problem for the ordinary differential operator L_ϵ.

The interest of one of us (F.Z.) in the study of the asymptotic behavior of the spectrum of the Fokker–Planck operators arose in the study of a method for global optimization based on the use of suitable stochastic differential equations.[11]

In Sec. II the eigenvalue problem for L_ϵ is reduced to an eigenvalue problem for a suitable Schrödinger Hamiltonian H_ϵ. The particular Schrödinger Hamiltonian obtained when f is a fourth-degree polynomial with two minimizers is studied in detail. In Sec. III some approximating Hamiltonians that will be used later are introduced and studied. In Sec. IV all the basic estimates needed to prove our main results are proved.

In Sec. V a theorem concerning the behavior as $\epsilon \to 0$ of the difference between the resolvent of H_ϵ and the resolvent of the approximating Hamiltonian is proved. Moreover, the asymptotic behavior as $\epsilon \to 0$ of the spectrum of H_ϵ, and as a consequence of the spectrum of L_ϵ, is considered. In Sec. VI, using the Rayleigh–Ritz principle for H_ϵ, a particularly simple asymptotic formula for the first nonzero eigenvalue of L_ϵ is obtained. Finally, in Sec. VII the case when f is given by a general smooth function is considered formally, and some conclusions are drawn.

[a)] Permanent address: Istituto Matematico, Universitá di Salerno, 84100 Salerno, Italy.

 0022-2488/85/040678-14$02.50

II. FROM THE FOKKER–PLANCK EQUATION TO THE SCHRÖDINGER EQUATION

Let us consider the eigenvalue problem

$$L_\epsilon(u) = \lambda u\,, \quad \lambda \in \mathbb{C}\,, \quad x \in \mathbb{R}\,, \tag{2.1}$$

where L_ϵ is given by (1.4).

Let us consider the change of variables

$$y = (\sqrt{2}/\epsilon)x\,, \tag{2.2}$$

$$v(y) = c_\epsilon^{-1/2} e^{f_\epsilon(y)/2} u((\epsilon/\sqrt{2})\,y)\,, \tag{2.3}$$

where c_ϵ is a normalization constant and

$$f_\epsilon(y) = (2/\epsilon^2) f((\epsilon/\sqrt{2})\,y)\,. \tag{2.4}$$

The eigenvalue problem (2.1) becomes

$$H_\epsilon v = -\lambda v\,, \quad \lambda \in \mathbb{C}\,, \quad y \in \mathbb{R}\,, \tag{2.5}$$

where

$$H_\epsilon = -\frac{d^2}{dy^2} + W_\epsilon(y) \tag{2.6}$$

and

$$W_\epsilon(y) = \frac{1}{2}\left(\frac{1}{2}\left(\frac{df_\epsilon}{dy}\right)^2 - \frac{d^2 f_\epsilon}{dy^2}\right). \tag{2.7}$$

Let us note that H_ϵ is a Schrödinger Hamiltonian. It is easy to verify that

$$v_0(y) = c_\epsilon^{1/2} e^{-f_\epsilon(y)/2}\,, \quad y \in \mathbb{R} \tag{2.8}$$

is a solution of (2.5) when $\lambda = 0$. Corresponding to $v_0(y)$ we have

$$u_0(x) = c_\epsilon e^{-(2/\epsilon^2) f(x)}\,, \quad x \in \mathbb{R}\,, \tag{2.9}$$

the solution of (2.1) when $\lambda = 0$. Since we would like to interpret $u_0(x)$ as the probability density of a random variable we will assume that

$$\int_{-\infty}^{+\infty} e^{-(2/\epsilon^2) f(x)}\, dx < \infty\,, \quad \forall \epsilon \neq 0\,, \tag{2.10}$$

and we will choose

$$c_\epsilon = \left(\int_{-\infty}^{+\infty} e^{-(2/\epsilon^2) f(x)}\, dx\right)^{-1}, \tag{2.11}$$

so that

$$\int_{-\infty}^{+\infty} u_0(x) dx = 1\,. \tag{2.12}$$

Condition (2.12) means that $u_0(x) \in L^1(\mathbb{R})$. This implies that $v_0(y) \in L^2(\mathbb{R})$, where $L^p(\mathbb{R})$ is the Lebesgue space of index p, so that it is natural to study the spectrum of H_ϵ in $L^2(\mathbb{R})$.

In this paper we will consider the case when $f(x)$ is given by

$$f_1(x) = (x^2 - \alpha^2)^2\,, \quad \alpha > 0\,, \quad x \in \mathbb{R}\,, \tag{2.13}$$

or by

$$f_2(x) = x^2(ax^2 + bx + c)\,, \quad x \in \mathbb{R}\,, \tag{2.14}$$

where $\alpha > 0$, a, b, c, are real constants and

$$a > 0\,, \tag{2.15}$$

$$b^2 - 4ac < 0\,, \tag{2.16}$$

$$9b^2 - 32ac > 0\,, \tag{2.17}$$

$$b < 0\,. \tag{2.18}$$

Since the spectrum of H_ϵ is invariant with respect to adding a constant to f, to making translation on the x axis, or to changing x into $-x$, $f_1(x)$ represents the most general fourth-degree polynomial with two global minimizers (Fig. 1), and $f_2(x)$ represents the most general fourth-degree polynomial with one global minimizer and one local minimizer. Let us remark the following: (2.15) and (2.16) imply that $f_2(x) \geqslant 0$, $\forall x \in \mathbb{R}$, with $f_2(x) = 0 \Leftrightarrow x = 0$; (2.17) implies that $f_2'(x) = 0$ has three real roots $0, x_1, x_2$ and that $f_2''(x) = 0$ has two real roots [that is, x_1 is a maximizer of f_2 and x_2 is a minimizer of $f_2(x)$]; finally (2.18) implies that $0 < x_1 < x_2$ (Fig. 2).

A straightforward computation gives

$$\begin{aligned} V_\epsilon(y) &= \frac{1}{2}\left(\frac{1}{2}\left(\frac{df_{1\epsilon}}{dy}\right)^2 - \frac{d^2 f_{1\epsilon}}{dy^2}\right) \\ &= \epsilon^4 y^6 - 4\alpha^2\epsilon^2 y^4 + (4\alpha^4 - 3\epsilon^2)y^2 + 2\alpha^2\,, \end{aligned} \tag{2.19}$$

$$\begin{aligned} U_\epsilon(y) &= \frac{1}{2}\left(\frac{1}{2}\left(\frac{df_{2\epsilon}}{dy}\right)^2 - \frac{d^2 f_{2\epsilon}}{dy^2}\right) \\ &= \tfrac{1}{4} y^2(2a\epsilon^2 y^2 + (3b\epsilon/\sqrt{2})\,y + 2c)^2 - \tfrac{1}{2}(6a\epsilon^2 y^2 \\ &\quad + (6b\epsilon/\sqrt{2})\,y + 2c) \\ &= a^2\epsilon^4 y^6 + (3ab/\sqrt{2})\epsilon^3 y^5 + (\tfrac{9}{8} b^2 + 2ac)\epsilon^2 y^4 \\ &\quad + (3bc/\sqrt{2})\epsilon y^3 + (c^2 - 3a\epsilon^2)\,y^2 - (3b\epsilon/\sqrt{2})\,y - c\,. \end{aligned} \tag{2.20}$$

In order to understand intuitively the behavior as $\epsilon \to 0$ of the spectrum of H_ϵ when the potential $W_\epsilon(y)$ is given by V_ϵ or U_ϵ, let us analyze the behavior of V_ϵ and U_ϵ when $\epsilon \to 0$.

Proposition 2.1: Let $V_\epsilon(y)$ be given by (2.19). Then $V_\epsilon(y)$ is an even sixth-degree polynomial. There exists $\epsilon_0 > 0$ such that for $0 < \epsilon < \epsilon_0$ we have the following.

(i) The equation

$$\frac{dV_\epsilon}{dy}(y) = 0 \tag{2.21}$$

has five real roots

$$y = 0\,, \quad y = \pm\frac{1}{\epsilon}\left(\frac{4\alpha^2 \pm \sqrt{4\alpha^4 + 9\epsilon^2}}{3}\right)^{1/2}.$$

That is, V_ϵ has a local minimizer at $y = 0$, two global minimizers at $y = \pm(1/\epsilon)((4\alpha^2 + \sqrt{4\alpha^4 + 9\epsilon^2})/3)^{1/2}$, and two local maximizers at $y = \pm 1/\epsilon((4\alpha^2 - \sqrt{4\alpha^4 + 9\epsilon^2})/3)^{1/2}$.

(ii) Since $V_\epsilon(y)$ is even let us consider only $y > 0$. By explicit computation it is easy to obtain Table I.

(iii) $V_\epsilon(y)$ is bounded below by a constant independent of ϵ.

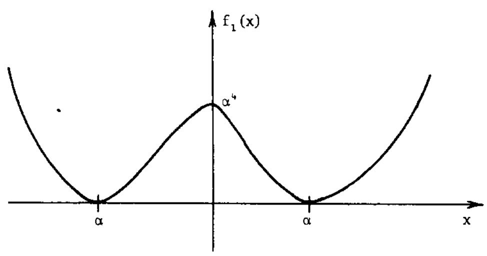


FIG. 1. The case of two symmetric wells.

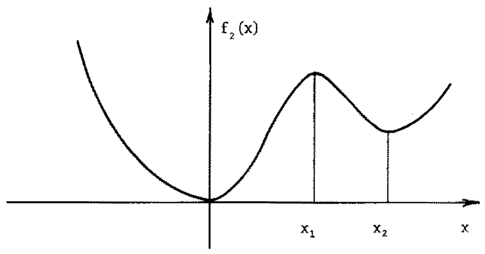


FIG. 2. The case of two nonsymmetric wells.

(iv) $V_\epsilon(y)$ is given by Fig. 3.

Proposition 2.2: Let $U_\epsilon(y)$ be the sixth-degree polynomial given by (2.20). There exists $\epsilon_0 > 0$ such that for $0 < \epsilon < \epsilon_0$ we have the following.

(i) We can consider the points

$$y = 0, \quad y_1 = (\sqrt{2}/\epsilon)x_1, \quad y_2 = (\sqrt{2}/\epsilon)x_2,$$

where $x_{1,2} = (-3b \mp \sqrt{9b^2 - 32ac})/8a$ are such that $(df_2/dx)(x_{1,2}) = 0$, and the points

$$\eta_1 = (\sqrt{2}/\epsilon)\xi_1, \quad \eta_2 = (\sqrt{2}/\epsilon)\xi_2,$$

where $\xi_{1,2} = (-3b \mp \sqrt{9b^2 - 24ac})/12a$ are such that $(d^2f_2/dx^2)(\xi_{1,2}) = 0$. Let us remark that (2.15), (2.16), and (2.17) imply that $\xi_{1,2}$ are real (i.e., $9b^2 - 24ac > 0$). Moreover, $0 < \xi_1 < x_1 < \xi_2 < x_2$ so that $0 < \eta_1 < y_1 < \eta_2 < y_2$.

(ii) We have

$$U'_\epsilon(y) = \tfrac{1}{2}(f'_{2\epsilon} f''_{2\epsilon} - f'''_{2\epsilon}), \tag{2.22}$$

$$U''_\epsilon(y) = \tfrac{1}{2}[(f''_{2\epsilon})^2 + f'_{2\epsilon} f'''_{2\epsilon}] - \tfrac{1}{2} f^{(\text{iv})}_{2\epsilon}, \tag{2.23}$$

where the primes mean differentiation.

(iii) By explicit computation from (ii) it is easy to obtain Table II. Here, $c_1 = \frac{1}{2}(d^2f_2/dx^2)(x_1) < 0$, $c_2 = \frac{1}{2}(d^2f_2/dx^2)(x_2) > 0$. Moreover $c = 2ax_1x_2$, $c_1 = 2ax_1(x_1 - x_2)$, and $c_2 = 2ax_2(x_2 - x_1)$.

(iv) From Table II we can deduce that the equation

$$\frac{dU_\epsilon}{dy} = 0$$

has five real roots so that $U_\epsilon(y)$ has three minimizers and two maximizers.

(v) $U_\epsilon(y)$ is bounded below by a constant independent of ϵ.

(vi) $U_\epsilon(y)$ is given by Fig. 4.

From Proposition 2.1 and Fig. 3 it follows that as $\epsilon \to 0$, $V_\epsilon(y)$ approaches three independent harmonic oscillator potentials, one with vertex at $y = 0$ and equation $4\alpha^4 y^2 + 2\alpha^2$ and two with vertices at $y = \pm(\sqrt{2}/\epsilon)\alpha$ and equations $16\alpha^4(y \mp (\sqrt{2}/\epsilon)\alpha)^2 - 4\alpha^2$.

Let H_ϵ be given by (2.6) and $W_\epsilon(y) = 4\alpha^4 y^2 + 2\alpha^2$. Then the eigenvalues in (2.5) are given by

$$-\lambda_n^{(1)} = 4\alpha^2(n+1), \quad n = 0,1,2,\dots. \tag{2.24}$$

The eigenvalues corresponding to the remaining two harmonic oscillators are

$$-\lambda_n^{(2)} = 8\alpha^2 n, \quad n = 0,1,2,\dots, \tag{2.25}$$

$$-\lambda_n^{(3)} = 8\alpha^2 n, \quad n = 0,1,2,\dots. \tag{2.26}$$

In Sec. V we will prove that the eigenvalues of

$$M_\epsilon = -\frac{d^2}{dy^2} + V_\epsilon(y), \quad y \in \mathbb{R} \tag{2.27}$$

approach (2.24), (2.25), and (2.26) when $\epsilon \to 0$. In particular, we will show that the first eigenvalue $\lambda_0 = 0$ as $\epsilon \to 0$ has asymptotically multiplicity 2 [i.e., $\lambda_1(\epsilon) \to 0$ when $\epsilon \to 0$] as can be seen from (2.25) and (2.26) when $n = 0$. Moreover,

$$\lim_{\epsilon \to 0} -\lambda_{2+4n}(\epsilon) = 4\alpha^2(2n+1), \quad n = 0,1,2, \tag{2.28}$$

as can be seen from (2.24), and

$$\lim_{\epsilon \to 0} -\lambda_{3+4n}(\epsilon) = \lim_{\epsilon \to 0} -\lambda_{4+4n}(\epsilon) = \lim_{\epsilon \to 0} -\lambda_{5+4n}(\epsilon) = 8\alpha^2(n+1),$$
$$n = 0,1,2,\dots, \tag{2.29}$$

as can be seen from (2.24), (2.25), and (2.26). Therefore, M_ϵ as $\epsilon \to 0$ has eigenvalues with multiplicity 1 [i.e., the eigenvalues coming from (2.28)] and eigenvalues with asymptotic multiplicity 3 [i.e., the eigenvalues coming from (2.29)].

From Proposition 2.2 and Fig. 4 it follows that as $\epsilon \to 0$ $U_\epsilon(y)$ approaches three independent harmonic oscillator potentials, one with vertex at $y = 0$ and equation $c^2y^2 - c$, one with vertex at $y = y_1$ and equation $c_1^2(y - y_1)^2 - c_1$ $(c_1 < 0)$, and one with vertex at $y = y_2$ and equation $c_2^2(y - y_2)^2 - c_2$ $(c_2 > 0)$.

Let H_ϵ be given by (2.6) and $W_\epsilon(y) = c^2y^2 - c$. Then the eigenvalues in (2.5) are given by

$$-\bar{\lambda}_n^{(1)} = (2n+1)c - c, \quad n = 0,1,2,\dots \quad (c > 0). \tag{2.30}$$

The eigenvalues corresponding to the remaining two harmonic oscillators are

$$-\bar{\lambda}_n^{(2)} = (2n+1)|c_1| - c_1, \quad n = 0,1,2,\dots, \quad (c_1 < 0), \tag{2.31}$$

$$-\bar{\lambda}_n^{(3)} = (2n+1)c_2 - c_2, \quad n = 0,1,2,\dots. \quad (c_2 > 0). \tag{2.32}$$

In Sec. V we will prove that the eigenvalues of

$$N_\epsilon = -\frac{d^2}{dy^2} + U_\epsilon(y), \quad y \in \mathbb{R} \tag{2.33}$$

TABLE I. The potential $V_\epsilon(y)$. Here the primes mean differentiation with respect to y.

y	0	$\sqrt{\dfrac{4\alpha^2 - \sqrt{4\alpha^4 + 9\epsilon^2}}{3\epsilon^2}}$	$\dfrac{\alpha\sqrt{6}}{3\epsilon}$	$\dfrac{\alpha\sqrt{2}}{\epsilon}$	$\sqrt{\dfrac{4\alpha^2 + \sqrt{4\alpha^4 + 9\epsilon^2}}{3\epsilon^2}}$
$V_\epsilon(y)$	$2\alpha^2$	$\dfrac{2}{27\epsilon^2}(8\alpha^6 - 27\alpha^2\epsilon^2 + (4\alpha^4 + 9\epsilon^2)^{3/2})$	$\dfrac{32\alpha^6}{27\epsilon^2}$	$-4\alpha^2$	$\dfrac{2}{27\epsilon^2}(8\alpha^6 - 27\alpha^2\epsilon^2 - (4\alpha^4 + 9\epsilon^2)^{3/2})$
$V'_\epsilon(y)$	0	0	$-2\alpha\epsilon\sqrt{6}$	$-6\alpha\epsilon\sqrt{2}$	0
$V''_\epsilon(y)$	$8\alpha^4 - 6\epsilon^2$	$\frac{8}{3}(4\alpha^4 + 9\epsilon^2 - 4\alpha^2\sqrt{4\alpha^4 + 9\epsilon^2})$	$-\frac{32}{3}\alpha^4 - 6\epsilon^2$	$32\alpha^4 - 6\epsilon^2$	$\frac{8}{3}((4\alpha^4 + 9\epsilon^2) + 4\alpha^2\sqrt{4\alpha^4 + 9\epsilon^2})$

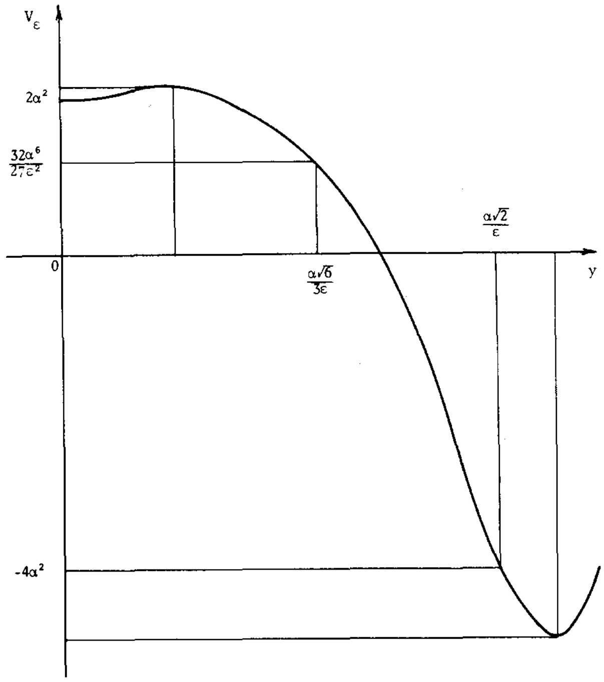


FIG. 3. The potential $V_\epsilon(y)$.

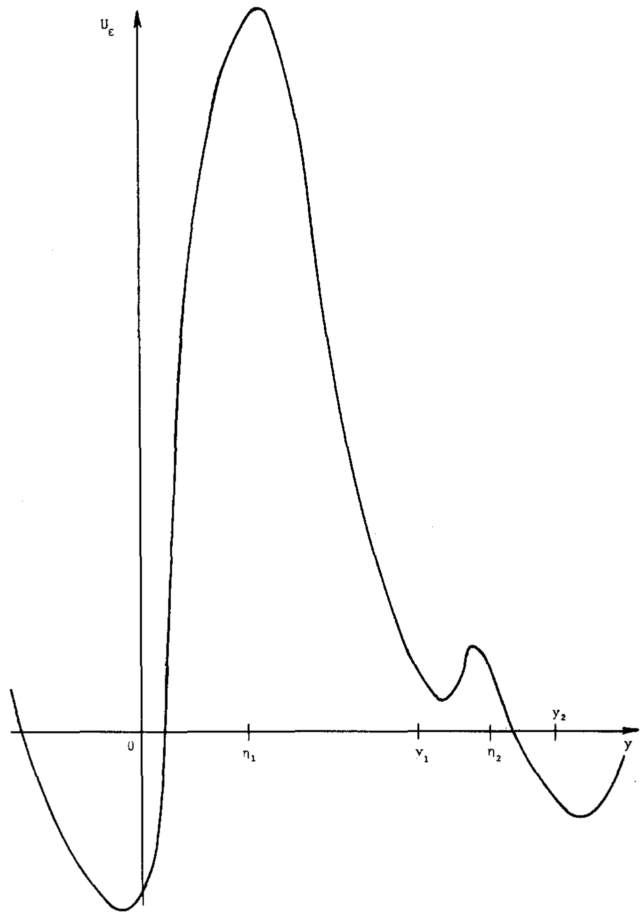


FIG. 4. The potential $U_\epsilon(y)$.

approach (2.30), (2.31), and (2.32) when $\epsilon \to 0$. In particular, we will show that the first eigenvalue $\lambda_0 = 0$ as $\epsilon \to 0$ has asymptotically multiplicity 2 [i.e., $\lambda_1(\epsilon) \to 0$ when $\epsilon \to 0$]. The remaining eigenvalues, since c, c_1, c_2 can be expressed in terms of a, x_1, x_2 as shown in Proposition 2.2 (iii), have asymptotically multiplicity 1 if x_1/x_2 is irrational and have asymptotically multiplicity 1 or 3 if x_1/x_2 is rational.

III. THE APPROXIMATING HAMILTONIANS

Let $\mathscr{C}_0^\infty(\mathbb{R})$ be the space of the infinitely differentiable functions of compact support. Let $h_0:\mathscr{D}(h_0) \subset L^2(\mathbb{R}) \to L^2(\mathbb{R})$ denote the self-adjoint extension of $-\partial^2/\partial y^2$ and let $\mathscr{D}(y^m)$ denote the domain of the self-adjoint multiplication operator y^m.

The Schrödinger Hamiltonians M_ϵ, N_ϵ as operators on $L^2(\mathbb{R})$ possess the following properties.

Theorem 3.1: For any $\epsilon \in \mathbb{R}$ with $\epsilon \neq 0$, we have the following.

(i) M_ϵ is essentially self-adjoint on $\mathscr{C}_0^\infty(\mathbb{R})$ and is self-adjoint on $\mathscr{D}(h_0) \cap \mathscr{D}(y^6)$.

(ii) M_ϵ has compact resolvent.

(iii) The eigenvalues of M_ϵ are nondegenerate.

(iv) The eigenfunctions alternate parity and the one corresponding to the smallest eigenvalue is even.

Proof: See Refs. 6 and 12.

Theorem 3.2: For any $\epsilon \in \mathbb{R}$ with $\epsilon \neq 0$, we have the following.

(i) N_ϵ is essentially self-adjoint on $\mathscr{C}_0^\infty(\mathbb{R})$ and is self-adjoint on $\mathscr{D}(h_0) \cap \mathscr{D}(y^6)$.

(ii) N_ϵ has compact resolvent.

(iii) The eigenvalues of N_ϵ are nondegenerate.

Proof: See Refs. 6 and 12.

Let $A_+ = \{ y \mid y > \alpha\sqrt{6}/3\epsilon \}$, $A_0 = \{ y \mid |y| < \alpha\sqrt{6}/3\epsilon \}$, $A_- = \{ y \mid y < -\alpha\sqrt{6}/3\epsilon \}$ and define $V_{2\epsilon}$ as follows:

TABLE II. The potential $U_\epsilon(y)$. Here the primes mean differentiation.

y	0	η_1	y_1	η_2	y_2
$U_\epsilon(y)$	$-c$	$\frac{1}{2\epsilon^2}(f_1'(\xi_1))^2$	$-c_1$	$\frac{1}{2\epsilon^2}(f_2'(\xi_2))^2$	$-c_2$
$U_\epsilon'(y)$	$-\frac{3b\epsilon}{\sqrt{2}}$	$-\frac{1}{2}\frac{\epsilon}{\sqrt{2}}f_2'''(\xi_1)$	$-\frac{1}{2}\frac{\epsilon}{\sqrt{2}}f_2'''(x_1)$	$-\frac{1}{2}\frac{\epsilon}{\sqrt{2}}f_2'''(\xi_2)$	$-\frac{1}{2}\frac{\epsilon}{\sqrt{2}}f_2'''(x_2)$
$U_\epsilon''(y)$	$2(c^2 - 3a\epsilon^2)$	$\frac{1}{2}\left[f_2'(\xi_1)f_2'''(\xi_1) - \frac{\epsilon^2}{2}f_2^{(iv)}(\xi_1)\right]$	$2c_1^2 - \frac{\epsilon^2}{4}f_2^{(iv)}(x_1)$	$\frac{1}{2}\left[f_2'(\xi_1)f_2'''(\xi_2) - \frac{\epsilon^2}{2}f_2^{(iv)}(\xi_2)\right]$	$2c_2^2 - \frac{\epsilon^2}{4}f_2^{(iv)}(x_2)$

$$V_{2\epsilon}(y)=\begin{cases}4\alpha^4\left(y-\frac{\alpha\sqrt{6}}{3\epsilon}-\frac{1}{2\nu}\left(y-\frac{\alpha\sqrt{6}}{3\epsilon}\right)^{-1}\right)^2-4\alpha^2, & \text{when } y\in A_+,\\ \frac{V_0}{\cos^2\beta y}-V_0+2\alpha^2, & \text{when } y\in A_0,\\ 4\alpha^4\left(y+\frac{\alpha\sqrt{6}}{3\epsilon}-\frac{1}{2\nu}\left(y+\frac{\alpha\sqrt{6}}{3\epsilon}\right)^{-1}\right)^2-4\alpha^2, & \text{when } y\in A_-\end{cases}\tag{3.1}$$

(see Fig. 5), where

$$\frac{1}{2\nu}=\left(\frac{\alpha\sqrt{2}}{\epsilon}-\frac{\alpha\sqrt{6}}{3\epsilon}\right)^2=\frac{2\alpha^2}{\epsilon^2}\left(\frac{3-\sqrt{3}}{3}\right)^2,\tag{3.2}$$

$$\beta=(\pi/2)(3\epsilon/\alpha\sqrt{6}),\tag{3.3}$$

$$V_0=(32/3)(\alpha^6/\pi^2\epsilon^2).\tag{3.4}$$

The function $V_{2\epsilon}$ as $\epsilon\to 0$ is an approximation to V_ϵ. In particular, $V_{2\epsilon}$ approaches three independent harmonic oscillator potentials, one with vertex at $y=0$ and equation $4\alpha^4y^2+2\alpha^2$ and two with vertices at $y=\pm\alpha\sqrt{2}/\epsilon$ and equations $16\alpha^4(y\mp(\alpha\sqrt{2}/\epsilon))^2-4\alpha^2$.

Let $0<\bar\eta_1(\epsilon)<\alpha\sqrt{6}/3\epsilon$, $0<\bar\eta_2(\epsilon)<1/\sqrt{2\nu}$ with

$$\lim_{\epsilon\to 0}\bar\eta_1(\epsilon)=\lim_{\epsilon\to 0}\bar\eta_2(\epsilon)=\infty,\tag{3.5}$$

$$\lim_{\epsilon\to 0}\epsilon\bar\eta_1(\epsilon)=\lim_{\epsilon\to 0}\epsilon\bar\eta_2(\epsilon)=0.\tag{3.6}$$

Given $\bar\eta_1(\epsilon)$ we choose $\bar\eta_2(\epsilon)$ to be the smallest solution of

$$V_{2\epsilon}(\bar\eta_1(\epsilon))=V_{2\epsilon}((\alpha\sqrt{2}/\epsilon)-\bar\eta_2(\epsilon)).\tag{3.7}$$

A straightforward computation shows that (3.7) can be solved and that $\bar\eta_1(\epsilon)$ should be of the same order of $\bar\eta_2(\epsilon)$ for $\epsilon\to 0$.

Let

$$I_1^{(\epsilon)}=\{y\in\mathbb{R}\,|\,\bar\eta_1(\epsilon)<y<(\alpha\sqrt{2}/\epsilon)-\bar\eta_2(\epsilon)\},\tag{3.8}$$

$$I_2^{(\epsilon)}=\{y\in\mathbb{R}\,|\,-(\alpha\sqrt{2}/\epsilon)+\bar\eta_2(\epsilon)<y<-\bar\eta_1(\epsilon)\}.\tag{3.9}$$

We define

$$V_{1\epsilon}(y)=\begin{cases}V_{2\epsilon}(y), & \text{when } y\notin I_1^{(\epsilon)}\cup I_2^{(\epsilon)},\\ V_{2\epsilon}(\bar\eta_1(\epsilon)), & \text{when } y\in I_1^{(\epsilon)}\cup I_2^{(\epsilon)}\end{cases}\tag{3.10}$$

(see Fig. 6). Note that $V_{1\epsilon}$ is a continuous function because of Eq. (3.7), and as $\epsilon\to 0$, $V_{1\epsilon}$ is an approximation to V_ϵ in the same sense as $V_{2\epsilon}$.

Let us now consider the operators

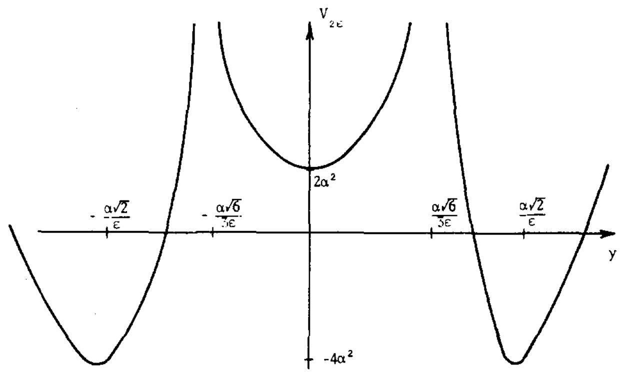


FIG. 5. The potential $V_{2\epsilon}(y)$.

$$M_\epsilon^{(2)}=-\frac{d^2}{dy^2}+V_{2\epsilon},\quad y\in\mathbb{R},\tag{3.11}$$

$$M_\epsilon^{(1)}=-\frac{d^2}{dy^2}+V_{1\epsilon},\quad y\in\mathbb{R}.\tag{3.12}$$

We will use them to approximate M_ϵ.

The eigenvalue problem for $M_\epsilon^{(2)}$

$$M_\epsilon^{(2)}v=\lambda v,\quad y\in\mathbb{R},\quad v\in L^2(\mathbb{R}),\tag{3.13}$$

can be reduced to the following eigenvalue problems:

$$M_\epsilon^{(2)}v=\lambda v,\quad y\in A_+,\quad v\in L^2(A_+),\tag{3.14}$$

$$M_\epsilon^{(2)}v=\lambda v,\quad y\in A_0,\quad v\in L^2(A_0),\tag{3.15}$$

$$M_\epsilon^{(2)}v=\lambda v,\quad y\in A_-,\quad v\in L^2(A_-).\tag{3.16}$$

The eigenvalue problems (3.14), (3.15), and (3.16) can be solved explicitly. In fact the eigenvalues and eigenfunctions of (3.14) and (3.16) are given by[13,14]

$$\lambda_{n\epsilon}^{\pm}=4\alpha^2[2n+\gamma-\alpha^2/\nu],\quad n=0,1,2,\ldots,\tag{3.17}$$

$$\phi_{n\epsilon}^{\pm}=N_{n\epsilon}[2\alpha^2(y\mp(\alpha\sqrt{6}/3\epsilon))^2]^{(2\gamma+1)/4}\times\exp\left[-\alpha^2\left(y\mp\frac{\alpha\sqrt{6}}{3\epsilon}\right)^2\right]L_n^{(\gamma)}\left[2\alpha^2\left(y\mp\frac{\alpha\sqrt{6}}{3\epsilon}\right)^2\right],\tag{3.18}$$

where $N_{n\epsilon}$ is a normalization constant, $L_n^{(\gamma)}$ are the generalized Laguerre polynomials, $\phi_{n\epsilon}^{+}$ is defined for $y>\alpha\sqrt{6}/3\epsilon$, $\phi_{n\epsilon}^{-}$ is defined $y<-\alpha\sqrt{6}/3\epsilon$, and

$$\gamma=(1/2\nu)\sqrt{4\alpha^4+\nu^2}.\tag{3.19}$$

The eigenvalues and eigenfunctions of (3.15) are given by[15]

$$\lambda_{n\epsilon}^{0}=\beta^2[n^2+\delta(2n+1)]+2\alpha^2,\quad n=0,1,\ldots,\tag{3.20}$$

$$\phi_{n\epsilon}^{0}=\begin{cases}\cos^\delta\beta y F\left(\delta+\frac{n}{2},\frac{-n}{2},\frac{1}{2},\sin^2\beta y\right), & \text{when } n \text{ is even},\\ \cos^\delta\beta y\sin\beta y F\left(\delta+\frac{n+1}{2},-\frac{n-1}{2},\frac{3}{2},\sin^2\beta y\right), & \text{when } n \text{ is odd},\end{cases}\tag{3.21}$$

where $F(x_1,x_2,x_3,z)$ is the hypergeometric function and δ is defined by the equation

$$V_0=\beta^2\delta(\delta-1),\quad \delta>1.\tag{3.22}$$

The eigenvalues of (3.13) are given by $\lambda_{n\epsilon}^{0}$ and $\lambda_{n\epsilon}^{\pm}$, $n=0,1,2,\ldots$. The eigenvalues $\lambda_{n\epsilon}^{\pm}$ have multiplicity 2. Moreover as $\epsilon\to 0$, $\lambda_{n\epsilon}^{0}$, $\lambda_{n\epsilon}^{\pm}$ approach the eigenvalues (2.24), (2.25), and (2.26) of the three harmonic oscillators considered before.

The eigenfunctions of (3.14) satisfy $\phi_{n\epsilon}^{+}(\alpha\sqrt{6}/3\epsilon)=(d\phi_{n\epsilon}^{+}/dy)(\alpha\sqrt{6}/3\epsilon)=0$, so that corresponding eigenfunctions of (3.13) can be obtained, extending $\phi_{n\epsilon}^{+}(y)$ with zero for $y\notin A_+$. Similar statements hold for the eigenfunc-

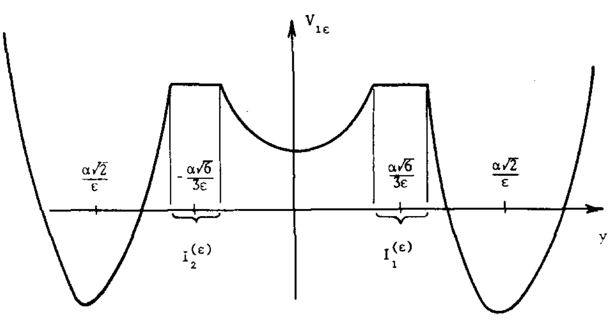


FIG. 6. The potential $V_{1\epsilon}(y)$.

tions of (3.15) and (3.16). Moreover, since the eigenfunctions of (3.15) are even or odd and the eigenvalues $\lambda^{\pm}_{n\epsilon}$ of (3.13) have multiplicity 2, the eigenfunction of (3.13) can be chosen to be even or odd.

Let $\mathscr{C}_0^\infty(\mathbb{R} - \{\pm\alpha\sqrt{6}/3\epsilon\}) = \{f \mid f \text{ is } \mathscr{C}^\infty$ and of compact support and is zero in a neighborhood of $y = +\alpha\sqrt{6}/3\epsilon$ and $y = -\alpha\sqrt{6}/3\epsilon\}$. We have the following.

Theorem 3.3: $M_\epsilon^{(2)}$ is essentially self-adjoint on $\mathscr{C}_0^\infty(\mathbb{R} - \{\pm\alpha\sqrt{6}/3\epsilon\})$.

Proof: It is a straightforward modification of Isaacson,[2] Appendix 2.

Theorem 3.4: $M_\epsilon^{(1)}$ is essentially self-adjoint on $\mathscr{C}_0^\infty(\mathbb{R})$.

Proof: It follows immediately from Theorem 10.23, p. 315 of Weidmann.[16]

Let

$$\bar{A}_1 = \{y \mid y > \eta_2\},$$

$$\bar{A}_0 = \{y \mid 2y_1 - \eta_2 < y < \eta_2\},$$

$$\bar{A}_- = \{y \mid y < 2y_1 - \eta_2\}$$

and define $U_{2\epsilon}$ as follows:

$$U_{2\epsilon} = \begin{cases} \dfrac{c_2^2}{4}\left(y - \eta_2 - \dfrac{1}{2\nu_2}\dfrac{1}{y - \eta_2}\right)^2 - c_2, \\ \qquad \text{when } y \in \bar{A}_+, \\ (\bar{V}_0/\cos^2\bar{\beta}(y - y_1)) - \bar{V}_0 - c_1, \quad \text{when } y \in \bar{A}_0, \\ \dfrac{c^2}{4}\left[y - (2y_1 - \eta_2) - \dfrac{1}{2\nu_1}\dfrac{1}{y - (2y_1 - \eta_2)}\right]^2 - c, \\ \qquad \text{when } y \in \bar{A}_- \end{cases} \tag{3.23}$$

(see Fig. 7), where

$$1/2\nu_2 = (y_2 - \eta_2)^2, \tag{3.24}$$

$$1/2\nu_1 = (2y_1 - \eta_2)^2, \tag{3.25}$$

$$\bar{\beta} = (\pi/2)[1/(\eta_2 - y_1)], \tag{3.26}$$

$$\bar{V}_0 = c_1^2/\bar{\beta}^2. \tag{3.27}$$

Let us remember that y_1, y_2, η_1, η_2 depend on ϵ (Proposition 2.2). It is easy to check by explicit computation that $2y_1 - \eta_2 > 0$ so that the function $U_{2\epsilon}$ (Fig. 7) as $\epsilon \to 0$ is an approximation to U_ϵ. In particular $U_{2\epsilon}$ approaches three independent harmonic oscillator potentials, one with vertex $y = 0$ and equation $c^2y^2 - c$, one with vertex at $y = y_1$ and equation $c_1^2(y - y_1)^2 - c_1$, and one with vertex at $y = y_2$ and equation $c_2^2(y - y_2)^2 - c_2$.

Let $\mu_1(\epsilon)$, $\mu_2(\epsilon)$, $\mu_3(\epsilon) > 0$ and let $J_1^{(\epsilon)} = \{y \in \mathbb{R} \mid \mu_1(\epsilon) < y < y_1 - \mu_2(\epsilon)\}$, $J_2^{(\epsilon)} = \{y \in \mathbb{R} \mid y_1 + \mu_2(\epsilon) < y < y_2 - \mu_3(\epsilon)\}$ be two intervals such that $2y_1 - \eta_2 \in J_1^{(\epsilon)}$ and $\eta_2 \in J_2^{(\epsilon)}$, such that

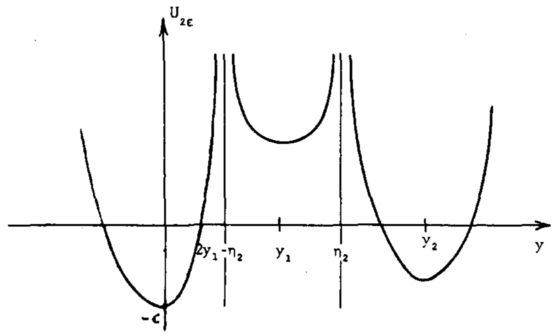


FIG. 7. The potential $U_{2\epsilon}(y)$.

$$U_{2\epsilon}(\mu_1(\epsilon)) = U_{2\epsilon}(y_1 - \mu_2(\epsilon)), \tag{3.28}$$

$$U_{2\epsilon}(y_1 + \mu_2(\epsilon)) = U_{2\epsilon}(y_2 - \mu_3(\epsilon)), \tag{3.29}$$

and $\bar{J}_1^{(\epsilon)} \cap \bar{J}_2^{(\epsilon)} = \{\phi\}$. Note that because of symmetry $U_{2\epsilon}(y_1 - \mu_2(\epsilon)) = U_{2\epsilon}(y_1 + \mu_2(\epsilon))$. Finally, we will later need

$$\lim_{\epsilon \to 0} U_{2\epsilon}(y_1 \mp \mu_2(\epsilon)) = \infty. \tag{3.30}$$

Let us define

$$U_{1\epsilon}(y) = \begin{cases} U_{2\epsilon}(y), & y \notin J_1^{(\epsilon)} \cup J_2^{(\epsilon)}, \\ U_{2\epsilon}(y_1 - \mu_2(\epsilon)), & y \in J_1^{(\epsilon)}, \\ U_{2\epsilon}(y_1 - \mu_2(\epsilon)), & y \in J_2^{(\epsilon)} \end{cases} \tag{3.31}$$

(see Fig. 8). Proceeding as before let us now consider

$$N_\epsilon^{(2)} = -\frac{d^2}{dy^2} + U_{2\epsilon}, \quad y \in \mathbb{R}, \tag{3.32}$$

$$N_\epsilon^{(1)} = -\frac{d^2}{dy^2} + U_{1\epsilon}, \quad y \in \mathbb{R}. \tag{3.33}$$

We will use them to approximate N_ϵ.

The eigenvalue problem for $N_\epsilon^{(2)}$ can be solved analogously to the eigenvalue problem for $M_\epsilon^{(2)}$. In particular as $\epsilon \to 0$ the eigenvalues of $N_\epsilon^{(2)}$ approach the eigenvalues (2.30), (2.31), and (2.32) of the three harmonic oscillators considered before.

Let $\mathscr{C}_0^\infty(\mathbb{R} - \{2y_1 - \eta_2\} - \{\eta_2\}) = \{f \mid f \text{ is } \mathscr{C}^\infty$ and of compact support and is zero in a neighborhood of $y = 2y_1 - \eta_2$ and $y = \eta_2\}$. We have the following.

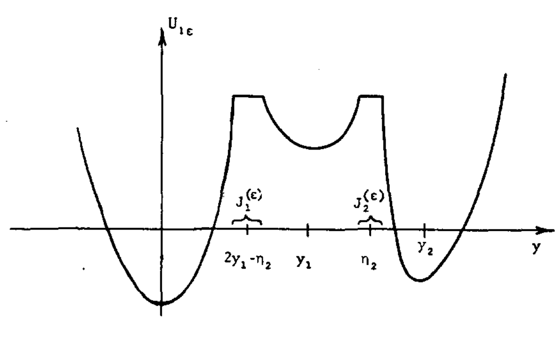


FIG. 8. The potential $U_{1\epsilon}(y)$.

Theorem 3.5: $N_\epsilon^{(2)}$ is essentially self-adjoint on $\mathscr{C}_0^\infty(\mathbb{R} - \{2y_1 - \eta_2\} - \{\eta_2\})$.

Proof: It is a straightforward modification of Isaacson,[2] Appendix 2.

Theorem 3.6: $N_\epsilon^{(1)}$ is essentially self-adjoint on $\mathscr{C}_0^\infty(\mathbb{R})$.

Proof: It follows immediately from Theorem 10.23, p. 315 of Weidmann.[16]

IV. THE BASIC ESTIMATES

We will prove here some estimates that will be used later.

Theorem 4.1: There exist constants $z_0 > 0$, $\epsilon_0 > 0$ such that when $z \geqslant z_0$ and $0 < \epsilon < \epsilon_0$ we have

$$(M_\epsilon + z)^2 \geqslant V_\epsilon^2, \quad \text{on } \mathscr{C}_0^\infty(\mathbb{R}) \times \mathscr{C}_0^\infty(\mathbb{R}), \tag{4.1}$$

$$(M_\epsilon^{(1)} + z)^2 \geqslant \tilde{\beta} V_{1\epsilon}^2, \quad \text{on } \mathscr{C}_0^\infty(\mathbb{R}) \times \mathscr{C}_0^\infty(\mathbb{R}), \tag{4.2}$$

$$(M_\epsilon^{(2)} + z)^2 \geqslant \tilde{\beta} V_{2\epsilon}^2, \quad \text{on } \mathscr{C}_0^\infty(\mathbb{R} - \{\pm \alpha\sqrt{6}/3\epsilon\}) \times \mathscr{C}_0^\infty(\mathbb{R} - \{\pm \alpha\sqrt{6}/3\epsilon\}), \tag{4.3}$$

where $0 < \tilde{\beta} < 1$.

Proof: Let us first prove (4.1) and let $p = i\, d/dy$. Then as a form on $\mathscr{C}_0^\infty(\mathbb{R}) \times \mathscr{C}_0^\infty(\mathbb{R})$ we have

$$(M_\epsilon + z)^2 = (p^2 + V_\epsilon + z)^2 = p^4 + V_\epsilon^2 + 2zV_\epsilon + z^2 + 2p(V_\epsilon + z)p - V_\epsilon''. \tag{4.4}$$

Since $V_\epsilon \geqslant$const independent of ϵ when $0 < \epsilon < \epsilon_0$,

$$p(V_\epsilon + z)p \geqslant 0, \quad \text{on } \mathscr{C}_0^\infty(\mathbb{R}) \times \mathscr{C}_0^\infty(\mathbb{R}), \tag{4.5}$$

for z large enough. From (4.4) and (4.5) we have

$$(M_\epsilon + z)^2 - V_\epsilon^2 \geqslant 2zV_\epsilon + z^2 - V_\epsilon'', \quad \text{on } \mathscr{C}_0^\infty(\mathbb{R}) \times \mathscr{C}_0^\infty(\mathbb{R}). \tag{4.6}$$

To prove (4.1) it will be enough to show that for $z \geqslant z_0$ and $0 < \epsilon < \epsilon_0$ we have

$$F_1(y^2) = 2zV_\epsilon + z^2 - V_\epsilon'' \geqslant 0, \quad y \in \mathbb{R}. \tag{4.7}$$

Let us define

$$t = y^2, \quad A = 2z\epsilon^4, \quad B = 8z\alpha^2\epsilon^2 + 30\epsilon^4,$$
$$C = 2z(4\alpha^4 - 3\epsilon^2) + 48\alpha^2\epsilon^2,$$
$$D = 4z\alpha^2 + z^2 - 8\alpha^4 + 6\epsilon^2.$$

A simple computation shows that

$$F_1(t) = t(At^2 - Bt + C) + D, \quad t \geqslant 0. \tag{4.8}$$

Let us first note that when $z > 2(\sqrt{3} - 1)\alpha^2$ and $0 < \epsilon < 2\alpha^2(\sqrt{3}/3)$ we have A,B,C,D positive. Consider now the parabola

$$At^2 - Bt + C. \tag{4.9}$$

Since $A > 0$, the parabola (4.9) will have a minimizer at $t_0 = B/2A$, where

$$At_0^2 - Bt_0 + C = (4AC - B^2)/4A = -(\epsilon^2/2z)(225\epsilon^2 + 24\alpha^2 z + 12z^2) \leqslant 0. \tag{4.10}$$

Moreover, the equation $At^2 - Bt + C = 0$ has two real roots

$$0 < t_1 = \frac{B - \sqrt{B^2 - 4AC}}{2A} < t_2 = \frac{B + \sqrt{B^2 - 4AC}}{2A} < \frac{B}{A}.$$

Therefore, $\forall t \geqslant 0$

$$F_1(t) \geqslant (B/A)(At_0^2 - Bt_0 + C) + D = z^2 - 20\alpha^2 z - 56\alpha^4 - 84\epsilon^2 - 630(\alpha^2\epsilon^2/z) - (3375/2)(\epsilon^4/z^2), \tag{4.11}$$

and this last expression can be made positive for $z > z_0$ and $0 < \epsilon < \epsilon_0$, choosing z_0 and ϵ_0. The estimate (4.1) is established.

Let us now prove (4.3). Let $0 < \tilde{\beta} < 1$. Proceeding as we have done in proving (4.1), we obtain as a form on $\mathscr{C}_0^\infty(\mathbb{R} - \{\pm \alpha\sqrt{6}/3\epsilon\}) \times \mathscr{C}_0^\infty(\mathbb{R} - \{\pm \alpha\sqrt{6}/3\epsilon\})$

$$(M_\epsilon^{(2)} + z)^2 - \tilde{\beta} V_{2\epsilon}^2 \geqslant (1 - \tilde{\beta}) V_{2\epsilon}^2 + 2zV_{2\epsilon} + z^2 - V_{2\epsilon}''. \tag{4.12}$$

To prove (4.3) it will be enough to show that for $z > z_0$, $0 < \epsilon < \epsilon_0$ we have

$$(1 - \tilde{\beta}) V_{2\epsilon}^2 + 2zV_{2\epsilon} + z^2 - V_{2\epsilon}'' \geqslant 0, \quad y \in \mathbb{R}. \tag{4.13}$$

For $y \in A_0$ formula (4.13) becomes

$$z^2 + 2(V_0 tg^2\beta y + 2\alpha^2)z + [(1 - \tilde{\beta})(V_0 tg^2\beta y + 2\alpha^2)^2 - 8\alpha^4(2\sin^2\beta y + 1)/\cos^4\beta y]. \tag{4.14}$$

When $|y| \leqslant \pi/4\beta$ we have $\cos^2\beta y \geqslant \frac{1}{2}$ and $\sin^2\beta y \leqslant \frac{1}{2}$, so that the expression (4.14) is greater than or equal to

$$z^2 + 4\alpha^2 z - (60 + 4\tilde{\beta})\alpha^4 \geqslant 0, \tag{4.15}$$

when $z \geqslant (\sqrt{64 + 4\tilde{\beta}} - 2)\alpha^2$ and $\forall \epsilon > 0$.

When $\pi/4\beta \leqslant |y| < \pi/2\beta$ we have $\sin^2\beta y \geqslant \frac{1}{2}$ and $\cos^2\beta y \leqslant \frac{1}{2}$ so that expression (4.14) is greater than or equal to

$$z^2 + 4\alpha^2 z + \frac{1}{\cos^4\beta y}[(1 - \tilde{\beta}) V_0^2 \sin^4\beta y - 8\alpha^4(2\sin^2\beta y + 1)] \geqslant z^2 + 4\alpha^2 z + 4[((1 - \tilde{\beta})/4)V_0^2 - 24\alpha^4] \geqslant 0, \tag{4.16}$$

since V_0 given (3.4) goes to infinity when $\epsilon \to 0$. The last inequality in (4.16) holds $\forall z > 0$, $0 < \epsilon < \frac{2}{3}(\alpha^2/\pi)(6(1 + \tilde{\beta}))^{1/4}$.

For $y \in A_+$ formula (4.13) becomes

$$z^2 + 2\left[4\alpha^4\left(y - \frac{\alpha\sqrt{6}}{3\epsilon} - \frac{1}{2\nu}\frac{1}{y - (\alpha\sqrt{6}/3\epsilon)}\right)^2 - 4\alpha^2\right]z + (1 - \tilde{\beta})\left[4\alpha^4\left(y - \frac{\alpha\sqrt{6}}{3\epsilon} - \frac{1}{2\nu}\frac{1}{y - (\alpha\sqrt{6}/3\epsilon)}\right)^2 - 4\alpha^2\right]^2 - 8\alpha^4\left(1 + \frac{3}{4\nu^2}\frac{1}{(y - (\alpha\sqrt{6}/3\epsilon))^4}\right) \geqslant 0. \tag{4.17}$$

With the substitution $t = 2\nu(y - (\alpha\sqrt{6}/3\epsilon))^2$ expression (4.17) becomes

$$t^2(z^2 - 8\alpha^2 z - 8\alpha^4) - 24\alpha^4 + (1 - \tilde{\beta})\left[\frac{4\alpha^4}{2\nu}(t - 1)^2 - 4\alpha^2 t\right]^2 \geqslant 0, \quad t \geqslant 0. \tag{4.18}$$

When $t \geqslant \frac{1}{2}$ and z such that $(z^2 - 8\alpha^2 z - 8\alpha^4)$ is positive, the left-hand side of (4.18) is greater than or equal to

$$\tfrac{1}{4}(z^2-8\alpha^2 z-8\alpha^4)-24\alpha^4\geqslant 0\,,$$
$$\text{for } z>(4+2\sqrt{30})\alpha^2\,,\quad \epsilon>0\,. \tag{4.19}$$

When $0<t<\frac{1}{2}$ and z such that $(z^2-8\alpha^2z-8\alpha^4)$ is positive, then the left-hand side of (4.18) is greater than or equal to

$$-24\alpha^4+(1-\tilde{\beta})[(4\alpha^4/2\nu)(t-1)^2-4\alpha^2 t]^2\,. \tag{4.20}$$

The expression (4.20) is positive for $0<\epsilon<\epsilon_0$ since ν given by (3.2) goes to zero as $\epsilon\to 0$.

The proof of (4.3) for $y\in A_-$ is analogous to the proof given for $y\in A_+$ and will be omitted.

The estimate (4.3) has been established.

Let us now prove (4.2). Let $0<\tilde{\beta}<1$. Proceeding as we have done proving (4.1), we obtain as a form on $\mathscr{C}_0^\infty(\mathbb{R})\times\mathscr{C}_0^\infty(\mathbb{R})$

$$(M_\epsilon^{(1)}+z)^2-\tilde{\beta}V_{1\epsilon}\geqslant(1-\tilde{\beta})V_{1\epsilon}^2+2zV_{1\epsilon}+z^2-V''_{1\epsilon}\,. \tag{4.21}$$

To prove (4.2) it will be enough to show that for $z>z_0$, $0<\epsilon<\epsilon_0$ we have

$$(1-\tilde{\beta})V_{1\epsilon}^2+2zV_{1\epsilon}+z^2-V''_{1\epsilon}\geqslant 0\,,\quad y\in\mathbb{R}\,. \tag{4.22}$$

Let $\chi_{I_1^{(\epsilon)}\cup I_2^{(\epsilon)}}$ be the characteristic function of $I_1^{(\epsilon)}\cup I_2^{(\epsilon)}$. We have

$$\begin{aligned} V''_{1\epsilon}=V''_{2\epsilon}(1-\chi_{I_1^{(\epsilon)}\cup I_2^{(\epsilon)}})-\bar{c}_1[\delta(y-\bar{\eta}_1)+\delta(y+\bar{\eta}_1)]\\ -\bar{c}_2[\delta(y-(\alpha\sqrt{2}/\epsilon)+\bar{\eta}_2)\\ +\delta(y+(\alpha\sqrt{2}/\epsilon)-\bar{\eta}_2)]\,,\end{aligned} \tag{4.23}$$

where $\delta(\cdot)$ is the Dirac's delta and

$$\bar{c}_1=2\beta V_0|\cos^{-3}\beta\bar{\eta}_1\sin\beta\bar{\eta}_1|\geqslant 0\,, \tag{4.24}$$

$$\bar{c}_2=\frac{8\alpha^4}{\sqrt{2\nu}}\left|1-\sqrt{2\nu}\bar{\eta}_1-\frac{1}{(1-\sqrt{2\nu}\bar{\eta}_1)^3}\right|\geqslant 0\,, \tag{4.25}$$

are the absolute values of the jumps at $y=\pm\bar{\eta}_1$ and $y=\pm((\alpha\sqrt{2}/\epsilon)-\bar{\eta}_2)$ of $V'_{1\epsilon}$.

Since $V_{1\epsilon}=V_{2\epsilon}$ when $y\in\mathbb{R}\setminus\{I_1^{(\epsilon)}\cup I_2^{(\epsilon)}\}$ we can rewrite equation (4.22) as follows:

$$\begin{aligned}(1-\chi_{I_1^{(\epsilon)}\cup I_2^{(\epsilon)}})\{(1-\tilde{\beta})V_{2\epsilon}^2+2zV_{2\epsilon}+z^2-V''_{2\epsilon}\}\\ +\chi_{I_1^{(\epsilon)}\cup I_2^{(\epsilon)}}\{(1-\tilde{\beta})V_{2\epsilon}^2(\bar{\eta}_1)+2zV_{2\epsilon}(\bar{\eta}_1)+z^2\}\\ +\bar{c}_1\{\delta(y-\bar{\eta}_1)+\delta(y+\bar{\eta}_1)\}\\ +\bar{c}_2\{\delta(y-(\alpha\sqrt{2}/\epsilon)+\bar{\eta}_2)\\ +\delta(y+(\alpha\sqrt{2}/\epsilon)-\bar{\eta}_2)\}\geqslant 0\,.\end{aligned} \tag{4.26}$$

In fact for $z>z_0>0$, $0<\epsilon<\epsilon_0$ we have $\{(1-\tilde{\beta})V_{2\epsilon}^2+2zV_{2\epsilon}+z^2-V''_{2\epsilon}\}\geqslant 0$. Moreover $\{(1-\tilde{\beta})V_{2\epsilon}^2(\bar{\eta}_1)+2zV_{2\epsilon}(\bar{\eta}_1)+z^2\}\geqslant 0$ and $\bar{c}_1\geqslant 0$, $\bar{c}_2\geqslant 0$.

The estimate (4.2) is established.

Let $\hat{c}$ be a constant such that

$$\hat{V}_\epsilon\equiv V_\epsilon+\hat{c}>0\,,\quad\text{and } (V_\epsilon+\hat{c})^2\geqslant V_\epsilon^2\,, \tag{4.27}$$

$$\hat{V}_{1\epsilon}\equiv V_{1\epsilon}+\hat{c}>0\,,\quad\text{and } (V_{1\epsilon}+\hat{c})^2\geqslant V_{1\epsilon}^2\,, \tag{4.28}$$

$$\hat{V}_{2\epsilon}\equiv V_{2\epsilon}+\hat{c}>0\,,\quad\text{and } (V_{2\epsilon}+\hat{c})^2\geqslant V_{2\epsilon}^2\,. \tag{4.29}$$

We define

$$\hat{M}_\epsilon=-\frac{d^2}{dy^2}+\hat{V}_\epsilon=M_\epsilon+\hat{c}\,, \tag{4.30}$$

$$\hat{M}_\epsilon^{(1)}=-\frac{d^2}{dy^2}+\hat{V}_{1\epsilon}=M_\epsilon^{(1)}+\hat{c}\,, \tag{4.31}$$

$$\hat{M}_\epsilon^{(2)}=-\frac{d^2}{dy^2}+\hat{V}_{2\epsilon}=M_\epsilon^{(2)}+\hat{c}\,. \tag{4.32}$$

Theorem 4.2: There exist $z_0>0$ and $\epsilon_0>0$ such that for $z\geqslant z_0$ and $0<\epsilon<\epsilon_0$ we have

$$(\hat{M}_\epsilon+z)^2\geqslant\hat{V}_\epsilon^2\,,\quad\text{on } \mathscr{C}_0^\infty(\mathbb{R})\times\mathscr{C}_0^\infty(\mathbb{R})\,, \tag{4.33}$$

$$(\hat{M}_\epsilon^{(1)}+z)^2\geqslant\tilde{\beta}\hat{V}_{1\epsilon}^2\,,\quad\text{on } \mathscr{C}_0^\infty(\mathbb{R})\times\mathscr{C}_0^\infty(\mathbb{R})\,, \tag{4.34}$$

$$\begin{aligned}(M_\epsilon^{(2)}+z)^2\geqslant\tilde{\beta}\hat{V}_{2\epsilon}^2\,,\quad\text{on } \mathscr{C}_0^\infty(\mathbb{R}-\{\pm\alpha\sqrt{6}/3\epsilon\}\\ \times\mathscr{C}_0^\infty(\mathbb{R}-\{\pm\alpha\sqrt{6}/3\epsilon\})\,,\end{aligned} \tag{4.35}$$

where $0<\tilde{\beta}<1$.

Proof: It follows from Theorem 4.1 since $\hat{V}''_\epsilon=V''_\epsilon$, $\hat{V}_\epsilon^2\geqslant V_\epsilon^2$, and $\hat{c}>0$ and from the similar statements for $\hat{V}_{1\epsilon}$, $V_{1\epsilon}$, $\hat{V}_{2\epsilon}$, $V_{2\epsilon}$.

Theorem 4.3: There exist $z_0>0$ and $\epsilon_0>0$ such that for $z\geqslant z_0$ and $0<\epsilon<\epsilon_0$ we have

$$\|(\hat{M}_\epsilon+z)^{-1}\psi\|\leqslant\|\hat{V}_\epsilon^{-1}\psi\|\,,\quad\forall\psi\in L^2(\mathbb{R})\,, \tag{4.36}$$

$$\|\hat{M}_\epsilon^{(1)}+z)^{-1}\psi\|\leqslant(1/\tilde{\beta}^{1/2})\|\hat{V}_{1\epsilon}^{-1}\psi\|\,,\quad\forall\psi\in L^2(\mathbb{R})\,, \tag{4.37}$$

$$\|(M_\epsilon^{(2)}+z)^{-1}\psi\|\leqslant(1/\tilde{\beta}^{1/2})\|\hat{V}_{2\epsilon}^{-1}\psi\|\,,\quad\forall\psi\in L^2(\mathbb{R})\,. \tag{4.38}$$

Proof: Note that (4.27), (4.28), and (4.29) imply $\hat{V}_\epsilon$, $\hat{V}_{1\epsilon}$, $\hat{V}_{2\epsilon}\geqslant\text{const}>0$ so that $\hat{V}_\epsilon^{-1}$, $\hat{V}_{1\epsilon}^{-1}$, $\hat{V}_{2\epsilon}^{-1}$ are bounded operators. The proof of Theorem 4.3 follows immediately from Theorem 2.21, p. 330 of Kato.[17]

Definition 4.4: Let P_1 be the projection on the subspace of the functions of $L^2(\mathbb{R})$ that have support on $I_1^{(\epsilon)}\cup I_2^{(\epsilon)}$. That is, P_1 is the multiplication operator given by $\chi_{I_1^{(\epsilon)}\cup I_2^{(\epsilon)}}$.

Definition 4.5: Let P_2 be the projection on the subspace of the functions of $L^2(\mathbb{R})$ that have support on $\mathbb{R}-U^{(\epsilon)}$, where

$$\begin{aligned}U^{(\epsilon)}=\{y\,|\,|y|<\bar{\eta}_1(\epsilon)\}\cup\{y\,|\,|y-(\alpha\sqrt{2}/\epsilon)|<\bar{\eta}_2(\epsilon)\}\\ \cup\{y\,|\,|y+(\alpha\sqrt{2}/\epsilon)|<\bar{\eta}_2(\epsilon)\}\,.\end{aligned}$$

That is, P_2 is the multiplication operator by $\chi_{\mathbb{R}-U^{(\epsilon)}}$. Let us now choose

$$\bar{\eta}_1(\epsilon)=\epsilon^{-\delta_1}\,,\quad 0<\delta_1<\tfrac{1}{3}\,. \tag{4.39}$$

Then $\bar{\eta}_2(\epsilon)$ will remain determined by Eq. (3.7).

Theorem 4.6: Let $\bar{\eta}_1(\epsilon)$ be given by (4.39) and $\bar{\eta}_2(\epsilon)$ be determined by (3.7). Then for $0<\epsilon<\epsilon_0$ we have the following estimates:

$$\|(\hat{V}_{2\epsilon}-\hat{V}_{1\epsilon})(I-P_1)\|=0\,, \tag{4.40}$$

$$\|\hat{V}_{2\epsilon}^{-1}(\hat{V}_{2\epsilon}-\hat{V}_{1\epsilon})P_1\|\leqslant\text{const}\,, \tag{4.41}$$

$$\|\hat{V}_{1\epsilon}^{-1}P_1\|\leqslant\text{const}\,\epsilon^{2\delta_1}\,, \tag{4.42}$$

$$\|(\hat{V}_\epsilon-\hat{V}_{1\epsilon})(I-P_2)\|\leqslant\text{const}\,\epsilon^{1-3\delta_1}\,, \tag{4.43}$$

$$\|\hat{V}_\epsilon^{-1}(\hat{V}_\epsilon-\hat{V}_{1\epsilon})P_2\|\leqslant\text{const}\,, \tag{4.44}$$

$$\|\hat{V}_{1\epsilon}^{-1}P_2\|\leqslant\text{const}\,\epsilon^{2\delta_1}\,, \tag{4.45}$$

where I is the identity on $L^2(\mathbb{R})$ and $\|\cdot\|$ is the operator norm induced by the L^2 norm.

Proof: The proof of (4.40) follows from the fact that on $\mathbb{R}\setminus(I_1^{(\epsilon)}\cup I_2^{(\epsilon)})$ we have $\hat{V}_{1\epsilon}=\hat{V}_{2\epsilon}$. The proof of (4.41) follows from the fact that

$$\hat{V}_{2\epsilon} > \hat{V}_{1\epsilon}, \quad \text{on } I_1^{(\epsilon)} \cup I_2^{(\epsilon)} \tag{4.46}$$

and

$$0 < \hat{V}_{2\epsilon}^{-1}(\hat{V}_{2\epsilon} - \hat{V}_{1\epsilon}) = 1 - (\hat{V}_{1\epsilon}/\hat{V}_{2\epsilon}) < 1. \tag{4.47}$$

The proof of (4.42) follows from the fact that on $I_1^{(\epsilon)} \cup I_2^{(\epsilon)}$ we have

$$\hat{V}_{1\epsilon}(y) = \hat{V}_{2\epsilon}(\bar{\eta}_1(\epsilon)) = V_0 \tan^2 \beta \bar{\eta}_1(\epsilon) + 2\alpha^2 + \hat{c}. \tag{4.48}$$

Therefore, since β and V_0 are given by (3.3) and (3.4) we have

$$\hat{V}_{1\epsilon}(y) > \text{const } \epsilon^{-2\delta_1}, \tag{4.49}$$

for $0 < \epsilon < \epsilon_0$.

Let us prove (4.43). Let us consider the function

$$F_\epsilon(y) = \hat{V}_\epsilon(y) - \hat{V}_{1\epsilon}(y). \tag{4.50}$$

Using the Taylor's formula at $y = 0$ we have

$$F_\epsilon(y) = -3\epsilon^2 y^2 + (F_\epsilon'''(\xi)/3!)\, y^3, \tag{4.51}$$

with ξ an intermediate point in the interval $(0, y)$. When $|y| < \bar{\eta}_1(\epsilon) = \epsilon^{-\delta_1}$ we have

$$|\hat{V}_\epsilon'''(y)| = |120\epsilon^4 y^3 - 96\alpha^2\epsilon^2 y| < 24\epsilon^{2-\delta_1}(5^{2-2\delta_1} + 4\alpha^2) \tag{4.52}$$

and

$$|\hat{V}_{1\epsilon}'''(y)| = 32\alpha^4\beta\, |\cos^{-5}\beta y|\, |\sin\beta y| (2 + \sin^2\beta y)$$
$$< \epsilon 24\sqrt{6}\alpha^3\pi \Big/ \left| \cos^5\left(\frac{\pi}{2} - \frac{3}{\alpha\sqrt{6}}\epsilon^{1-\delta_1}\right)\right|. \tag{4.53}$$

Therefore, when $|y| < \bar{\eta}_1(\epsilon) = \epsilon^{-\delta_1}$ from (4.51), (4.52), and (4.53) we have

$$|F_\epsilon(y)| < \text{const } \epsilon^{1-3\delta_1}. \tag{4.54}$$

When $|y - (\alpha\sqrt{2}/\epsilon)| < \bar{\eta}_2(\epsilon)$ we have $\hat{V}_{1\epsilon}(y) = \hat{V}_{2\epsilon}(y)$, so using the Taylor formula at $y = \alpha\sqrt{2}/\epsilon$ we have

$$F_\epsilon(y) = -6\alpha\sqrt{2}\epsilon(y - (\alpha\sqrt{2}/\epsilon)) - 3\epsilon^2(y - (\alpha\sqrt{2}/\epsilon))^2$$
$$+ (F_\epsilon'''(\xi)/3!)(y - (\alpha\sqrt{2}/\epsilon))^3, \tag{4.55}$$

with ξ an intermediate point in the interval $(\alpha\sqrt{2}/\epsilon, y)$. For $|y - (\alpha\sqrt{2}/\epsilon)| < \bar{\eta}_2(\epsilon)$ we have

$$|\hat{V}_\epsilon'''(y)| < 144\sqrt{2}\alpha^3\epsilon + 624\alpha^2\epsilon^2\bar{\eta}_2$$
$$+ 360\sqrt{2}\alpha\epsilon^3\bar{\eta}_2^2 + 120\epsilon^4\bar{\eta}_2^3 \tag{4.56}$$

and

$$|\hat{V}_{1\epsilon}'''(y)| < \frac{1}{4\nu^2}(96\alpha^4/|y - (\alpha\sqrt{6}/3\epsilon)|^5)$$
$$< 96\alpha^4(\sqrt{2\nu}/|1 - \bar{\eta}_2\sqrt{2\nu}|^5). \tag{4.57}$$

Since Eq. (3.7) implies that $\lim_{\epsilon\to 0} \bar{\eta}_1(\epsilon)/\bar{\eta}_2(\epsilon) = \text{const} \neq 0$ from (4.55), (4.56), and (4.39) we have

$$|F_\epsilon(y)| < \text{const } \epsilon^{1-3\delta_1}, \tag{4.58}$$

when $|y - (\alpha\sqrt{2}/\epsilon)| < \bar{\eta}_2(\epsilon)$. Reasoning in the same way it can be shown that

$$|F_\epsilon(y)| < \text{const } \epsilon^{1-3\delta_1}, \tag{4.59}$$

when $|y + (\alpha\sqrt{2}/\epsilon)| < \bar{\eta}_2(\epsilon)$. This establishes estimate (4.43).

Let us prove (4.44). From Proposition 2.1 (i) we know that $\hat{V}_\epsilon$ given by (4.27) has three minimizers $y = 0$, $y = \pm(1/\epsilon)((\alpha^2 + \sqrt{4\alpha^4 + 9\epsilon^2})/3)^{1/2}$ and two maximizers at $y = \pm(1/\epsilon)((4\alpha^2 - \sqrt{4\alpha^4 + 9\epsilon^2})/3)^{1/2}$. Moreover

$$\lim_{\epsilon\to 0} \epsilon^2 \hat{V}_\epsilon\left(\pm\frac{1}{\epsilon}\left(\frac{4\alpha^2 - \sqrt{4\alpha^4 + 9\epsilon^2}}{3}\right)^{1/2}\right) = \frac{32}{27}\alpha^6, \tag{4.60}$$

$$\lim_{\epsilon\to 0} \hat{V}_\epsilon\left(\pm\frac{1}{\epsilon}\left(\frac{4\alpha^2 + \sqrt{4\alpha^4 + 9\epsilon^2}}{3}\right)^{1/2}\right) = -4\alpha^2 + \hat{c}, \tag{4.61}$$

and for $0 < \epsilon < \epsilon_0$

$$\frac{1}{\epsilon}\left(\frac{4\alpha^2 - \sqrt{4\alpha^4 + 9\epsilon^2}}{3}\right)^{1/2} \in I_1^{(\epsilon)},$$
$$-\frac{1}{\epsilon}\left(\frac{4\alpha^2 - \sqrt{4\alpha^4 + 9\epsilon^2}}{3}\right)^{1/2} \in I_2^{(\epsilon)}, \tag{4.62}$$

and

$$\frac{1}{\epsilon}\left(\frac{4\alpha^2 + \sqrt{4\alpha^4 + 9\epsilon^2}}{3}\right)^{1/2} \notin I_1^{(\epsilon)},$$
$$-\frac{1}{\epsilon}\left(\frac{4\alpha^2 + \sqrt{4\alpha^4 + 9\epsilon^2}}{3}\right)^{1/2} \notin I_2^{(\epsilon)}. \tag{4.63}$$

Let $y \in I_1^{(\epsilon)} \cup I_2^{(\epsilon)}$. Then $\hat{V}_{1\epsilon}(y) = V_{2\epsilon}(\bar{\eta}_1(\epsilon)) + \hat{c}$ so that

$$|\hat{V}_\epsilon^{-1}(\hat{V}_\epsilon - \hat{V}_{1\epsilon})| = \left|1 - \frac{\hat{V}_{1\epsilon}}{\hat{V}_\epsilon}\right| < 1 + \frac{V_{2\epsilon}(\bar{\eta}_1(\epsilon)) + \hat{c}}{m(\epsilon)}, \tag{4.64}$$

where

$$m(\epsilon) = \min_{y \in I_1^{(\epsilon)} \cup I_2^{(\epsilon)}} \hat{V}_\epsilon(y) = \min_{y \in I_1^{(\epsilon)}} \hat{V}_\epsilon(y)$$
$$= \min\{\hat{V}_\epsilon(\bar{\eta}_1(\epsilon)), \hat{V}_\epsilon((\alpha\sqrt{2}/\epsilon) - \bar{\eta}_2(\epsilon))\}. \tag{4.65}$$

Equation (4.65) follows from the fact that $\hat{V}_\epsilon$ is even, and from (4.62) and (4.63).

An elementary computation now shows that

$$|\hat{V}_\epsilon^{-1}(\hat{V}_\epsilon - \hat{V}_{1\epsilon})| < \text{const}, \quad \text{for } 0 < \epsilon < \epsilon_0,$$
$$\text{when } y \in I_1^{(\epsilon)} \cup I_2^{(\epsilon)}. \tag{4.66}$$

Let $y > (\alpha\sqrt{2}/\epsilon) + \bar{\eta}_2(\epsilon)$. We have $V_{1\epsilon}(y) = V_{2\epsilon}(y)$. Define $y' = y - (\alpha\sqrt{6}/3\epsilon)$. We therefore have $y' > \bar{\eta}_2(\epsilon) + (1/\sqrt{2\nu})$ and

$$V_{1\epsilon}(y) = 4\alpha^4(y' - (1/2\nu y'))^2 - 4\alpha^2$$
$$= 4\alpha^4\left(y' - \frac{1}{\sqrt{2\nu}}\right)^2\left(1 + \frac{1}{y'\sqrt{2\nu}}\right)^2 - 4\alpha^2, \tag{4.67}$$

since, when $y' > \bar{\eta}_2(\epsilon) + (1/\sqrt{2\nu})$ we have $(1 + (1/y'\sqrt{2\nu}))^2 < 4$. It follows

$$V_{1\epsilon}(y) < 16\alpha^4(y - (\alpha\sqrt{2}/\epsilon))^2 - 4\alpha^2$$
$$\text{when } y > (\alpha\sqrt{2}/\epsilon) + \bar{\eta}_2(\epsilon). \tag{4.68}$$

Moreover

$$V_\epsilon(y) = -4\alpha^2 - 6\sqrt{2}\alpha\epsilon\left(y - \frac{\alpha\sqrt{2}}{\epsilon}\right)$$
$$+ (16\alpha^4 - 3\epsilon^2)\left(y - \frac{\alpha\sqrt{2}}{\epsilon}\right)^2$$
$$+ 24\sqrt{2}\alpha^3\epsilon\left(y - \frac{\alpha\sqrt{2}}{\epsilon}\right)^3 + 26\alpha^2\epsilon^2\left(y - \frac{\alpha\sqrt{2}}{\epsilon}\right)^4$$
$$+ 6\sqrt{2}\alpha\epsilon^3\left(y - \frac{\alpha\sqrt{2}}{\epsilon}\right)^5 + \epsilon^4\left(y - \frac{\alpha\sqrt{2}}{\epsilon}\right)^6. \tag{4.69}$$

From (4.68) and (4.69) when $0<\epsilon<\epsilon_0$ and $y>(\alpha\sqrt{2}/\epsilon)+\bar{\eta}_2(\epsilon)$ we have

$$\left|\frac{V_{1\epsilon}}{V_\epsilon}\right|\leqslant\left(16\alpha^4+\frac{4}{\bar{\eta}_2^2(\epsilon)}\right)\times\left(\left|-\frac{4\alpha^2}{\bar{\eta}_2^2(\epsilon)}-\frac{6\sqrt{2}\alpha\epsilon}{\bar{\eta}_2(\epsilon)}+16\alpha^2-3\epsilon^2\right|\right)^{-1}, \tag{4.70}$$

so that $|V_{1\epsilon}/V_\epsilon|\leqslant$const when $0<\epsilon<\epsilon_0$. That is,

$$|\hat{V}_\epsilon^{-1}(\hat{V}_\epsilon-\hat{V}_{1\epsilon})|\leqslant\text{const},\quad \text{when } 0<\epsilon<\epsilon_0,\quad y>(\alpha\sqrt{2}/\epsilon)+\eta_2(\epsilon). \tag{4.71}$$

Reasoning in the same way it can be shown that

$$|\hat{V}_\epsilon^{-1}(\hat{V}_\epsilon-\hat{V}_{1\epsilon})|<\text{const},\quad \text{when } 0<\epsilon<\epsilon_0,\quad y<-(\alpha\sqrt{2}/\epsilon)-\bar{\eta}_2(\epsilon). \tag{4.72}$$

The equations (4.66), (4.71), and (4.72) establish (4.44).

Let us prove (4.45). When $y\in I_1^{(\epsilon)}\cup I_2^{(\epsilon)}$ we have

$$\hat{V}_{1\epsilon}(y)=V_{2\epsilon}(\bar{\eta}_1(\epsilon))+c \tag{4.73}$$

and

$$\lim_{\epsilon\to 0}\bar{\eta}_1^{-2}(\epsilon)V_{2\epsilon}(\bar{\eta}_1(\epsilon))=\text{const}\neq 0. \tag{4.74}$$

From (4.39) it follows that

$$|\hat{V}_{1\epsilon}^{-1}|\leqslant\text{const }\epsilon^{2\delta_1},\quad \text{when } 0<\epsilon<\epsilon_0,\quad y\in I_1^{(\epsilon)}\cup I_2^{(\epsilon)}. \tag{4.75}$$

Moreover

$$V_{1\epsilon}(y)=V_{2\epsilon}(y)\geqslant V_{2\epsilon}((\alpha\sqrt{2}/\epsilon)+\bar{\eta}_2(\epsilon)),\quad \text{when } 0<\epsilon<\epsilon_0,\quad y>(\alpha\sqrt{2}/\epsilon)+\bar{\eta}_2(\epsilon) \tag{4.76}$$

and

$$\lim_{\epsilon\to 0}\bar{\eta}_2^{-2}(\epsilon)V_{2\epsilon}((\alpha\sqrt{2}/\epsilon)+\bar{\eta}_2(\epsilon))=\text{const}\neq 0. \tag{4.77}$$

Since $\bar{\eta}_1(\epsilon)$, $\bar{\eta}_2(\epsilon)$ are of the same order as $\epsilon\to 0$ from (4.76) and (4.77) it follows

$$|\hat{V}_{1\epsilon}^{-1}|<\text{const }\epsilon^{2\delta_1},\quad \text{when } 0<\epsilon<\epsilon_0,\quad y>(\alpha\sqrt{2}/\epsilon)+\bar{\eta}_2(\epsilon). \tag{4.78}$$

Reasoning in the same way it can be shown that

$$|\hat{V}_{1\epsilon}^{-1}|<\text{const }\epsilon^{2\delta_1},\quad \text{when } 0<\epsilon<\epsilon_0,\quad y<-(\alpha\sqrt{2}/\epsilon)-\bar{\eta}_2(\epsilon). \tag{4.79}$$

The equations (4.75), (4.78), and (4.79) establish (4.45).

This completes the proof of Theorem 4.6.

Theorem 4.7: There exist constants $z_0>0$, $\epsilon_0>0$ such that when $z>z_0$ and $0<\epsilon<\epsilon_0$ we have

$$(N_\epsilon+z)^2\geqslant U^2,\quad \text{on } \mathscr{C}_0^\infty(\mathbb{R})\times\mathscr{C}_0^\infty(\mathbb{R}), \tag{4.80}$$

$$(N_\epsilon^{(1)}+z)^2\geqslant\tilde{\beta}U_{1\epsilon}^2,\quad \text{on } \mathscr{C}_0^\infty(\mathbb{R})\times\mathscr{C}_0^\infty(\mathbb{R}), \tag{4.81}$$

$$(N_\epsilon^{(2)}+z)^2\geqslant\tilde{\beta}U_{2\epsilon}^2,\quad \text{on } \mathscr{C}_0^\infty(\mathbb{R}-\{2y_1-\eta_2\}-\{\eta_2\})\times\mathscr{C}_0^\infty(\mathbb{R}-\{2y_1-\eta_2\}-\{\eta_2\}), \tag{4.82}$$

where $0<\tilde{\beta}<1$.

Proof: Let us first prove (4.80). Proceeding as in the proof of (4.1) we can show that

$$(N_\epsilon+z)^2-U_\epsilon^2\geqslant 2zU_\epsilon+z^2-U_\epsilon''. \tag{4.83}$$

Therefore, to prove (4.80) it will be enough to show that for $z>z_0$, $0<\epsilon<\epsilon_0$ we have

$$F_2(y)=2zU_\epsilon+z^2-U_\epsilon''\geqslant 0,\quad y\in\mathbb{R}. \tag{4.84}$$

A simple computation shows that

$$\begin{aligned}F_2(y)={}&2a^2\epsilon^4zy^6+(6ab/\sqrt{2})\epsilon^3zy^5+\{2z(\tfrac{9}{8}b^2+2ac)\epsilon^2\\&-30a^2\epsilon^4\}y^4+\{(6/\sqrt{2})bc\epsilon z-(60ab/\sqrt{2})\epsilon^3\}y^3\\&+\{2z(c^2-3a\epsilon^2)-12(\tfrac{9}{8}b^2+2ac)\epsilon^2\}y^2\\&-\{(6b/\sqrt{2})\epsilon z+(18bc/\sqrt{2})\epsilon\}y-2zc-2c^2\\&+6a\epsilon^2+z^2.\end{aligned} \tag{4.85}$$

Let $t=\epsilon y$. Rearranging the terms in (4.85) we have

$$\begin{aligned}F_2\left(\frac{t}{\epsilon}\right)={}&2z\left\{\left[\frac{1}{\epsilon^2}\frac{1}{4}t^2\left(2at^2+\frac{3b}{\sqrt{2}}t+2c\right)^2-\left(3at^2+\frac{3b}{\sqrt{2}}t+c\right)+\frac{z}{2}\right]\right.\\&-\frac{1}{2z}\left[30a^2t^4+\frac{60ab}{\sqrt{2}}t^3+12\left(\frac{9}{8}b^2+2ac\right)t^2\right.\\&\left.\left.+\frac{18}{\sqrt{2}}bct+2c^2-6a\epsilon^2\right]\right\}.\end{aligned} \tag{4.86}$$

For $z>z_0$ and $0<\epsilon<\epsilon_0$ the expression (4.86) will be positive for any $t\in\mathbb{R}$. This proves (4.80).

The proof of (4.81) and (4.82) can be obtained from the proof of (4.2) and (4.3) with only minor changes and will be omitted.

Let $\hat{c}_*$ be a constant such that

$$\hat{U}_\epsilon=U_\epsilon+\hat{c}_*,\quad \text{and } (U_\epsilon+\hat{c}_*)^2\geqslant U_\epsilon^2, \tag{4.87}$$

$$\hat{U}_{1\epsilon}=U_{1\epsilon}+\hat{c}_*,\quad \text{and } (U_{1\epsilon}+\hat{c}_*)^2\geqslant U_{1\epsilon}^2, \tag{4.88}$$

$$\hat{U}_{2\epsilon}=U_{2\epsilon}+\hat{c}_*,\quad \text{and } (U_{2\epsilon}+\hat{c}_*)^2U_{2\epsilon}^2. \tag{4.89}$$

We define

$$\hat{N}_\epsilon=-\frac{d^2}{dy^2}+\hat{U}_\epsilon=N_\epsilon+\hat{c}_*, \tag{4.90}$$

$$\hat{N}_{1\epsilon}=-\frac{d^2}{dy^2}+\hat{U}_{1\epsilon}=N_{1\epsilon}+\hat{c}_*, \tag{4.91}$$

$$\hat{N}_{2\epsilon}=-\frac{d^2}{dy^2}+\hat{U}_{2\epsilon}=N_{2\epsilon}+\hat{c}_*. \tag{4.92}$$

Theorem 4.8: There exist $z_0>0$ and $\epsilon_0>0$ such that for $z>z_0$ and $0<\epsilon<\epsilon_0$ we have

$$(\hat{N}_\epsilon+z)^2\geqslant\hat{U}_\epsilon^2,\quad \text{on } \mathscr{C}_0^\infty(\mathbb{R})\times\mathscr{C}_0^\infty(\mathbb{R}), \tag{4.93}$$

$$(\hat{N}_\epsilon^{(1)}+z)^2\geqslant\tilde{\beta}\hat{U}_{1\epsilon}^2,\quad \text{on } \mathscr{C}(\mathbb{R})\times\mathscr{C}_0^\infty(\mathbb{R}), \tag{4.94}$$

$$(N_\epsilon^{(2)}+z)^2\geqslant\tilde{\beta}\hat{U}_{2\epsilon}^2,\quad \text{on } \mathscr{C}_0^\infty(\mathbb{R}-\{2y_1-\eta_2\}-\{\eta_2\})\times\mathscr{C}_0^\infty(\mathbb{R}-\{2y_1-\eta_2\}-\{\eta_2\}), \tag{4.95}$$

where $0<\tilde{\beta}<1$.

Proof: It follows from Theorem 4.7 since $\hat{U}_\epsilon''=U_\epsilon''$, $\hat{U}_\epsilon^2\geqslant U_\epsilon^2$, and $\hat{c}_*>0$ and from the similar statements for $\hat{U}_{1\epsilon}$, $U_{1\epsilon}$, $\hat{U}_{2\epsilon}$, $U_{2\epsilon}$.

Theorem 4.9: There exist $z_0>0$ and $\epsilon_0>0$ such that for $z\geqslant z_0$ and $0<\epsilon<\epsilon_0$ we have

$$\|(\hat{N}_\epsilon+z)^{-1}\psi\|\leqslant\|\hat{U}_\epsilon^{-1}\psi\|,\quad \forall\psi\in L^2(\mathbb{R}), \tag{4.96}$$

$$\|(\hat N_\epsilon^{(1)}+z)^{-1}\psi\| \leqslant \frac{1}{\tilde\beta^{1/2}}\|\hat U_{1\epsilon}^{-1}\psi\|, \quad \forall\psi\in L^2(\mathbb{R}), \tag{4.97}$$

$$\|(\hat N_\epsilon^{(2)}+z)^{-1}\psi\| \leqslant \frac{1}{\tilde\beta^{1/2}}\|\hat U_{2\epsilon}^{-1}\psi\|, \quad \forall\psi\in L^2(\mathbb{R}), \tag{4.98}$$

where $0<\tilde\beta<1$.

Proof: It follows immediately from Theorem 2.21, p. 330 of Kato.[17]

Definition 4.10: Let P_1^* be the projection on the subspace of the functions of $L^2(\mathbb{R})$ that have support on $J_1^{(\epsilon)}\cup J_2^{(\epsilon)}$.

Definition 4.11: Let P_2^* be the projection on the subspace of the functions of $L^2(\mathbb{R})$ that have support on $\mathbb{R}\setminus U_*^{(\epsilon)}$, where

$$U_*^{(\epsilon)} = \{y \,|\, |y|<\mu_1(\epsilon)\}\cup\{y \,|\, |y-y_1|<\mu_2(\epsilon)\} \cup\{y \,|\, |y-y_2|<\mu_3(\epsilon)\}.$$

Let us now choose

$$\mu_1(\epsilon)=\epsilon^{-\delta_1}, \quad 0<\delta_1<\tfrac{1}{3}. \tag{4.99}$$

Then $\mu_2(\epsilon)$ and $\mu_3(\epsilon)$ will remain determined by Eqs. (3.28) and (3.29).

Theorem 4.12: Let $\mu_1(\epsilon)$ be given by (4.99) and $\mu_2(\epsilon)$, $\mu_3(\epsilon)$ be determined by (3.28) and (3.29). Then for $0<\epsilon<\epsilon_0$ we have the following estimates:

$$\|(\hat U_{2\epsilon}-\hat U_{1\epsilon})(I-P_1^*)\| = 0, \tag{4.100}$$

$$\|\hat U_{2\epsilon}^{-1}(\hat U_{2\epsilon}-\hat U_{1\epsilon})P_1^*\| \leqslant \text{const}, \tag{4.101}$$

$$\|\hat U_{1\epsilon}^{-1}P_1^*\| \leqslant \text{const}\,\epsilon^{2\delta_1}, \tag{4.102}$$

$$\|(\hat U_\epsilon-\hat U_{1\epsilon})(I-P_2^*)\| \leqslant \text{const}\,\epsilon^{1-3\delta_1}, \tag{4.103}$$

$$\|\hat U_\epsilon^{-1}(\hat U_\epsilon-\hat U_{1\epsilon})P_2^*\| \leqslant \text{const}, \tag{4.104}$$

$$\|\hat U_{1\epsilon}^{-1}P_2^*\| \leqslant \text{const}\,\epsilon^{2\delta_1}. \tag{4.105}$$

Proof: The estimates (4.100), (4.101),...,(4.105) can be proved as the corresponding estimates (4.40), (4.41),...,(4.45) of Theorem 4.6.

V. THE BEHAVIOR AS $\epsilon\to 0$ OF EIGENVALUES AND EIGENVECTORS OF M_ϵ, N_ϵ

Let us first make precise in which sense $\hat M_\epsilon$ is approximated by $\hat M_\epsilon^{(1)}$, $\hat M_\epsilon^{(2)}$ and $\hat N_\epsilon$ is approximated by $\hat N_\epsilon^{(1)}$, $\hat N_\epsilon^{(2)}$.

Theorem 5.1: There exist constants A, $z_0>0$, $\epsilon_0>0$, $\delta_1^*>0$ such that for $z>z_0$, $0<\epsilon<\epsilon_0$ we have

$$\|(\hat M_\epsilon^{(2)}+z)^{-1}-(\hat M_\epsilon+z)^{-1}\| \leqslant A\epsilon^{\delta_1^*}, \tag{5.1}$$

$$\|(\hat M_\epsilon^{(2)}+z)^{-1}-(\hat M_\epsilon^{(1)}+z)^{-1}\| \leqslant A\epsilon^{\delta_1^*}, \tag{5.2}$$

$$\|(\hat M_\epsilon^{(1)}+z)^{-1}-(\hat M_\epsilon+z)^{-1}\| \leqslant A\epsilon^{\delta_1^*}, \tag{5.3}$$

$$\|\hat N_\epsilon^{(2)}+z)^{-1}-(\hat N_\epsilon+z)^{-1}\| \leqslant A\epsilon^{\delta_1^*}, \tag{5.4}$$

$$\|(\hat N_\epsilon^{(2)}+z)^{-1}-(\hat N_\epsilon^{(1)}+\epsilon)^{-1}\| \leqslant A\epsilon^{\delta_1^*}, \tag{5.5}$$

$$\|(\hat N_\epsilon^{(1)}+z)^{-1}-(\hat N_\epsilon+z)^{-1}\| \leqslant A\epsilon^{\delta_1^*}. \tag{5.6}$$

Proof: The proof of (5.1), (5.2), and (5.3) follows from Theorems 4.3 and 4.6, reasoning as in Isaacson,[2] Theorem 3.1. Similarly, the proof of (5.4), (5.5), and (5.6) follows from Theorems 4.9 and 4.12.

Let us remark that (5.1) and (5.4) say that the resolvent of M_ϵ converges to the resolvent of $M_\epsilon^{(2)}$ and the resolvent of N_ϵ converges to the resolvent of $N_\epsilon^{(2)}$ as $\epsilon\to 0$. In Sec. III we have studied the eigenvalues and eigenfunctions of $M_\epsilon^{(2)}$ and $N_\epsilon^{(2)}$; here we will see the consequences of (5.1) and (5.4) on the eigenvalues of M_ϵ, N_ϵ.

Let $P_\epsilon(S)$ and $P_\epsilon^{(2)}(S)$ be the spectral projectors of M_ϵ and $M_\epsilon^{(2)}$ associated with the Borel set $S\subset\mathbb{C}$.

The eigenvalues of $M_\epsilon^{(2)}$, (3.17) and (3.20), when $\epsilon\to 0$, are given by

$$\lambda_{n\epsilon}^{\pm} = 8\alpha^2 n + \mathcal{O}(\epsilon^2), \quad n=0,1,2,\ldots, \tag{5.7}$$

$$\lambda_{n\epsilon}^{0} = 4\alpha^2(n+1) + \mathcal{O}(\epsilon^2), \quad n=0,1,2,\ldots \tag{5.8}$$

(see Fig. 9). We remark here that $\lambda_{n\epsilon}^{\pm}$ has multiplicity 2 and $\lambda_{n\epsilon}^{0}$ has multiplicity 1. Let

$$C_k(r) = \{z \,|\, |z-4\alpha^2 k| = r\}, \quad k=0,1,2,\ldots, \tag{5.9}$$

and

$$D_k = \{z \,|\, |z-4\alpha^2 k| \leqslant r\}, \quad k=0,1,2,\ldots, \tag{5.10}$$

with $r\leqslant\alpha^2$, and let

$$P_\epsilon^{(2)}(D_k) = \frac{1}{2\pi i}\oint_{C_k(r)} (z-M_\epsilon^{(2)})^{-1}\,dz. \tag{5.11}$$

Then

$$P_\epsilon^{(2)}(D_0) = P_\epsilon^{(2)}(\{\lambda_{0\epsilon}^{\pm}\}), \quad \text{for } \epsilon \text{ small enough}, \tag{5.12}$$

$$P_\epsilon^{(2)}(D_k) = \begin{cases} P_\epsilon^{(2)}(\{\lambda_{k-1,\epsilon}^{0}\}), & k \text{ odd}, \\ P_\epsilon^{(2)}(\{\lambda_{k-1,\epsilon}^{0}\}\cup\{\lambda_{k/2,\epsilon}^{\pm}\}), & k \text{ even}, \end{cases} \tag{5.13}$$

for $\epsilon\leqslant\bar\epsilon_k$ (see Fig. 9). We remark that $\bar\epsilon_k$ cannot be chosen independent of k.

Theorem 5.2: There exists $\bar\epsilon_k>0$ such that for all $z\in C_k(r)$ and all $0<\epsilon<\bar\epsilon_k$

$$(z-M_\epsilon)^{-1} \tag{5.14}$$

exists, and

$$\sup_{z\in C_k(r)} \|(z-M_\epsilon)^{-1}-(z-M_\epsilon^{(2)})^{-1}\| \leqslant \text{const}\,\epsilon^{\delta_1^*/2}. \tag{5.15}$$

Proof: It follows from Theorem 5.1 and the known properties of the spectrum of $M_\epsilon^{(2)}$, rearranging the proof of Theorem 4.1 of Isaacson.[2]

Theorem 5.3: For $k=0,1,2,\ldots$, we have

$$\lim_{\epsilon\to 0}\|P_\epsilon(D_k)-P_\epsilon^{(2)}(D_k)\| = 0.$$

Moreover for all ϵ sufficiently small M_ϵ possesses the following.

(i) Two distinct eigenvalues $\mu_0(\epsilon)\equiv 0$, $\mu_0'(\epsilon)>0$ such that

$$\lim_{\epsilon\to 0}\mu_0'(\epsilon) = \mu_0(\epsilon)\equiv 0. \tag{5.16}$$

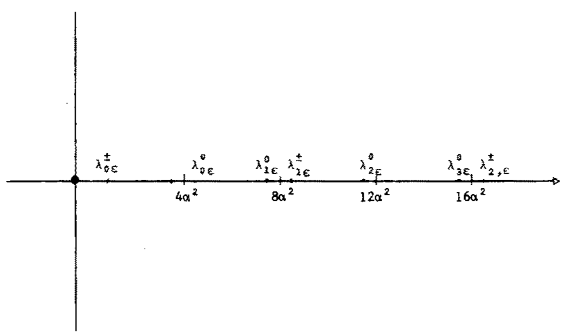


FIG. 9. The spectrum of $M_\epsilon^{(2)}$.

(ii) When k is odd, one eigenvalue $\mu_k(\epsilon)$ such that

$$\lim_{\epsilon\to 0}\mu_k(\epsilon) = 4\alpha^2 k\,, \quad k = 1,3,\dots\,. \tag{5.17}$$

(iii) When k is even, three distinct eigenvalues $\mu_k(\epsilon)$, $\mu'_k(\epsilon)$, $\mu''_k(\epsilon)$ such that

$$\lim_{\epsilon\to 0}\mu_k(\epsilon) = \lim_{\epsilon\to 0}\mu'_k(\epsilon) = \lim_{\epsilon\to 0}\mu''_k(\epsilon) = 4\alpha^2 k\,, \quad k = 2,4,\dots\,. \tag{5.18}$$

Proof: From (5.15) of Theorem 5.2 we have

$$\|P_\epsilon(D_k) - P_\epsilon^{(2)}(D_k)\| = \left\| \left| \frac{1}{2\pi i}\oint_{C_k(r)} [(z-M_\epsilon)^{-1} - (z - M_\epsilon^{(2)})^{-1}]dz \right| \right\| \leqslant \text{const}\; r\epsilon^{\delta \dagger/2}\,. \tag{5.19}$$

Therefore, for ϵ sufficiently small

$$\dim P_\epsilon(D_k) = \dim P_\epsilon^{(2)}(D_k)\,. \tag{5.20}$$

The remaining part of Theorem 5.3 follows from (5.12), (5.13), (5.7), and (5.8).

Let us now establish the results announced in Sec. II.

Theorem 5.4: Let $0 \equiv -\lambda_0(\epsilon) < -\lambda_1(\epsilon) < -\lambda_2(\epsilon) < \dots$ be the eigenvalues of M_ϵ. Then

$$\lim_{\epsilon\to 0} -\lambda_1(\epsilon) = 0\,, \tag{5.21}$$

$$\lim_{\epsilon\to 0} -\lambda_{2+4n}(\epsilon) = 4\alpha^2(2n+1)\,, \quad n = 0,1,2,\dots\,, \tag{5.22}$$

$$\lim_{\epsilon\to 0} -\lambda_{3+4n}(\epsilon) = \lim_{\epsilon\to 0} -\lambda_{4+4n}(\epsilon) = \lim_{\epsilon\to 0} -\lambda_{5+4n}(\epsilon) = 8\alpha^2(n+1)\,, \quad n = 0,1,2,\dots\,. \tag{5.23}$$

Proof: Let

$$S_k = \{z = x + iy |\, -1 \leqslant x \leqslant 4\alpha^2 k + 2\alpha^2,\, -1 \leqslant y \leqslant -1\}\,,\quad k = 0,1,\dots\,.$$

By estimates analogous to the ones of Theorem 5.2 it is possible to show that

$$\lim_{\epsilon\to 0}\|P_\epsilon(S_k) - P_\epsilon^{(2)}(S_k)\| = 0\,.$$

That is, for ϵ sufficiently small

$$\dim P_\epsilon(S_k) = \dim P_\epsilon^{(2)}(S_k)\,.$$

Theorem 5.4 follows now from Theorem 5.3.

A straightforward computation shows that the eigenvalues of $N_\epsilon^{(2)}$, when $\epsilon\to 0$, are given by

$$-\bar\lambda_{n\epsilon}^{(1)} = c(2n+1) - c + \mathscr{O}(\epsilon^2)\,, \quad n = 0,1,2,\dots\,, \tag{5.24}$$

$$-\bar\lambda_{n\epsilon}^{(2)} = |c_1|(2n+1) - c_1 + \mathscr{O}(\epsilon^2)\,, \quad n = 0,1,2,\dots\,, \tag{5.25}$$

$$-\bar\lambda_{n\epsilon}^{(3)} = c_2(2n+1) - c_2 + \mathscr{O}(\epsilon^2)\,, \quad n = 0,1,2,\dots\,, \tag{5.26}$$

where $c = 2ax_1x_2$, $c_1 = 2ax_1(x_1 - x_2) < 0$, $c_2 = 2ax_2(x_2 - x_1)$ and x_1, x_2 are given in (i) of Proposition 2.2.

Let $\{-\bar\lambda_n\}_{n=0}^\infty$ be the set obtained reordering the numbers of $E_1 = \{c(2n+1) - c\}_{n=0}^\infty$, $E_2 = \{|c_1|(2n+1) - c_1\}_{n=0}^\infty$, $E_3 = \{c_2(2n+1) - c_2\}_{n=0}^\infty$ in such a way that $-\bar\lambda_n \leqslant -\bar\lambda_{n+1}$, $n = 0,1,\dots$. Moreover if a number appears in more than one E_i, $i = 1,2,3$ it will appear a corresponding number of times in $\{-\bar\lambda_n\}_{n=0}^\infty$. In particular, since zero appears in E_1 and E_3 we will have $-\bar\lambda_0 = -\bar\lambda_1 = 0$.

Theorem 5.5: Let $0 \equiv -\bar\lambda_0(\epsilon) < -\bar\lambda_1(\epsilon) < \dots$ be the eigenvalues of N_ϵ. Then

$$\lim_{\epsilon\to 0} -\bar\lambda_n(\epsilon) = -\bar\lambda_n\,. \tag{5.27}$$

Proof: The proof can be obtained from (5.24), (5.25), and (5.26) rearranging the proofs of Theorems 5.2, 5.3, and 5.4.

We remark that when a certain value appears more than once in $\{-\bar\lambda_n\}_{n=0}^\infty$ this corresponds to asymptotic eigenvalue degeneracy for N_ϵ.

Since $-\bar\lambda_0 = -\bar\lambda_1 = 0$, we have

$$\lim_{\epsilon\to 0} -\bar\lambda_1(\epsilon) = -\bar\lambda_0(\epsilon) \equiv 0\,.$$

All the remaining $\{-\bar\lambda_n\}_{n=2}^\infty$ are distinct if x_1/x_2 is irrational; if x_1/x_2 is rational $\{-\bar\lambda_n\}_{n=2}^\infty$ contains values that appear only once and values that appear three times.

That is, there are eigenvalues of N_ϵ that remain isolated when $\epsilon\to 0$ and eigenvalues that have asymptotic multiplicity 3 when $\epsilon\to 0$. We have already observed this phenomenon in the study of M_ϵ.

VI. THE ESTIMATE OF THE FIRST NONZERO EIGENVALUE OF M_ϵ AND N_ϵ

In Sec. V it has been shown that

$$-\lambda_0(\epsilon) = -\bar\lambda_0(\epsilon) \equiv 0\,, \quad \forall \epsilon \neq 0\,, \tag{6.1}$$

$$\lim_{\epsilon\to 0} -\lambda_1(\epsilon) = \lim_{\epsilon\to 0} -\bar\lambda_1(\epsilon) = 0\,, \tag{6.2}$$

where $-\lambda_0(\epsilon)$, $-\lambda_1(\epsilon) > 0$ are the first two eigenvalues of M_ϵ and $-\bar\lambda_0(\epsilon)$, $-\bar\lambda_1(\epsilon) > 0$ are the first two eigenvalues of N_ϵ.

In Sec. II it has been shown that the eigenfunctions corresponding to $-\lambda_0(\epsilon)$ and $-\bar\lambda_0(\epsilon)$ are, respectively,

$$v_0(y) = d_\epsilon^{1/2} e^{-f_{1\epsilon}/2} \tag{6.3}$$

and

$$\bar v_0(y) = \bar d_\epsilon^{1/2} e^{-f_{2\epsilon}/2}\,, \tag{6.4}$$

where f_1, f_2 are given by (2.13) and (2.14), $f_{1\epsilon}, f_{2\epsilon}$ by (2.4),

$$d_\epsilon = \left(\int_{-\infty}^{+\infty} e^{-f_{1\epsilon}/2}\,dy\right)^{-1} = \frac{\epsilon}{\sqrt{2}}\,c_\epsilon\,, \tag{6.5}$$

$$\bar d_\epsilon = \left(\int_{-\infty}^{+\infty} e^{-\bar f_{2\epsilon}/2}\,dy\right)^{-1} = \frac{\epsilon}{\sqrt{2}}\,\bar c_\epsilon \tag{6.6}$$

are normalization constants such that $\|v_0\|_{L^2(\mathbf{R})} = \|\bar v_0\|_{L^2(\mathbf{R})} = 1$, and $c_\epsilon, \bar c_\epsilon$ are given by (2.11).

Using the Rayleigh–Ritz principle (see Ref. 6, p. 78, Theorem XIII.2) we want to estimate the quantities $-\lambda_1(\epsilon) + \lambda_0(\epsilon)$ and $-\bar\lambda_1(\epsilon) + \bar\lambda_0(\epsilon)$ as $\epsilon\to 0$, that is, the first nonzero eigenvalue of M_ϵ and N_ϵ.

The same problem for the Fokker–Planck operators corresponding to M_ϵ and N_ϵ and for some more general Fokker–Planck operators has been considered by Matkowsky–Schuss in Ref. 10.

Matkowsky–Schuss[10] used the technique of matching asymptotic expansions. The results obtained here using the Rayleigh–Ritz principle are contained in the ones obtained by Matkowsky–Schuss,[10] but are derived in a more elementary way.

Theorem 6.1: Let $-\lambda_0(\epsilon)$, $-\lambda_1(\epsilon)$, M_ϵ be as above. Then as $\epsilon \to 0$ we have

$$0 < -\lambda_1(\epsilon) + \lambda_0(\epsilon) \equiv -\lambda_1(\epsilon) \leqslant \text{const}\, e^{-(2/\epsilon^2)\alpha^4}. \quad (6.7)$$

Proof: From the Rayleigh–Ritz principle (see Ref. 6, p. 78, Theorem XIII.2) we have

$$0 < -\lambda_1(\epsilon) + \lambda_0(\epsilon) \equiv -\lambda_1(\epsilon) \leqslant \frac{\langle g, M_\epsilon g\rangle_{L^2(\mathbb{R})}}{\langle g, g\rangle_{L^2(\mathbb{R})}}, \quad (6.8)$$

where $g \in L^2(\mathbb{R})$ is any function orthogonal to v_0 [given by (6.3)] that belongs to the domain of M_ϵ as a form.

Since v_0 is an even function let us choose

$$g = uv_0, \quad (6.9)$$

where $u(y) = -u(-y)$ is an odd function such that $u \in L^\infty(\mathbb{R})$ and $du/dy \in L^\infty(\mathbb{R})$, where du/dy is the distributional derivative of u.

The function g is orthogonal to v_0 and belongs to the form domain of M_ϵ.

We have

$$\begin{aligned}\langle g, M_\epsilon g\rangle_{L^2(\mathbb{R})} &= \int_{-\infty}^{+\infty} uv_0\left(-\frac{d^2}{dy^2} + V_\epsilon\right)uv_0\, dy \\ &= \int_{-\infty}^{+\infty} \left\{\left[\frac{d}{dy}(uv_0)\right]^2 + V_\epsilon u^2 v_0^2\right\}dy \\ &= \int_{-\infty}^{+\infty} \left\{\left(\frac{du}{dy}\right)^2 v_0^2 + u^2\left(\frac{dv_0}{dy}\right)^2 \right. \\ &\quad \left. + 2u\frac{du}{dy}v_0\frac{dv_0}{dy} + V_\epsilon u^2 v_0^2\right\}dy \\ &= \int_{-\infty}^{+\infty} \left(\frac{du}{dy}\right)^2 v_0^2\, dy, \end{aligned} \quad (6.10)$$

since

$$\begin{aligned}\int_{-\infty}^{+\infty} u^2\left(\frac{dv_0}{dy}\right)^2 dy &= -\int_{-\infty}^{+\infty} v_0 \frac{d}{dy}\left(u^2\frac{dv_0}{dy}\right)dy \\ &= -\int_{-\infty}^{+\infty}\left\{2u\frac{du}{dy}v_0\frac{dv_0}{dy} \right. \\ &\quad \left. + u^2 v_0\frac{d^2 v_0}{dy^2}\right\}dy, \end{aligned} \quad (6.11)$$

and $M_\epsilon v_0 = -d^2v_0/dy^2 + V_\epsilon v_0 = 0$. Therefore

$$0 < -\lambda_1(\epsilon) + \lambda_0(\epsilon) \equiv -\lambda_1(\epsilon) \leqslant \frac{\int_{-\infty}^{+\infty}(du/dy)^2 v_0^2\, dy}{\int_{-\infty}^{+\infty} u^2 v_0^2\, dy}. \quad (6.12)$$

Let us choose

$$u(y) = \begin{cases} 1, & y > 1, \\ y, & |y| < 1, \\ -1, & y < -1. \end{cases}$$

Equation (6.12) becomes

$$0 < -\lambda_1(\epsilon) + \lambda_0(\epsilon) = -\lambda_1(\epsilon) \leqslant \frac{\int_{-1}^{1} v_0^2\, dy}{\int_{-\infty}^{+\infty} u^2 v_0^2\, dy}. \quad (6.13)$$

Moreover

$$\begin{aligned}\int_{-1}^{1} v_0^2\, dy &= d_\epsilon \int_{-1}^{+1} e^{-f_{1\epsilon}(y)}\, dy = \frac{\epsilon}{\sqrt{2}} c_\epsilon \int_{-1}^{1} e^{-f_{1\epsilon}(\epsilon)}\, dy \\ &\leqslant \frac{\epsilon}{\sqrt{2}} c_\epsilon 2e^{-f_{1\epsilon}(1)} \\ &= \frac{\epsilon}{\sqrt{2}} c_\epsilon 2e^{+2\alpha^2} e^{-\epsilon^2/2} e^{-(2/\epsilon^2)\alpha^4}. \end{aligned} \quad (6.14)$$

It can be easily shown that

$$\lim_{\epsilon\to 0} \epsilon c_\epsilon = \sqrt{2/\pi}\, \alpha, \quad (6.15)$$

and that, since $x = (\epsilon/\sqrt{2})y$,

$$\begin{aligned}&\lim_{\epsilon\to 0}\int_{-\infty}^{+\infty} v_0^2 u^2\, dy \\ &= \lim_{\epsilon\to 0}\int_{-\infty}^{+\infty} c_\epsilon e^{-(2/\epsilon^2)f_1(x)} u^2\left(\frac{\sqrt{2}}{\epsilon}x\right)dx = 1. \end{aligned} \quad (6.16)$$

In fact, in the sense of distribution

$$\lim_{\epsilon\to 0} c_\epsilon e^{-(2/\epsilon^2)f_1(x)} = \tfrac{1}{2}(\delta(x-\alpha) + \delta(x+\alpha)), \quad (6.17)$$

where $\delta(\cdot)$ is the Dirac's delta. Theorem 6.1 now follows from (6.13), (6.14), (6.15), and (6.16).

We remark that since $\alpha^4 = f_1(0) - f_1(\alpha)$ the estimate (6.13) agrees with the one of Matkowsky–Schuss.[10]

Theorem 6.2: Let $-\bar{\lambda}_0(\epsilon)$, $-\bar{\lambda}_1(\epsilon)$, N_ϵ be as above. Then as $\epsilon \to 0$ we have

$$\begin{aligned}0 &< -\bar{\lambda}_1(\epsilon) + \bar{\lambda}_0(\epsilon) \equiv \lambda_1(\epsilon) \\ &\leqslant \text{const}\exp(-(2/\epsilon^2)(f_2(x_1) - f_2(x_2))), \end{aligned} \quad (6.18)$$

where x_1 and x_2 are given in Proposition 2.2 (i) (see Fig. 2).

Proof: Reasoning as in the proof of Theorem 6.1 we have

$$0 < -\bar{\lambda}_1(\epsilon) + \bar{\lambda}_0(\epsilon) \equiv -\bar{\lambda}_1(\epsilon) \leqslant \frac{\int_{-\infty}^{+\infty}(dh/dy)^2 \bar{v}_0^2\, dy}{\int_{-\infty}^{+\infty} h^2 \bar{v}_0^2\, dy}, \quad (6.19)$$

where $g = h\bar{v}_0 \in L^2(\mathbb{R})$ is a function orthogonal to $\bar{v}_0$ such that $h \in L^\infty(\mathbb{R})$ and $dh/dy \in L^\infty(\mathbb{R})$.

Let us choose

$$h = \bar{u} - \langle \bar{u}\bar{v}_0, \bar{v}_0\rangle_{L^2(\mathbb{R})}, \quad (6.20)$$

where

$$\bar{u}(y) = \begin{cases} 1, & y > y_1 + 1, \\ y - y_1, & |y - y_1| < 1, \\ -1, & y < y_1 - 1, \end{cases} \quad (6.21)$$

where $y_1 = (\sqrt{2}/\epsilon)x_1$.

Reasoning as in Theorem 6.1 it can be shown that

$$\int_{-\infty}^{+\infty}\left(\frac{dh}{dy}\right)^2 \bar{v}_0^2\, dy \leqslant \text{const}\, e^{-(2/\epsilon^2)f_2(x_1)}. \quad (6.22)$$

Moreover

$$\int_{-\infty}^{+\infty} h^2\bar{v}_0^2\, dy = \int_{-\infty}^{+\infty} \bar{u}^2\bar{v}_0^2\, dy - \left(\int_{-\infty}^{+\infty} \bar{u}\bar{v}_0^2\, dy\right)^2, \quad (6.23)$$

and

$$\lim_{\epsilon\to 0}\int_{-\infty}^{+\infty} \bar{u}^2\bar{v}_0^2\, dy = 1, \quad (6.24)$$

$$\lim_{\epsilon\to 0} \frac{\int_{-\infty}^{+\infty} \bar{u}\bar{v}_0^2\, dy + 1}{e^{-2/\epsilon^2 f_2(x_2)}} = \sqrt{\frac{f_2''(0)}{f_2''(x_2)}}\,. \tag{6.25}$$

Theorem 6.2 now follows from (6.19), (6.22), (6.23), (6.24), and (6.25).

VII. CONCLUSIONS

Let $f(x)\in\mathscr{C}^3(\mathbb{R})$ be such that $e^{-(2/\epsilon^2)f(x)}\in L^1(\mathbb{R})$, $\forall\epsilon\neq 0$ and suppose that

$$f'(x) = 0 \tag{7.1}$$

has n roots $\xi_1,\xi_2,\ldots,\xi_n$ such that

$$f''(\xi_i) = \alpha_i \neq 0\,, \quad i = 1,2,\ldots,n\,. \tag{7.2}$$

That is, $\xi_1,\xi_2,\ldots,\xi_n$ are nondegenerate minimizers or maximizers of f.

Let

$$L_\epsilon(\,\cdot\,) = \frac{\epsilon^2}{2}\frac{\partial^2\cdot}{\partial x^2} + \frac{\partial}{\partial x}\left(\frac{df}{dx}\cdot\right), \tag{7.3}$$

the Fokker–Planck operator associated to f.

Proceeding as in Sec. II, the study of the spectrum of (7.3) can be reduced to the study of the spectrum of

$$H_\epsilon = -\frac{d^2}{dy^2} + W_\epsilon(y) \tag{7.4}$$

on $L^2(\mathbb{R})$, where $W_\epsilon(y)$ is given by (2.7).

Let $y_i = (\sqrt{2}/\epsilon)\xi_i$, $i = 1,2,\ldots,k$. A straightforward computation shows that as $\epsilon\to 0$, $W_\epsilon(y)$ approaches n decoupled harmonic oscillator potentials $\frac{1}{4}\alpha_i^2(y-y_i)^2 - \frac{1}{2}\alpha_i$.

Therefore, we expect the spectrum of H_ϵ to approximate the spectrum of n decoupled harmonic oscillators $\lambda_k^i = \frac{1}{2}|\alpha_i|(2k+1) + \frac{1}{2}\alpha_i$, $i = 1,2,\ldots,n$ and $k = 0,1,2,\ldots$.

In particular, if f has m ($<n$) minimizers, that is, $\alpha_{i_j} > 0$, $j = 1,2,\ldots,m$ we expect the eigenvalue zero of H_ϵ (or L_ϵ) to have asymptotically multiplicity m when $\epsilon\to 0$.

ACKNOWLEDGMENT

The research reported in this document has been made possible through the support and sponsorship of the United States Government through its European Research Office of the United States Army under Contract No. DAJA-37-81-C-0740.

[1]C. J. Thompson and M. Kac, Stud. Appl. Math. **48**, 257 (1969).

[2]D. Isaacson, Commun. Pure Appl. Math. **29**, 531 (1976).

[3]D. Isaacson and D. Marchesin, Commun. Pure Appl. Math. **31**, 659 (1978).

[4]E. Harrell, Commun. Math. Phys. **60**, 73 (1978).

[5]E. Harrell, Commun. Math. Phys. **75**, 239 (1980).

[6]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1978), Vol. IV.

[7]J. Glimm and A. Jaffe, *Quantum Physics: A Functional Integral Point of View* (Springer, New York, 1981).

[8]Z. Schuss, SIAM Rev. **22**, 119 (1980).

[9]S. Chandresekhar, "Stochastic problems in physics and astronomy," in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York, 1954).

[10]B. J. Matkowsky and Z. Schuss, SIAM J. Appl. Math. **40**, 242 (1981).

[11]F. Aluffi-Pentini, V. Parisi, and F. Zirilli, "Global optimization and stochastic differential equations," to appear in J. Optim. Theory Appl.

[12]B. Simon, Ann. Phys. (NY) **58**, 76 (1970).

[13]F. Zirilli, J. Math. Phys. **15**, 1202 (1974).

[14]E. C. Titchmarsh, *Eigenfunction Expansions* (Oxford U. P., London, 1962), Part I.

[15]S. Flügge, *Rechenmethoden der Quantentheorie: 1 Teil Elementare Quantenmechanik* (Springer–Verlag, Berlin, 1947).

[16]J. Weidmann, *Linear Operators in Hilbert Spaces* (Springer, New York, 1980).

[17]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).

Stochastic difference equations for a spin system

H. Falk
Department of Physics, City College of the City University of New York, New York, New York 10031

(Received 1 June 1984; accepted for publication 14 September 1984)

An N-spin model is given with a discrete-time evolution specified by a system of stochastic difference equations. A Markov chain associated with the evolution is decomposed into two, nonhomogeneous, absorbing Markov chains. Analysis of each chain yields the probability, given a specific initial state, of ultimate absorption into a specific state. As time $t \to \infty$ the spin model will, with probability equal to 1, have all spins up, all spins down, or will oscillate between two antiferromagnetic (Néel) states. The time-dependent correlation functions $\langle s_i(t) s_j(t)\rangle$ are also obtained.

I. INTRODUCTION

Stochastic spin models are often closely related to a model which Glauber introduced. Such models are typically written in terms of a master equation[1] describing the (continuous) time evolution of the probability of a state of an N-spin system. The master equation contains specified transition rates for spin flips—usually single or double spin flips.[2] The transition rates are functions of a few spins which constitute the local environment of a given spin.

Similar models have been studied in a discrete-time[3] framework as finite Markov chains. In that context the Glauber model may be viewed as a generalization of the Ehrenfest urn model.[4]

Work on models of the above types and on models containing many-spin transitions extends into areas outside of conventional physics. There are voter models, cell-growth models, neural-net models, etc., some of which appear to have originated in Russian journals on cybernetics and information theory. A subset of these models is reviewed in a very approachable monograph by Kindermann and Snell[5] and in a review article by Durrett.[6] A recent review article by Wolfram[7] contains extensive computer simulation results directed towards a classification of the behavior of cellular automata.

The purpose of the present paper is to describe the behavior of a model involving a row of N Ising spins (± 1 variables). The evolution of the spin system is given by a discrete-time, nonlinear system of stochastic difference equations. At each time step many spins may flip.

The behavior of the spin system is expressed in terms of a finite Markov chain.[8] The Markov chain is decomposed into two Markov chains operating in disjoint space-time sublattices. Each of the latter two Markov chains is an absorbing Markov chain with two ferromagnetic absorbing states. Given any state of the spins on the independent sublattices, we find the probabilities of ultimate absorption into the absorbing states. In other words, given the initial state of the N spins, we have the probabilities of the spin system ultimately achieving a ferromagnetic state of all spins down, of ultimately achieving a ferromagnetic state of all spins up, or of oscillating forever between two antiferromagnetic (Néel) states; no other behavior is possible.

Additionally, we have the two-spin, time-dependent correlation functions $\langle s_i(t) s_j(t)\rangle$.

Aspects of these exact, analytic results have been tested by computer simulations. For $N = 20$, large variations in the absorption time are observed.

II. SYSTEM OF STOCHASTIC DIFFERENCE EQUATIONS

Consider a one-dimensional lattice of sites labeled by the integers $1,2,\ldots,N$ so that site j has left neighbor $j-1$ and right neighbor $j+1$, for $j = 2,3,\ldots,N-1$. Site 1 has site 2 as its only neighbor and site N has site $N-1$ as its only neighbor. N is taken to be an even integer not less than 4.

Associate with each site a time-dependent spin variable $s_j(t)$ restricted to the values ± 1. The time parameter t is discrete:

$$s_j(t) = \pm 1, \quad \text{for } t = 0,1,\ldots . \tag{2.1}$$

The evolution of the spin variables is given by the following stochastic process: at time t, spin $s_j(t)$ ($j \geqslant 2$) "looks" at its two neighbor spins $s_{j-1}(t)$ and $s_{j+1}(t)$. If the two neighbor spins $s_{j-1}(t)$ and $s_{j+1}(t)$ are parallel, then $s_j(t+1)$ assumes the value of $s_{j-1}(t)$ [or equivalently, $s_{j+1}(t)$]. If the two neighbor spins are antiparallel, then $s_j(t)$ "tosses" a coin and $s_j(t+1)$ assumes the value $+1$ if the coin displays "heads" and -1 if the coin displays "tails." Each end spin at time $t+1$ assumes the value its neighbor had at time t. This process may be expressed in terms of the following nonlinear system of stochastic difference equations:

$$\begin{aligned} s_1(t+1) &= s_2(t) \\ s_j(t+1) &= \tfrac{1}{2}[s_{j-1}(t) + s_{j+1}(t)] \\ &\quad + \tfrac{1}{2}[1 - s_{j-1}(t) s_{j+1}(t)]\theta_j(t), \\ s_N(t+1) &= s_{N-1}(t), \end{aligned} \tag{2.2}$$

where $j = 2,3,\ldots, N-1$ and $t = 0,1,\ldots$.

The random variables $\theta_j(t)$ for $j = 2,3,\ldots,N-1$ and $t = 0,1,\ldots, m^*$ are statistically independent and identically distributed according to the prescription

$$\begin{aligned} \theta_j(t) &= +1, \quad \text{with probability } \tfrac{1}{2} + \epsilon, \\ \theta_j(t) &= -1, \quad \text{with probability } \tfrac{1}{2} - \epsilon, \end{aligned} \tag{2.3}$$

where $|\epsilon| < \frac{1}{2}$ and m^* is any positive integer. The coin is not necessarily "fair"; consequently, ϵ is not necessarily zero.

Next consider a Markov chain associated with the above system of stochastic difference equations.

 0022-2488/85/040692-06$02.50

III. ASSOCIATED MARKOV CHAIN

To lighten the notation, write

$$s_j' = s_j(t+1),\qquad s_j = s_j(t), \tag{3.1}$$

and use the relation

$$\delta(s,s'') = \tfrac{1}{2}(1+ss''), \tag{3.2}$$

for the Kronecker delta $\delta(s,s'')$ involving any spin variables s,s'' restricted to the values ± 1. Then the stochastic process described in the preceding section can be associated with a probability $p(s_1',...,s_N'|s_1,...,s_N)$ of the spin system realizing the state $(s_1',...,s_N')$ at time $t+1$, given that the system was in state $(s_1,...,s_N)$ at time t:

$$\begin{aligned} p(s_1',...,s_N'|s_1,...,s_N) &= \frac{1}{2}(1+s_1's_2)\Bigg\{\prod_{j=2}^{N-1}\Bigg[\frac{1}{2}(1+s_{j-1}s_{j+1})\\ &\quad\times\tfrac{1}{2}(1+s_js_{j+1})\delta(s_j',s_j)\\ &\quad+\tfrac{1}{2}(1+s_{j-1}s_{j+1})\tfrac{1}{2}(1-s_js_{j+1})\delta(s_j',-s_j)\\ &\quad+\tfrac{1}{2}(1-s_{j-1}s_{j+1})\big[(\tfrac{1}{2}+\epsilon)\delta(s_j',1)\\ &\quad+(\tfrac{1}{2}-\epsilon)\delta(s_j',-1)\big]\Bigg]\Bigg\}\tfrac{1}{2}(1+s_N's_{N-1}). \end{aligned} \tag{3.3}$$

This one-step transition matrix defines a Markov chain associated with the system of stochastic difference equations (2.2).

IV. DECOMPOSITION OF THE MARKOV CHAIN

When one looks at a two-dimensional lattice (one space dimension and one time dimension) shown in Fig. 1, one sees that the system (2.2) decomposes into two systems. The two-dimensional space-time lattice decomposes into two disjoint sublattices. This decomposition enables one to factor the Markov chain transition probability (3.3). The latter factorization significantly simplifies the problem and is accomplished in the following way.

Agree to let s_j relate to time $2t$, s_j' to time $2t+1$, and s_j'' to time $2t+2$. Then introduce new variables (for $j=1,2,...,N/2$)

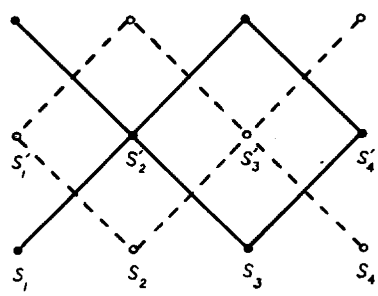


FIG. 1. The sublattice decomposition for a system of $N=4$ spins evolving according to the stochastic difference equations (2.2).

$$\begin{aligned} u_j &= s_{2j}, & u_j' &= s_{2j-1}',\\ v_j &= s_{2j-1}, & v_j' &= s_{2j}'; \end{aligned} \tag{4.1}$$

$$\begin{aligned} v_j' &= s_{2j}', & v_j'' &= s_{2j-1}'',\\ u_j' &= s_{2j-1}', & u_j'' &= s_{2j}''. \end{aligned} \tag{4.2}$$

From Fig. 1 one sees that the u variables refer to the sublattice with the dashed-line bonds and the v variables refer to the sublattice with solid-line bonds.

The transition probability for the time step $2t$ to $2t+1$ is

$$\begin{aligned} p(s_1',...,s_N'|s_1,...,s_N) &= p_1(u_1',...,u_{N/2}'|u_1,...,u_{N/2})\\ &\quad\times p_2(v_1',...,v_{N/2}'|v_1,...,v_{N/2}), \end{aligned} \tag{4.3}$$

and for the time step $2t+1$ to $2t+2$

$$\begin{aligned} p(s_1'',...,s_N''|s_1',...,s_N') &= p_2(u_1'',...,u_{N/2}''|u_1',...,u_{N/2}')\\ &\quad\times p_1(v_1'',...,v_{N/2}''|v_1',...,v_{N/2}'), \end{aligned} \tag{4.4}$$

where

$$\begin{aligned} p_1(x_1',...,x_{N/2}'|x_1,...,x_{N/2}) &= \frac{1}{2}(1+x_1'x_1)\Bigg\{\prod_{j=2}^{N/2}\Bigg[\frac{1}{2}\Big[1+\frac{1}{2}(x_{j-1}+x_j)x_j'\Big]\\ &\quad+\frac{1}{2}(1-x_{j-1}x_j)x_j'\epsilon\Bigg]\Bigg\}, \end{aligned} \tag{4.5}$$

and

$$\begin{aligned} p_2(x_1',...,x_{N/2}'|x_1,...,x_{N/2}) &= \Bigg\{\prod_{j=1}^{N/2-1}\Bigg[\frac{1}{2}\Big[1+\frac{1}{2}(x_j+x_{j+1})x_j'\Big]\\ &\quad+\frac{1}{2}(1-x_jx_{j+1})x_j'\epsilon\Bigg]\Bigg\}\\ &\quad\times\tfrac{1}{2}(1+x_{N/2}'x_{N/2}), \end{aligned} \tag{4.6}$$

in terms of arbitrary spin variables x_j, $x_j' = \pm 1$ for $j=1,2,...,N/2$.

Let P denote the $2^N\times 2^N$ matrix with elements $p(s_1',...,s_N'|s_1,...,s_N)$. Here, P is a stochastic matrix since each of its elements is in the interval [0,1], and the elements in each column sum to 1.

Similary, let P_1 denote the $2^{N/2}\times 2^{N/2}$ stochastic matrix with elements $p_1(u_1',...,u_{N/2}'|u_1,...,u_{N/2})$; let P_2 denote the $2^{N/2}\times 2^{N/2}$ stochastic matrix with elements $p_2(u_1'',...,u_{N/2}''|u_1',...,u_{N/2}')$. To help distinguish the two sublattices, let Q_1 denote the stochastic matrix with elements $p_1(v_1'',...,v_{N/2}''|v_1',...,v_{N/2}')$ and let Q_2 denote the stochastic matrix with elements $p_2(v_1',...,v_{N/2}'|v_1,...v_{N/2})$.

With the above definitions one has the direct-product factorization

$$\begin{aligned} P &= P_1\otimes Q_2 \text{ for the step } 2t\to 2t+1,\\ P &= P_2\otimes Q_1 \text{ for the step } 2t+1\to 2t+2, \end{aligned} \tag{4.7}$$

and for $t = 0,1,2,...,$

$$p^{2t} = (P_2P_1)^t \otimes (Q_1Q_2)^t,$$
$$p^{2t+1} = [P_1(P_2P_1)^t] \otimes [Q_2(Q_1Q_2)^t]. \quad (4.8)$$

This factorization, which provides a decomposition into two nonhomogeneous Markov chains, clearly introduces no statistical connection between the two sublattices. Consequently, if spins on one sublattice are initially statistically independent with respect to spins on the other, then the statistical independence will persist for all times $t = 0,1,2,...$. In the following, we restrict the choices of initial probability so as to gain that statistical independence. One is then able to obtain results for the N-spin system by studying the $N/2$ spins on one sublattice.

V. SYMMETRY PROPERTIES OF THE TRANSITION MATRICES

For definiteness, focus attention on the $N/2$ spins $u_1,...,u_{N/2}$ located on the sublattice with dashed bonds [see Fig. 1 and Eqs. (4.1) and (4.2)]. The transition matrices relating to those spins are P_1 and P_2, with respective matrix elements $p_1(u_1',...,u_{N/2}'|u_1,...,u_{N/2})$ and $p_2(u_1'',...,u_{N/2}''|u_1',...,u_{N/2}')$ defined by Eqs. (4.5) and (4.6).

Let $\mathbf{u}$ denote the $N/2$-tuple $(u_1,...,u_{N/2})$ and denote the matrix elements of P_1 and P_2 by

$$p_k^{(\epsilon)}(\mathbf{u}'|\mathbf{u}) = p_k(u_1',...,u_{N/2}'|u_1,...,u_{N/2}), \quad (5.1)$$

where $k = 1,2$. Note that ϵ, which now appears explicitly on the left side, was partly suppressed in the defining equations (4.5) and (4.6).

Here are some important, easily established properties of the transition matrices. Matrix elements will be written as in (5.1) and also in the Dirac notation form as in (5.3) below.

Property I:

$$p_k^{(\epsilon)}(\mathbf{u}'|\mathbf{u}) = p_k^{(-\epsilon)}(-\mathbf{u}'|-\mathbf{u}). \quad (5.2)$$

In terms of the matrix elements, such as

$$\langle \mathbf{u}'|(P_2P_1)|\mathbf{u}\rangle^{(\epsilon)} = \sum_{\mathbf{u}''} p_2^{(\epsilon)}(\mathbf{u}'|\mathbf{u}'')p_1^{(\epsilon)}(\mathbf{u}''|\mathbf{u}), \quad (5.3)$$

etc., the latter property leads to

Property II:

$$\langle \mathbf{u}'|(P_2P_1)^t|\mathbf{u}\rangle^{(\epsilon)} = \langle -\mathbf{u}'|(P_2P_1)^t|-\mathbf{u}\rangle^{(-\epsilon)}, \quad (5.4)$$

and

Property III:

$$\langle \mathbf{u}'|P_1(P_2P_1)^t|\mathbf{u}\rangle^{(\epsilon)} = \langle -\mathbf{u}'|P_1(P_2P_1)^t|-\mathbf{u}\rangle^{(-\epsilon)}, \quad (5.5)$$

for $t = 0,1,2,...$.

To gain a measure of the importance of the above properties, consider the time-dependent expectation

$$\langle u_j(2t)\rangle_{\mathbf{u}^0;\epsilon} = \sum_{\mathbf{u}',\mathbf{u}} u_j' \langle \mathbf{u}'|[(P_2P_1)^t]|\mathbf{u}\rangle^{(\epsilon)}\delta(\mathbf{u},\mathbf{u}^0), \quad (5.6)$$

subject to the initial sublattice spin configuration $\mathbf{u}^0$. Then

$$\langle u_j(2t)\rangle_{\mathbf{u}^0;\epsilon} = \langle u_j(2t)\rangle_{-\mathbf{u}^0;-\epsilon},$$
$$\text{for } j = 1,...,N/2; \quad t = 0,1,..., \quad (5.7)$$

and the result also obtains for $2t$ replaced by $2t + 1$. So one has

$$\langle u_j(t)\rangle_{\mathbf{u}^0;\epsilon} = -\langle u_j(t)\rangle_{\mathbf{u}^0;-\epsilon},$$
$$\text{for } j = 1,...,N/2; \quad t = 0,1,... . \quad (5.8)$$

Also,

$$\langle u_j(t)u_m(t)\rangle_{\mathbf{u}^0;\epsilon} = \langle u_j(t)u_m(t)\rangle_{-\mathbf{u}^0;-\epsilon}. \quad (5.9)$$

Two additional properties of the transition matrices are expressed in terms of the permutation matrix R with matrix elements

$$\langle \mathbf{u}'|R|\mathbf{u}\rangle = \tfrac{1}{2}(1 + u_{N/2}'u_1)\tfrac{1}{2}(1 + u_{N/2-1}'u_2)\cdots\tfrac{1}{2}(1 + u_1'u_{N/2}). \quad (5.10)$$

Notice that

$$\langle \mathbf{u}'|R|\mathbf{u}\rangle = \langle \mathbf{u}|R|\mathbf{u}'\rangle, \quad (5.11)$$
$$\langle \mathbf{u}'|R|\mathbf{u}\rangle = \langle -u'|R|-\mathbf{u}\rangle, \quad (5.12)$$
$$R^2 = I. \quad (5.13)$$

Since R is real, symmetric, and satisfies the last equation, the eigenvalues of R are ± 1.

From the definitions of P_1, P_2, and R one has Property IV.

Property IV:

$$R^{-1}P_1R = P_2. \quad (5.14)$$

Thus, P_1 and P_2 have the same characteristic equation and the same spectrum.

Additionally, since $R = R^{-1}$, one has Property V.

Property V:

$$(P_2P_1)^t = (RP_1)^{2t}, \quad \text{for } t = 0,1,... . \quad (5.15)$$

VI. THE LIMIT $t\to\infty$

P_1, P_2, and R are all stochastic matrices, so any product of those matrices is a stochastic matrix. Of particular interest is the stochastic matrix P_2P_1, since the time dependence of the Markov chain for the sublattice spins is essentially determined by $(P_2P_1)^t$. For $0 \leqslant |\epsilon| < \frac{1}{2}$, the matrix P_2P_1 is a transition matrix for a Markov chain with only two absorbing states. One absorbing state has all (sublattice) spins $u_j = 1$ $(j = 1,...,N/2)$. Thus, for $t\to\infty$ the probability of all sublattice spins down added to the probability of all sublattice spins up yields 1. It follows that, irrespective of the initial state of the one-dimensional N-spin system (see the last paragraph of Sec. IV), the two-dimensional space-time lattice will ultimately (with probability unity) have both sublattices parallel up (all spins $+1$), both sublattices parallel down (all spins -1), or the sublattices mutually antiparallel. Therefore, the N-spin system ultimately (with probability unity) has all spins up, all spins down, or it oscillates between two antiferromagnetic (Néel type) states.

To establish the above assertions, which were made for $0 \leqslant |\epsilon| < \frac{1}{2}$, let $\mathbf{u}^+$ denote the $(N/2)$-tuple $(1,...,1)$, and let $\mathbf{u}^-$ denote $(-1,..., -1)$. Then from (4.5), (4.6), and (5.1)

$$\langle \mathbf{u}'|P_2P_1|\mathbf{u}\rangle^{(\epsilon)} = \sum_{\mathbf{u}''} p_2^{(\epsilon)}(\mathbf{u}'|\mathbf{u}'')p_1^{(\epsilon)}(\mathbf{u}''|\mathbf{u})$$
$$= \delta(\mathbf{u}',\mathbf{u}^+)\delta(\mathbf{u},\mathbf{u}^+) + \delta(\mathbf{u}',\mathbf{u}^-)\delta(\mathbf{u},\mathbf{u}^-)$$
$$+ [1-\delta(\mathbf{u},\mathbf{u}^+)][1-\delta(\mathbf{u},\mathbf{u}^-)]$$
$$\times \sum_{\mathbf{u}''\neq\mathbf{u}^\pm} p_2^{(\epsilon)}(\mathbf{u}'|\mathbf{u}'')p_1^{(\epsilon)}(\mathbf{u}''|\mathbf{u}). \tag{6.1}$$

For $0<|\epsilon|<\frac{1}{2}$ the matrix element

$$\langle \mathbf{u}'|P_2P_1|\mathbf{u}\rangle^{(\epsilon)} = \delta(\mathbf{u}',\mathbf{u}^+)\delta(\mathbf{u},\mathbf{u}^+) + \delta(\mathbf{u}',\mathbf{u}^-)\delta(\mathbf{u},\mathbf{u}^-), \tag{6.2}$$

for $\mathbf{u}=\mathbf{u}^+$ or $\mathbf{u}=\mathbf{u}^-$.

For any $\mathbf{u}\neq\mathbf{u}^+$ or $\mathbf{u}^-$

$$\sum_{\mathbf{u}'\neq\mathbf{u}^\pm} \langle \mathbf{u}'|P_2P_1|\mathbf{u}\rangle < 1. \tag{6.3}$$

This establishes that P_2P_1 is a transition matrix for an absorbing Markov chain with $\mathbf{u}^+$ and $\mathbf{u}^-$ as the only absorbing states.

Now consider the initial sublattice spin configuration $\mathbf{u}^0 = (u_{1;0},\ldots,u_{N/2;0})$. Given $\mathbf{u}^0$, what is the probability $p_+(\mathbf{u}^0)$ of ultimate absorption into the state $\mathbf{u}^+$? [The associated probability $p_-(\mathbf{u}^0) = 1 - p_+(\mathbf{u}^0)$, since the Markov chain has $\mathbf{u}^+$ and $\mathbf{u}^-$ as the only absorbing states.] The calculation of the probabilities $p_+(\mathbf{u}^0)$, $p_-(\mathbf{u}^0)$ generally involves the difficult task of obtaining a so-called fundamental matrix.[8] Here we have a way around this difficulty when $\epsilon=0$. Since we do not have a rigorous proof of the validity of the method, the word "conjecture" will serve as a hedge.

Given the initial sublattice state $\mathbf{u}^0$, the probability that spin j has orientation u_j at time t is[1]

$$p(u_j|\mathbf{u}^0;t) = \tfrac{1}{2}[1 + u_j\langle u_j(t)\rangle], \tag{6.4}$$

where $\langle u_j(t)\rangle$ is the time-dependent average computed by averaging (2.2) over all appropriately weighted sample paths and using (4.1) to relate s_j to the u_i. In the above notation for $\langle u_j(t)\rangle$ the dependence on $\mathbf{u}^0$ is implicit.

Since the two absorbing states have all spins up or all spins down, we conjecture that

$$\begin{aligned} p_+(\mathbf{u}^0) &= t\to\infty \quad \text{limit of } p(1|\mathbf{u}^0;t),\\ p_-(\mathbf{u}^0) &= t\to\infty \quad \text{limit of } p(-1|\mathbf{u}^0;t), \end{aligned} \tag{6.5}$$

where

$$p_-(\mathbf{u}^0) + p_-(\mathbf{u}^0) = 1. \tag{6.6}$$

To compute the right side of (6.4) one uses (2.2). The latter equation provides a means of expressing $s_j(t)$ in terms of the statistically independent quantities $\theta_2(0)$, ..., $\theta_{N-1}(0)$, $\theta_2(1)$, ..., $\theta_{N-1}(1)$, ..., $\theta_2(t-1)$, ..., $\theta_{N-1}(t-1)$; consequently, $\langle s_i(t)\theta_j(t)\rangle = \langle s_i(t)\rangle\langle\theta_j(t)\rangle$ and we have

$$\begin{aligned} \langle s_1(t+1)\rangle &= \langle s_2(t)\rangle,\\ \langle s_j(t+1)\rangle &= \tfrac{1}{2}[\langle s_{j-1}(t)\rangle + \langle s_{j+1}(t)\rangle]\\ &\quad + \tfrac{1}{2}[1 - \langle s_{j-1}(t)s_{j+1}(t)\rangle]\langle\theta_j(t)\rangle,\\ \langle s_N(t+1)\rangle &= s_{N-1}(t)\rangle, \end{aligned} \tag{6.7}$$

where $j = 2,3,\ldots,N-1$, and $t = 0,1,\ldots$.

The system of equations (6.7) which can alternatively be obtained via the Markov chain transition matrix (3.3) is only the beginning of a hierarchy containing single-spin averages $\langle s_j(t)\rangle$, two-spin averages $\langle s_j(t)s_k(t)\rangle$, etc.

For the special case of $\epsilon=0$, however,

$$\langle\theta_j(t)\rangle = 0 \quad (\text{for } \epsilon=0), \tag{6.8}$$

and (6.5) becomes a closed, linear system involving single-spin averages:

$$\langle \mathbf{s}(t+1)\rangle = A\,\langle \mathbf{s}(t)\rangle \quad (\text{for } \epsilon=0), \tag{6.9}$$

where $\mathbf{s}(t)$ is a column vector with transpose $[s_1(t),\ldots,s_N(t)]$, and A is the $N\times N$, nonsymmetric matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & & & 0\\ \frac12 & 0 & \frac12 & & & 0\\ 0 & \frac12 & 0 & \frac12 & & \\ & \vdots & \vdots & \vdots & \vdots & \\ & & \frac12 & 0 & \frac12 & 0\\ 0 & & & \frac12 & 0 & \frac12\\ 0 & & & 0 & 1 & 0 \end{bmatrix}, \tag{6.10}$$

with the spectral representation[3] ($\tilde{\mathbf{w}}_k$ is the transpose of the vector $\mathbf{w}_k$)

$$A = \sum_{k=0}^{N-1} e_k \frac{\mathbf{w}_k\tilde{\mathbf{w}}_k S^{-2}}{\tilde{\mathbf{w}}_k S^{-2}\mathbf{w}_k}, \tag{6.11}$$

where the eigenvalues

$$e_k = \cos(a_k), \tag{6.12}$$

$$a_k = k\pi/(N-1) \quad (k=0,1,\ldots,N-1) \tag{6.13}$$

and corresponding eigenvectors $\mathbf{w}_k$ with components $(\mathbf{w}_k)_j$,

$$(\mathbf{w}_k)_j = w_{jk} = w_{lk}\cos[(j-1)a_k]. \tag{6.14}$$

S denotes a diagonal matrix

$$S = \mathrm{diag}(1,2^{-1/2},2^{-1/2},\ldots,2^{-1/2},1), \tag{6.15}$$

with inverse

$$S^{-1} = \mathrm{diag}(1,2^{1/2},2^{1/2},\ldots,2^{1/2},1). \tag{6.16}$$

The solution of (6.7) is (for $\epsilon=0$)

$$\langle \mathbf{s}(t)\rangle = \sum_{k=0}^{N-1} e_k^t \frac{\mathbf{w}_k\tilde{\mathbf{w}}_k S^{-2}}{\tilde{\mathbf{w}}_k S^{-2}\mathbf{w}_k}\langle \mathbf{s}(0)\rangle. \tag{6.17}$$

Notice that the scale factor w_{lk} drops out of the solution, and for $t\to\infty$ the only surviving terms are the one with $k=0$ (projecting ferromagnetic alignment) and the one with $k=N-1$ (projecting antiferromagnetic alignment). Looking at the limit on even time values, one has (for $\epsilon=0$)

$$t\to\infty \text{ limit of } \langle s_j(2t)\rangle = \sum_{i=1}^{N}\left\{\left(\frac{\mathbf{w}_0\tilde{\mathbf{w}}_0S^{-2}}{\tilde{\mathbf{w}}_0S^{-2}\mathbf{w}_0}\right)_{ij} s_i^0 = \left(\frac{\mathbf{w}_{N-1}\tilde{\mathbf{w}}_{N-1}S^{-2}}{\tilde{\mathbf{w}}_{N-1}S^{-2}\mathbf{w}_{N-1}}\right)_{ij} s_i^0\right\}. \tag{6.18}$$

By using (6.18), (6.4), (6.5), and (4.1) (to relate the s_j to the u_i) one now has

$$\begin{aligned} p_+(\mathbf{u}^0) &= \tfrac12[1 + \langle u_j(\infty)\rangle],\\ p_-(\mathbf{u}^0) &= \tfrac12(1 - \langle u_j(\infty)\rangle), \end{aligned} \tag{6.19}$$

where (for $\epsilon=0$)

$$\langle u_j(\infty)\rangle = t\to\infty \quad \text{limit of } \langle u_j(2t)\rangle$$
$$= \frac{1}{N-1} u^0_{N/2} + \frac{2}{N-1}\sum_{n=1}^{N/2-1} u^0_n, \qquad (6.20)$$
where $u_j^0 = \langle u_j(0)\rangle$.

Equations (6.19) and (6.20) enable one to express the probability $p_+(\mathbf{u}^0)$ of ultimate absorption into the state $\mathbf{u}^+$ for the sublattice starting in the state $\mathbf{u}^0$. The probability of ultimate absorption into the state $\mathbf{u}^-$ for the sublattice starting in the state $\mathbf{u}^0$ is $p_-(\mathbf{u}^0) = 1 - p_+(\mathbf{u}^0)$.

For the other sublattice, with spins labeled with v_j ($j = 1,...,N/2$) according to (4.1), the probability of ultimate absorption into the state $\mathbf{v}^+$ (all sublattice spins up) starting from state $\mathbf{v}^0$ is $p_+(R\mathbf{v}^0)$, where R is the permutation matrix defined in (5.10). Similarly, the probability of ultimate absorption into the state $\mathbf{v}^-$ (all sublattice spins down) starting from state $\mathbf{v}^0$ is $p_-(R\mathbf{v}^0) = 1 - p_+(R\mathbf{v}^0)$. Thus, (6.19) and (6.20) can be utilized for the v sublattice if $\mathbf{u}^0$ is replaced by $R\mathbf{v}^0$.

Since the two sublattices are independent, one can immediately write the probabilities for the four possible limiting spin configurations on the two-dimensional (space-time) lattice; viz.,

$$p_+(\mathbf{u}^0)p_+(R\mathbf{v}^0), \quad p_+(\mathbf{u}^0)p_-(R\mathbf{v}^0),$$
$$p_-(\mathbf{u}^0)p_+(R\mathbf{v}^0), \quad p_-(\mathbf{u}^0)p_-(R\mathbf{v}^0).$$

VII. CORRELATION FUNCTIONS ($\epsilon = 0$)

The term $\frac{1}{2}[1 - s_{j-1}(t)s_{j+1}(t)]\theta_j(t)$ in (2.2) contributes the term

$$\text{term I} = \tfrac{1}{2}(1 - s_{j-1}s_{j+1})[(\tfrac{1}{2}+\epsilon)\delta(s_j',1) + (\tfrac{1}{2}-\epsilon)\delta(s_j',-1)] \qquad (7.1)$$

to the Markov chain transition matrix (3.3). If in (2.2) one replaces $\frac{1}{2}[1 - s_{j-1}(t)s_{j+1}(t)]\theta_j(t)$ by $\frac{1}{2}[s_{j+1}(t) - s_{j-1}(t)]\theta_j(t)$, then term I in the Markov chain transition matrix is replaced by

$$\text{term II} = \tfrac{1}{4}(1 - s_{j-1}s_{j+1})[\delta(s_j',1) + \delta(s_j',-1)] + \tfrac{1}{4}(s_{j+1} - s_{j-1})[2\epsilon\delta(s_j',1) - 2\epsilon\delta(s_j',-1)]. \qquad (7.2)$$

Notice that

$$\text{term I} = \text{term II}, \quad \text{for } \epsilon = 0. \qquad (7.3)$$

This justifies replacing (2.2) by

$$s_1(t+1) = s_2(t),$$
$$s_j(t+1) = \tfrac{1}{2}[s_{j-1}(t) + s_{j+1}(t)] + \tfrac{1}{2}[s_{j+1}(t) - s_{j-1}(t)]\theta_j(t), \qquad (7.4)$$
$$s_N(t+1) = s_{N-1}(t),$$

where $j = 2,3,...,N-1$, $t = 0,1,...$, and $\epsilon = 0$. In that sense the linear stochastic system (7.4) for $\epsilon = 0$ is equivalent to the original nonlinear stochastic system (2.2) for $\epsilon = 0$.

Using vector notation for $\mathbf{s}(t)$ and the matrix A, defined in the preceding section, enables one to write (7.4) in the form

$$\mathbf{s}(t+1) = [A + B(t)]\mathbf{s}(t), \qquad (7.5)$$

where the $N\times N$ matrix A is given by (6.10) and the $N\times N$ random matrix $B(t)$ is

$$B(t) = \tfrac{1}{2}\begin{bmatrix} 0 & 0 & 0 & & & 0 \\ -\theta_2(t) & 0 & \theta_2(t) & & & 0 \\ 0 & -\theta_3(t) & 0 & \theta_3(t) & & \\ & \vdots & \vdots & \vdots & & \\ & & & -\theta_{N-1}(t) & 0 & \theta_{N-1}(t) \\ 0 & 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}, \qquad (7.6)$$

where

$$\theta_j(t) = \pm 1, \quad \text{with probabilities } \tfrac{1}{2}. \qquad (7.7)$$

In the preceding section we used the fact that $\langle B(t)\mathbf{s}(t)\rangle = \langle B(t)\rangle\langle \mathbf{s}(t)\rangle = 0$ to find $\langle \mathbf{s}(t)\rangle$ by solving (6.9).

Now we want to calculate correlations $\langle s_i(t)s_j(t)\rangle$ so we consider the formal solution of (7.5):

$$\mathbf{s}(t+1) = [A + B(t)]\cdots[A + B(1)][A + B(0)]\mathbf{s}(0), \qquad (7.8)$$

and write the direct product

$$\mathbf{s}(t+1)\otimes\mathbf{s}(t+1)$$
$$= \{[A + B(t)]\cdots[A + B(1)][A + B(0)]\mathbf{s}(0)\} \otimes \{[A + B(t)]\cdots[A + B(1)][A + B(0)]\mathbf{s}(0)\} \qquad (7.9)$$
$$= \{[A + B(t)]\otimes[A + B(t)]\}\cdots\{[A + B(0)] \otimes [A + B(0)]\}\{\mathbf{s}(0)\otimes\mathbf{s}(0)\}. \qquad (7.10)$$

The correlation matrix

$$\langle \mathbf{s}(t+1)\otimes\mathbf{s}(t+1)\rangle = \langle [A + B(t)]\otimes[A + B(t)]\rangle\cdots\langle [A + B(0)] \otimes [A + B(0)]\rangle[\mathbf{s}(0)\otimes\mathbf{s}(0)]. \qquad (7.11)$$

But

$$\langle [A + B(t)]\otimes[A + B(t)]\rangle = A\otimes A + \langle B(t)\otimes B(t)\rangle, \qquad (7.12)$$

since $\langle B(t)\rangle = 0$, for $\epsilon = 0$. Furthermore, $\langle B(t)\otimes B(t)\rangle$ is independent of t. This follows from the definition in Sec. II of the independent, identically distributed random variables $\theta_j(t)$. Using $\langle \theta_i(t)\theta_j(t)\rangle = \delta(i,j)$ one finds the explicit result

$$\langle [B(t)\otimes B(t)]_{ik;jm}\rangle = \langle B_{ij}(t)B_{km}(t)\rangle$$
$$= \tfrac{1}{4} z_i z_k \delta(i,k)[-\delta(i,j+1) + \delta(i,j-1)] \times[-\delta(k,m+1) + \delta(k,m-1)], \qquad (7.13)$$

where

$$z_i = \begin{cases} 0, & \text{for } i = 1 \text{ or } N, \\ 1, & \text{for } i = 2,\dots,N-1, \end{cases} \tag{7.14}$$

so one can replace $\frac{1}{4}z_i z_k \delta(i,k)$ by $\frac{1}{4}z_i\delta(i,k)$ in (7.13). With (7.14) the elements of A may be written as

$$A_{jm} = \tfrac{1}{2} z_j [\delta(j,m+1) + \delta(j,m-1)] + \delta(j,1)\delta(m,2) + \delta(j,N)\delta(m,N-1). \tag{7.15}$$

Introduce the time-independent $N^2 \times N^2$ matrix

$$C = \langle B(t) \otimes B(t) \rangle, \tag{7.16}$$

with elements given by (7.13). Then from (7.11)

$$\langle \mathbf{s}(t) \otimes \mathbf{s}(t) \rangle = [A \otimes A + C]^t [\mathbf{s}(0) \otimes \mathbf{s}(0)], \tag{7.17}$$

or, equivalently,

$$\langle \mathbf{s}(t+1) \otimes \mathbf{s}(t+1) \rangle = [A \otimes A + C] \langle \mathbf{s}(t) \otimes \mathbf{s}(t) \rangle, \tag{7.18}$$

i.e.,

$$\langle s_i(t+1) s_k(t+1) \rangle = \sum_{j,m} [(A \otimes A)_{ik;jm} + C_{ik;jm}] \langle s_j(t) s_m(t) \rangle = \sum_{j,m} [A_{ij} A_{km} + C_{ik;jm}] \langle s_j(t) s_m(t) \rangle, \tag{7.19}$$

where

$$\langle s_i(t) s_i(t) \rangle = 1, \quad \text{for } t = 0,1,\dots; \quad i = 1,\dots,N, \tag{7.20}$$

and

$$\langle s_j(0) s_m(0) \rangle = s_j(0) s_m(0). \tag{7.21}$$

From the discussion of the decomposition (Sec. IV) one sees that spins on different space-time sublattices are mutually uncorrelated (initially, and therefore for all time). Consequently, for spins s_i and s_k on different space-time sublattices

$$\langle s_i(t) s_k(t) \rangle = \langle s_i(t) \rangle \langle s_k(t) \rangle, \tag{7.22}$$

where $\langle s_i(t) \rangle$ is given by (6.17).

For spins s_i and s_k on the same space-time sublattice, one solves the nonhomogeneous, linear system (7.19) subject to (7.20) and (7.21). The result is

$$\langle s_j(t) s_m(t) \rangle = 1 + \sum_{k=0}^{N-1} c_k \sin[a_k((j+m)/2 - 1)] \times \sin[a_k(j-m)/2)] (\cos a_k)^t, \tag{7.23}$$

where $j, m = 1,2,\dots,N$, and $j+m$ is an even integer. The symbol a_k is defined in (6.13) and the coefficient c_k is found from (7.23) for $t=0$ by using the initial values (7.21). Notice that

$$t \to \infty \quad \text{limit of } \langle s_j(t) s_m(t) \rangle = 1, \tag{7.24}$$

consistent with the sublattice limiting behavior found in Sec. VI.

VIII. REMARKS

For $\epsilon = 0$ the above time-dependent model does not entirely "forget" its $t = 0$ condition for $t \to \infty$. Such long-range correlation in a space-time lattice which is infinite in only one dimension does not violate the Perron–Frobenius theorem. One is studying a system which is very similar to the zero-temperature case of Baxter's eight-vertex model,[9,10] where the latter model is taken in the form without four-spin interactions. The eight-vertex model (toroidal boundary conditions) then decomposes (see Fig. 10.4 of Baxter's book[11]) into two independent, square-lattice Ising models. At zero temperature there is, of course, long-range correlation in even the one-dimensional Ising model.

Clearly the time evolution of one-dimensional Ising spin systems is linked to the equilibrium behavior of two-dimensional Ising spin systems and related vertex models. This was evident, for example, when Felderhof[12] diagonalized the evolution operator for the Glauber model. His method is closely connected with the Fermi-operator technique used to diagonalize the transfer matrix for the two-dimensional Ising model.[13,14]

ACKNOWLEDGMENTS

I would like to acknowledge the kind, active interest of N. G. van Kampen and the comments of P. W. Kasteleyn. The latter part of this work was done during the summer of 1983 while I was a visiter at the Institut Lorentz, and I would like to thank H. W. Capel for helping arrange my pleasant visit there.

[1]R. J. Glauber, J. Math. Phys. **4**, 294 (1963).
[2]K. Kawasaki, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1972), Vol. 2.
[3]H. Falk, Physica A **119**, 580 (1983); **117**, 561 (1983); **104**, 475 (1980).
[4]H. Falk, Physica A **104**, 459 (1980).
[5]R. Kinderman and J. L. Snell, *Markov Random Fields and their Applications*, Contemporary Mathematics Series, Vol. 1 (Am. Math. Soc., Providence, RI, 1980).
[6]R. Durrett, Stoch. Proc. Applic. **11**, 109 (1981).
[7]S. Wolfram, Rev. Mod. Phys. **55**, 601 (1983).
[8]J. G. Kemeny and J. L. Snell, *Finite Markov Chains* (Van Nostrand, Englewood Cliffs, NJ, 1960).
[9]P. W. Kasteleyn, in *Fundamental Problems in Statistical Mechanics*, edited by E. G. D. Cohen (North–Holland, Amsterdam, 1975), Vol. III.
[10]E. H. Lieb and F. Y. Wu, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1972), Vol. 1.
[11]R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, London, 1982).
[12]B. U. Felderhof, Rep. Math. Phys. **1**, 215 (1971); **2**, 151 (1972).
[13]T. D. Schultz, D. C. Mattis, and E. H. Lieb, Rev. Mod. Phys. **36**, 856 (1964).
[14]M. Suzuki, Prog. Theor. Phys. **46**, 1337 (1971).

Minimization of the energy functional of a one-dimensional fermionic system in the large-N limit

I. Andrić and A. Mikelić
Rudjer Bošković Institute, POB 1016, 41001 Zagreb, Yugoslavia

(Received 25 April 1984; accepted for publication 7 September 1984)

A one-dimensional system of N nonrelativistic fermions in the confining potential is studied in the large-N limit where a classical limit appears.

I. INTRODUCTION

The analysis of many quantum theories with a large number of degrees of freedom surprisingly simplifies with increasing number of degrees of freedom.[1] In the limit $N \to \infty$, where N measures the number of degrees of freedom, it is possible to obtain a new type of classical limit of the quantum system. This is mainly due to the fact that quantum fluctuations of suitably chosen operators vanish in the limit when the number of degrees of freedom increases. It was shown[2] that one can find a classical phase space to define a consistent Poisson bracket and a classical Hamiltonian. In this paper we have reduced the problem of finding average values and the ground-state energy to the problem of minimizing the corresponding classical Hamiltonian.

A particularly efficient method for extracting the large-N behavior of a quantum theory is a collective-field approach used by Jevicki and Sakita,[3] which represents a generalization of the Bohm–Pines treatment. It was shown[4] how to extend this approach to include nonrelativistic fermions in one dimension. Choosing the density as a collective field, the theory in the large-N limit, where N is the number of fermions, results in the Thomas–Fermi theory plus correction terms. This is in analogy with the result for the three-dimensional case, as stated by Lieb.[5]

It is interesting that the same type of functional appears in the large-N limit of the SU(N) invariant random matrix model.[6] This is a consequence of the fact that the N-particle Fermi gas can be restated in terms of the matrix problem mentioned. The same type of functional appears in the large-N limit of the Calogero–Moser system with specific values of the coupling constant.[7]

In this paper we solve the problem of the N fermions in one dimension. The fermions interact via the confining potential $V(x,y) = 2g|x-y|$, arising in the large-N limit of the one-dimensional quantum chromodynamics.[8] This problem also appears in Witten's analysis of barions,[9,10] concerning symmetric ground states and antisymmetric excited states. In Witten's further analysis barions are similar to solitons. In our one-dimensional analysis we have found that particles concentrate only on the interval of finite length. If we introduce the density of fermions $\rho(x)$ as a collective field, then the resulting energy functional[4] is

$$J(\phi) = \frac{\pi^2}{6}\int_{\mathbb{R}} \phi^3\,dx + g\int_{\mathbb{R}}\int_{\mathbb{R}} |x-y|\phi(x)\phi(y)dx\,dy,$$

where we have taken the mass as $m = 1$. Solving the minimization problem should give us the minimal energy and the distribution of particles in this state.

In Sec. II we give a mathematical formulation of the problem and show that this problem is nonconvex. We also derive the virial theorem. In Sec. III we construct a stationary point, and in Sec. IV we show the uniqueness of this point up to the translation. In Sec. V we prove that the stationary point constructed is the point of the minimum. In Sec. VI we briefly compare our results with the three-dimensional case.

II. FORMULATION OF THE PROBLEM AND SOME AUXILIARY RESULTS

Let $L^p = L^p(\mathbb{R})$, $1 \leqslant p < \infty$, be a Banach space consisting of real measurable functions on $\mathbb{R}$ with the norm

$$\|f\|_p = \left(\int_{\mathbb{R}} |f(x)|^p\,dx\right)^{1/p}.$$

For a given $N \in \mathbb{N}$, let K be a convex set defined by

$$K = \left\{ z \in L^1 \cap L^3 : z \geqslant 0, \|z\|_1 = N, \int_{\mathbb{R}} |x| z\,dx < \infty \right\}. \quad (2.1)$$

Assume that g is a given positive number. Let us define the functional $J: K \to \mathbb{R}$ as

$$J(\phi) = \frac{\pi^2}{6}\int_{\mathbb{R}} \phi^3\,dx + g\int_{\mathbb{R}}\int_{\mathbb{R}} |x-y|\phi(x)\phi(y)dx\,dy. \quad (2.2)$$

We consider the following problem:

$$\inf\{J(\phi): \phi \in K\}. \quad (2.3)$$

The first term in J is obviously convex. Let us investigate the properties of the second term in J.

Proposition 2.1: Let us define a functional $I: K \to \mathbb{R}$ by

$$I(\phi) = \int_{\mathbb{R}}\int_{\mathbb{R}} |x-y|\phi(x)\phi(y)dx\,dy. \quad (2.4)$$

Then I is a concave function on K. Let ϕ be an even function from K. Then

$$N\int_{\mathbb{R}} |x|\phi(x)dx \leqslant I(\phi) = 2\int_{\mathbb{R}} |x|\phi(x)\int_{-|x|}^{|x|}\phi(y)dy$$
$$\leqslant 2N\int_{\mathbb{R}} |x|\phi(x)dx.$$

Proof: Let us first prove the second assertion. We use the equalities

$$I(\phi) = 2\int_0^{+\infty}\int_0^{+\infty} \{|x-y| + |x+y|\}\phi(x)\phi(y)dx\,dy$$
$$= 2\int_0^{+\infty}\int_0^{x} 2x\phi(x)\phi(y)dx\,dy$$

 0022-2488/85/040698-07$02.50

$$+2\int_0^{+\infty}\int_x^{+\infty} 2y\phi(x)\phi(y)dx\,dy$$
$$=8\int_0^{+\infty} x\phi(x)\int_0^x \phi(y)dy$$
$$=2\int_{\mathbb{R}} |x|\phi(x)\int_{-|x|}^{|x|}\phi(y)dy\,dx.$$

Because of $\|\phi\|_1 = N$, we have $I(\phi) \leqslant 2N\int_{\mathbb{R}}|x|\phi(x)$. Using the inequality $|x-y|+|x+y|\geqslant 2x$, we obtain

$$I(\phi)\geqslant N\int_{\mathbb{R}}|x|\phi(x)dx.$$

Now let us prove the first assertion. Let $u_1, u_2 \in K$, $u(t) = (1-t)u_1 + tu_2$, and $v = u_1 - u_2$. We use the equality

$$I(u(t)) = (1-t)I(u_1) + tI(u_2) - t(1-t)I(v).$$

We have $v \in L^1 \cap L^3$, $\int_{\mathbb{R}} v(x)dx = 0$, and $\int_{\mathbb{R}}|x||v(x)|dx < +\infty$. Let $\hat{v}$ denote a Fourier transform of v. Then

$$\hat{v}\in L^\infty,\quad \dot{\hat{v}}\in L^\infty,\quad \|\hat{v}\|_\infty < C/\sqrt{2\pi},\quad \hat{v}(0)=0. \tag{2.5}$$

Now we conclude that

$$|\hat{v}(p)/p| < C/\sqrt{2\pi}. \tag{2.6}$$

Because of $|x| = -2\wp(1/p^2)$, (2.5), and (2.6), for $w \in C_0^1$, $\dot{w} = v$, we have

$$I(v) = -2\int_{\mathbb{R}}|\hat{w}(p)|^2\,dp = -2\int_{\mathbb{R}} p^{-2}|\hat{v}(p)|^2\,dp,$$

i.e., $I(v) < 0$ for $u_1 \neq u_2$. We have already proved

$$I(u(t)) \geqslant (1-t)I(u_1) + tI(u_2);$$

hence I is a concave function.

Corollary 2.2: Problem (2.3) is a nonconvex minimization problem.

Theorem 2.3 (virial theorem): Let ϕ_0 be a solution for (2.3), $C_1 = (\pi^2/6)\int_{\mathbb{R}}\phi_0^3$ and $C_2 = gI(\phi_0)$. Then $2C_1 = C_2$.

Proof: Let us define a function ϕ by

$$\phi(x) = (C_2/2C_1)^{1/3}\phi_0(x(C_2/2C_1)^{1/3}).$$

Then we have

$$J(\phi) = 3/4^{1/3}C_1^{1/3}C_2^{2/3} \geqslant J(\phi_0) = C_1 + C_2. \tag{2.7}$$

On the other hand, it is well known that

$$C_1 + C_2 \geqslant 3/4^{1/3}C_1^{1/3}C_2^{2/3},\quad \text{for every } C_1, C_2 > 0, \tag{2.8}$$

with equality in the case $2C_1 = C_2$. The statement of this Theorem follows from (2.7) and (2.8).

III. CONSTRUCTION OF THE STATIONARY POINT

The main aim of this section is to construct a stationary point for the minimization problem (2.3). In the following lemma we obtain the necessary conditions for (2.3).

Lemma 3.1: Let ϕ_0 be a solution for (2.3). Then there exists a positive number λ (Lagrange multiplier) such that

$$\frac{\pi^2}{2}\phi_0^2(x) + 2g\int_{\mathbb{R}}|x-y|\phi_0(y)dy = \lambda$$

on the set $\{x\in\mathbb{R}:\phi_0(x) > 0 \quad (\text{a.e.})\}$,

$$2g\int_{\mathbb{R}}|x-y|\phi_0(y)dy \geqslant \lambda$$

on the set $\{x\in\mathbb{R}:\phi_0(x) = 0 \quad (\text{a.e.})\}$. (3.1)

The expression $\lambda = 1/N\,\frac{7}{3}J(\phi_0)$ is also valid.

Proof: Let $\phi \in K$ and $t\in(0,1)$. Then $\phi_t = (1-t)\phi_0 + t\phi \in K$. We also have $t^{-1}[J(\phi_t) - J(\phi_0)] \geqslant 0$. Taking the limit $t\to 0$, from the preceding inequality we obtain

$$\int_{\mathbb{R}}\left\{\frac{\pi^2}{2}\phi_0^2(x) + 2g\int_{\mathbb{R}}|x-y|\phi_0(y)dy\right\}(\phi-\phi_0)dx \geqslant 0. \tag{3.2}$$

Arguing along the same lines as in Ref. 10, from Eq. (3.2) we obtain the existence of λ such that (3.1) is valid. λ is obviously positive. Let us multiply the first equation in (3.1) by a ϕ_0 and integrate. We obtain

$$\frac{\pi^2}{2}\int_{\mathbb{R}}\phi_0^3 + 2gI(\phi_0) = \lambda\int_{\mathbb{R}}\phi_0 = N. \tag{3.3}$$

Because of Theorem 2.3, Eq. (3.3) implies $\lambda = (1/N)\frac{7}{3}J(\phi_0)$.

By using the scaling arguments, we can show that the solution of (2.3) has a form

$$\phi_{N,g}(x) = N^{2/3}g^{1/3}\phi(x(g/N)^{1/3}),$$

where ϕ is the solution for the case $N=1, g=1$. Then

$$J(\phi_{N,g}) = N^{7/3}g^{2/3}J(\phi).$$

If λ is a Lagrange multiplier in the case $N=1, g=1$, then $\lambda_{N,g} = g^{2/3}N^{4/3}\lambda$.

In the following we consider the case $N=1, g=1$. Because of the scaling arguments, we can easily construct a solution for the general case if we know the solution for the case $N=1$, $g=1$. Let us consider the following problem: Find $\lambda\in\mathbb{R}$ and a positive function $\phi\in L^1\cap L^3$, such that

$$\frac{\pi^2}{2}\phi^2(x) + 2\int_{\mathbb{R}}|x-y|\phi(y)dy = \lambda$$

on a set $\{x:\phi(x) > 0 \quad \text{a.e. on } \mathbb{R}\}$,

$$2\int_{\mathbb{R}}|x-y|\phi(y)dy \geqslant \lambda$$

on a set $\{x:\phi(x) = 0 \quad \text{a.e. on } \mathbb{R}\}$,

$$\int_{\mathbb{R}}\phi(y)dy = 1. \tag{3.4}$$

Every function ϕ which satisfies problem (3.4) we call a stationary point for (2.3), and λ is a Lagrange multiplier. Conditions (3.4) are the well-known Kuhn–Tucker conditions for problem (2.3). For convex minimization problems, the Kuhn–Tucker conditions are necessary and sufficient conditions for the extremum, but for nonconvex problems they are only necessary conditions.

Our aim is to obtain a solution for (3.4). Let us look for an even solution of problem (3.4) such that $\{x:\phi(x)>0\} = (-a,a)$, for some $a>0$. By differentiating and introducing a new function $u = \phi^2$, we obtain the following differential equation:

$$-u'' = (8/\pi^2)\sqrt{u}\quad \text{on } (-a,a). \tag{3.5}$$

Because of the even properties of the solution, we have

$$u(-a) = u(a) = 0, \quad u'(0) = 0, \quad \int_{-a}^{a} \sqrt{u}\, dx = 1. \quad (3.6)$$

By multiplying (3.5) by u' and integrating, it follows that

$$u'^2(x) = u'^2(a) - (32/3\pi^2)u^{3/2}(x),$$
$$u'^2(x) = u'^2(-a) - (32/3\pi^2)u^{3/2}(x), \quad \forall x \in (-a,a). \quad (3.7)$$

Because of the positivity of the solution, from (3.7) we conclude that

$$u'(a) = -u'(-a). \quad (3.8)$$

Integrating Eq. (3.5), we have

$$-u'(a) + u'(-a) = 8/\pi^2. \quad (3.9)$$

Equations (3.8) and (3.9) therefore imply

$$u'(a) = -4/\pi^2 = -u'(-a). \quad (3.10)$$

Because of the positivity of u, we conclude that u satisfies the following conditions:

$$u' = 4/\pi^2(1 - (2\pi^2/3)u^{3/2})^{1/2} \quad \text{on } (-a,0),$$
$$u' = -4/\pi^2(1 - (2\pi^2/3)u^{3/2})^{1/2} \quad \text{on } (0,a), \quad (3.11)$$
$$u(-a) = u(a) = 0, \quad u'(-a) = 4/\pi^2 = -u'(a),$$
$$u'(0) = 0.$$

From (3.11) it follows that

$$u(0) = (3/2\pi^2)^{2/3}. \quad (3.12)$$

From (3.11) we can also determine the length of the interval

$$\int_{u(0)}^{u(a)} \frac{du}{(1 - (2\pi^2/3)u^{3/2})^{1/2}} = -\frac{4}{\pi^2}a. \quad (3.13)$$

Let us use the expression

$$\int_0^1 \frac{dy}{(1 - y^{3/2})^{1/2}} = 2\int_0^1 \frac{z\, dz}{(1 - z^3)^{1/2}} = \frac{2\sqrt{3}}{\pi\, {}^3\!\sqrt{4}} \left\{\Gamma\left(\frac{2}{3}\right)\right\}^3, \quad (3.14)$$

(see Ref. 11).

By using (3.14), from (3.13) we conclude that

$$a = \tfrac{3}{4}(\sqrt{3}/2\pi)^{1/3}\{\Gamma(\tfrac{2}{3})\}^3. \quad (3.15)$$

Let us mention that Cauchy's problems

$$v' = -4/\pi^2(1 - (2\pi^2/3)v^{3/2})^{1/2},$$
$$v(a) = 0 \quad \text{on } (0,a), \quad (3.16)$$
$$v' = 4/\pi^2(1 - (2\pi^2/3)v^{3/2})^{1/2},$$
$$v(-a) = 0 \quad \text{on } (-a,0), \quad (3.17)$$

have a unique solution, with a given by (3.15). Therefore we may conclude that the solutions of Eqs. (3.16) and (3.17) take the same value $(3/2\pi^2)^{2/3}$ at zero. Hence we have constructed a C^1 function v, which satisfies (3.16) and (3.17). Furthermore, we have $v \in C^2(-a,a)$.

Let us now prove that $\phi = \sqrt{v}$ is a solution of problem (3.4).

From (3.16) and (3.17) we have

$$v'^2(x) = (16/\pi^4)(1 - (2\pi^2/3)v^{3/2}(x)). \quad (3.18)$$

By differentiating, it follows that

$$2v'v'' = -(16/\pi^2)v'\sqrt{v} \quad \text{on } (-a,a). \quad (3.19)$$

Zero is the only null point of v', hence we obtain

$$\pi^2/2(\phi^2)'' + 4\phi = 0 \quad \text{on } (-a,a). \quad (3.20)$$

Integrating (3.20), we have

$$\frac{\pi^2}{2}\{(\phi^2)'(x) - (\phi^2)'(-a)\} + 4\int_{-a}^{x} \phi(y)dy = 0, \quad (3.21)$$

$$-\frac{\pi^2}{2}\{(\phi^2)'(a) - (\phi^2)'(x)\} - 4\int_{x}^{a} \phi(y)dy = 0. \quad (3.22)$$

Adding up Eqs. (3.21) and (3.22), we obtain

$$\frac{\pi^2}{2}2(\phi^2)'(x) + 4\int_{-a}^{a} \frac{x-y}{|x-y|}\phi(y)dy = 0. \quad (3.23)$$

Integrating (3.23), we have

$$\frac{\pi^2}{2}\phi^2(x) + 2\int_{-a}^{a} |x-y|\phi(y)dy = \frac{\pi^2}{2}\phi^2(0) + 2\int_{-a}^{a} |y|\phi(y)dy. \quad (3.24)$$

Let us define the Lagrange multiplier λ as

$$\lambda = \frac{\pi^2}{2}\phi^2(0) + 2\int_{-a}^{a} |y|\phi(y)dy. \quad (3.25)$$

From (3.20) we have

$$\int_{-a}^{a} |y|\phi(y) = -\frac{\pi^2}{4}\int_0^a v''\cdot y = a - \frac{\pi^2}{4}\phi^2(0). \quad (3.26)$$

Therefore,

$$\lambda = 2a. \quad (3.27)$$

Let us now evaluate $\int_{-a}^{a}\phi(y)dy$. Calculating as in the equality (3.26), we obtain

$$\int_{-a}^{a} \phi(y)dy = \int_{-a}^{a} \sqrt{u}\, dy = -\frac{4}{\pi^2}[u'(a) - u'(-a)] = 1. \quad (3.28)$$

Also, let $|x| > a$. Then

$$2\int_{-a}^{a} |x-y|\phi(y)dy \geqslant 2|x|\int_{-a}^{a}\phi(y)dy \geqslant 2|x| \geqslant 2a = \lambda. \quad (3.29)$$

Hence the pair $\{\lambda,\sqrt{v}\}$, where λ is given by (3.27) and v by (3.16) and (3.17), is a solution of problem (3.4). Because of the virial theorem, we have

$$J(\phi) = \tfrac{3}{4}\lambda = \tfrac{3}{2}a. \quad (3.30)$$

The numerical value of a number a is

$$a = \tfrac{3}{4}(\sqrt{3}/2\pi)^{1/3}\{\Gamma(\tfrac{2}{3})\}^3 = 1.211\,8677. \quad (3.31)$$

From (3.31) we have $\lambda = 2.423\,9354$ and $J(\phi) = 1.038\,8295$. The value of a solution at zero is

$$\phi(0) = (3/2\pi^2)^{1/3} = 0.533\,659. \quad (3.32)$$

Values of the function $\phi = \sqrt{v}$ on the interval $(0,a)$ are determined by numerical integration of Cauchy's problem (3.16) using the Runge–Kutta method with a step $h = -a/1000$. The results are plotted in Fig. 1.

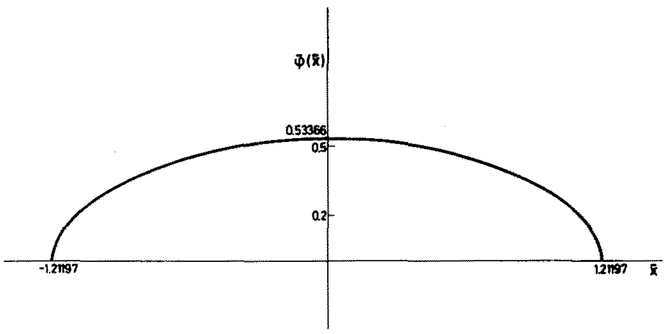


FIG. 1. Density distribution $\phi(x)$. Plotted is the rescaled function $\tilde{\phi}(\tilde{x}) = N^{-2/3}g^{-1/3}\phi(x)$, $\tilde{x} = x(g/N)^{1/3}$.

IV. UNIQUENESS OF THE SOLUTION

Because of translational invariance, we obviously have infinitely many solutions for problem (3.4). However, if we specify a solution such that $\int_0^{+\infty}\phi = \int_{-\infty}^{0}\phi$, we shall be able to prove the uniqueness in this case. If we choose the condition at another point z instead of at zero, the solution will be fixed in respect to that point. Hence we shall prove the uniqueness of the solution up to the translation.

In order to show the uniqueness of the solution, we start with a related simpler problem: Find a positive number b and $u \in C^1(0,b)$ such that

$$u' = -4/\pi^2(1-(2\pi^2/3)u^{3/2})^{1/2} \quad \text{on } (0,b), \qquad u(0) = (3/2\pi^2)^{2/3}, \quad u(b) = 0. \tag{4.1}$$

Proposition 4.1: Problem (4.1) has a unique solution $\{a,v\} \in \mathbb{R}\times C^1(-a,a)$, where a is given by (3.15) and v is a solution of Cauchy's problem (3.16).

Proof: The uniqueness of the solution for problem (4.1) follows from the construction of a and v.

Let us now extend problem (4.1) to the following problem: Find a positive number b and $u \in C^1(-b,b)$ such that

$$\begin{aligned} &u' = -(4/\pi^2)(1-(2\pi^2/3)u^{3/2})^{1/2} \quad \text{on } (0,b),\\ &u' = (4/\pi^2)(1-(2\pi^2/3)u^{3/2})^{1/2} \quad \text{on } (-b,0),\\ &u(0) = (3/2\pi^2)^{3/2}, \quad u(-b) = u(b) = 0. \end{aligned} \tag{4.2}$$

Proposition 4.2: Problem (4.2) has a unique solution $\{a,v\} \in \mathbb{R}\times C^1(-a,a)$, where a is given by (3.15) and v is a solution of Cauchy's problems (3.16) and (3.17).

Now we are able to show the uniqueness of the following auxiliary problem: Find positive numbers b and c, and a positive function $u \in C^2(-b,c)$ such that

$$\begin{aligned} &-u'' = (8/\pi^2)\sqrt{u} \quad \text{on } (-b,c),\\ &\int_{-b}^{c}\sqrt{u} = 1, \quad u(-b) = u(c) = 0, \quad \int_{-b}^{0}\sqrt{u} = \int_0^c\sqrt{u}. \end{aligned} \tag{4.3}$$

Lemma 4.3: Let $\{a,v\}$ be as in Proposition 4.2.Then $\{a,a,v\}$ is a unique solution for a in (4.3).

Proof: We have already proved that $\{a,a,v\}$ is a solution for (4.3). It remains to prove that $\{a,a,v\}$ is a unique solution for (4.3). Let $\{b,c,z\}$ be another solution for (4.3). Then z satisfies the equation

$$-z'' = 8/\pi^2\sqrt{z} \text{ on } (-b,c). \tag{4.4}$$

Integrating Eq. (4.4), we obtain

$$-z'(c) + z'(-b) = 8/\pi^2. \tag{4.5}$$

Multiplying Eq. (4.4) by $2z'$ and integrating the resulting expression, we obtain

$$\begin{aligned} &-z'^2(x) + z'^2(-b) = (32/3\pi^2)z^{3/2}(x),\\ &z'^2(x) - z'^2(c) = -(32/3\pi^2)z^{3/2}(x). \end{aligned} \tag{4.6}$$

Adding these two equations, and because of (4.5), we obtain

$$z'(-b) = -z'(c) = 4/\pi^2. \tag{4.7}$$

From (4.6) and (4.7) we derive a differential equation of the first order for z:

$$z'^2(x) = (16/\pi^4)(1-(2\pi^2/3)z^{3/2}(x)) \text{ on } (-b,c).$$

Let us define x_s as a point from $(-b,c)$ such that $z'(x_s) = 0$. Obviously, there is a unique $x_s \in (-b,c)$ such that $z'(x_s) = 0$. We also have $z(x_s) = (3/2\pi^2)^{2/3}$. Then we have

$$\begin{aligned} &z' = (4/\pi^2)(1-(2\pi^2/3)z^{3/2})^{1/2}, \quad z(-b) = 0,\\ &z(x_s) = (3/2\pi^2)^{2/3} \quad \text{on } (-b,x_s)\\ &\text{and}\\ &z' = -(4/\pi^2)(1-(2\pi^2/3)z^{3/2})^{1/2}, z(x_s) = (3/2\pi^2)^{2/3},\\ &z(c) = 0 \quad \text{on } (x_s,c). \end{aligned} \tag{4.8}$$

Therefore we have the equalities

$$\begin{aligned} x_s + b &= \frac{\pi^2}{4}\int_{z(-b)}^{z(x_s)}\frac{dy}{(1-(2\pi^2/3)y^{3/2})^{1/2}}\\ &= \frac{\pi^2}{4}\int_{z(x_s)}^{z(c)}\frac{dy}{(1-(2\pi^2/3)y^{3/2})^{1/2}}\\ &= c - x_s. \end{aligned} \tag{4.9}$$

From (4.8) we conclude that $x_s = (-b+c)/2$. By performing a translation for x_s in (4.8), we obtain problem (4.2). Because of Propositions (4.1) and (4.2), it follows from (4.8) that $z(x-x_s) = v(x)$, and $c - x_s = a = -b - x_s = (b+c)/2$. Now we have that z is symmetric around x_s. Hence from $\int_{-b}^{0}\sqrt{z} = \int_0^c\sqrt{z}$, we conclude that $b = c$. Therefore $x_s = 0$ and problem (4.3) has a unique solution.

Lemma 4.4: Let $\{\lambda,\phi\}$ be a solution for (3.4). Then ϕ has a compact support.

Proof: We have the inequality

$$\lambda = \frac{\pi^2}{2}\phi^2(x) + 2\int_{\mathbb{R}}|x-y|\phi(y)dy \geqslant \frac{\pi^2}{2}\phi^2(x) + 2|x| + c$$

on the set $\{x:\phi(x) > 0\}$.

The statement of Lemma 4.4 follows directly from this inequality.

Theorem 4.5: Let a and v be as in Proposition 4.2. Let λ be given by $\lambda = 2a$ and ϕ by $\phi = \sqrt{v}$. Then $\{\lambda,\phi\}$ is a unique solution for (3.4) such that ϕ is continuous and $\int_0^{+\infty}\phi(x)dx = \int_{-\infty}^{0}\phi(x)dx$.

Proof: We have already proved that $\{2a,\sqrt{v}\}$ is a solution for (3.4).

Let $\{\mu,\psi\}$ be another solution for (3.4) such that ψ is continuous. Them $\{\psi > 0\}$ is an open set. Let (b_1,b_2) be a subset of $\{\psi > 0\}$ such that $\psi(b_1) = \psi(b_2) = 0$. By differentiating, we obtain

$$-(\psi^2)'' = (8/\pi^2)\psi \quad \text{on } (b_1,b_2), \quad \psi(b_1) = \psi(b_2) = 0,$$
$$\int_{b_1}^{b_2} \psi = C > 0. \tag{4.10}$$

From (4.10) we conclude in a similar way as in Lemma 4.3 that $\psi'(b_1) = -\psi'(b_2) = (4/\pi^2)C$. Let us suppose that supp ψ contains more than one part. Let $(A,B) \subset \{\psi > 0\}$ be such that $\psi(A) = \psi(B) = 0$ and $0 \leqslant A \leqslant B$. Then we have

$$2\int_{\mathbf{R}} |A-y|\psi(y)dy = \mu, \quad 2\int_{\mathbf{R}} |B-y|\psi(y)dy = \mu. \tag{4.11}$$

We can write (4.11) in the form

$$\int_{-A}^{A} (A-y)\psi + \int_{-\infty}^{-A} (A-y)\psi + \int_{A}^{+\infty} (y-A)\psi = \frac{\mu}{2}, \tag{4.12}$$
$$\int_{-B}^{B} (B-y)\psi + \int_{-\infty}^{-B} (B-y)\psi + \int_{B}^{+\infty} (y-B)\psi = \frac{\mu}{2}.$$

From (4.12) we obtain

$$A\int_{-A}^{A} \psi + \int_{A}^{+\infty} y\psi - \int_{-\infty}^{-A} y\psi + A\left\{\int_{-\infty}^{-A} \psi - \int_{A}^{+\infty} \psi\right\} = \frac{\mu}{2} + \int_{-A}^{A} y\psi, \tag{4.13}$$

$$B\left\{\int_{-A}^{A} \psi + 2\int_{A}^{B} \psi\right\} + \int_{A}^{+\infty} y\psi - \int_{-\infty}^{-A} y\psi - 2\int_{A}^{B} y\psi + B\left\{\int_{-\infty}^{-A} \psi - \int_{A}^{+\infty} \psi\right\} = \frac{\mu}{2} + \int_{-A}^{A} y\psi. \tag{4.14}$$

Substituting (4.13) into (4.14) and using the expression $\int_A^B y\psi = (B+A)/2\int_A^B \psi$, we have

$$(A-B)\left\{\int_{-A}^{A} \psi + \int_{-\infty}^{-A} \psi - \int_{A}^{+\infty} \psi + \int_{A}^{B} \psi\right\} = 0. \tag{4.15}$$

From $\psi > 0$ on (A,B), $\int_0^{+\infty} \psi = \int_{-\infty}^0 \psi$, and (4.15) we conclude that $A = B$. Now we have $0 \in \{\psi > 0\}$, and the set $\{\psi > 0\}$ consists only of one interval, i.e., there exist numbers $c,b > 0$ such that $(-b,c) = \{\psi > 0\}$. Now from Lemma 4.3 it follows that $b = a$ and $\psi = \sqrt{v}$.

V. SOLUTION OF THE MINIMIZATION PROBLEM

In this section we prove the existence of a solution of the minimization problem (2.3). The results obtained in the preceding sections are crucial for finding a solution of this problem. Obviously, it is enough to treat a case $g = 1, N = 1$.

For a $u:\mathbf{R}\to[0,\infty)$, let u^* be a symmetric decreasing rearrangement defined as in Refs. 12 and 13. For the positive symmetric decreasing g and non-negative functions $f \in L^p$ and $h \in L^q$, we have

$$\int_{\mathbf{R}}\int_{\mathbf{R}} f(x)g(x-y)h(y)dx\,dy \leqslant \int_{\mathbf{R}}\int_{\mathbf{R}} f^*(x)g(x-y)h^*(x)dx\,dy, \tag{5.1}$$

where f^* and h^* are symmetric decreasing rearrangements of f and h, respectively. This is the well-known Riesz theorem (see Refs. 12 and 13). We also have $(f^\alpha)^* = (f^*)^\alpha$, pointwise for every measurable f and positive number α, and $\|f\|_p = \|f^*\|_p$ for $p \in [1,\infty)$.

Lemma 5.1: Let $u \in K$. Then

$$\int_{\mathbf{R}}\int_{\mathbf{R}} u(x)|x-y|u(y)dx\,dy \geqslant \int_{\mathbf{R}}\int_{\mathbf{R}} u^*(x)|x-y|u^*(y)dx\,dy. \tag{5.2}$$

Proof: Without loss of generality, we may suppose that u has a compact support. Then u^* also has a compact support. Hence there exists $C > 0$, such that

$$\iint_{|x-y|>C} [|x-y| - C]u(x)u(y)dx\,dy = 0$$

and

$$\iint_{|x-y|>C} [|x-y| - C]u^*(x)u^*(y)dx\,dy = 0.$$

Let us define the function g by $g(x) = |x|$ for $x \leqslant C$ and $g(x) = C$ elsewhere. Then we have

$$\begin{aligned}
&\int_{\mathbf{R}}\int_{\mathbf{R}} |x-y|u(x)u(y)dx\,dy - \int_{\mathbf{R}}\int_{\mathbf{R}} |x-y|u^*(x)u^*(y)dx\,dy \\
&= -\int_{\mathbf{R}}\int_{\mathbf{R}} [C-g(x-y)]u(x)u(y)dx\,dy + \int_{\mathbf{R}}\int_{\mathbf{R}} [C-g(x-y)]u^*(x)u^*(y)dx\,dy \\
&\quad - C\left[\int_{\mathbf{R}} u(x)dx\right]^2 + C\left[\int_{\mathbf{R}} u^*(x)dx\right]^2 \\
&\quad + \int_{\mathbf{R}}\int_{\mathbf{R}} \{|x-y| - g(x-y)\}u(x)u(y)dx\,dy - \int_{\mathbf{R}}\int_{\mathbf{R}} \{|x-y| - g(x-y)\}u^*(x)u^*(y)dx\,dy \\
&= -\int_{\mathbf{R}}\int_{\mathbf{R}} [C-g(x-y)]u(x)u(y)dx\,dy + \int_{\mathbf{R}}\int_{\mathbf{R}} [C-g(x-y)]u^*(x)u^*(y)dx\,dy \geqslant 0,
\end{aligned}$$

because of (5.1). Thus (5.2) is proved.

In order to prove the existence of a solution of problem (2.3), we start with a related auxiliary problem. Let m be a positive real number $m \geqslant 2$, and

$$K_m = \left\{ u \in L^1(-m,m) \cap L^3(-m,m) \,|\, u \geqslant 0, \int_{-m}^{m} |x|u(x)dx < +\infty, \int_{-m}^{m} u(x)dx = 1, u \text{ even} \right\}.$$

We are looking for a solution of the following problem: Find $u \in K_m$, such that

$$\frac{\pi^2}{6}\int_{-m}^{m} u^3(x)dx + \int_{-m}^{m}\int_{-m}^{m} |x-y|u(x)u(y)dx\,dy \leqslant \frac{\pi^2}{6}\int_{-m}^{m} v^3(x)dx + \int_{-m}^{m}\int_{-m}^{m} |x-y|v(x)v(y)dx\,dy, \tag{5.3}$$
$$\forall v \in K_m.$$

Lemma 5.2: Let ϕ be a solution for (3.4). Then ϕ is a unique solution for (5.3).

First step: Proof: We first prove that (5.3) has at least one solution. Let $\{u_n\} \subset K_m$ be a minimizing sequence for (5.3). Because of Lemma 5.1, $\{u_n^*\} \subset K_m$ is also a minimizing sequence. Because of Proposition 2.1, we also have

$$\|u_n^*\|_{L^3(-m,m)} \leqslant C, \tag{5.4}$$

$$\|u_n^*\|_{L^1(-m,m)} = 1, \tag{5.5}$$

$$\int_{-m}^{m}\int_{-m}^{m} |x-y| u_n^*(x) u_n^*(y) dx\, dy$$
$$= 2\int_{-m}^{m} |x| u_n^*(x) \int_{-|x|}^{|x|} u_n^*(y) dy\, dx \leqslant C. \tag{5.6}$$

Hence there exists a symmetric decreasing $u_0 \in L^1(-m,m) \cap L^3(-m,m)$ and a subsequence of $\{u_n^*\}$, again denoted by the same symbol $\{u_n^*\}$, such that

$$u_n^* \to u_0, \quad \text{weakly in } L^2(-m,m). \tag{5.7}$$

Let us define a sequence $\{\Phi_n\}$ by

$$\Phi_n(x) = \int_{-|x|}^{|x|} u_n^*(y) dy. \tag{5.8}$$

Because of (5.4), (5.5), and (5.8), we conclude that

$$\|\Phi_n\|_{H^1(-m,m)} \leqslant C. \tag{5.9}$$

Hence there exists an even $\Phi_0 \in H^1(-m,m)$ and a subsequence of $\{\Phi_n\}$, again denoted by the same symbol $\{\Phi_n\}$, such that

$$\begin{aligned} &\Phi_n \to \Phi_0, \quad \text{weakly in } H^1(-m,m), \\ &\Phi_n \to \Phi_0, \quad \text{strongly in } L^2(-m,m). \end{aligned} \tag{5.10}$$

It is obvious that

$$\Phi_0(x) = \int_{-|x|}^{|x|} u_0(y) dy, \quad \text{(a.e.) on } [-m,m].$$

Because of $1 \in L^2(-m,m)$, it is easy to prove $u_0 \in K_m$. We also have

$$\underline{\lim} \int_{-m}^{m} (u_n^*)^3(x) dx \geqslant \int_{-m}^{m} u_0^3(x) dx. \tag{5.11}$$

Now let us prove the equality

$$\lim \int_{-m}^{m}\int_{-m}^{m} |x-y| u_n^*(x) u_n^*(y) dx\, dy$$
$$= \int_{-m}^{m}\int_{-m}^{m} |x-y| u_0(x) u_0(y) dx\, dy. \tag{5.12}$$

We have

$$\left|\int_{-m}^{m}\int_{-m}^{m} |x-y| \left[u_n^*(x)u_n^*(y) - u_0(x)u_0(y)\right] dx\, dy\right|$$
$$= 2\left|\int_{-m}^{m} |x| \left[u_n^*(x)\Phi_n(x) - u_0(x)\Phi_0(x)\right] dx\right|$$
$$\leqslant 2\left|\int_{-m}^{m} |x| u_n^*(x) [\Phi_n - \Phi_0](x) dx\right|$$
$$+ 2\left|\int_{-m}^{m} |x| (u_0 - u_n^*)(x) \Phi_0(x) dx\right|$$
$$\leqslant 2\||x|u_n^*\|_{L^2(-m,m)} \|\Phi_n - \Phi_0\|_{L^2(-m,m)}$$
$$+ 2\left|\int_{-m}^{m} |x| \Phi_0(x)(u_0 - u_n^*)(x) dx\right|. \tag{5.13}$$

Because of (5.4), (5.5), and $|x|\Phi_0 \in L^2(-m,m)$, (5.13) implies (5.12). We can therefore conclude that u_0 is a solution for (5.3).

Second step: Let us obtain necessary conditions for (5.3). Let u_0 be a solution for (5.3). Arguing along the same lines as in the proof of Lemma 3.1, we obtain that there exists a positive number λ_m (Lagrange multiplier) such that

$$\begin{aligned} &\frac{\pi^2}{2} u_0^2(x) + 2\int_{-m}^{m} |x-y| u_0(y) dy = \lambda_m \\ &\quad \text{on the set } \{-m \leqslant x \leqslant m : u_0(x) > 0, \text{(a.e.)}\}, \\ &\int_{-m}^{m} |x-y| u_0(y) dy \geqslant \lambda_m \\ &\quad \text{on the set } \{-m \leqslant x \leqslant m : u_0(x) = 0, \text{(a.e.)}\}, \\ &\int_{-m}^{m} u_0(y) dy = 1. \end{aligned} \tag{5.14}$$

Let $\{\lambda, \phi\}$ be a solution for (3.4), defined in Sec. II. Because of $m \geqslant 2$, $\{\lambda, \phi\}$ is also a solution of (5.14). However, arguing as in Sec. III, we easily obtain that (5.14) has only one solution, $\{\lambda, \phi\}$. We therefore may conclude that $\lambda_m = \lambda$ and $u_0 = \phi$. Hence ϕ is a unique solution for (5.3).

Theorem 5.3: Let $\{\lambda, \phi\}$ be a solution for (3.4), defined in Sec. II. Then ϕ is a solution for (2.3) and $J(\phi) = \frac{3}{5}\lambda$.

Proof: Let $u \in K$ and $J(u) < +\infty$. We wish to prove $J(u) \geqslant J(\phi)$. For $u \in K$, we have $u^* \in K$ and $J(u) \geqslant J(u^*)$. Because of Proposition 2.1, we have $u^* \in L^1 \cap L^3$ and $\int_{\mathbf{R}} |x| u^*(x) dx < +\infty$. Let us define a Banach space B by

$$B = \left\{ z \in L^1 \cap L^3 \,\middle|\, \int_{\mathbf{R}} |x| |z(x)| dx < \infty \right\}. \tag{5.15}$$

The norm on B is defined by

$$\|z\|_B = \|(1 + |x|) z\|_1 + \|z\|_3. \tag{5.16}$$

J is continuous on B.

For $\forall \epsilon > 0$, there is an even $u_\epsilon \in B$ with a compact support, such that

$$\|u_\epsilon - u^*\|_B \leqslant \epsilon. \tag{5.17}$$

Therefore, for every $\epsilon > 0$, there are even $u_\epsilon \in K$ and $m > 2$, such that supp $u_\epsilon \subset [-m,m]$

$$|J(u_\epsilon) - J(u^*)| \leqslant \epsilon. \tag{5.18}$$

Because of Lemma 5.2, we have $J(u_\epsilon) > J(\phi)$; hence, we conclude that

$$J(u) \geqslant J(u^*) \geqslant J(\phi) - \epsilon,$$

for every $\epsilon > 0$. The function ϕ is, therefore, a solution of (2.3). Lemma 3.1 implies $J(\phi) = \frac{3}{5}\lambda$.

Corollary 5.4: The function ϕ is a unique solution for (2.3) up to the translation.

VI. CONCLUSION

In this paper we have analyzed a one-dimensional system of nonrelativistic fermions interacting via confining potential. In the limit where the number of particles N increases, we can introduce the local density as a collective variable. In the leading order in N, we obtain the Thomas–Fermi approximation and the corresponding minimal energy is

$$E^{\mathrm{TF}} = \tfrac{9}{28}(\sqrt{3}/\pi)^{1/3}[\Gamma(\tfrac{2}{3})]^3 N^{7/3}(2g)^{2/3}$$

$$= 0.654 N^{7/3}(2g)^{2/3}.$$

Let us note that this result, obtained in the Thomas–Fermi approximation, is consistent with the lower bound[14] of the Dyson–Lenard type, which holds for the Schrödinger equation for fermions

$$E_{\min} \geqslant \frac{\pi^{2/3} 3^{5/3}}{2^{7/3} 5} (2g)^{2/3} N^{7/3}$$

$$= 0.531 N^{7/3}(2g)^{2/3}.$$

We also want to emphasize that the solution of the fermionic problem has a compact support and is unique. Comparison shows that our results for the one-dimensional case have the same qualitative properties as those for the solutions of the three-dimensional Thomas–Fermi theory,[5,10] in spite of the fact that the one-dimensional problem is nonconvex.

ACKNOWLEDGMENT

Part of this work was done during the stay of one of the authors (I.A.) at the ZIF, University of Bielefeld, BRD, with the support of the KFA, Jülich, BRD.

[1]See, for example, S. Coleman, Erice Lecture Notes, 1979.
[2]L. Yaffe, Rev. Mod. Phys. **54**, 407 (1982).
[3]A. Jevicki and B. Sakita, Nucl. Phys. B. **165**, 511 (1980).
[4]I. Andrić, V. Bardek, and A. Mikelić, Phys. Lett. A **101**, 376 (1984).
[5]E. H. Lieb, Rev. Mod. Phys. **53**, 603 (1981).
[6]E. Brezin, C. Itzykson, G. Parisi, and J. B. Zuber, Commun. Math. Phys. **59**, 35 (1978).
[7]I. Andrić, A. Jevicki, and H. Levine, Nucl. Phys. B **215**, 307 (1983).
[8]G. 't Hooft, Nucl. Phys. B **72**, 461 (1974); B **75**, 461 (1974).
[9]E. Witten, Nucl. Phys. B **160**, 57 (1979); B **223**, 433 (1983).
[10]H. Brézis, in *Variational Inequalities and Complementarity Problems*, edited by R. J. Cottle, F. Giannessi, and J-L. Lions (Wiley, New York, 1980).
[11]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals Series and Products* (Academic, New York, 1965), p. 229. Actually, in this reference, formulas 3.139.5 and 3.139.6 are misprints; the correct result is given by our formula (3.14).
[12]G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities* (Cambridge U.P., Cambridge, 1952).
[13]E. H. Lieb, Stud. Appl. Math. **57**, 93 (1977).
[14]I. Andrić and V. Bardek Phys. Rev. A **30**, 3319 (1984).

Fractional approximations to the Bessel function $J_0(x)$

Pablo Martín and Antonio L. Guerrero
Departamento de Física, Universidad Simon Bolivar, Apartado 80659, Caracas 1080-A, Venezuela

(Received 1 June 1984; accepted for publication 19 October 1984)

A method to obtain fractional approximations to the Bessel function $J_0(x)$ is reported here. This method improves a recently published one principally in that all the parameters are uniquely determined by linear equations. Our approximations give fairly good accuracy for all real, positive values. The maximum absolute error for the first-degree approximation is about 0.0035, and for the fourth-degree one, about 2.8×10^{-6}.

I. INTRODUCTION

Recently, a new method for the approximation of widely used functions in physics was introduced.[1] This method is based on the simultaneous expansion of the function for large and small values of the independent variable. This method yields good approximations to the Coulomb function. However, some difficulties appear, since the approximations depend on one phase parameter δ, which cannot be obtained from algebraic equations. This parameter must be found by trial.

A different method is presented here, keeping the basic idea of simultaneous use of large and small values of the independent variable. This latter method gives an excellent, though simple, approximation to the Bessel function $J_0(x)$ which can prove useful in most areas of physics (e.g., quantum mechanics, electrodynamics, plasma physics, geophysics, etc.) where $J_0(x)$ appears. The method is reminiscent of other two-point Padé approximations[2] such as the phase-amplitude method (widely used in nuclear physics[3]).

The method outlined here represents an improvement over that developed in earlier papers[1,4,5] in that (a) all the parameters of the approximation are determined from linear equations, (b) these equations and the parameters are real, (c) there is no need for finding by trial and error the phase parameter on which the old method depends, (d) there is a substantial improvement in accuracy as we shall show presently, and (e) fewer parameters are needed.

II. PROCEDURE

The differential equation for $y = J_0(x)$ is

$$\frac{d^2y}{dx^2}+\frac{1}{x}\frac{dy}{dx}+y=0\,.$$

In this equation it is possible, in principle, to substitute the function y by a quotient of two polynomials and, after rationalizing, to equate the lowest powers of the resulting polynomial. Thus, we could obtain the parameters of the approximation in a way similar to the Padé method. By doing so, however, the approximation would fail for large values of x. In order to obtain approximations for large values of x we could use equations arising from equating coefficients of the highest powers of the variable x, but then the values of the parameters so obtained would turn out to be zero, and no additional information would be obtained. To prevent this, transformations in the original equation are necessary before the polynomial quotient is substituted. The functions $\exp(ix)$ and $x^{-1/2}$ must be factored out of the dependent variable. Now, the factor $x^{-1/2}$ is singular at $x=0$. Since we want a regular solution at the origin we use the factor $(1+x)^{-1/2}$ instead.

The solution $y=J_0(x)$ to our equation becomes, after the transformations,

$$\begin{aligned}J_0(x) &= (1+x)^{-1/2}[w(x)\exp(ix)+w^*(x)\exp(-ix)]\\ &= (1+x)^{-1/2}[u(x)\cos x+v(x)\sin x]\,,\end{aligned}$$

where $w(x)\equiv\frac{1}{2}[u(x)-iv(x)]$ and satisfies

$$\begin{aligned}&x(1+x)^2w''+(1+x)(2ix^2+2ix+1)w'\\ &\quad+[(\tfrac{1}{4}+i)x+(-\tfrac{1}{2}+i)]w=0\,.\end{aligned}$$

The asymptotic expansion for $w(x)$ can be found and, since the leading term gives information about the real and imaginary parts of $w(x)$, we can thus determine the asymptotic expansions for $u(x)$ and $v(x)$. The respective ascending series for $u(x)$ and $v(x)$ are not known, however, for these functions do not satisfy separately the differential equation. Nevertheless, it is enough to know the ascending series of $u(x)\cos x+v(x)\sin x$, as a whole, in order to obtain the coefficients of the fractional approximation.

In approximating $u(x)$ and $v(x)$ by a polynomial quotient we require that the equations which yield the coefficients be real and linear to avoid multivaluation and make the calculations convenient. The simplest approximation (i.e., first-degree polynomial) would be

$$J_0(x)\cong(1+x)^{-1/2}\left(\frac{P_0+P_1x}{1+q_1x}\cos x+\frac{p_0+p_1x}{1+q_1x}\sin x\right),$$

subjected to the conditions

$$\begin{aligned}&\frac{P_0+P_1x}{1+q_1x}\cos x+\frac{p_0+p_1x}{1+q_1x}\sin x\\ &\quad\cong u(x)\cos x+v(x)\sin x\\ &\quad=(1+x)^{1/2}J_0(x)\\ &\quad=\begin{cases}\displaystyle\sum_{n=0}^{\infty}\left[\sum_{k=0}^{[n/2]}\frac{(-1)^k}{2^{2k}(k!)^2}\binom{\frac{1}{2}}{n-2k}\right]x^n, & x\ll 1,\\[2ex] \displaystyle\left[\frac{1}{\sqrt{r}}+O_1\left(\frac{1}{x}\right)\right]\cos x+\left[\frac{1}{\sqrt{r}}+O_2\left(\frac{1}{x}\right)\right]\sin x, & \\ x\gg 1, & \end{cases}\end{aligned}$$

where only the first three terms of the ascending series have been taken. Explicit expressions for $O_1(1/x)$ and $O_2(1/x)$ for

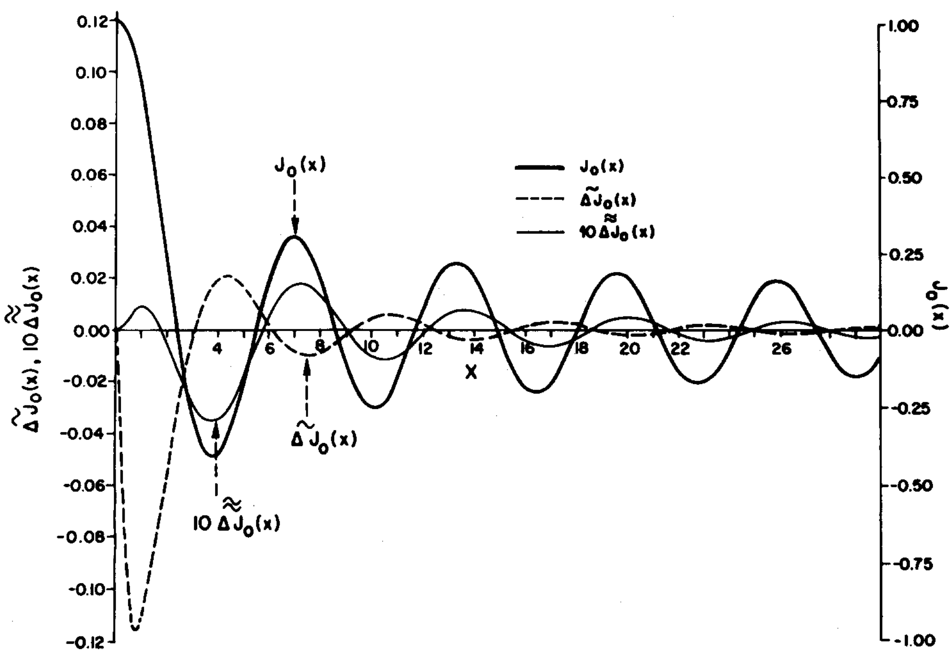


FIG. 1. The error of our latest approximation $\tilde{\tilde{J}}_0(x) - J_0(x)$ (multiplied by ten) compared with the error $\tilde{J}_0(x) - J_0(x)$ of the approximation obtained by a method recently published (see Ref. 1). The exact function $J_0(x)$ is also plotted for reference.

any order n of the approximation have been obtained from the differential equation.

However, for the present case $n = 1$ only the leading terms $1/\sqrt{\pi}$ are needed.

The above conditions fix the values of q_1, P_0, P_1, p_0, and p_1, so that we finally get

$$J_0(x) \cong (1+x)^{-1/2}\left[\frac{(8-4\sqrt{\pi})+x}{(8-4\sqrt{\pi})+\sqrt{\pi}x}\cos x + \frac{(3-\sqrt{\pi})+x}{(8-4\sqrt{\pi})+\sqrt{\pi}x}\sin x\right].$$

This approximation, even if evaluated to only three significant figures,

$$\tilde{\tilde{J}}_0(x) = (1+x)^{-1/2}\left(\frac{1.00+1.10x}{1+1.95x}\cos x + \frac{1.35+1.10x}{1+1.95x}\sin x\right),$$

gives a maximum error

$$\Delta\tilde{\tilde{J}}_0(x) \equiv \tilde{\tilde{J}}_0(x) - J_0(x)\,,$$

of 3.5×10^{-3} at $x = 3.7$, which represents a maximum relative error of 8.8×10^{-3}. The first 10 zeros of the exact function $J_0(x)$ and those of $\tilde{\tilde{J}}_0(x)$ are listed in Table I together with their percent relative error.

Higher-order approximations have also been obtained. For $n = 4$ the maximum error occurs at $x = 2.4$ and is less than 2.8×10^{-6}. Details of higher-order approximations to Bessel functions of integer order will be given in a subsequent paper.

A regular-size plot of $J_0(x)$ together with $\tilde{\tilde{J}}_0(x)$ shows no difference between these functions. Hence, we present in Fig. 1 the plot of 10 times the difference $\Delta\tilde{\tilde{J}}_0(x)$ together with the exact function $J_0(x)$ for easy reference. In Fig. 1 we also show, for comparison purposes, $\Delta\tilde{J}_0(x)$, which is the difference $\tilde{J}_0(x) - J_0(x)$, where

$$\tilde{J}_0(x) = (1+x)^{-1/2}\exp(ix) \times\left[\frac{(0.0162-0.226i)+(0.282-0.282i)x}{(0.156-0.404i)+x}\right] + \text{c.c.}$$

is the approximation to $J_0(x)$ obtained by the earlier method referred to at the beginning of this article [i.e., by that method (which is being published) applied to the Coulomb function[1,6]]. The figure shows that the maximum error of the approximation obtained by our new method is about 30 times lower than that of the approximation $J_0^{\sim}(x)$ obtained by the other one. Thus, the method we outline here is not only more accurate, but in addition no phase finding is needed.

Finally, we wish to stress the advantages of using a simple fractional approximation to $J_0(x)$ like ours. Although it is true that the convergence radius of $J_0(x)$ is infinite and any required accuracy may be obtained by use of the potential series taking a sufficiently large number of terms, this procedure could become fairly cumbersome if good accuracy is required for large x. Moreover, it is often found in physics that $J_0(x)$ appears under the integral sign frequently com-

TABLE I. Comparison between the zeros of the exact function $J_0(x)$ and those of our latest first-order approximation $\tilde{\tilde{J}}_0(x)$.

Zeros $J_0(x)$	Zeros $\tilde{\tilde{J}}_0(x)$	Relative error (percent)
2.4048	2.4020	0.12
5.5201	5.5219	0.033
8.6537	8.6557	0.024
11.7915	11.7933	0.016
14.9309	14.9325	0.011
18.0711	18.0725	0.0076
21.2116	21.2129	0.0061
24.3525	24.3536	0.0045
27.4935	27.4945	0.0037
30.6346	30.6355	0.0031

bined with other functions. Use of the potential series in this case may alter the region of convergence. Our approximation is an explicit and simple expression easy to handle in a variety of applications.

[1]E. Chalbaud and P. Martín, J. Math. Phys. **24**, 1268 (1984).
[2]W. J. Thron, "Two-Point Padé Tables, *T*-Fractions, and Sequences of Schur," in *Padé and Rational Approximation*, edited by E. B. Saff and R. S. Varga (Academic, New York, 1977), pp. 215–226.
[3]J. Zinn-Justin, Phys. Rep. C **1**, 57 (1971).
[4]P. Martín and A. L. Guerrero, Bull. Am. Phys. Soc. **27**, 874 (1982).
[5]P. Martín, in *Proceedings of the International Conference on Plasma Physics*, edited by H. Wilhelmsson and J. Weiland (Chalmers University of Technology, Goteborg, Sweden, 1982), p. 325.
[6]This approximation $\tilde{J}_0(x)$ had already been found by us before the method was applied to the Coulomb function. Its publication, however, was postponed until now due to the continual improvements of the method which finally led to $\tilde{\tilde{J}}_0(x)$.

The weak-Painlevé property as a criterion for the integrability of dynamical systems

A. F. Rañada
Departamento de Fisica Téorica, Universidad Complutense, 28040 Madrid, Spain
A. Ramani
Centre de Physique Théorique, Ecole Polytechnique, 91128 Palaiseau, France
B. Dorizzi and B. Grammaticos
Département de Mathématiques (M.T.I.), Centre National d'Etudes des Télécommunications, 92131 Issy Les Moulineaux, France

(Received 25 July 1984; accepted for publication 9 November 1984)

We investigate the validity of the weak-Painlevé property as an integrability criterion. We present an example of a time-dependent Hamiltonian system which possesses a weak-Painlevé type expansion, while presenting a chaotic behavior. However, this system presents also critical fixed singularities. The importance of the latter, as far as integrability is concerned, is discussed here.

I. INTRODUCTION

The singularity analysis has been resurrected in the past few years as an integrability criterion. Introduced a century ago, this method, usually associated to the name of Painlevé,[1] has been initially used in order to investigate integrability of nonlinear first- and second-order ordinary differential equations (ODE's). The recent use of this method concerned the integrability of nonlinear partial differential equations (PDE's).[2] However, as the original formulation of the Ablowitz–Ramani–Segur (ARS) conjecture dealt with ODE reductions of the PDE's, it was most natural to use this singularity analysis as a tool for the investigation of the integrability of dynamical systems described by ODE's. In that context, the most natural extension of the ARS conjecture would read like this: "A system of coupled nonlinear ODE's is integrable whenever it possesses the Painlevé property, i.e., the only movable singularities of the solutions in the complex t plane are poles." Several works have been devoted to the study of dynamical systems using the Painlevé property.[3,4,5,6,7] New integrable systems have thus been discovered and confirmed the particular usefulness of the Painlevé criterion. Whenever a system exhibits the Painlevé property it is integrable (although the precise meaning of integrability must be specified).

The reciprocal proposition seems less well-established. Starting from trivial examples, (e.g., Hamiltonian systems in one dimension), one can convince oneself that integrability can sometimes exist independently of a "nice" singularity structure. During the course of our investigations, we have discovered that some systems possess a particular intermediate status.[8]

They are integrable and, although they do not possess the full Painlevé property, they exhibit a simple singularity expansion in powers of $(t-t_0)^{1/r}$, with r an integer. We have called this property "weak-Painlevé." Several integrable systems have been discovered ranging from the initial 2-D Hamiltonians to N-dimensional systems[9] and even PDE's.[10]

However, recent findings, by one of us[11] make mandatory the examination of the weak-Painlevé property as integrability criterion. Namely the question we address ourselves to in this paper is whether the weak-Painlevé property is *always* sufficient for integrability.

II. PAINLEVÉ AND WEAK-PAINLEVÉ PROPERTIES ASSOCIATED TO INTEGRABILITY

In our initial work,[8,12,13] we have concentrated on autonomous systems (in fact, two-dimensional time-independent Hamiltonian systems). These systems present only movable singularities, and no fixed ones. For these systems, we believe that the weak Painlevé property suffices for integrability (although it is not always necessary).[14] When we turn now to time-dependent systems, two situations can arise: Either the fixed singularities are "nice," or they are not. The latter case is far from being an abstract one.

Consider the very simple case of the Riccati equation

$$\dot{x} = x^2 + f(t). \tag{1}$$

The movable singularities of this equation are pure poles. However, if $f(t)$ has singularities at finite values t_i of t, x has *fixed* singularities at these values, these singularities depending on the behavior of f near t_i. One can easily choose f in order that these singularities be critical (i.e., not poles). It is enough for f to have double poles:

$$f(t) = \alpha/t^2,$$

α not of the form $n(n-1)$ with n integer.

However, whatever f is, this equation can be reduced to the second-order linear equation

$$\ddot{y} = f(t)\,y, \tag{2}$$

by $x = \dot{y}/y$.

This equation is considered integrable because it is linear, independent of what the singularities of f are. In general, one cannot express y explicitly (except for very special choices of f), even if f has no singularities at finite t_i's, but still this is considered as an integrable case.

In fact, linearization can even accommodate critical singularities which are movable in some sense. Consider a time-independent system where one integration is explicitly possible. This reduces the original system to a new one with one less degree of freedom and a possible explicit time dependence.

In an earlier paper,[15] we have presented such a system starting from

$$\begin{aligned} \dot{x} &= -x^2 + axy + \alpha x + \beta y + \lambda, \\ \dot{y} &= -y^2 + bxy + \gamma x + \delta y + \mu. \end{aligned} \tag{3}$$

 0022-2488/85/040708-03$02.50

We are interested in $a = 0, \beta = 0$.

In this case the first equation for x separates:

$$\dot{x} = -x^2 + \alpha x + \lambda. \tag{4}$$

It is of Riccati type and, of course, it has the full Painlevé property.[1]

Integrating it, we obtain

$$x(t) = (r_1 + cr_2 e^{(r_2 - r_1)t})/(1 + ce^{(r_2 - r_1)t}),$$

with r_1, r_2 solutions of $r^2 - \alpha r - \lambda = 0$.

Choosing a solution for x, we can write the second equation as

$$\dot{y} = -y^2 + (bx(t) + \delta)y + (\gamma x(t) + \mu). \tag{5}$$

This equation is again a Riccati for y and its movable singularities at given x are poles. However, the Riccati for y could *a priori* have a "fixed" singularity which is worse than a pole. But this "fixed" singularity is really a movable singularity of the original system (3) because the pole of x is movable.

In Ref. 15, we have presented a detailed analysis of the conditions for the system (3) to possess the Painlevé property. However, what is clear from what we said above is that this is by no means essential for integrability: the two Riccati equations can be integrated in cascade, through the usual local linearization procedure one applies to the Riccati equations. So here we see a case where critical singularities that are fixed or even movable in the original system (although fixed in the reduced one) do not hinder integrability.

As a matter of fact, we do not know of any case of systems of nonlinear ODE's which possesses fixed critical singularities and is integrable otherwise than through a linearization.

Again let us recall that, according to the currently accepted definition of integrability, a linear ODE with variable coefficients is considered as integrable even if it presents critical fixed singularities. However, this does not necessarily mean that fixed critical singularities are not revelant for integrability. They may well be acceptable only whenever the system is linearizable.

In a recent work,[11] one of us has investigated the singularity structure of one-degree-of-freedom nonautonomous systems. One-degree-of-freedom, Hamiltonian, time-dependent systems fall in the class examined in detail by Painlevé and Gambier. There, the full-Painlevé property leads to integrability (although, sometimes at the expense of introducing new transcendents).

How about the weak Painlevé? The study of a system due to Sitnikov[16] has revealed that the weak-Painlevé property does not preclude chaotic behavior of the system. The equations of motion of the Sitnikov case are

$$\ddot{z} = -z/[z^2 + \tfrac{1}{4}(1 - \epsilon \cos\ 2t)^2]^{3/2}. \tag{6}$$

The singularity expansion of z can be written as $(\tau = t - t_0)$

$$z = z_0 + \sum_{k=4}^{\infty} \alpha_k \tau^{k/5}, \tag{7}$$

with $z_0 = \pm (i/2)(1 - \epsilon \cos 2t_0)$, t_0 free, and a_6 free (associated to the two resonances -1 and 6). A calculation of the first terms of the series yields

$$a_4 = (625/128\, z_0)^{1/5}, \quad a_5 = \pm i\epsilon \sin 2t_0, \quad a_7 = 0.$$

The above analysis concerns the movable singularities of (6).

We turn now to the fixed singularities.

It is clear that z in (6) can have a singular behavior whenever

$$\epsilon \cos 2t_0 = 1. \tag{8}$$

Expanding $1 - \epsilon \cos 2t_0$ around t_0, we obtain

$$(1 - \epsilon \cos 2t)^2 = 4B\tau^2 + \dots,$$

with $B = \epsilon^2 - 1$.

The leading behavior is, in this case,

$$z = \beta\tau^{2/3},$$

with $\beta^3 = \frac{9}{2}$.

Looking for the resonances we find them at -2 and $\frac{2}{3}$. The compatibility condition at $\frac{2}{3}$ is *not* satisfied. Thus, a logarithmic term enters the expansion at $n = \frac{2}{3}$. However, this critical singularity is fixed since t_0 is not free but given by (8).

III. CONCLUSION

So we are in the presence of an equation of motion (one dimensional, Hamiltonian, nonautonomous) which has a weak-Painlevé expansion around a movable singularity and which possesses a fixed critical singularity. Moreover, the solutions of this equation are known to exhibit chaotic behavior which makes them incompatible with integrability.

This could mean one of the two following things: Either allowing fractional powers is too weak a criterion in order to ensure integrability, or fixed singularities must also be taken into account. Our findings do not allow us to draw a clear conclusion at this stage. We can remark however that the predictive power of the weak-Painlevé property for time-independent systems (where fixed singularities do not arise) has been well established to date. On the other hand, up to second order, the full Painlevé property, i.e., movable poles only, does ensure integrability even in the presence of fixed critical singularities (e.g., Riccati), but then integrability is obtained through linearization. For higher order, however, it has not been proved yet that movable poles lead to integrability in the presence of fixed critical singularities.

It might turn out that even fixed critical singularities are not compatible with integrability for higher-order equations. One must acknowledge, at this point, Painlevé's powerful intuition. In his initial project[17] (always motivated by integrability), he was interested in equations with no critical singularities at all, although he devoted the major part of his work to equations with just no movable critical singularities.

[1] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1926).
[2] M. Ablowitz, A. Ramani, and H. Segur, Lett. Nuovo Cimento **23**, 333 (1978).
[3] T. Bountis, H. Segur, F. Vivaldi, Phys. Rev. A **25**, 1257 (1982).
[4] J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983).
[5] B. Grammaticos, B. Dorizzi, and R. Padjen, Phys. Lett. **89**, 111 (1982).
[6] J. Hietarinta, Phys. Rev. A **28**, 3670, (1983).
[7] C. R. Menyuk, H. H. Chen, and Y. C. Lee, Phys. Rev. A **27**, 1597 (1983).
[8] A. Ramani, B. Dorizzi, and B. Grammaticos, Phys. Rev. Lett. **49**, 1539, (1982).

[9]M. Lakshmanan and R. Sahadevan, Phys. Lett. **101**, 189 (1984).
[10]J. Weiss, J. Math. Phys. **24**, 1405 (1983).
[11]A. F. Ranada, in *Libro Homenaje a A. Duran*, edited by C. Sanchez del Rio (Univ. Complutense, Madrid, 1983).
[12]B., Dorizzi, B. Grammaticos, and A. Ramani, J. Math. Phys. **24**, 2282 (1983).
[13]B. Grammaticos, B. Dorizzi, and A. Ramani, J. Math. Phys. **24**, 2289 (1983).
[14]B. Dorizzi, B. Grammaticos, and A. Ramani, J. Math. Phys. **25**, 481 (1984).
[15]A. Ramani, B. Dorizzi, and B. Grammaticos, J. Math. Phys. **25**, 878 (1984).
[16]K. Sitnikov, Dokl. Akad. Nauk. USSR **133**, 303 (1960).
[17]P. Painlevé, Acta Math. **25**, 1 (1902).

Geometrical methods for the elasticity theory of membranes

Mark A. Peterson
Mount Holyoke College, South Hadley, Massachusetts 01075

(Received 13 January 1984; accepted for publication 24 August 1984)

The elasticity theory of curved membranes is developed using geometrical methods. The strain tensor is shown to be the Lie derivative of the metric tensor with respect to a flow. This formulation is coordinate-free, and can be expressed in any convenient coordinate system. A useful fixed coordinate system for the neighborhood of any given membrane surface is constructed. The formalism is used to expand the curvature free energy of a membrane about its minimum and to find the induced flow which minimizes shear dissipation in a curved membrane when the membrane is deformed at constant density. The shear free energy is expanded about its minimum for deformations of the shape at constant density. The curvature free energy is explicitly expanded about its minimum for shape changes of a closed membrane which conserve area and interior volume.

I. INTRODUCTION

The use of geometrical methods in theoretical continuum mechanics is now well established.[1,2] The utility of such methods is perhaps nowhere so immediate as in the elasticity theory of (curved) membranes, for in this case one is necessarily confronted with geometrical complexity right from the start. In practice, however, experiments on real membranes have been interpreted in terms of geometrically oversimplified models. One of the most interesting experiments on the red blood cell membrane, that of Lennon and Brochard on the Brownian shape fluctuations of red blood cells, has been interpreted by modeling the red blood cell as two infinite plane parallel membranes, separated by the approximate thickness of the cell.[3] Other experiments on membranes[4,5] have made similarly drastic simplifying assumptions about the geometry of the membrane. The picture which emerges is not consistent: estimates of the bending modulus of biological membranes in the references cited above differ from each other by more than an order of magnitude.

The theoretical uncertainty about analyses of existing experiments can be removed completely. Modern geometrical methods lend themselves very well to real computations. The methods described in this paper are intended to bridge the gap which presently exists between theory and experiment in membrane mechanics. The motivation was the above-mentioned puzzle about biological membranes, but the methods and results are considerably more general. They have already led to exact infinitesimal stability results for the red blood cell shape[6] and exact results for Brownian shape fluctuations of red cells.[7]

The basic idea is to formulate elasticity theory in coordinate-free form, and then to choose coordinates adapted to the problem at hand to simplify actual computations—a familiar strategy, of course. Our basic coordinate-free tools will be the calculus of variations and the Lie derivative. An example of a coordinate-free statement in elasticity theory is that the Lie derivative of the Euclidean metric is twice the strain.[1,2,8] This beautiful and useful result, rederived in Sec. II, is perhaps not as well known as it should be. In Sec. III a very advantageous coordinate system for membrane problems is described. It is a natural choice and has been used extensively in shell theory,[9] but is characterized here for completeness. In Sec. IV familiar results of membrane theory are shown to follow immediately from the results of Secs. II and III. In Sec. V a new result is obtained, the expansion of the Helfrich curvature free-energy[10] about a reference configuration. In Sec. VI a similar result is obtained for the membrane shear free-energy. In Sec. VII these free-energy expansions are restricted to the case of deformations which preserve the volume and surface area of a cell. (This last is really the case of interest in, say, Brownian shape fluctuations of red blood cells.)

II. STRAIN AS A LIE DERIVATIVE

Let a vector field X in $\mathbf{R}^3$, given in local coordinates by

$$X = a^i(x)\frac{\partial}{\partial x^i}, \tag{1}$$

generate a flow

$$x^i(0) \to x^i(s), \tag{2}$$

defined by the solution to the ordinary differential equations

$$\frac{dx^i}{ds} = a^i(x), \tag{3}$$

with initial condition $x^i = x^i(0)$. Let g be the usual Euclidean metric on $\mathbf{R}^3$. Then the first-order strain U associated with the flow of X is just

$$U = \tfrac{1}{2} L_X\, g, \tag{4}$$

where L_X is the Lie derivative[11] with respect to X. To prove Eq. (4), evaluate in Cartesian coordinates:

$$\begin{aligned} 2U_{jk} &= 2U\left(\frac{\partial}{\partial x^j},\frac{\partial}{\partial x^k}\right) = (L_X\, g)\left(\frac{\partial}{\partial x^j},\frac{\partial}{\partial x^k}\right) \\ &= Xg\left(\frac{\partial}{\partial x^j},\frac{\partial}{\partial x^k}\right) + g\left(\left[\frac{\partial}{\partial x^j},X\right],\frac{\partial}{\partial x^k}\right) \\ &\quad + g\left(\frac{\partial}{\partial x^j},\left[\frac{\partial}{\partial x^k},X\right]\right) \\ &= g_{ji}\frac{\partial a^i}{\partial x^k} + g_{ki}\frac{\partial a^i}{\partial x^j} = \frac{\partial a_j}{\partial x^k} + \frac{\partial a_k}{\partial x^j}, \end{aligned} \tag{5}$$

in agreement with, say, Landau and Lifschitz.[12] Of course,

 0022-2488/85/040711-07$02.50

Eq. (4) is coordinate-free, which is its main advantage for this paper.

If M, representing a membrane, is a two-dimensional surface smoothly embedded in $\mathbb{R}^3$

$$i{:}M \rightarrow \mathbb{R}^3, \tag{6}$$

and we now imagine M to be carried along with the flow generated by X, then the first-order metric strain U_M in M is just the restriction of U to M, namely,

$$U_M = i^* U. \tag{7}$$

Higher order strains can be handled in the same way, by going to higher order in the Lie–Taylor series for the metric tensor, but this is unnecessary for statics, linear hydrodynamics, or infinitesimal stability studies.

III. ADAPTED COORDINATES

Since Eqs. (4) and (7) for the strain induced in a membrane by the flow X are coordinate-free, one can evaluate them in coordinates especially chosen to simplify computations (and exposition). This section introduces a fixed coordinate system with advantages for membrane problems, which will then be used, where necessary, in the remaining sections of the paper. Let P be a point of M. A coordinate system (x^1, x^2, x^3) will be called "adapted at $P \in M$" if it has the following properties: (1) it is defined in some three-dimensional neighborhood $U \subset \mathbb{R}^3$ of $P \in M \subset \mathbb{R}^3$; (2) $M \cap U$ is the locus $x^1 = 0$; (3) the Euclidean metric takes the form $g = \mathrm{diag}(1, g_{22}, g_{33})$ in U; (4) at P we have $g_{22} = g_{33} = 1$ and $g_{22,2} = g_{33,3} = 0$; and (5) the level curves $x^2 = \text{const}$ and $x^3 = \text{const}$ in $M \cap U$ have as tangents the principal curvature directions of M.

Let c_2 and c_3 denote the principal curvatures at $P \in M$. If $c_2 = c_3$, P is called an *umbilic*.

Theorem: If $P \in M$ is not an umbilic, then there exists a coordinate system adapted to M at P.

The proof is given in Appendix A.

Theorem: If $P \in M$ is an umbilic, and if an entire neighborhood of $P \in M$ consists of umbilic points, then there exists a coordinate system adapted to M at P.

Proof: By a theorem of Hilbert,[13] a surface which consists entirely of umbilic points is either (1) spherical or (2) planar. In the first case, spherical polar coordinates with $x^1 = r$ and P coordinatized as a point on the equator of the unit sphere are adapted at $P \in M$. In the second case, Cartesian coordinates are adapted.

If P is an umbilic, but every neighborhood of $P \in M$ contains nonumbilic points, there may not exist a coordinate system adapted at $P \in M$. The existence theorems above, though, are sufficient for the purposes of this paper. The result of any computation in adapted coordinates can be extended to the umbilic points by continuity. All results will have a coordinate-free significance and will be given a coordinate-free expression (valid also at the umbilics). The umbilic points are, in short, coordinate singularities of the coordinate system we choose to work in, but they are in no sense physical singularities.

If membrane hydrodynamics involved only the intrinsic geometry of M, the simplest coordinate system to use on M in a neighborhood of P would be a geodesic coordinate system. All Christoffel symbols would vanish at P. Adapted coordinates simplify the discussion of extrinsic geometry—the way M is embedded in $\mathbb{R}^3$. The price paid for this is a slightly more complicated description of the intrinsic geometry. In adapted coordinates the following Christoffel symbols are, as a rule, nonzero at P:

$$\begin{aligned}
\begin{Bmatrix} 2 \\ 23 \end{Bmatrix} &= \begin{Bmatrix} 2 \\ 32 \end{Bmatrix} = \frac{1}{2} g^{22} g_{22,3}, \\
\begin{Bmatrix} 2 \\ 33 \end{Bmatrix} &= -\frac{1}{2} g^{22} g_{33,2}, \\
\begin{Bmatrix} 3 \\ 32 \end{Bmatrix} &= \begin{Bmatrix} 3 \\ 23 \end{Bmatrix} = \frac{1}{2} g^{33} g_{33,2}, \\
\begin{Bmatrix} 3 \\ 22 \end{Bmatrix} &= -\frac{1}{2} g^{33} g_{22,3}.
\end{aligned} \tag{8}$$

IV. DILATION AND SHEAR OF MEMBRANES

A. Tangential motion

Take coordinates adapted at $P \in M$ and consider a flow generated by a vector field X, which moves the material of the membrane M but leaves the external shape of M fixed. In a neighborhood of P, then, X has the form

$$X = u^2(x) \frac{\partial}{\partial x^2} + u^3(x) \frac{\partial}{\partial x^3}. \tag{9}$$

What happens off the membrane is irrelevant to the membrane strain, so impose the additional condition

$$\frac{\partial u^2}{\partial x^1} = \frac{\partial u^3}{\partial x^1} = 0. \tag{10}$$

The more compact notation $u^2{}_{,1} = u^3{}_{,1} = 0$ will also be used. In fact one may now drop all reference to the coordinate x^1. All expressions in this section refer to the intrinsic two-dimensional geometry of M. Index sums go over the values 2 and 3 only.

Let $\tilde{g}$ be the restriction of the Euclidean metric g to the surface M, also called the first fundamental form of the surface. By choice of coordinates, $\tilde{g} = \mathrm{diag}(g_{22}, g_{33})$. Then, by the arguments of Sec. II, the symmetric tensor

$$U = \tfrac{1}{2} L_X \tilde{g} \tag{11}$$

is just the first-order strain associated with X. A computation in adapted coordinates verifies this:

$$U_{ij} = \tfrac{1}{2} (u_{i;j} + u_{j;i}), \tag{12}$$

where the semicolon denotes the covariant derivative using the metric connection on M. [See Eq. (8).]

It is natural to break U into a trace, which measures dilation strain, and a traceless part, shear strain. One finds, not surprisingly,

$$\mathrm{Tr}\, U = \tilde{g}^{ij} U_{ij} = u^i{}_{;i} = \mathrm{div}\, u, \tag{13}$$

$$S_{ij} = U_{ij} - \tfrac{1}{2} (\mathrm{Tr}\, U) \tilde{g}_{ij}, \tag{14}$$

as dilation and shear strain, respectively, for purely tangential motions.

B. Normal motion

Now consider motion normal to the membrane. In particular, consider the vector field

$$X = n(x^2,x^3)\frac{\partial}{\partial x^1}, \tag{15}$$

which is normal to M. The flow generated by X carries M into a new surface M', and in fact any small shape deformation of M may be represented this way. (Tangential flows may redistribute the material within the shape, but such flows do not contribute to a change of shape. For the moment consider this purely normal flow.)

As the flow along X proceeds from parameter value $s = 0$ to $s = 1$ [see Eq. (2)], the points of the membrane move according to

$$(0,x^2,x^3) \mapsto (n(x^2,x^3),x^2,x^3), \tag{16}$$

and the Euclidean metric components at the location of the membrane change according to

$$\mathrm{diag}(1,g_{22},g_{33}) \mapsto \mathrm{diag}(1, g_{22} + 2nc_2\, g_{22} + n^2c_2{}^2\, g_{22}, \\ g_{33} + 2nc_3\, g_{33} + n^2c_3{}^2\, g_{33}), \tag{17}$$

where c_2 and c_3 are the principal curvatures on M. This result is derived in Appendix A.

One half the Lie derivative of the metric with respect to X gives, as always, the first-order metric strains associated with the flow. In particular, restricting the strain tensor to the membrane $x^1 = 0$ gives the strains in the membrane in case the material really does follow this flow. One finds in this way the strain induced by the flow of Eq. (15)

$$U = \tfrac{1}{2}\mathscr{L}_X\, g\,|_{x^1=0} = nC, \tag{18}$$

where $C_{22} = c_2\, g_{22}$, $C_{33} = c_3\, g_{33}$, $C_{32} = C_{23} = 0$, in adapted coordinates. (C is the second fundamental form of the surface.)

In particular, there is dilation strain

$$\mathrm{Tr}(nC) = n\,\tilde{g}^{ij}\, C_{ij} = n(c_2 + c_3) = 2nH, \tag{19}$$

where $H = (c_2 + c_3)/2$ is the mean curvature, and shear strain

$$S = nC - nH\,\tilde{g}. \tag{20}$$

C. General motion

The general first-order dilation strain and shear strain is the superposition of Eq. (13) for tangential flow and Eq. (19) for normal flow (by linearity of the Lie derivative)

$$\mathrm{Tr}(U) = \mathrm{div}\, u + 2nH, \tag{21}$$

$$S_{ij} = \tfrac{1}{2}(u_{i;j} + u_{j;i}) - \tfrac{1}{2}(\mathrm{div}\, u)\,\tilde{g}_{ij} + n(C_{ij} - H\,\tilde{g}_{ij}). \tag{22}$$

These are standard results of membrane theory.

D. Free energies

Associated to the metric strains is a free energy density F and a dissipation function D. The assumption of linear hydrodynamics is that, with respect to an equilibrium state which is not strained, one has

$$F = \tfrac{1}{2} A^{ijkl}\, U_{ij}\, U_{kl}, \tag{23}$$

$$D = \tfrac{1}{2} B^{ijkl}\, V_{ij}\, V_{kl}, \tag{24}$$

where V_{ij} has the same form as U_{ij} but with displacements u_i, n replaced by their time derivatives $v_i = \dot{u}_i$, $\dot{n}$. The coefficient tensors A and B have the symmetries implied by Eq. (23). In an isotropic material, standard phenomenology reduces Eqs. (23) and (24) to

$$F = \tfrac{1}{2}\sigma(\mathrm{div}\, u + 2nH)^2 + \tfrac{1}{2}\mu\, S^{ij} S_{ij}, \tag{25}$$

$$D = \tfrac{1}{2}\rho(\mathrm{div}\, v + 2\dot{n}H)^2 + \tfrac{1}{2}\eta\, \dot{S}^{ij} \dot{S}_{ij}, \tag{26}$$

where σ is the surface tension, μ is the shear modulus, ρ is the dilation viscosity, and η is the shear viscosity.

In addition to dilation and shear free energies, there is still another membrane free energy, namely the free energy of bending. There exists a widely accepted expression for this free energy, first given by Helfrich,[10] but what one really needs in many applications is its expansion to second-order around an equilibrium configuration. [This would cast it in the same form as Eq. (23).] That expansion is carried out in the next section, where Lie derivatives and adapted coordinates play an essential role in the derivation.

V. BENDING FREE ENERGY

It has been pointed out by Helfrich[10] that the most general bending free energy density of degree 2 or less in the principal curvatures for a homogeneous, isotropic membrane is, up to a divergence,

$$F_c = \tfrac{1}{2} k_c (2H - c_0)^2, \tag{27}$$

where k_c is the bending modulus, H is the mean curvature, and c_0 is a constant (tending to bias the mean curvature). In what follows, c_0 could actually be an arbitrary function.

For hydrodynamics, one wishes to know the variation of F_c with respect to a displacement of the membrane away from the equilibrium shape, or, what is equivalent, the variation of H with respect to the flow X of Eq. (15). To handle the general case of an equilibrium configuration which, because of constraints, does not satisfy $2H \equiv c_0$, one needs both the first and second variations.

This problem, in turn, reduces to finding (to second order in n) the mean curvature H' of the surface M', described after Eq. (15). The simplest way to do that is to find the dilation strain on M' associated to a flow generated by the unit normal of M'. That strain would be $2H'$, according to Eq. (19).

A unit vector field normal to M' (through second order in n) is, from Eqs. (16) and (17),

$$X' = a\frac{\partial}{\partial x^1} + b\frac{\partial}{\partial x^2} + c\frac{\partial}{\partial x^3}, \tag{28}$$

where

$$\begin{aligned} a &= 1 - \tfrac{1}{2}(g^{22}\, n_{,2}\, n_{,2} + g^{33}\, n_{,3}\, n_{,3}), \\ b &= -g^{22}\, n_{,2}(1 - 2x^1c_2), \\ c &= -g^{33}\, n_{,3}(1 - 2x^1c_3). \end{aligned} \tag{29}$$

Suitable tangent vector fields to the surface M', on which to evaluate the strain tensor, are

$$\begin{aligned} T_2 &= \frac{\partial}{\partial x^2} + \frac{\partial n}{\partial x^2}\frac{\partial}{\partial x^1}, \\ T_3 &= \frac{\partial}{\partial x^3} + \frac{\partial n}{\partial x^3}\frac{\partial}{\partial x^1}. \end{aligned} \tag{30}$$

A straightforward computation in adapted coordinates then yields (see Appendix B)

$$H' = H_0 + H_1 + H_2, \tag{31}$$

where

$$2H_0 = 2H = c_2 + c_3, \tag{32}$$

$$2H_1 = -\nabla^2 n - (c_2{}^2 + c_3{}^2)n, \tag{33}$$

$$2H_2 = (c_2{}^3 + c_3{}^3)n^2 + 2C^{ij}nn_{,i;j} + (C^{ij} - H\tilde{g}^{ij})n_{,i}\,n_{,j} + 2(C^{ij} - H\tilde{g}^{ij})_{;i}nn_{,j}. \tag{34}$$

Here, ∇^2 is the Laplace–Beltrami operator on M. It is also useful to know the area element on M' to second order in n:

$$\sqrt{\det \tilde{g}} \rightarrow \sqrt{\det \tilde{g}}\,(1 + n(c_2 + c_3) + \tfrac{1}{2}|\nabla n|^2 + c_2 c_3 n^2). \tag{35}$$

Putting in the above expansions and collecting terms, one finds the Helfrich curvature free energy to second order in n is

$$\mathscr{F}_c = \frac{k_c}{2}\int [F_0 + F_1 + F_2]\sqrt{\det \tilde{g}}\, dA, \tag{36}$$

where

$$F_0 = (2H - c_0)^2, \tag{37}$$

$$F_1 = A_1 \nabla^2 n + B_1 n, \tag{38}$$

$$F_2 = (\nabla^2 n)^2 + A_2^{ij} n_{,i}\, n_{,j} + B_2 n^2, \tag{39}$$

and

$$A_1 = -2(2H - c_0), \tag{40}$$

$$B_1 = -(c_2 - c_3)^2(c_2 + c_3) - 4c_0 c_2 c_3 + c_0{}^2(c_2 + c_3), \tag{41}$$

$$A_2^{ij} = -2(2H - c_0)(C^{ij} - H\tilde{g}^{ij}) + \tilde{g}^{ij}[-\tfrac{3}{2}(c_2{}^2 + c_3{}^2) + c_2 c_3 - c_0(c_2 + c_3) + \tfrac{1}{2}c_0{}^2], \tag{42}$$

$$B_2 = c_2{}^4 + c_3{}^4 - (c_2 c_3{}^3 + c_3 c_2{}^3) + c_0{}^2 c_2 c_3 + \nabla^2(c_2{}^2 + c_3{}^2) + 2[(C^{ij} - H\tilde{g}^{ij})(2H - c_0)_{,i}]_{;j}. \tag{43}$$

The corresponding dissipation function would have the same form, with n replaced by $\dot{n}$.

VI. SHEAR FREE ENERGY AND SHEAR FLOW

In a general distortion of a membrane, one may expect the various free energy terms to fall into a hierarchy, with (dilation energy)$\gg$(shear energy)$\gg$(bending energy). This seems to be true, at least, for red blood cell membranes.[14] It may happen, however, that some of these terms are in fact absent. A distortion at constant density has no dilation energy, and if the shear modulus is zero (fluid membrane) there is no shear energy, so that the bending energy becomes the dominant term.

It was pointed out in Eqs. (19) and (20) that, without an induced tangential flow in the membrane, bending implies both dilation and shear. In actuality, because of the above-mentioned hierarchy of energies, membrane material will move to keep the density constant and minimize shear energy. In a fluid membrane, the material will move to keep the density constant and minimize entropy production in the shear dissipation function, which is formally the same problem.

In either case, associated to a bending distortion

$$X = n\frac{\partial}{\partial x^1}, \tag{44}$$

which does not change the area, we seek an induced flow

$$Y = v^2\frac{\partial}{\partial x^2} + v^3\frac{\partial}{\partial x^3}, \tag{45}$$

such that if S is the membrane shear strain associated with $X + Y$, then

$$\delta\int S^{ij} S_{ij}\sqrt{\tilde{g}}\, dA = 0, \tag{46}$$

subject to the constraint

$$\operatorname{div} Y + 2nH = 0. \tag{47}$$

Introduce the one-form v by

$$v = v_2\, dx^2 + v_3\, dx^3, \tag{48}$$

where $v_i = g_{ij} v^j$. This one-form can always be represented in terms of scalars α and β as

$$v = d\alpha + *\, d\beta, \tag{49}$$

where $*$ is the Hodge star operator. In fact,

$$\operatorname{div} v = *\, d *v = \nabla^2\alpha, \tag{50}$$

$$\operatorname{curl} v = *\, dv = \nabla^2\beta, \tag{51}$$

so that α and β are essentially determined by div v and curl v, respectively.

Assume for simplicity that M has no boundary, or that $v = 0$ on the boundary. Then α is determined up to an irrelevant constant by the constraint Eq. (47)

$$\nabla^2\alpha = -2nH. \tag{52}$$

The integrability condition $\int nH\sqrt{\tilde{g}}\, dA = 0$ is satisfied because, by assumption the total surface area is constant.

The constraint on α is incorporated into the variational problem Eq. (46) by varying β only. The coordinate form of Eq. (49) is

$$v_i = \alpha_{,i} + e_i{}^j\beta_{,j}, \tag{53}$$

where e_{ij} is the fully antisymmetric tensor

$$e_{ij} = \sqrt{\det \tilde{g}}\,\epsilon_{ij}, \tag{54}$$

and $\epsilon_{12} = -\epsilon_{21} = 1$, $\epsilon_{22} = \epsilon_{11} = 0$. Varying β in Eq. (46) leads to

$$S_{ij;k;l}\, g^{jk} e^{il} = 0, \tag{55}$$

a fourth-order partial differential equation for β.

This equation can be made more explicit using standard results of tensor analysis in two dimensions, in particular,

$$\begin{aligned} &g_{\alpha\beta;\gamma} = 0,\\ &e_{\alpha\beta;\gamma} = 0,\\ &A^{\alpha}{}_{;\beta;\gamma} - A^{\alpha}{}_{;\gamma;\beta} = R^{\alpha}{}_{\delta\beta\gamma}A^{\delta},\\ &R_{\alpha\beta\gamma\delta} = c_2 c_3(g_{\alpha\delta}g_{\beta\gamma} - g_{\alpha\gamma}g_{\beta\delta}). \end{aligned} \tag{56}$$

The resulting equation for β is

$$\nabla^4\beta + 2(c_2 c_3 \beta_{,i})_{;j}\, g^{ij} + 2\alpha_{,i}(c_2 c_3)_{,j}e^{ij} + [n(C^{ij} - H g^{ij})]_{;i;k}\, e_j{}^k = 0. \tag{57}$$

The corresponding homogeneous equation for β has nontrivial solutions if M is invariant under a rigid motion, like rotation about a symmetry axis. To this extent the solution to Eq. (57) is not unique.

It is noteworthy that the induced vorticity $\nabla^2\beta$ is not zero as a rule. Equations (52) and (57) together constitute an explicit equation for it. Thus, membrane vorticity and shape change are coupled.

The functions α and β, and the displacement v, which they define by Eq. (53), might arise in two different ways. The displacement v is, in one view, the tangential flow induced by the shape change n [Eq. (44)] which conserves area and minimizes shear dissipation in a fluid membrane. In a different interpretation, the same flow minimizes the increase in the shear free energy of a solid membrane under the area-preserving shape change [Eq. (44)], keeping the density fixed. Evaluating the shear free energy [Eq. (25)] for this choice of tangential displacement represents the free energy of a solid incompressible membrane as a quadratic form in n. This quadratic form, representing the shear free energy of an incompressible membrane, must be left implicit, since it requires the solution of Eqs. (52) and (57). By contrast, one can write the curvature free energy [Eq. (27)] as an explicit quadratic form in n, as in Eqs. (36)–(43).

VII. CONSTRAINTS: CONSTANT VOLUME AND SURFACE AREA

Consider membranes which bound a finite volume—topological spheres, for example. One may well be interested only in motions which preserve the total area and interior volume of the membrane. The purely tangential flows described in Sec. III A certainly satisfy this constraint. The normal flow of Sec. III B does not satisfy the constraint unless one puts additional conditions on the function $n(x^2, x^3)$ in Eq. (15). In this section we find the consequences of this constraint for the expansion of the Helfrich curvature free energy and the shear free energy. If one thinks of these expansions as quadratic forms in n, then what we are finding in this section is the restriction of the quadratic form to the subspace (really a Banach submanifold) which satisfies the global constraint.

The idea is familiar from perturbation theory: represent the normal deformation $n(x^2, x^3)$ as

$$n = n_1 + n_2, \tag{58}$$

where n_1 is such as to leave the volume and surface area unchanged to first order in n_1, and where $n_2 = \mathcal{O}(n_1^2)$ is chosen to enforce the constraint also to second order. Thus the deformation can be parametrized by n_1, but in the expansion of free energies there are quadratic terms not only from the second order in n_1 but also from first order in n_2. In the absence of the constraint, the latter terms would not be necessary.

The changes in volume and surface area are

$$\delta V = \int n_1\sqrt{g}\, dA + \left[\int n_2\sqrt{g}\, dA + \int n_1^2 H\sqrt{g}\, dA\right], \tag{59}$$

$$\delta A = 2\int n_1 H\sqrt{g}\, dA + 2\int n_2 H\sqrt{g}\, dA + \int\left[\frac{1}{2}|\nabla n_1|^2 + K n_1^2\right]\sqrt{g}\, dA, \tag{60}$$

where K is the Gauss curvature. In order for the volume and surface area to be constant to second order in n_1, require

$$0 = \int n_1\sqrt{g}\, dA = \int n_1 H\sqrt{g}\, dA, \tag{61}$$

and take

$$n_2 = a + bH, \tag{62}$$

where

$$a = [-\langle H^2\rangle\langle n_1^2 H\rangle + \tfrac{1}{2}\langle H\rangle\langle \tfrac{1}{2}|\nabla n_1|^2 + n_1^2 K\rangle]/S, \tag{63}$$

$$b = [\langle H\rangle\langle n_1^2 H\rangle - \tfrac{1}{2}\langle \tfrac{1}{2}|\nabla n_1|^2 + n_1^2 K\rangle]/S, \tag{64}$$

and $S = \langle H^2\rangle - \langle H\rangle^2$. In the above expressions $\langle\ \ \rangle$ means average over the surface.

Substituting n of the form of Eq. (58), where n_1 and n_2 obey Eqs. (61)–(64), into Eqs. (36)–(43) gives the Helfrich curvature free energy as a quadratic form in n_1 for constrained motions about equilibrium, and similarly for the shear free energy. The expansion of the Helfrich curvature free energy for constrained motion about an extremum is

$$F_c = \frac{1}{2}k_c\int_M [(\nabla^2 n_1)^2 + A^{ij} n_{,i} n_{,j} + B n_1^2]\sqrt{g}\, dA, \tag{65}$$

where

$$A^{ij} = -2(2H - c_0)(C^{ij} - Hg^{ij}) + g^{ij}[-6H^2 + 4K - 2c_0 H - (Q - R\langle H\rangle)/4S], \tag{66}$$

$$B = 16H^4 - 20H^2K + 4K^2 + 2\nabla^2(2H^2 - K) + 4[(C^{ij} - g^{ij})H_{,i}]_{,j} + Q(\langle H\rangle H - K/2)/S + R(\langle H^2\rangle H - \langle H\rangle K/2)/S. \tag{67}$$

Here

$$Q = \langle 4|\nabla H|^2 - 8H^4 + 4HK(2H - c_0)\rangle, \tag{68}$$

$$R = \langle 8H^3 - 4K(2H - c_0)\rangle, \tag{69}$$

$$S = \langle H^2\rangle - \langle H\rangle^2, \tag{70}$$

$$H = (c_2 + c_3)/2, \tag{71}$$

$$K = c_2 c_3. \tag{72}$$

A numerical check of this expression has been carried out in the course of a stability study of the red blood cell shape.[6] The basis of the check is that there are motions which preserve area and volume but do not affect the curvature free-energy at all, namely rigid translations and rotations in $\mathbb{R}^3$. These motions should emerge from a numerical analysis as null vectors of the quadratic form [Eq. (65)], and they do, to an accuracy of many figures. See Ref. 6 for details.

APPENDIX A: EXISTENCE OF ADAPTED COORDINATES NEAR A NONUMBILIC POINT

At each point of M imagine a short line segment normal to M. Let the x^1 coordinate of any point near M be its Euclidean distance from M measured along this line (positive on one side and negative on the other). Clearly M itself will be the locus $x^1 = 0$. Since M is smoothly embedded, its principal curvatures are bounded and this construction is valid for $|x|^1$ sufficiently small.

Let Y be a vector field on $U \cap M$ tangent to the principal curvature direction of the smaller principal curvature, and Z a vector field on $U \cap M$ tangent to the principal curvature direction of the larger principal curvature. (We assume $U \cap M$ is small enough to contain no umbilic points, so that smaller and larger are unambiguous.) Their solution curves are everywhere orthogonal in $U \cap M$. Let x^2 be any continuous monotonic labeling of the solution curves of Y, and x^3 any continuous monotonic labeling of the solution curves of Z. Then x^2 and x^3 are a coordinate system in $U \cap M$. By coordinate changes of the form $x^2 \to \tilde{x}^2(x^2)$, $x^3 \to \tilde{x}^3(x^3)$ it is possible to make $g_{22} = g_{33} = 1$ and $g_{22,2} = g_{33,3} = 0$ at P. Since these are just different monotonic labels, assume x^2 and x^3 were chosen that way to begin with.

Extend x^2 and x^3 to the whole of U by taking them to be constant along the normal line segments parameterized by x^1.

Computation of the Euclidean metric components in these coordinates now completes the proof, since every condition is seen to be satisfied except (3). In fact, the metric does have the required form $g = \mathrm{diag}(1, g_{22}, g_{33})$ at points in $M \subset \mathbb{R}^3$. Take any point $Q \in U \cap M$ and assume, without loss of generality, that its coordinates are (0,0,0). Let y^i be coordinates in the tangent space to $\mathbb{R}^3$ at Q such that

$$y^i\left(\frac{\partial}{\partial x^j}\right) = \delta^i{}_j,$$

and identify that tangent space with $\mathbb{R}^3$ in such a way that Q is the origin for the y coordinates. One has

$$g\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j}\right) = [\mathrm{diag}(1, g_{22}, g_{33})]_{ij}$$

everywhere in U. A Taylor expansion of the y coordinates around $x^2 = x^3 = 0$ gives, to first order in x^2, x^3,

$$y^1 \doteq x^1,$$
$$y^2 \doteq x^2(1 + c_2 x^1),$$
$$y^3 \doteq x^3(1 + c_3 x^1),$$

where c_2, c_3 are the principal curvatures at Q. The Jacobian matrix of this approximate coordinate transformation at $x^2 = x^3 = 0$ is actually exact, since the terms omitted were second-order. Using it to find the components of g in the x coordinates at points over Q gives

$$g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = [\mathrm{diag}(1, g_{22}{}^*, g_{33}{}^*)]_{ij},$$

where

$$g_{22}{}^* = g_{22}(1 + c_2 x^1)^2,$$
$$g_{33}{}^* = g_{33}(1 + c_3 x^1)^2.$$

Since there is nothing special about the point Q, the result must hold generally in U. (Of course c_2 and c_3 then refer to the "base point" obtained by projecting back to M.)

This completes the proof that adapted coordinates exist. We have proved slightly more: the exact form of the metric components off the membrane will be useful in Sec. IV.

APPENDIX B: VARIATION OF MEAN CURVATURE

First find the strain tensor to second order in n. [See Eqs. (4), (6), and (28)–(30).] One has

$$\begin{aligned} X'g(T_2, T_2) &= 2g_{22}\, c_2 + g_{22}\, 2nc_2{}^2 \\ &\quad - g_{22}\, c_2(n_{,2}{}^{,2} + n_{,3}{}^{,3}) \\ &\quad - 2n_{,2}\, nc_{2,2} + 2n_{,3} \begin{Bmatrix} 3 \\ 22 \end{Bmatrix} \\ &\quad + 2nc_3\, n_{,3}\, g^{33}\, g_{22,3} - 2g^{33}\, n_{,3}\, n(g_{22}c_2)_{,3}. \end{aligned} \tag{B1}$$

Also,

$$\begin{aligned} [T_2, X'] &= a_{,2} \frac{\partial}{\partial x^1} + b_{,2} \frac{\partial}{\partial x^2} + c_{,2} \frac{\partial}{\partial x^3} \\ &\quad + n_{,2}\, b_{,1} \frac{\partial}{\partial x^2} + n_{,2}\, c_{,1} \frac{\partial}{\partial x^3} \\ &\quad - bn_{,22} \frac{\partial}{\partial x^1} - cn_{,23} \frac{\partial}{\partial x^1}, \end{aligned} \tag{B2}$$

so that

$$2g(T_2, [T_2, X']) = -2n_{,22} + 4n_{,2}\, nc_{2,2} + 4c_2\, n_{,2}\, n_{,2}. \tag{B3}$$

The sum of Eqs. (B1) and (B3) is twice the (2,2) component of strain:

$$\begin{aligned} 2U_{22} &= 2c_2\, g_{22} + 2nc_2{}^2\, g_{22} - 2n_{,2;2} \\ &\quad - c_2\, g_{22}(n_{,2}{}^{,2} + n_{,3}{}^{,3}) \\ &\quad + 2n_{,2}\, nc_{2,2} - 2g^{33}\, n_{,3}\, n(g_{22}c_2)_{,3} \\ &\quad + g^{33}\, n_{,3}\, 2nc_3\, g_{22,3} + 4n_{,2}\, n_{,2}c_2. \end{aligned} \tag{B4}$$

To find $4H' = g^{ij}\, 2U_{ij}$, multiply Eq. (B4) by

$$g^{22}(1 - 2nc_2 + 3n^2c_2{}^2 - n_{,2}\, n^{,2}), \tag{B5}$$

and add $(2 \leftrightarrow 3)$, the corresponding term with subscripts 2 and 3 interchanged:

$$\begin{aligned} 4H' &= 2(c_2 + c_3) - 2n(c_2{}^2 + c_3{}^2) - 2\nabla^2 n \\ &\quad + 2n^2(c_2{}^3 + c_3{}^3) + 4C^{ij}\, nn_{,i;j} + (c_2 - c_3)n_{,2}\, n^{,2} \\ &\quad + (c_3 - c_2)n_{,3}\, n^{,3} + 2g^{22}\, n_{,2}n(c_2 - c_3)_{,2} \\ &\quad + 2g^{33}\, n_{,3}n(c_3 - c_2)_{,3} \\ &\quad + 2g^{33}\, n_{,3}n(c_3 - c_2)2 \begin{Bmatrix} 2 \\ 32 \end{Bmatrix} \\ &\quad + 2g^{22}\, n_{,2}n(c_2 - c_3)2 \begin{Bmatrix} 3 \\ 23 \end{Bmatrix}. \end{aligned} \tag{B6}$$

Now the second fundamental form C_{ij}, first introduced in Eq. (18), is a tensor. Its traceless part $G_{ij} = C_{ij} - Hg_{ij}$ has components $G_2{}^2 = -G_3{}^3 = \frac{1}{2}(c_2 - c_3)$, $G_2{}^3 = G_3{}^2 = 0$ in adapted coordinates. The last four terms in Eq. (B6) can be recognized as $4n^{,i}\, nG^j{}_{i;j}$, as one sees by expanding this expression out.

This completes the derivation of Eqs. (32)–(34).

[1]J. E. Marsden and T. J. R. Hughes, "Topics on the Mathematical Foundations of Elasticity," in *Nonlinear Analysis and Mechanics, Herriot-Watt Symposium*, edited by R. J. Knops (Pitman, London, 1978), Vol. II.
[2]J. E. Marsden and T. J. R. Hughes, *Mathematical Foundations of Elasticity* (Prentice-Hall, Englewood Cliffs, NJ, 1983).
[3]F. Brochard and J. F. Lennon, J. Phys. (Paris) **36**, 1035 (1975).
[4]R. M. Servuss, W. Harbich, and W. Helfrich, Biochem. Biophys. Acta **436**, 900 (1976).
[5]E. A. Evans, Biophys. J. **43**, 27 (1983).
[6]M. A. Peterson, to be published in J. Appl. Phys.
[7]M. A. Peterson, submitted to *Proceedings of the Tenth International Liquid Crystal Conference*, edited by G. Gray.
[8]C. A. Truesdell and R. Toupin, "The Classical Field Theories," in *Handbuch der Physik* (Springer-Verlag, Berlin, 1960), Vol. III, p. 1.
[9]V. V. Novozhilov, *Thin Shell Theory* (Russian translation) (Noordhoff, Groningen, The Netherlands, 1964).
[10]W. Helfrich, Z. Naturforsch. C **28**, 693 (1973).
[11]B. Schultz, *Geometrical Methods of Mathematical Physics* (Cambridge U. P., Cambridge, 1980).
[12]L. D. Landau and E. M. Lifschitz, *Theory of Elasticity* (Pergamon, New York, 1959), p. 3.
[13]T. J. Willmore, *An Introduction to Differential Geometry* (Clarendon, Oxford, 1966), p. 128.
[14]E. A. Evans and R. Skalak, *Mechanics and Thermodynamics of Biomembranes* (CRC, Boca Raton, FL, 1980), p. 142.

Conservation laws in the Dirac theory

Roger Boudet
Université de Provence, Place V. Hugo, 13331 Marseille Cedex 3, France

(Received 18 June 1984; accepted for publication 2 November 1984)

It is well known that the Dirac equations determine three conservation laws, for current, for energy momentum, and for angular momentum. Dividing the Dirac equations in two parts, one, D_I, which does not contain the density ρ, the other, D_{II}, which contains ρ, it is shown that D_I together with the three conservation equations, determine D_{II}. From this result, a model of formulation of the Dirac theory is proposed, following a scheme similar to that of classical mechanics, in which principles regarding the motion of one particle are associated with conservation theorems concerning statistical ensembles of particles.

I. INTRODUCTION

In the first place, our paper aims to investigate the following question. It is well known[1,2] that the Dirac equation determines three conservation laws: (1) for probability or particle number, (2) for energy momentum, and (3) for angular momentum. The question is as follows: To what degree do the conservation laws determine the Dirac equation?

In the second place, starting from the answer we give to this question, which is the proof of a theorem, our paper will propose a new model of formulation of the Dirac theory, following a scheme similar to that of classical mechanics, in which principles regarding the motion of one particle are associated with conservation theorems concerning statistical ensembles of particles.

A. The main theorem

(a) We know[3–6] that the Dirac wave function ψ may be decomposed into eight real scalar parameters. On one side, the unit four-vector velocity v, the unit four-vector spin s, orthogonal to v ($v^\mu v_\mu = 1, s^\mu s_\mu = -1, v^\mu s_\mu = 0$), the phase χ, and the so-called Yvon–Takabayasi angle β (see Sec. II). We will call these seven parameters the "proper parameters" of the Dirac particle. On the other, the particle number density ρ.

(b) As we shall see in Sec. II, the Dirac equations may be divided into two intrinsic four-vector equations, one, (D_I), which involves the four-vector potential A and only the "proper parameters," and so does not contain ρ, the other, (D_{II}), which contains ρ.

(c) We know that the three conservation relations

$$\partial_\mu(\rho c v^\mu) = 0, \quad \partial_\mu(\rho T^{\mu\alpha}) = \rho f^\alpha,$$
$$\rho(T^{\alpha\beta} - T^{\beta\alpha}) = -\partial_\mu(\rho(\hbar c/2)S^{\mu\alpha\beta}) \tag{1.1}$$

(c is the light velocity, h is Planck's constant, $\hbar = h/2\pi$), where ρT is the Tetrode energy-momentum tensor, ρf is the Lorentz force density, and $-\rho\hbar c S/2$ is the spin density are implied by the Dirac equations.

S depends on v and s, and, as we shall see in Sec. III, T depends on A and the seven proper parameters.

Our theorem is as follows: (D_{II}) is implied by (D_I) and the three conservation relations (1.1).

In other words, the part of the Dirac equations which involves the density ρ expresses nothing more than the conservation laws.

As exposed in Sec. VI, this theorem helps to justify some well known—but till now unexplained—features of the Dirac theory.

N.B. The number of equations of the system (D_I) [Eq. (2.13)] plus (1.1) is greater than those of the system (D_I) [(2.13)] plus (D_{II}) [Eq. (5.14)]. But some of the equations of (1.1) are not independent and the two systems are to be considered as equivalent.

B. A model of formulation of the Dirac theory

We will consider the seven proper parameters as relative to one particle, and (D_I) as some quantum dynamical law—to be taken as a first principle—bounding the proper parameters and the potential A.

But the set of equations (D_I), i.e., four scalar equations with seven scalar unknowns, is an incomplete system which does not allow us to determine the motion of the particle. Let us now consider a fluid $\mathscr{P}$ which may be equally regarded: either, in agreement with the Copenhagen probabilistic interpretation, as a continuous probability fluid formed by different events of the same real particle, or (in a quite different meaning), in accordance with the statistical interpretation of the quantum mechanics, in which the wave-particle duality is excluded[7–9] as a discrete "fictive" fluid. Such a fluid may be defined as formed by different real particles, considered one by one in successive but similar experiments, and joined together by thought in a sole fictive experiment.

The conservation laws, applied to the fluid $\mathscr{P}$, will be interpreted in the following ways.

(1) The conservation of the current of the fluid is to be regarded simply as the definition of the density function associated to the ensemble $\mathscr{P}$.

(2) We will consider—as a second principle—each particle of the ensemble as being subjected to the same Lorentz force as for a classical particle, but, in the conservation law of the energy momentum, we will replace the tensor $\rho K^{\alpha\beta} = \rho m c^2 v^\alpha v^\beta$, which would have been considered if the particles of the ensemble $\mathscr{P}$ had been classical, by the tensor $\rho T^{\alpha\beta}$. The definition of ρT as the energy tensor of $\mathscr{P}$ is to be taken as a principle. However, as we shall see in Sec. III the replacement of the tensor K by the tensor T is partly justified by a relation between Tr K and Tr T which exists in (D_I) (and which concerns each particle of $\mathscr{P}$).

 0022-2488/85/040718-07$02.50

(3) Considering $-\hbar cS/2$ as corresponding to the proper angular momentum of the particle, we will apply, as a theorem, the third conservation law, in the same way as in classical mechanics.

So, by application of the main theorem, it is then possible to obtain a complete system of eight scalar equations (D_I) and (D_{II}), with eight scalar unknowns, i.e., the proper parameters and ρ.

In other words, the problem of one particle cannot be solved by the dynamical law (D_I), but the union of (D_I) and of the conservations laws (1.1) allows us to solve the problem of a statistical ensemble of particles.

If one adopts the statistical interpretation point of view, the part played by the function ψ is then reduced to a simple auxiliary mathematical function which allows us to solve a correctly defined mathematical problem, and the notion of wave disappears.

However, in both interpretations, our model remains indeterminist. The model is exposed in Secs. II–IV; the theorem is proved in Sec. V.

II. THE QUANTUM DYNAMICAL LAW AND THE PROPER PARAMETERS

A. Notation

We use the same notations as in the geometric algebra developed by D. Hestenes.[6,10] Let E be the space-time, considered as a real vector space of signature $+ - - -$, and $\overset{p}{\wedge}$ the vector space of the antisymmetrical tensors of rank p. We will note $A \wedge B$ and $A \cdot B$ the outer and inner products of $A \in \overset{p}{\wedge} E$ and $B \in \overset{p}{\wedge} E$: for example, if $a,b,c, \in E$, the symbol $(a \wedge b)\cdot c$ will indicate the vector of components $(a^ib^\mu - b^ia^\mu)c_\mu$, $(i = 0,1,2,3)$.

As in Ref. 10, we define the associative Clifford product AB by using the rules

$$aA = a\cdot A + a \wedge A, \quad Aa = A \cdot a + A \wedge a,$$
$$\text{if} \quad a \in E, \quad A \in \overset{p}{\wedge} E. \tag{2.1}$$

Let $\{\gamma_0, \gamma_1, \gamma_2, \gamma_3\}$ be an orthonormal frame of E:

$$\gamma_i \cdot \gamma_j = 0 \quad \text{if} \quad i \neq j, \quad \gamma_0 \cdot \gamma_0 = 1,$$
$$\gamma_k \cdot \gamma_k = -1 \quad (k = 1,2,3). \tag{2.2}$$

N.B. There exists a representation of the γ_j by the Dirac matrices[6] and that is why the symbols used are the same. But here, the γ_j have only to be considered as vectors of E (they may be used as operators by means of the Clifford product).

We define the gradient operator $\nabla = \gamma^\mu \partial_\mu$, where $\gamma^i \in E$, $\gamma^i \cdot \gamma_j = \delta^i_j$.

Using the rules (2.1), it is easy to verify, in particular, that

$$\gamma_i \gamma_j = \gamma_i \cdot \gamma_j + \gamma_i \wedge \gamma_j = \gamma_i \wedge \gamma_j$$
$$= -\gamma_j \wedge \gamma_i = -\gamma_j \gamma_i, \quad \text{if} \quad i \neq j, \tag{2.3}$$

and then, denoting $i = \gamma_0 \wedge \gamma_1 \wedge \gamma_2 \wedge \gamma_3$, that $i = \gamma_0\gamma_1\gamma_2\gamma_3$, $i^2 = -1$, and $ia = -ai$ if $a \in E$.

Note that the dual of a simple p vector $B = b_1 \wedge \cdots \wedge b_p$ may be written iB.

B. An intrinsic form of the Dirac equation

Our first task is the separation of what does not involve the density ρ in the Dirac equations from the rest.

The simplest way seems to use Eq. (6) that we established in Ref. 11. Let us recall how we obtained this equation. We started from the Hestenes formulation of the Dirac theory[6] where the ψ function is written (in Clifford algebra)

$$\psi = (\rho e^{i\beta})^{1/2} R \tag{2.4}$$

and has the following properties.

(a) ρ is the density.

(b) R is a Lorentz rotation such that

$$v = R\gamma_0 \widetilde{R}, \quad n_k = R\gamma_k \widetilde{R} \quad (k = 1,2),$$
$$s = R\gamma_3 \widetilde{R} \quad (R\widetilde{R} = \widetilde{R}R = 1), \tag{2.5}$$

where $\mathscr{R} = \{v, n_1, n_2, s\}$ is an orthonormal frame (the "proper frame" of the particle) such that cv is the space-time velocity of the particle, s is the unit "spin vector," and n_1, n_2 are two orthonormal vectors of the "spin plane." (The phase χ of ψ is the parameter needed to fix the direction of n_1 in the spin plane.) We will use the two bivectors

$$\sigma = n_1 \wedge n_2 = n_1 n_2, \quad \hat{\sigma} = -i\sigma = v \wedge s = vs. \tag{2.6}$$

(c) β is the so-called Takabayasi angle,[4] first introduced by Yvon[3] and then, independently and in a quite different way, by Hestenes.[6,12]

Consider the Dirac–Hestenes equation [Ref. 6, Eq. (5-1)]

$$\hbar c\gamma^\mu \partial_\mu \psi + (mc^2 \psi\gamma_0 + eA\psi)\gamma_1\gamma_2 = 0, \tag{2.7}$$

where m is the mass of the particle, e is the charge, and $A \in E$ is the vector potential.

We define as in Refs. 11, 13, or 14,

$$\Omega_j = 2(\partial_j R)\widetilde{R} \quad (j = 0,1,2,3). \tag{2.8}$$

Here, $\Omega_j \in \overset{2}{\wedge} E$ is the bivector which expresses the infinitesimal rotation of the frame $\mathscr{R}$ due to an infinitesimal displacement in the γ_j direction:

$$\partial_j v = \Omega_j \cdot v, \quad \partial_j n_k = \Omega_j \cdot n_k \quad (k = 1,2), \quad \partial_j s = \Omega_j \cdot s. \tag{2.9}$$

Developing $\partial_j \psi$ in (2.4), noting that we can write

$$2\partial_j R = 2(\partial_j R)\widetilde{R}R = \Omega_j R, \quad \gamma^j \exp(i\beta/2) = \exp(-i\beta/2)\gamma^j, \tag{2.10}$$

multiplying (2.7) by $1/\sqrt{\rho}$, then, on the left, by $\exp(i\beta/2)$, and, on the right, by $\widetilde{R}$, and since $R\gamma_1\gamma_2\widetilde{R} = n_1 n_2 = \sigma$, we obtain the equation[11]

$$(\hbar/2)c(\nabla(\ln \rho) + \gamma^\mu \Omega_\mu + (\nabla\beta)i)$$
$$+ (mc^2(\cos\beta v + \sin\beta iv) + eA)\sigma = 0, \tag{2.11}$$

which is strictly equivalent to the Dirac equation.

C. The quantum dynamical law

Taking the pseudovector part of (2.11), then multiplying on the left by i, denoting

$$u = i(\gamma^\mu \wedge \Omega_\mu) \in E, \tag{2.12}$$

and since $i(v \wedge \sigma) = s$, $i(A \wedge \sigma) = A \cdot (v \wedge s)$, we obtain the following vector equation [$(\mathrm{D_I})$] which does not contain ρ:

$$mc^2 \cos\beta s + (\hbar c/2)(u + \nabla\beta) + eA \cdot (v \wedge s) = 0. \quad (2.13)$$

The unknowns of this equation are the six kinematical real scalar parameters which define the situation of the frame $\mathscr{R}$ at each point $x \in E$, and the angle β. We will call these seven scalar parameters the proper parameters of the particle, and, since (2.13) does not contain the density ρ, we will suppose that these parameters are relative to *one* particle.

N.B. The vector equation $(\mathrm{D_I})$ [(2.13)] is equivalent to the set of equations (3′), (4′), and (12′) of Takabayasi.[4]

See also the article by Hestenes.[14] [The equation $(\mathrm{D_I})$ does not appear explicitly in this article, but a large number of intrinsic equations, some of them directly related to $(\mathrm{D_I})$, and their physical interpretations are mentioned in this paper.]

III. DEFINITION OF TENSORS AND ENERGIES ASSOCIATED TO ONE PARTICLE

Let us now define three tensors associated to these parameters.

(1) The situation tensor of the spin plane is

$$C\colon\ n \in E \to C(n) = n \cdot (v \wedge \sigma) \in \wedge^2 E. \quad (3.1)$$

Its values are

$$C(v) = \sigma = n_1 \wedge n_2,\quad C(n_1) = v \wedge n_2,$$
$$C(n_2) = n_1 \wedge v,\quad C(s) = 0. \quad (3.2)$$

(2) The rotation tensor of the spin plane is

$$L\colon\ n \in E \to L(n) = (\Omega_\mu \cdot (i(n \wedge s)))\gamma^\mu \in E. \quad (3.3)$$

Its values are

$$L(v) = \omega = (\Omega_\mu \cdot \sigma)\gamma^\mu = ((\Omega_\mu \cdot n_1) \cdot n_2)\gamma^\mu$$
$$= (\partial_\mu n_1 \cdot n_2)\gamma^\mu = -(\partial_\mu n_2 \cdot n_1)\gamma^\mu, \quad (3.4)$$

$$L(n_1) = (\Omega_\mu \cdot (v \wedge n_2))\gamma^\mu,$$
$$L(n_2) = (\Omega_\mu \cdot (n_1 \wedge v))\gamma^\mu,\quad L(s) = 0. \quad (3.5)$$

The vector ω represents the rotation of the spin plane on itself.[11] Here, $L(n_1)$ and $L(n_2)$ correspond to the rotation of the direction of the spin plane, and are related to the bend of the curve which is tangent, at each point $x \in E$, to the vector $v(x)$, according to

$$L(n_1) \cdot v = \Gamma \cdot n_2,\quad L(n_2) \cdot v = -\Gamma \cdot n_1,$$
$$\text{where } \Gamma = (v \cdot \nabla)v. \quad (3.6)$$

We can write (see Appendix A)

$$\mathrm{Tr}\, L = u \cdot s = (\omega + \nabla \cdot \sigma) \cdot v. \quad (3.7)$$

(3) We also have the following tensor (whose meaning is more obscure than that of the others):

$$M\colon\ n \in E \to M(n) = (s \cdot n)\nabla\beta \in E. \quad (3.8)$$

A. Gauge invariance of the equation $(\mathrm{D_I})$

One can prove (see Appendix A)

$$u = \hat{\sigma} \cdot (\omega + \nabla \cdot \sigma) + \sigma \cdot (\overline{\omega} + \nabla \cdot \hat{\sigma}), \quad (3.9)$$

where $\overline{\omega} = (\Omega_\mu \cdot \hat{\sigma})\gamma^\mu$ is the proper rotation vector of the (v, s) plane.

So $(\mathrm{D_I})$ takes the form

$$mc^2 \cos\beta s + (v \wedge s) \cdot ((\hbar c/2)\omega - eA)$$
$$+ (\hbar c/2)(\hat{\sigma} \cdot (\nabla \cdot \sigma) + \sigma \cdot (\overline{\omega} + \nabla \cdot \hat{\sigma}) + \nabla\beta) = 0, \quad (3.10)$$

where we note the presence of the vector

$$p = (\hbar c/2)\omega - eA, \quad (3.11)$$

which corresponds to the energy-momentum vector of the particle.

The gauge invariance of p, and so of (2.13), after replacing the vector potential A by $A' + \nabla\phi$, is verified by adding a scalar φ to the phase χ, because ω is then replaced by $\omega' = \omega + \nabla\varphi$, and by taking $\varphi = 2e\phi/(\hbar c)$ (see Refs. 6 and 13).

B. Energies associated to the particle

We consider three energy tensors (associated to *one* particle): the kinetic energy tensor $K\colon n \in E \to K(n) = mc^2(v \cdot n)v \in E$; the quantum energy tensor $Q\colon Q = \hbar c(L + M)/2$; and the potential energy tensor $P\colon n \in E \to P(n) = -e(v \cdot n)A \in E$.

We define $E = mc^2 = \mathrm{Tr}\, K$, $\epsilon_Q = \mathrm{Tr}\, Q$, $\epsilon_P = \mathrm{Tr}\, P$, as the proper kinetic, quantum, and potential energies of the particle.

The part of ϵ_Q due to the rotation of the spin plane is

$$\epsilon_R = (\hbar c/2)(\mathrm{Tr}\, L). \quad (3.12)$$

As follows from (3.7), ϵ_R can be decomposed into a part due to the rotation of the spin plane on itself (on a displacement in the v direction), $\hbar\omega \cdot cv/2$, and a part due to the rotation in space-time of the direction of the spin plane, $\hbar(\nabla \cdot \sigma) \cdot cv/2$.

The scalar $\hbar\omega \cdot c\gamma_0/2 = p \cdot \gamma_0 + eV$, where $V = A \cdot \gamma_0$, is considered as the energy of the particle relative to the Galilean frame $\{\gamma_0, \gamma_1, \gamma_2, \gamma_3\}$.

Note that for a plane wave, the spin plane direction is constant, $\beta = 0$ everywhere, and so ϵ_Q is reduced to $\hbar\omega \cdot cv/2$ and expresses an oscillatory phenomenon due to the rotation of the spin plane on itself. (For the case of the central potential see Refs. 15 and 16.)

C. The relation between the tensors K and T

We can deduce from $(\mathrm{D_I})$ a relation between E and $\epsilon_Q + \epsilon_P$. Considering the inner product of (2.13) by the spin vector s, we obtain

$$mc^2 \cos\beta = (\hbar c/2)(s \cdot u + s \cdot \nabla\beta) - eA \cdot v. \quad (3.13)$$

Writing

$$T = Q + P, \quad (3.14)$$

it follows from (3.7) and (3.8) that

$$(\mathrm{Tr}\, K)\cos\beta = \mathrm{Tr}\, T, \quad (3.15)$$

or

$$E\cos\beta = \epsilon_Q + \epsilon_P. \quad (3.16)$$

IV. CONSTRUCTION OF THE CONSERVATION LAWS IN RELATIVISTIC QUANTUM MECHANICS

Let us consider a fluid $\mathscr{P}$ that we will equally interpret as formed either by different events of the same particle, or, as explained in Sec. I, by different similar particles (considered one by one in successive experiments and joined by thought in a sole fictive experiment).

(1) We consider here relativistic particles, and it is not possible to take the mean of different space-time velocities cv. So, instead of considering, at each point $x \in E$, different velocities, we suppose that all the particles whose trajectories pass through x have, at point x, the same velocity $v(x)$. In this way, we do not define the most general (fictive) fluid that it is possible to consider in a given field, but only a partial fluid. In actual fact, there is an infinity of different fluids, defined in this way, to consider. So, it is not surprising that the equations we will obtain admit solutions depending on constants which may take an infinity of different values.

Let $\rho(x)$ be the density of the fluid at $x \in E$. By definition of ρ we can write

$$\nabla \cdot (\rho v) = 0. \tag{4.1}$$

(2) If the particles were classical, the conservation law of the energy momentum would be

$$\iiint_{\Sigma} \rho K(n) d\sigma = \iiiint_{V} \rho f \, d\tau, \tag{4.2}$$

where V is a hypervolume of E bounded by a hypersurface Σ, n is a unit space-time vector normal to Σ and oriented towards the exterior of V, $f = (\nabla \wedge A) \cdot v$ is the Lorentz force acting on each particle, and $\rho K(n) = mcv(\rho cv \cdot n)$ is the flux of the classical momenta mcv of the particles whose trajectories cross a unit hyperarea normal to n.

Now, we replace in (4.2) the tensor K by the tensor T:

$$\iiint_{\Sigma} \rho T(n) d\sigma = \iiiint_{V} \rho f \, d\tau. \tag{4.3}$$

This replacement is (partly) justified by Eq. (3.15) because, in the case where $\beta = 0$, Tr K and Tr T are equal for each particle.

As explained in Appendix B, ρT is the so-called Tetrode tensor.[1] We deduce

$$\iiiint_{V} \partial_\mu (\rho T^\mu) d\tau = \iiiint_{V} \rho f \, d\tau, \quad T^\mu = T(\gamma^\mu). \tag{4.4}$$

Considering this equation for all arbitrary small volume surrounding each point $x \in E$, we write (as a principle),

$$\partial_\mu (\rho T^\mu) = \rho f, \quad f = (\nabla \wedge A) \cdot v. \tag{4.5}$$

(3) The third conservation law may be constructed as in classical mechanics. Interpreting $\rho\, U(n) = -\rho \hbar c C(n)/2$, where C is the situation tensor of the spin plane, defined in Sec. III as the flux of the proper angular momentum, we can write

$$\iiint_{\Sigma} \rho (U(n) + x \wedge T(n)) d\sigma = \iiiint_{V} \rho x \wedge f \, d\tau, \tag{4.6}$$

hence,

$$-\nabla \cdot (\rho (\hbar c/2) v \wedge \sigma) + \rho \gamma_\mu \wedge T^\mu + x \wedge \partial_\mu (\rho T^\mu) = \rho x \wedge f, \tag{4.7}$$

and from (4.5), we deduce (as a theorem)

$$\rho T^\mu \wedge \gamma_\mu = -\nabla \cdot (\rho (\hbar c/2) v \wedge \sigma). \tag{4.8}$$

N.B. Let g be some scalar or vector mechanical quantity associated with a relativistic particle, and a conservation law similar to (4.2), as follows

$$\iiint_{\Sigma} g\rho (v \cdot n) d\sigma = \iiiint_{V} \rho q \, d\tau. \tag{4.9}$$

One can deduce from (4.1)

$$\rho q = \partial_\mu (\rho g v^\mu) = \partial_\mu (\rho v^\mu) g + \rho (\partial_\mu g) v^\mu = \rho (\partial_\mu g) v^\mu \tag{4.10}$$

or

$$q = (\partial_\mu g) v^\mu. \tag{4.11}$$

Here, ρ is eliminated from the conservation law [but not in (4.5)], because the flux $\rho g(v \cdot n)$ is null for n orthogonal to v [but not $\rho T(n)$].

This difference between the classical and quantum fluxes (to be or not to be null for n orthogonal to v) could explain the reason why it is not possible to deduce rigorously or as some approximation, for example, the Weyssenhoff motion from the Dirac theory (see Ref. 14, No. 7).

Now, the last task for ending our model of formulation of the Dirac equations is the proof of the main theorem.

V. PROOF OF THE MAIN THEOREM

The proof we give is long (we could not find a simpler one, but perhaps it exists), so we only indicate the main steps of the proof, the rest is in the appendices.

Considering the outer product of (2.13) by s, we obtain the bivector equation

$$((\hbar c/2)(u + \nabla\beta) - e(s \cdot A) v) \wedge s = 0. \tag{5.1}$$

We can write (see Appendix C)

$$\begin{aligned} &L^\mu \wedge \gamma_\mu = i(s^\mu \Omega_\mu) + i((\gamma^\mu \cdot \Omega_\mu) \wedge s), \\ &L^j = L(\gamma^j), \quad s^j = s \cdot \gamma^j. \end{aligned} \tag{5.2}$$

$$T^\mu \wedge \gamma_\mu = (\hbar/2) c (L^\mu \wedge \gamma_\mu + (\nabla\beta) \wedge s) + ev \wedge A, \tag{5.3}$$

$$-\nabla \cdot S = i(s^\mu \Omega_\mu) - u \wedge s, \quad \text{where} \quad S = v \wedge \sigma. \tag{5.4}$$

$$\nabla \cdot (\rho S) = \nabla\rho \cdot S + \rho \nabla \cdot S = i(\nabla\rho \wedge s) + \rho \nabla \cdot S. \tag{5.5}$$

It follows from (5.1), after elimination of $i(s^\mu \Omega_\mu)$ and $u \wedge s$, that

$$\begin{aligned} -(\hbar/2) c \nabla \cdot (\rho S) = \rho T^\mu \wedge \gamma_\mu + \rho e (A + (A \cdot s) s) \wedge v \\ - (\hbar/2) c i ((\nabla\rho + \rho \gamma^\mu \cdot \Omega_\mu) \wedge s). \end{aligned} \tag{5.6}$$

We deduce from (4.8), and because $i((A + (A \cdot s) s) \wedge v) = (A \cdot \sigma) \wedge s$,

$$(eA \cdot \sigma + (\hbar/2) c (\nabla(\ln \rho) + \gamma^\mu \cdot \Omega_\mu)) \wedge s = 0. \tag{5.7}$$

Moreover, we can write the following relation, that we have already established in Ref. 16:

$$\begin{aligned} \rho T^\mu \wedge \gamma_\mu = \rho ((\hbar/2) c ((\nabla\beta) \wedge s - i((\nabla \ln \rho) \wedge s) \\ + i(s^\mu \Omega_\mu)) - e(A \cdot s) v \wedge s). \end{aligned} \tag{5.8}$$

In other respects, we deduce from (3.15)

$$-mc^2 \sin\beta(\nabla\beta) = \nabla(\mathrm{Tr}\, T). \tag{5.9}$$

We calculate, respectively (see Appendix D),

(a) $\rho\nabla(\mathrm{Tr}\, M) - \partial_\mu(\rho M^\mu)$,

(b) $\rho\nabla(\mathrm{Tr}\, P) - \partial_\mu(\rho P^\mu)$,

(c) $\rho\nabla(\mathrm{Tr}\, L) - \partial_\mu(\rho L^\mu)$,

and we replace the term $\nabla\cdot(\rho v \wedge \sigma)$ which appears in the third calculation, by its expression deduced from (4.8) and (5.8). Then we deduce from (5.9), (4.1), and (4.5),

$$\begin{aligned}\rho(\hbar/2)c\,[\,(((i(\partial_\nu\Omega_\mu - \partial_\mu\Omega_\nu))\cdot\gamma^\mu + \Omega_\mu\cdot((i\Omega_\nu)\cdot\gamma^\mu)\\ - (i\Omega_\nu)\cdot(\Omega_\mu\cdot\gamma^\mu))\cdot s)\gamma^\nu\,]\\ + (\rho mc^2\sin\beta - (\hbar/2)c\nabla\cdot(\rho s))(\nabla\beta) = 0.\end{aligned} \tag{5.10}$$

We can deduce from the relations[13,14]

$$\partial_j\Omega_k - \partial_k\Omega_j + \tfrac{1}{2}(\Omega_k\Omega_j - \Omega_j\Omega_k) = 0, \tag{5.11}$$

that the bracket of Eq. (5.10) is null. So, if β is not reduced to a constant, we obtain [$(\mathrm{D_{II}})$]

$$\rho\, mc^2\sin\beta - (\hbar/2)c\nabla\cdot(\rho s) = 0, \tag{5.12}$$

which may be written

$$(mc^2\sin\beta s + (\hbar/c)c(\nabla(\ln\rho) + \gamma^\mu\cdot\Omega_\mu))\cdot s = 0. \tag{5.13}$$

Equations (5.13) and (5.7) may be united in a sole vector equation

$$mc^2\sin\beta s + eA\cdot\sigma + (\hbar c/2)(\nabla(\ln\rho) + \gamma^\mu\cdot\Omega_\mu) = 0, \tag{5.14}$$

which is just the vector part of Eq. (2.11). So the theorem is proved.

VI. CONCLUSION

The previous propositions, especially the relation (5.9), explain the presence of $\sin\beta$ in the Dirac equations (but not that of $\cos\beta$). In particular, they help to justify the well-known[3] relation (5.12) which expresses the nonconservativity of the spin vector density.

Our theorem mathematically confirms the correctness of identifying the Tetrode tensor with the energy-momentum tensor, because it does not seem possible to obtain a similar result with the other tensors which have been proposed.[14]

Furthermore, in our scheme, the specificity of the Dirac equations, the quantum particularities of the Dirac theory, appear as if practically concentrated inside the equation $(\mathrm{D_I})$ [(2.13)] (and the definition of the tensor T), because the rest of the equations may be deduced [taking into account the property of $T(n)$ to be non-null for n orthogonal to v] from properties similar to those of the classical mechanics. In particular, we can notice that the exterior field is used in two ways. First, in $(\mathrm{D_I})$, in a form which implies the electromagnetic gauge (we have seen in Sec. III what is the incidence of the gauge on the proper parameter χ), and second, in the right side of the energy-momentum equation, without the intervention of the gauge, as in the relativistic classical mechanics.

However, our paper does not offer answers to many questions which may be asked; for example, we have the following.

—Why is the rotation tensor L of the spin plane changed into a flux tensor when it is multiplied by $\hbar c/2$?

—Why is the tensor K replaced, in the conservation laws, by T and not by $T/\cos\beta$, in complete agreement with Eq. (3.15)?

—What part is played by the "mysterious" angle β? etc...

Some answers could be imagined by using more detailed theories.

For example, the interpretation of the spin as a correction term, which accounts for the orbital motion about some mean position of the particle[9,17] could give an explanation to the first question, by interpreting $\hbar cL(n)/2$ as the flux of energy due to this orbital motion.

It seems, from Ref. 18, that β could be relevant to some general gauge theory. But these questions lie outside the scope of our paper.

We said that our model satisifes both the probabilistic and statistical interpretations. However, it seems to us that it brings arguments in favor of the second interpretation.

For example, in classical mechanics, especially in the kinetic theory of gases, the conservation laws are relative to *discrete* ensembles of material particles, and cannot be constructed without supposing that each particle is a localized system. If the probabilistic interpretation were correct, it would be necessary to suppose that the basic presence in the Dirac equations of very similar conservation laws, relative to a *continuous* probability fluid, would be a fortuitous coincidence.

Furthermore—this argument seems to us very convincing—it is easy to imagine that a force (the Lorentz force) may act on a real particle, but not on the "event" of a particle. So the fluid $\mathscr{P}$ of our model might be interpreted rather as a statistical ensemble than as a pure probability fluid.

But on the other hand, the harmonious union of $(\mathrm{D_I})$ and $(\mathrm{D_{II}})$ in Eq. (2.11) is an argument for considering the Dirac equations as a whole, and for giving a physical meaning to the ψ function, as in the probabilistic interpretation.

APPENDIX A: PROOF OF THE RELATIONS (2.13), (3.10), (3.13), (5.1)

We can write

$$i = vn_1n_2s = v\sigma s, \quad \sigma^2 = n_1n_2n_1n_2 = -n_1^2n_2^2 = -1, \tag{A1}$$

hence,

$$i(v\wedge\sigma) = iv\sigma = v\sigma s v\sigma = -v^2\sigma^2 s = s. \tag{A2}$$

It follows from (2.1)

$$\begin{aligned}i(A\wedge\sigma) &= i(A\sigma + \sigma A)/2 = (-Ai\sigma + i\sigma A)/2\\ &= -A\cdot(i\sigma) = A\cdot(v\wedge s).\end{aligned} \tag{A3}$$

Note that $i\nabla\beta i = -i^2\nabla\beta = \nabla\beta$.

We write

$$s\cdot s = -1, \qquad A\cdot(v\wedge s) = (A\cdot v)s - (A\cdot s)v,$$

$$\begin{cases}(A\cdot(v\wedge s))\cdot s = -A\cdot v,\\ (A\cdot(v\wedge s))\wedge s = -(A\cdot s)v\wedge s.\end{cases} \tag{A4}$$

We obtain

$$\begin{aligned}\operatorname{Tr} L = L^{\mu} \cdot \gamma_{\mu} &= \Omega_{\nu} \cdot (i(\gamma^{\mu} \wedge s))(\gamma^{\nu} \cdot \gamma_{\mu}) \\ &= \Omega_{\mu} \cdot (i(\gamma^{\mu} \wedge s)) = (i\Omega_{\mu}) \cdot (\gamma^{\mu} \wedge s) \\ &= ((i\Omega_{\mu}) \cdot \gamma^{\mu}) \cdot s = (i(\gamma^{\mu} \wedge \Omega_{\mu})) \cdot s = u \cdot s, \end{aligned} \tag{A5}$$

because if $U, V \in \overset{2}{\wedge} E$,

$$U \cdot (iV) = [UiV]_{\text{sc.}} = [iUV]_{\text{sc.}} = (iU) \cdot V, \tag{A6}$$

and

$$\begin{aligned}(i\Omega_{\mu}) \cdot \gamma^{\mu} &= (i\Omega_{\mu}\gamma^{\mu} - \gamma^{\mu} i\Omega_{\mu})/2 = i(\Omega_{\mu}\gamma^{\mu} + \gamma^{\mu}\Omega_{\mu})/2 \\ &= i(\gamma^{\mu} \wedge \Omega_{\mu}). \end{aligned} \tag{A7}$$

In other respects,

$$\begin{aligned}\operatorname{Tr} L &= L(v) \cdot v + L(-n_1) \cdot n_1 + L(-n_2) \cdot n_2 + L(-s) \cdot s \\ &= \omega \cdot v + ((\Omega_{\mu} n_1^{\mu}) \cdot n_2) \cdot v - ((\Omega_{\mu} n_2^{\mu}) \cdot n_1) \cdot v \\ &= \omega \cdot v + (\nabla \cdot (n_1 \wedge n_2)) \cdot v = (\omega + \nabla \cdot \sigma) \cdot v. \end{aligned} \tag{A8}$$

We write

$$\begin{aligned}\gamma^{\mu}\Omega_{\mu} i &= \gamma^{\mu}\Omega_{\mu} v n_1 n_2 s = \gamma^{\mu}\Omega_{\mu} n_1 n_2 \hat{\sigma} \\ &= \gamma^{\mu}((\Omega_{\mu} \cdot n_1) + \Omega_{\mu} \wedge n_1) n_2 \hat{\sigma} \\ &= \gamma^{\mu}((\partial_{\mu} n_1) n_2 + (\Omega_{\mu} \wedge n_1) \cdot n_2 + \Omega_{\mu} \wedge n_1 \wedge n_2)\hat{\sigma} \\ &= \gamma^{\mu}((\partial_{\mu} n_1) \cdot n_2 + (\partial_{\mu} n_1) \wedge n_2 + n_1 \wedge (\Omega_{\mu} \cdot n_2))\hat{\sigma} \\ &\quad + \gamma^{\mu}(((\Omega_{\mu} \wedge n_1 \wedge n_2) \cdot v) \cdot s) \\ &= \gamma^{\mu}((\partial_{\mu} n_1) \cdot n_2 + (\partial_{\mu} n_1) \wedge n_2 + n_1 \wedge (\partial_{\mu} n_2))\hat{\sigma} \\ &\quad + \gamma^{\mu}(((\Omega_{\mu} \cdot v) \wedge \sigma) \cdot s) \\ &= (\omega + \nabla\sigma)\hat{\sigma} + \gamma^{\mu}((\partial_{\mu} v) \cdot s)\sigma = (\omega + \nabla\sigma)\hat{\sigma} + \overline{\omega}\sigma, \end{aligned} \tag{A9}$$

hence

$$\begin{aligned}u = i(\gamma^{\mu} \wedge \Omega_{\mu}) &= [i\gamma^{\mu}\Omega_{\mu}]_{\text{vec.}} = [-\gamma^{\mu}\Omega_{\mu} i]_{\text{vec.}} \\ &= -(\omega \cdot \hat{\sigma} + (\nabla \cdot \sigma) \cdot \hat{\sigma} + (\nabla \wedge \sigma) \cdot \hat{\sigma} + \overline{\omega} \cdot \sigma), \end{aligned} \tag{A10}$$

From $\nabla \wedge \sigma = i(\nabla \cdot (i\sigma)) = -i(\nabla \cdot \sigma)$, and if $a \in E$, $U \in \overset{2}{\wedge} E$, $(ia) \cdot U = -a \cdot (iU)$, we deduce

$$(\nabla \wedge \sigma) \cdot \hat{\sigma} = -(i(\nabla \cdot \hat{\sigma})) \cdot \hat{\sigma} = (\nabla \cdot \hat{\sigma}) \cdot (i\hat{\sigma}) = (\nabla \cdot \hat{\sigma}) \cdot \sigma, \tag{A11}$$

and hence (3.9)

APPENDIX B: THE TETRODE TENSOR

We showed in a previous paper[16] that $T = Q + P$ is equal to T_0/ρ, where T_0 is the so-called Tetrode tensor. Let us recall the proof of this result. We started from the Hestenes form of T_0 [Ref. 14, Eq. (2-3)] (with a change of sign due to the convention adopted on the orientation of the spin plane):

$$T_0\text{:}\ \ T_0(\gamma_j) = -\hbar c \gamma^{\nu}[\gamma_j(\partial_{\nu}\psi) i\gamma_3\widetilde{\psi}]_{\text{sc.}} - \rho e v_j A. \tag{B1}$$

Developing $\partial_{\nu}\psi$ we obtain

$$\begin{aligned}&[\gamma_j(\partial_{\nu}\psi) i\gamma_3\widetilde{\psi}]_{\text{sc.}} \\ &= (\rho/2)[\gamma_j(i\Omega_{\nu} - \partial_{\nu}\beta + i\partial_{\nu}(\ln\rho))s]_{\text{sc.}} \\ &= (\rho/2)(-\Omega_{\nu} \cdot (i(\gamma_j \wedge s)) - (\gamma_j \cdot s)\partial_{\nu}\beta), \end{aligned} \tag{B2}$$

and so [Ref. 16, Eq. (1)],

$$\begin{aligned}T(n) &= (\hbar c/2)((\Omega_{\nu} \cdot (i(n \wedge s)))\gamma^{\nu} + (n \cdot s)\nabla\beta) \\ &\quad - e(n \cdot v)A = Q(n) + P(n). \end{aligned} \tag{B3}$$

APPENDIX C: PROOF OF THE RELATION (5.7)

We will use the following relations which may be easily proved:

$$(B \wedge a) \cdot b = (a \cdot b)B - (B \cdot b) \wedge a, \quad (i(B)) \cdot a = i(a \wedge B),$$

$$B \cdot a = -i((i(B)) \wedge a), \quad a, b \in E, \quad B \in \overset{2}{\wedge} E, \quad \text{etc.}$$

Let $\overline{L}$ be the transposed tensor of L. Then we have

$$\begin{aligned}L(n) \cdot N &= \Omega_{\mu} \cdot (i(n \wedge s))(N \cdot \gamma^{\mu}) = (i(\Omega_{\mu} N^{\mu})) \cdot (n \wedge s) \\ &= -((i(\Omega_{\mu} N^{\mu})) \cdot s) \cdot n = \overline{L}(N) \cdot n, \end{aligned}$$

hence

$$\begin{aligned}\overline{L}(n) &= -(i(\Omega_{\mu} n^{\mu})) \cdot s, \\ L^{\mu} \wedge \gamma_{\mu} &= \gamma^{\mu} \wedge \overline{L}_{\mu} = -\gamma^{\mu} \wedge ((i\Omega_{\mu}) \cdot s) \\ &= (\gamma^{\mu} \cdot s)(i\Omega_{\mu}) - ((i\Omega_{\mu}) \wedge \gamma^{\mu}) \cdot s \\ &= i(s^{\mu}\Omega_{\mu}) + (i(\gamma^{\mu} \cdot \Omega_{\mu})) \cdot s \\ &= i(s^{\mu}\Omega_{\mu}) + i((\gamma^{\mu} \cdot \Omega_{\mu}) \wedge s). \end{aligned} \tag{C1}$$

In other respects,

$$\begin{aligned}-\nabla \cdot (v \wedge \sigma) &= \nabla \cdot (is) = \gamma^{\mu} \cdot \left(i\left(\frac{\partial s}{\partial x^{\mu}}\right)\right) = \gamma^{\mu} \cdot (i(\Omega_{\mu} \cdot s)) \\ &= \gamma^{\mu} \cdot ((i\Omega_{\mu}) \wedge s) = i(s^{\mu}\Omega_{\mu}) - ((i\Omega_{\mu}) \cdot \gamma^{\mu}) \wedge s \\ &= i(s^{\mu}\Omega_{\mu}) - (i(\gamma^{\mu} \wedge \Omega_{\mu})) \wedge s \\ &= i(s^{\mu}\Omega_{\mu}) - u \wedge s. \end{aligned} \tag{C2}$$

Hence,

$$\begin{aligned}-\nabla \cdot (v \wedge \sigma) &= L^{\mu} \wedge \gamma_{\mu} - i((\gamma^{\mu} \cdot \Omega_{\mu}) \wedge s) \\ &\quad - (i(\gamma^{\mu} \wedge \Omega_{\mu})) \wedge s. \end{aligned} \tag{C3}$$

Writing

$$m_0 = mc/\hbar, \quad e_0 = e/(\hbar c), \quad \hbar c = 1,$$

we deduce from (5.1) and (C3) after elimination of $u \wedge s$,

$$\begin{aligned}-\nabla \cdot (v \wedge \sigma) &= L^{\mu} \wedge \gamma^{\mu}_{\mu} + (\nabla\beta) \wedge s - e_0 2(s \cdot A)(v \wedge s) \\ &\quad - i((\gamma^{\mu} \cdot \Omega_{\mu}) \wedge s). \end{aligned} \tag{C4}$$

We write

$$\nabla \cdot (\rho v \wedge \sigma) = \nabla\rho \cdot (v \wedge \sigma) + \rho\nabla \cdot (v \wedge \sigma), \tag{C5}$$

so, because

$$T^{\mu} \wedge \gamma_{\mu} = (L^{\mu} \wedge \gamma_{\mu} + (\nabla\beta) \wedge s)/2 + P^{\mu} \wedge \gamma_{\mu}$$

and

$$P^{\mu} \wedge \gamma_{\mu} = -e_0 A \wedge v,$$

thus

$$-\tfrac{1}{2}\nabla\cdot(\rho v\wedge\sigma)=\rho T^{\mu}\wedge\gamma_{\mu}+\rho e_0(A+(A\cdot s)s)\wedge v$$
$$-(i/2)((\nabla\rho+\rho\gamma^{\mu}\cdot\Omega_{\mu})\wedge s)=0. \quad \text{(C6)}$$

Now,

$$i((A+(A\cdot s)s)\wedge v)$$
$$=-i(((A\cdot n_1)n_1+(A\cdot n_2)n_2)v)$$
$$=((A\cdot n_1)n_2-(A\cdot n_2)n_1)\wedge s=(A\cdot\sigma)\wedge s.$$

We deduce from (C6) and (4.8)

$$(e_0A\cdot\sigma+\tfrac{1}{2}(\nabla\ln\rho+\gamma^{\mu}\cdot\Omega_{\mu}))\wedge s=0, \quad \text{(C7)}$$

i.e., the relation (5.7).

From (C1), (C4), and (C5)

$$\tfrac{1}{2}\nabla\cdot(\rho i(s))$$
$$=\rho T^{\mu}\wedge\gamma_{\mu}=\rho\,(\tfrac{1}{2}((\nabla\beta)\wedge s-i((\nabla\log\rho)\wedge s)$$
$$+i(s^{\mu}\Omega_{\mu}))-e_0(A\cdot s)v\wedge s). \quad \text{(C8)}$$

APPENDIX D: PROOF OF THE RELATION (5.13)

From (3.13),

$$m_0\cos\beta=\mathrm{Tr}\,(\tfrac{1}{2}(L+M)+P),$$

one deduces

$$\nabla(\mathrm{Tr}(L+M)/2+P)+m_0\sin\beta(\nabla\beta)=0. \quad \text{(D1)}$$

We can write

(a) $\rho\nabla(\mathrm{Tr}\,M)=\rho(\partial_{\mu}s\cdot\nabla\beta)\gamma^{\mu}+\rho(s\cdot\nabla)(\nabla\beta),$

$$\partial_{\mu}(\rho M^{\mu})=(\nabla\cdot(\rho s))(\nabla\beta)+\rho\,(s\cdot\nabla)(\nabla\beta),$$
$$\rho\nabla(\mathrm{Tr}\,M)=\partial_{\mu}(\rho M^{\mu})+\rho((\Omega_{\nu}\cdot s)\cdot\nabla\beta)\gamma^{\nu}$$
$$-(\nabla\cdot(\rho s))(\nabla\beta). \quad \text{(D2)}$$

(b) $(\nabla\wedge A)\cdot v=(v\cdot\partial_{\mu}A)\gamma^{\mu}-(v\cdot\nabla)A.$

Then,

$$\rho\nabla(\mathrm{Tr}\,P)=-\rho e_0(\partial_{\mu}v\cdot A+v\cdot\partial_{\mu}A)\gamma^{\mu}$$
$$=-\rho e_0(((\Omega_{\mu}\cdot v)\cdot A)\gamma^{\mu}$$
$$+(\nabla\wedge A)\cdot v+(v\cdot\nabla)A),$$
$$\partial_{\mu}(\rho P^{\mu})=-e_0(\partial_{\mu}(\rho v^{\mu})A+\rho(v\cdot\nabla)A)=-\rho e_0(v\cdot\nabla)A,$$

because of (4.1). Then

$$\rho\nabla(\mathrm{Tr}\,P)=\partial_{\mu}(\rho P^{\mu})-\rho e_0(\Omega_{\nu}\cdot(v\wedge A))\gamma^{\nu}$$
$$-e_0(\nabla\wedge A)\cdot v. \quad \text{(D3)}$$

(c) $\mathrm{Tr}\,L=\Omega_{\mu}\cdot(i(\gamma^{\mu}\wedge s)),$

$$L^{\mu}=(\Omega_{\nu}\cdot(i(\gamma^{\mu}\wedge s)))\gamma^{\nu}.$$
$$\rho\nabla(\mathrm{Tr}\,L)=\rho(\partial_{\nu}\Omega_{\mu}\cdot(i(\gamma^{\mu}\wedge s)))\gamma^{\nu}+\rho(\Omega_{\mu}$$
$$\cdot(i(\gamma^{\mu}\wedge(\Omega_{\nu}\cdot s))))\gamma^{\nu},$$
$$\partial_{\mu}(\rho L^{\mu})=\rho(\partial_{\mu}\Omega_{\nu}\cdot(i(\gamma^{\mu}\wedge s)))\gamma^{\nu}$$
$$-(\Omega_{\nu}\cdot(\gamma^{\mu}\cdot(\partial_{\mu}(\rho is))))\gamma^{\nu}.$$
$$\rho\nabla(\mathrm{Tr}\,L)=\partial_{\mu}(\rho L^{\mu})+\rho((\partial_{\nu}\Omega_{\mu}-\partial_{\mu}\Omega_{\nu})$$
$$\cdot(i(\gamma^{\mu}\wedge s)))\gamma^{\nu}+\rho(\Omega_{\mu}\cdot(i(\gamma^{\mu}\wedge(\Omega_{\nu}$$
$$\cdot s))))\gamma^{\nu}+I, \quad \text{(D4)}$$

where

$$I=(\Omega_{\nu}\cdot(\nabla\cdot(\rho is)))\gamma^{\nu}. \quad \text{(D5)}$$

From (D1), (D5) and (C8) one obtains, after elimination of $\Omega_{\nu}\cdot(s\wedge(\nabla\beta))$,

$$\partial_{\mu}(\rho T^{\mu})+\rho((-(i/2)(\nabla\ln\rho\wedge s)+(i/2)(s^{\mu}\Omega_{\mu})$$
$$+e_0(A+(A\cdot s)s)\wedge v)\cdot\Omega_{\nu})\gamma^{\nu}+\rho(J/2)$$
$$+(\rho m_0\sin\beta-\tfrac{1}{2}\nabla\cdot(\rho s)(\nabla\beta))=\rho e_0(\nabla\wedge A)\cdot v, \quad \text{(D6)}$$

where ρJ is the sum of the second and the third terms on the right-hand side of (4).

From (4.5) and (C7) one deduces

$$(\rho/2)[J+((i(s^{\mu}\Omega_{\mu})+i((\gamma^{\mu}\cdot\Omega_{\mu})\wedge s))\cdot\Omega_{\nu})\gamma^{\nu}]$$
$$+(\rho m_0\sin\beta-\tfrac{1}{2}\nabla\cdot(\rho s))(\nabla\beta)=0. \quad \text{(D7)}$$

Calling the bracket K and developing, one obtains

$$K=((i(\partial_{\nu}\Omega_{\mu}-\partial_{\mu}\Omega_{\nu}))\cdot\gamma^{\mu}$$
$$+\Omega_{\mu}\cdot((i\Omega_{\nu})\cdot\gamma^{\mu})-((i\Omega_{\nu})\cdot(\Omega_{\mu}\cdot\gamma^{\mu}))\cdot s)\gamma^{\nu}. \quad \text{(D8)}$$

Let us recall that the relation (5.11) may be deduced from $2\partial_j R=\Omega_j R,$

$$2\,\partial_k\,\partial_j\,R=(\partial_k\Omega_j)R+\Omega_j(\partial_k R)\tilde{R}R=(\partial_k\Omega_j+\tfrac{1}{2}\,\Omega_j\Omega_k)R$$

and from

$$\partial_k\partial_j\,R=\partial_j\,\partial_k R.$$

Writing $\gamma^{\mu}=a$, and $B=\partial_{\nu}\Omega_{\mu}-\partial_{\mu}\Omega_{\nu}$, $C=\Omega_{\mu}$, $D=\Omega_{\nu}$, so $B+\tfrac{1}{2}(CD-DC)=0$, we obtain

$$K=(((iB)\cdot a+C\cdot((iD)\cdot a)-(iD)\cdot(C\cdot a))\cdot s)\gamma^{\nu}$$
$$=((\tfrac{1}{2}(iBa-aBi)+\tfrac{1}{4}\,(C(iDa-aiD)-(iDa-aiD)C$$
$$-iD(Ca-aC)+(Ca-aC)iD))\cdot s)\gamma^{\nu}$$
$$=((\tfrac{1}{2}\,i(B+\tfrac{1}{2}(CD-DC))a-ai(B$$
$$+\tfrac{1}{2}(CD-DC)))\cdot s)\gamma^{\nu}=0, \quad \text{(D9)}$$

and so (5.12). Equation (5.13) is deduced from (5.12) by

$$\nabla\cdot s=\partial^{\mu}\cdot(\gamma_{\mu}s)=\gamma^{\mu}\cdot(\Omega_{\mu}\cdot s)=\Omega_{\mu}\cdot(s\wedge\gamma^{\mu})$$
$$=-(\Omega_{\mu}\cdot\gamma^{\mu})\cdot s=(\gamma^{\mu}\cdot\Omega_{\mu})\cdot s.$$

[1] H. Tetrode, Z. Phys. **49**, 858 (1928).
[2] O. Costa de Beauregard, *La Théorie de la Relativité Restreinte* (Masson, Paris, 1949).
[3] J. Yvon, J. Phys. Radium **8**, 18 (1940).
[4] T. Takabayasi, Prog. Theor. Phys. Suppl. **4**, 1 (1957).
[5] F. Halbwacs, *Théorie Relativiste des Fluides á Spin* (Gauthier–Villars, Paris, 1960).
[6] D. Hestenes, J. Math. Phys. **8**, 798 (1967).
[7] A. Landé, Am. J. Phys. **33**, 123 (1965).
[8] L. Ballentine, Rev. Mod. Phys. **42**, 358 (1970).
[9] D. Hestenes, Am. J. Phys. **47**, 399 (1979).
[10] D. Hestenes, *Space-Time Algebra* (Gordon and Breach, New York, 1966).
[11] R. Boudet, C. R. Acad. Sci. Paris A **272**, 767 (1971).
[12] G. Casanova, C. R. Acad. Sci. Paris B **266**, 1551 (1968).
[13] R. Boudet and P. Quilichini, C. R. Acad. Sci. Paris A **268**, 724 (1969).
[14] H. Hestenes, J. Math. Phys. **14**, 893 (1973).
[15] P. Quilichini, C. R. Acad. Sci. Paris B **273**, 829 (1971).
[16] R. Boudet, C. R. Acad. Sci. Paris A **278**, 1063 (1974).
[17] R. Deck, Am. J. Phys. **49**, 495 (1981).
[18] D. Hestenes, Found. Phys. **12**, 153 (1982).

A rigged Hilbert space for the free radiation field

M. Gadella[a)]
Center for Particle Theory, The University of Texas at Austin, Austin, Texas 78712 and Facultad de Ciencias de Santander, Avenida los Castros s/n, Santander, Spain

(Received 30 March 1983; accepted for publication 22 June 1984)

We here construct a rigged Hilbert space suitable for the description of the free radiation field in a box and in the Coulomb gauge. We prove the continuity of the main observables on this structure.

I. INTRODUCTION

The formulation of quantum mechanics (QM) in terms of Hilbert spaces is the classical manner in which this theory is presented. QM was formalized first by Dirac.[1] Later, von Neumann intended to give a rigorous mathematical foundation structure to the Dirac formalization.[2] This structure, which uses Hilbert spaces as its basis, does not account for some of the most interesting features of the Dirac theory. In particular, given any observable A, we do not have an eigenfunction f of A in the Hilbert space for any point in the spectrum of A, as claimed by the Dirac theory. Also, A cannot be expanded in terms of elementary projections associated with these functions (since they do not exist).

In order to really implement the Dirac formalization, Gelfand has introduced the structure of rigged Hilbert space (RHS). We recall that a RHS is a triplet of spaces

$$\Phi \subset \mathscr{H} \subset \Phi^{\times}, \tag{1}$$

where $\mathscr{H}$ is a Hilbert space and Φ is a complete nuclear locally convex space, densely and continuously embedded into $\mathscr{H}$. $\Phi^{\times}$ is the antidual of Φ: the vector space of all continuous antilinear functionals on Φ. Dirac's theory requires eigenfunction expansions, as mentioned above. The corresponding results have been obtained by Gelfand[3] and Maurin.[4,5] For the existence of an eigenfunction expansion for an observable A, we require the stability of Φ under A, i.e., $A\Phi \subset \Phi$. We also need the continuity of A under the nuclear topology on Φ. Nuclearity plays an important role in the proof of the existence of the eigenfunction expansion and it cannot be dropped.

The general formulation of QM in terms of RHS has been mainly proposed by Böhm,[6,7] Roberts,[8,9] Antoine,[10,11] and Melsheimer.[12,13] It has been used in textbooks such as the one by Böhm.[14] Furthermore, RHS's allow an elegant and coherent description of resonances, decaying states and virtual states as generalized eigenvectors or as pairs of generalized eigenvectors of the total Hamiltonian H in a decaying scattering process. Research in this field has been performed by Baumgärtel,[15] Böhm,[14,16] the author,[17] and others.[18]

Although the general theory can be considered as already constructed, we note a lack of explicit examples in the literature. The free particle in Cartesian coordinates is easy to formulate by using $\Phi = \mathscr{S}(R^n)$ ($n = 1$, 2, or 3 being the dimension of the space in which the particle lives). The same space works for the harmonic oscillator.[19] The aim of the present paper is to provide one more example: that of the electromagnetic field in the Coulomb gauge. For that, we need first mention the idea of π-topology on tensor products.

Let $\Phi_1 \subset \mathscr{H}_1 \subset \Phi_1^{\times}$ and $\Phi_2 \subset \mathscr{H}_2 \subset \Phi_2^{\times}$ be two RHS's. Consider the (topological) tensor product $\mathscr{H}_1 \otimes \mathscr{H}_2$. We look for a topology τ_π on the (algebraic) tensor product $\Phi_1 \otimes \Phi_2$, such that (1) its completion under τ_π is a nuclear locally convex space which we call $\Phi_1 \otimes_\pi \Phi_2$; (2) $\Phi_1 \otimes_\pi \Phi_2$ is dense on $\mathscr{H}_1 \otimes \mathscr{H}_2$. The canonical embedding i: $\Phi_1 \otimes_\pi \Phi_2 \to \mathscr{H}_1 \otimes \mathscr{H}_2$ is continuous; (3) if A_2 is a continuous operator on Φ_2 and A_2 is on Φ_2, $A_1\Phi_1 \subset \Phi_1$, and $A_2\Phi_2 \subset \Phi_2$, then $A_1 \otimes I$ and $I \otimes A_2$ are continuous operators on $\Phi_1 \otimes_\pi \Phi_2$ and both leave this space invariant. I denotes the identity on both Φ_1 and on Φ_2.

For a discussion of the construction of τ_π see Ref. 12. Following this procedure, we obtain a compounded RHS $\Phi_1 \otimes \Phi_2 \subset \mathscr{H}_1 \otimes \mathscr{H}_2 \subset (\Phi_1 \otimes_\pi \Phi_2)^{\times}$, such that any continuous observable on Φ_1 or Φ_2 can be extended into a continuous observable on $[\Phi_1 \otimes_\pi \Phi_2]^{\times}$. Hereafter, we call this compounded RHS the π-tensor product of the two single RHS's.

II. THE FREE RADIATION FIELD IN THE COULOMB GAUGE

As is well known, the free radiation field in the Coulomb gauge under periodic boundary conditions in a cubic box of side L is equivalent to a countable infinite set of independent and uncoupled harmonic oscillators.[20] Based on this idea, we shall find a RHS, $\Phi \subset \mathscr{H} \subset \Phi^{\times}$ such that the nth creation and annihilation operators, the total number operator and the renormalized Hamiltonian are well-defined continuous operators on Φ.

To construct the Hilbert space $\mathscr{H}$ used here, consider the set T of all linear combinations of tensor products of the form[21]

$$v = v_1 \otimes v_2 \otimes \cdots \otimes v_p \otimes \phi_0 \otimes \phi_0 \otimes \cdots, \tag{2}$$
$$v_i \in L^2(R),$$

where only the p first terms in the tensor product are different from $\phi_0 = (w/\pi\hbar)^{1/4} e^{-x^2/2\hbar}$, the ground state of the harmonic oscillator. The scalar product of v with

$$w = w_1 \otimes w_2 \otimes \cdots \otimes w_q \otimes \phi_0 \otimes \phi_0 \otimes \cdots \tag{3}$$

if, for instance $p \geqslant q$, is

$$\langle v|w\rangle = \langle v_1|w_1\rangle\langle v_2|w_2\rangle \cdots \langle v_q|w_q\rangle \cdots \langle v_p|\phi_0\rangle. \tag{4}$$

$\mathscr{H}$ will be the completion of T under this scalar product. $\mathscr{H}$ is isomorphic to the Fock space $\mathscr{G} = \otimes_{n=1}^{\infty} [\otimes^n L^2(R)]$, but it has some advantages over $\mathscr{G}$. First, we have a nice representation of the vacuum as $\psi_0 = \phi_0 \otimes \phi_0 \otimes \phi_0 \otimes \cdots$, the state for which all the oscillators

[a)] Permanent address: Facultad de Ciencias, Departamento de Física Teórica, Prado de la Magdalena, s/n, Valladolid, Spain.

 0022-2488/85/040725-03$02.50

are in their ground state. Second, $\mathscr{H}$ is appropriate because of the idea that the free radiation field (under these boundary conditions) can be regarded as a set of uncoupled harmonic oscillators. An orthonormal basis for $\mathscr{H}$ is given by

$$\phi_{n_1} \otimes \phi_{n_2} \otimes \cdots \otimes \phi_{n_k} \otimes \phi_0 \otimes \phi_0 \cdots, \tag{5}$$

where ϕ_{n_i} is the n_ith Hermite function in $L^2(R)$ and k is an arbitrary number.

The action of the kth position Q_k or momentum P_k operators on T is defined as usual. For instance,

$$\begin{aligned} Q_k v &= Q_k(v_1 \otimes v_2 \otimes \cdots \otimes v_k(x) \otimes \cdots \otimes v_p \otimes \phi_0 \otimes \cdots) \\ &= v_1 \otimes v_2 \otimes \cdots \otimes x v_k(x) \otimes \cdots \otimes v_p \otimes \phi_0 \otimes \cdots . \end{aligned} \tag{6}$$

We can also define the kth annihilation a_k and creation a_k^+ operators, where $a_k = \sqrt{w_k/2\hbar}\; Q_k + i\sqrt{\hbar/w_k}\; P_k$ and $a_k^+ = \sqrt{w_k/2\hbar}\; Q_k - i\sqrt{\hbar/w_k}\; P_k$, on T. Q_k and P_k are extensible into bounded self-adjoint operators on $\mathscr{H}$ and $[Q_k, P_k] = i\hbar I$ on their common domain. They are a representation of the canonical commutation relations over $\mathscr{H}$,[22] with vacuum state ψ_0 as can be shown.

In order to complete the triplet, we need to find a Φ fulfilling the required conditions. To construct Φ, consider the sequence of spaces

$$\begin{aligned} \Phi_1 &= \mathscr{S}(R) \otimes \phi_0 \otimes \phi_0 \otimes \cdots \\ \Phi_2 &= \mathscr{S}(R) \otimes_\pi \mathscr{S}(R) \otimes \phi_0 \otimes \phi_0 \otimes \cdots \\ \Phi_3 &= \mathscr{S}(R) \otimes_\pi \mathscr{S}(R) \otimes_\pi \mathscr{S}(R) \otimes \phi_0 \otimes \phi_0 \otimes \cdots \\ &\;\;\vdots \\ \Phi_n &= \underbrace{\mathscr{S}(R) \otimes_\pi \mathscr{S}(R) \otimes_\pi \cdots \otimes_\pi \mathscr{S}(R)}_{n \text{ times}} \otimes \phi_0 \otimes \cdots \\ &= \Big[\mathop{\otimes}_{\pi}^{n} \mathscr{S}(R) \Big] \otimes \phi_0 \otimes \phi_0 \otimes \cdots . \end{aligned} \tag{7}$$

If $\phi \in \Phi_1$, $\phi = v(x) \otimes \phi_0 \otimes \phi_0 \otimes \cdots \in \mathscr{H}$, where $v(x) \in \mathscr{S}(R)$. $\phi \in \Phi_2$ if and only if $\phi = v(x_1, x_2) \otimes \phi_0 \otimes \phi_0 \otimes \cdots$, where $v(x_1,x_2) \in \mathscr{S}(R) \otimes_\pi \mathscr{S}(R)$, etc. Also note that Φ_1 is isomorphic to $\mathscr{S}(R)$, Φ_2 to $\mathscr{S}(R) \otimes_\pi \mathscr{S}(R)$, etc. If we transport the topology from $\otimes_\pi^n \mathscr{S}(R)$ into Φ_n through this isomorphism, we have endowed the Φ_n with a complete nuclear locally convex topology. Take now the subspaces

$$\begin{aligned} \mathscr{H}_1 &= L^2(R) \otimes \phi_0 \otimes \phi_0 \otimes \cdots \\ \mathscr{H}_2 &= L^2(R) \otimes L^2(R) \otimes \phi_0 \otimes \phi_0 \otimes \cdots \\ &\;\;\vdots \\ \mathscr{H}_n &= \underbrace{L^2(R) \otimes L^2(R) \otimes \cdots \otimes L^2(R)}_{n \text{ times}} \otimes \phi_0 \otimes \phi_0 \otimes \cdots \\ &= [\otimes^n L^2(R)] \otimes \phi_0 \otimes \phi_0 \otimes \cdots . \end{aligned} \tag{8}$$

Obviously, $\mathscr{H}_n$ is isomorphic (algebraically and topologically) to $\otimes^n L^2(R)$. Therefore, Φ_n is dense in $\mathscr{H}_n$ for all n and its canonical injection is continuous due to the properties of the π-tensor products.[12]

By definition, Φ is the inductive limit of the Φ_n. (Note that $\Phi_n \subset \Phi_{n+1}$ and the restriction on Φ_n of the topology on Φ_{n+1} gives back the topology on Φ_n.) From the set theoretical point of view, $\Phi = U\Phi_n$. A sequence $\{x_i\}_{i\in N}$ converges to $x \in \Phi$ if and only if there exists a natural number n such that $x_i \in \Phi_n$ for all i and $x_i \to x$ in Φ_n.[23]

It is readily seen that $\otimes_\pi^n \mathscr{S}(R)$ is a Frechet space because it is complete and its topology is generated by a countably infinite set of seminorms. Therefore, Φ_n also is a Frechet space and this is true for all n, so that Φ is an inductive limit of Frechet spaces, and indeed a strict one (LF-space).

Theorem 1: Φ is a complete nuclear locally convex space, such that $\Phi \subset \mathscr{H} \subset \Phi^\times$ is a RHS.

Proof: From the definition of inductive limit, Φ is locally convex.[23] Also, any strict inductive limit of Frechet spaces is complete.[24] The inductive limit of nuclear spaces is also nuclear.[25]

Now, we need to prove that: (1) Φ is dense in $\mathscr{H}$; (2) the canonical injection $i: \Phi \to \mathscr{H}$ is continuous.

(1) Let us pick $\mathscr{G} = \cup_{n=1}^\infty \mathscr{H}_n$, which is a subspace of $\mathscr{H}$. Since any vector in the orthonormal basis of $\mathscr{H}$ belongs to it, $\mathscr{G}$ is dense in $\mathscr{H}$. On the other hand, if $v \in \mathscr{G}$ there exists a natural n such that $v \in \mathscr{H}_n$. Φ_n is dense in $\mathscr{H}_n$. Φ_n is dense in $\mathscr{H}_n$ and, thus, there exists a sequence $(v_i)_{i\in N}$ with $v_1 \in \Phi_n \subset \Phi$ for all i such that $v_1 \to v$. This proves the denseness of Φ in $\mathscr{H}$.

(2) Since $\mathscr{H}_n$ is a topological subspace of $\mathscr{H}$, the injection $j_n: \mathscr{H}_n \to \mathscr{H}$ is continuous. So is the injection $k_n: \Phi_n \to \mathscr{H}_n$ as we have previously stated. Thus $i_n = j_n k_n: \Phi_n \to \mathscr{H}$ is continuous. i_n is the restriction to Φ_n of the injection i from Φ into $\mathscr{H}$, which is continuous if and only if all the i_n are continuous,[23] as is the case.

From all of the above, we conclude that $\Phi \subset \mathscr{H} \subset \Phi^\times$ is a RHS.

Theorem 2: All the creation a_k^+ and annihilation a_k operators as well as the total number operator $N = \sum_{k=1}^\infty a_k^+ a_k$ and the renormalized Hamiltonian $H_0 = \sum_{k=1}^\infty \hbar w_k a_k^+ a_k$ are well-defined continuous operators on Φ.

Before giving the proof, we observe that the LF character of Φ implies the following.

Proposition: Let A be a linear mapping from Φ into itself. A is continuous on Φ if and only if, given any convergent sequence $(x_i)_{i\in N}$ on Φ with $x_i \to x \in \Phi$, $Ax_i \to Ax$.[23]

Proof of Theorem 2: Let a be the annihilator operator for the one-dimensional harmonic oscillator. $a\mathscr{S}(R) \subset \mathscr{S}(R)$ and it is continuous under the metrizable topology on $\mathscr{S}(R)$. Let $n \geqslant k$ and $\tilde{a}_k = I \otimes \cdots \otimes a \otimes I \otimes \cdots \otimes I$, where a is placed at the kth position and the other operators in this tensor product are all equal to the identity on $\mathscr{S}(R)$. $\tilde{a}_k$ operates on $\otimes_\pi{}^n \mathscr{S}(R)$ and because of the properties of the π-topologies, is continuous on that space. $\otimes_\pi^n \mathscr{S}(R)$ is both topological and algebraically isomorphic to Φ_n. Under this isomorphism, $\tilde{a}_k$ corresponds to a_k. Therefore, $a_k \Phi_n \subset \Phi_n$ and a_k is continuous on Φ_n, $n \geqslant k$.

Let $(v_i)_{i\in N}$ be a sequence in Φ converging to $v \in \Phi$. Then, there exists an $n \geqslant k$ such that $v_i, v \in \Phi_n$ and $v_i \to v$ in Φ_n. Since a_k is continous on Φ_n, $a_k v_i \to v$ in Φ_n and thus $a_k v_i \to v$ in Φ. The continuity of a_k follows from the proposition. The same considerations and proof are valid for a_k^+.

To see that $N\Phi \subset \Phi$, take $\Phi_M = [\otimes_\pi^M \mathscr{S}(R)] \otimes \phi_0 \otimes \phi_0 \cdots$

$$N\Phi_M = \sum_{n=1}^\infty a_k^+ a_k([\otimes_\pi^M \mathscr{S}(R)] \otimes \phi_0 \otimes \phi_0 \otimes \cdots)$$

$$= \left[\sum_{n=1}^{M} \tilde{a}_k^+ \tilde{a}_k \otimes_\pi^M \mathscr{S}(R) \right] \otimes \phi_0 \otimes \phi_0 \otimes \cdots$$

$$= \left[\sum_{n=1}^{M} a_k^+ a_k \right] \Phi_M \subset \Phi_M \subset \Phi \Rightarrow N\Phi \subset \Phi.$$

The continuity of N on Φ follows from the continuity of $\Sigma_{k=1}^{n} a_k^+ a_k$ and the proposition. The proof of the theorem for H_0 is identical and deserves no further comments.

Corollary: One immediately sees that P_k and Q_k are also continuous on Φ for any k.

Remark 1: N, H_0, Q_k, and P_k are essentially self-adjoint on Φ. This comes more or less trivially from the essential self-adjointness of Q and P on $\mathscr{S}(R)$ and the following result[26]: A symmetric operator A is essentially self-adjoint on a domain $\mathscr{D}$ if and only if $(A \pm iI)\mathscr{D}$ are dense on $\mathscr{H}$, where I is the identity.

Remark 2: Note that the unrenormalized Hamiltonian $\overline{H}_0 = \Sigma_{k=1}^{\infty} (\hbar w_k + \frac{1}{2}) a_k^+ a_k$ is not even defined on Φ.

Remark 3: It should be noted that the space Φ as here constructed is isomorphic to the Borchers algebra $\mathscr{s} = \oplus_{n=0}^{\infty} \mathscr{S}(R^{4n})$ (Ref. 27). This fact easily follows from the isomorphism between $\mathscr{S}(R^{n+m})$ and $\mathscr{S}(R^m) \otimes_\pi \mathscr{S}(R^n)$.[28] However, it is important to remark that the general ideas used in this paper to construct our model for the free radiation field have been these of basic quantum mechanics in RHS and not those of the general theory of quantized fields.[29]

[1] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford U.P., Oxford, 1958).

[2] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer, Berlin, 1932).

[3] I. M. Gelfand and N. Y. Vilenkin, *Generalized Functions* (Academic, New York, 1964), Vol. IV.

[4] K. Maurin, Bull. Acad. Pol. Sci., Ser. Sci., Math., Astron. Phys. **VII**, 32 (1959); **VIII**, 381 (1959); K. Maurin, Math. Scand. **9**, 359 (1961).

[5] K. Maurin, Bull. Acad. Pol. Sci., Ser. Sci., Math., Astron. Phys. **VII**, 471 (1959).

[6] A. Böhm, *Boulder Lectures in Theoretical Physics* (Gordon and Breach, New York, 1967), Vol. 9a.

[7] A. Böhm, "The Rigged Hilbert Space and Quantum Mechanics," in *Lecture Notes in Physics* (Springer, New York, 1978), Vol. 78.

[8] J. E. Roberts, J. Math. Phys. **7**, 1097 (1966).

[9] J. E. Roberts, Commun. Math. Phys. **3**, 98 (1966).

[10] J. P. Antoine, J. Math. Phys. **10**, 53 (1969).

[11] J. P. Antoine, J. Math. Phys. **10**, 2276 (1969).

[12] O. Melsheimer, J. Math. Phys. **15**, 902 (1974).

[13] O. Melsheimer, J. Math. Phys. **15**, 917 (1974).

[14] A. Böhm, *Quantum Mechanics* (Springer, New York, 1979).

[15] H. Baumgärtel, "Group Theoretical Methods in Physics," Proceedings of the Austin Conference 1978, in *Lecture Notes in Physics* (Springer, Berlin, 1979), Vol. 94; H. Baumgärtel, Math. Nachr. **75**, 133 (1976).

[16] A. Böhm, "The Rigged Hilbert Space and Decaying States," in *Group Theoretical Methods in Physics, Proceedings of the Austin Conference*, 1978 (Springer, Berlin, 1979), Vol. 94; A. Böhm, J. Math. Phys. **21**, 1040 (1980); **22**, 2813 (1981); Lett. Math. Phys. **3**, 455 (1979).

[17] M. Gadella, J. Math. Phys. **24**, 1462 (1983); M. Gadella, J. Math. Phys. **24**, 2142 (1983).

[18] G. Parravicini, V. Gorini, and E. C. G. Sudarshan, J. Math. Phys. **21**, 2208 (1980).

[19] A construction of a RHS for the harmonic oscillator is given in Ref. 7.

[20] J. J. Sakurai, *Advanced Quantum Mechanics* (Addison, Reading, MA, 1967).

[21] For an introduction into Fock space techniques see for instance D. Kastler, *Introduction a l'electrodinamique quantique* (Dunod, Paris, 1961). This infinite tensor product is constructed for instance in M. Gadella and E. Santos, J. Math. Phys. **22**, 1651 (1981).

[22] G. E. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley, New York, 1972).

[23] Vo-Khac Khoan, *Distributions, Analyse de Fourier, Operateurs aux Derivees Partielles* (Viubert, Paris, 1972), Vol. 1.

[24] G. Köthe, *Topological Vector Spaces* (Springer, Berlin, 1969), p. 223.

[25] K. Maurin, *General Eigenfunction Expansions and Unitary Representations of Topological Groups* (Polish Scientific Publishers, Warsaw, 1968), p. 68.

[26] M. Reed and B. Simon, *Functional Analysis* (Academic, New York, 1972).

[27] H. J. Borchers, "Algebraic Aspects of Wightman Field Theory," in *Statistical Mechanics and Field Theory*, edited by R. N. Sen and C. Weyl (Helsted, New York, 1972).

[28] F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967), p. 530.

[29] See for instance, R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, Reading, MA, 1964).

Usual and unusual summation rules over j angular momentum

E. Elbaz
Institut de Physique Nucléaire (et IN2P3), Université Claude Bernard Lyon-1, 43, Bd du 11 Novembre 1918, 69622, Villeurbanne Cedex, France

(Received 5 June 1984; accepted for publication 26 October 1984)

After recalling the usual j-summation rule over an angular momentum we describe two unusual summation rules. The first one concerns sum of products of "$3nj$" and "$3j0$" coefficients. The second involves sum of products of "$3jm$" coefficients but with the $(2j+1)$ weighting factor missing.

I. INTRODUCTION

When dealing with angular momentum algebra (the so-called Racah algebra) the use of the orthogonality relation in the "$3jm$" Wigner coefficients leads to a "j-summation rule"[1,2] in which a $(2j+1)$ weighting factor always appears when the sum holds over an angular momentum j. This usual j-summation rule cannot be used if the j angular momentum appears simultaneously in "$3jm$" and "$3j0$" coefficients or if the $(2j+1)$ weighting factor is missing.

This paper is devoted to the solution of these problems. For the sake of simplicity we use the well-known graphical technique of the graphical spin algebra (GSA)[1,2] but the results here defined may be obtained in a pure analytical procedure.

In the second section we recall the usual j-summation rule in its graphical aspect and give an example of application. The third section is devoted to the "summation rule over marked poles" concerning summation over sum of products of $3jm$ and $3j0$ coefficients. Such a rule has already been given[2] but is not yet well-known. We give here the general summation rule and an example of application. The last section is devoted to the "unusual j-summation rule" in which the $(2j+1)$ weighting coefficient is missing. We define a general procedure and show how particular cases already given by Dunlap and Judd[3] are obtained. In order to make clear the difference between the three summation rules here presented, we have applied them on the same "$6j$" coefficient.

II. THE USUAL j-SUMMATION RULE

Let us consider a general expression, which is the sum of the product of the $3jm$ Wigner coefficients over different magnetic momenta and a j-angular momentum:

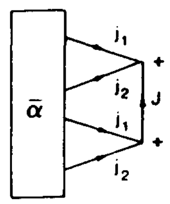
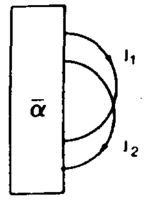
$$E=\sum_{m_1m_2m_1'm_2'}[j^2]\left(\bar{\alpha}\left|\begin{matrix} j_1 & j_2 & m_1' & m_2' \\ m_1 & m_2 & j_1 & j_2\end{matrix}\right.\right)\times\sum_{jm}\begin{pmatrix} m_1 & m_2 & m \\ j_1 & j_2 & j\end{pmatrix}\begin{pmatrix} j_1 & j_2 & j \\ m_1' & m_2' & m\end{pmatrix}. \tag{2.1}$$

The covariant Wigner's notation of $3jm$ coefficients

$$\begin{pmatrix} m & j_1 & j_2 \\ j & m_1 & m_2\end{pmatrix}=(-)^{j-m}\begin{pmatrix} j & j_1 & j_2 \\ -m & m_1 & m_2\end{pmatrix} \tag{2.2}$$

has been used while $[j^n]=(2j+1)^{n/2}$.

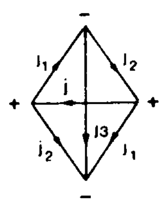
A graphical representation[1,2] of (2.1) shows in (2.3) the different angular momenta and the j summation. The usual j-summation rule is a graphical expression of orthogonality rule over the two $3jm$ coefficients and consists in erasing the j-summed line and linking the ends of the corresponding lines

$E=\sum_j [j^2]$  $=$  $\cdot$ (2.3)

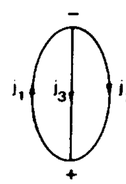
The last diagram reads analytically

$$E=\left(\bar{\alpha}\left|\begin{matrix} j_1 & j_2 & m_1 & m_2 \\ m_1 & m_2 & j_1 & j_2\end{matrix}\right.\right). \tag{2.4}$$

Let us take as an example, the summation over a j-angular momentum in a $6j$ coefficient:

$A=\sum_j [j^2]\begin{Bmatrix} j_1 & j_2 & j \\ j_1 & j_2 & j_3\end{Bmatrix}=\sum_j [j^2]$  $\cdot$ (2.5)

We erase the j summed line and link j_1 to j_1 and j_2 to j_2:

$A=(-)^{2j_2}$  $=(-)^{2j_3}\{j_1j_2j_3\}$. (2.6)

III. THE SUMMATION RULE OVER MARKED POLES

We consider now the following expression:

$$F=\sum_{m_1m_2m_3m_4}\left(\bar{\alpha}\left|\begin{matrix} l_1 & l_2 & l_3 & l_4 \\ m_1 & m_2 & m_3 & m_4\end{matrix}\right.\right)\times\int Y^*_{l_1m_1}(\Omega)Y^*_{l_2m_2}(\Omega)Y^*_{l_3m_3}(\Omega)Y^*_{l_4m_4}(\Omega)d\Omega, \tag{3.1}$$

which gives, after integration over the solid angle Ω,

$$F=\frac{[l_1l_2l_3l_4]}{4\pi}\sum_{m_1m_2m_3m_4}\left(\bar{\alpha}\left|\begin{matrix} l_1 & l_2 & l_3 & l_4 \\ m_1 & m_2 & m_3 & m_4\end{matrix}\right.\right)\times\sum_{LM}[L^2]\begin{pmatrix} 0 & 0 & L \\ l_1 & l_2 & 0\end{pmatrix}\begin{pmatrix} 0 & 0 & 0 \\ l_3 & l_4 & L\end{pmatrix}\times\begin{pmatrix} m_1 & m_2 & L \\ l_1 & l_2 & M\end{pmatrix}\begin{pmatrix} m_3 & m_4 & M \\ l_3 & l_4 & L\end{pmatrix}. \tag{3.2}$$

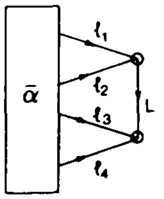
0022-2488/85/040728-04$02.50

The usual j-summation rule cannot be used since the L-summed angular momentum appears in the $3jm$ and in the $3j0$ coefficients.

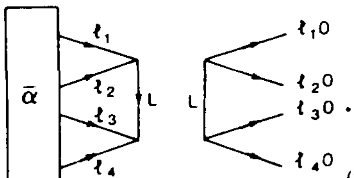
Graphically F is represented with marked poles representing the reduced matrix element of a spherical harmonic irreducible tensor operator:

$$\langle l_1 \| Y_{l_2} \| l_3 \rangle = \frac{[l_1 l_2 l_3]}{\sqrt{4\pi}} \begin{pmatrix} 0 & l_2 & l_3 \\ l_1 & 0 & 0 \end{pmatrix}. \quad (3.3)$$

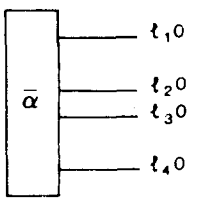
We thus obtain

$F = \sum_{L}$  , (3.4)

and if we express the two marked poles

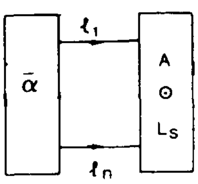
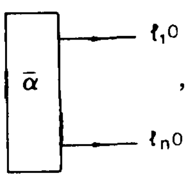
$F = \dfrac{[l_1 l_2 l_3 l_4]}{4\pi} \sum_{L} [L^2]$  . (3.5)

The well-known pinching rule of the (GSA)[1,2] gives, after summation over L,

$F = \dfrac{[l_1 l_2 l_3 l_4]}{4\pi}$ 

$$= \frac{[l_1 l_2 l_3 l_4]}{4\pi} \left(\bar{\alpha} \left| \begin{matrix} l_1 & l_2 & l_3 & l_4 \\ 0 & 0 & 0 & 0 \end{matrix} \right. \right). \quad (3.6)$$

The summation rule over marked poles thus consists in erasing the l-summed line and fixing the magnetic orbital momenta to their zero value. We can generalize this rule and obtain

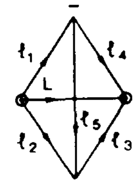
$\sum_{L_1 \cdots L_s}$  $= \dfrac{[l_1 \cdots l_n]}{(4\pi)^{n/2-1}}$  , (3.7)

which reads analytically

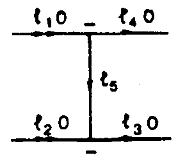
$$\sum_{L_1 \cdots L_s} \left(\bar{\alpha} \left| \begin{matrix} l_1 & \cdots & l_n \\ m_1 & \cdots & m_n \end{matrix} \right. \right) \left(\begin{matrix} m_1 & \cdots & m_n \\ l_1 & \cdots & l_n \end{matrix} \right| A \Big)_{L_s} \times \left(\begin{matrix} l_1 & \cdots & l_n \\ 0 & \cdots & 0 \end{matrix} \right| A \Big)_{L_s}$$
$$= \frac{[l_1 \cdots l_n]}{(4\pi)^{n/2-1}} \left(\bar{\alpha} \left| \begin{matrix} l_1 & \cdots & l_n \\ 0 & & 0 \end{matrix} \right. \right). \quad (3.8)$$

We can take a $6j$ with two marked poles as an example:

$$B = \frac{[l_1 l_2 l_3 l_4]}{4\pi} \times \sum_{L} [L^2] \begin{Bmatrix} l_1 & l_2 & L \\ l_3 & l_4 & l_5 \end{Bmatrix} \begin{pmatrix} l_1 & l_2 & L \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_3 & l_4 & L \\ 0 & 0 & 0 \end{pmatrix}$$

$= \sum_{L}$  . (3.9)

We erase the L-summed line and set to their zero value the magnetic momenta

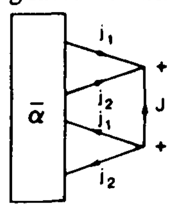
$B = \dfrac{[l_1 l_2 l_3 l_4]}{4\pi}$  . (3.10)

The following result comes analytically:

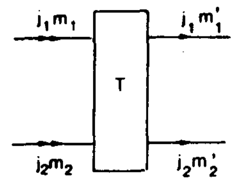
$$B = \frac{[l_1 l_2 l_3 l_4]}{4\pi} (-)^{l_1 + l_2} \times \begin{pmatrix} l_1 & l_4 & l_5 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_2 & l_3 & l_5 \\ 0 & 0 & 0 \end{pmatrix}. \quad (3.11)$$

IV. THE UNUSUAL j-SUMMATION RULE

Let us now consider a general expression in which the $(2j+1)$ weighting factor is missing,

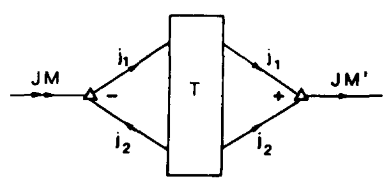
$G = \sum_{j}$  , (4.1)

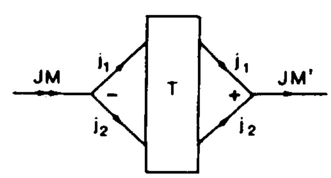
so that the usual j-summation rule (2.3) is not valid. In order to obtain a summation procedure, we first express the matrix element of a two-body operator in an uncoupled basis:

$\langle j_1 m_1 j_2 m_2 | T | j_1 m_1' j_2 m_2' \rangle =$  . (4.2)

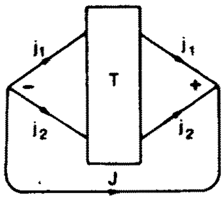
In a coupled basis the same matrix element becomes

$\langle (j_1 j_2) JM | T | (j_1 j_2) JM' \rangle$

$=$ 

$= [J_2] (-)^{2J}$  (4.3)

We use the pinching rule over two lines of the GSA[1,2] to get

$$\langle (j_1 j_2) JM | T | (j_1 j_2) JM' \rangle =$$ 

$$\delta_{MM'}. \quad (4.4)$$

The trace of this operator is identical in the two basis

$$\sum_{m_1 m_2} \langle j_1 m_1 j_2 m_2 | T | j_1 m_1 j_2 m_2 \rangle$$

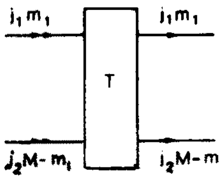
$$= \sum_{JM} \langle (j_1 j_2) JM | T | (j_1 j_2) JM \rangle. \quad (4.5)$$

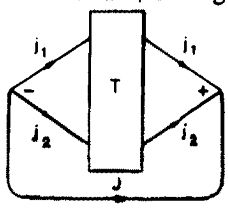
We can express the left-hand side with a summation over M since $M = m_1 + m_2$,

$$\sum_{m_1, M} \langle j_1 m_1 j_2 M - m_1 | T | j_1 m_1 j_2 M - m_1 \rangle$$

$$= \sum_{JM} \langle (j_1 j_2) JM | T | (j_1 j_2) JM \rangle. \quad (4.6)$$

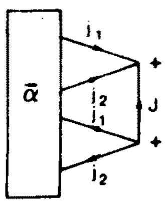
For any fixed value of M between $-J$ and $+J$ we get

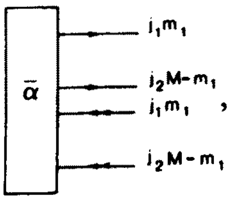
$$\sum_{m_1}$$ 

$$= \sum_J$$ 

$$. \quad (4.7)$$

We can express such a result in terms of a j-summation rule:

$$G = \sum_J$$ 

$$= \sum_{m_1}$$ 

$$, \quad (4.8)$$

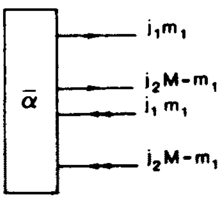
with $-J \leqslant M \leqslant J$ and $|j_1 - j_2| \leqslant J \leqslant j_1 + j_2$.

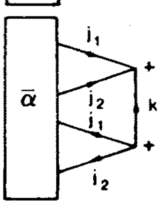
Before using this unusual j-summation rule some remarks apply.

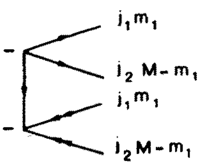
(1) If j_1 and j_2 are both half-integers, J is an integer and one can use the $M = 0$ value in (4.8). If only one of j_1 or j_2 is a half-integer, J is a half-integer and one can set $M = \frac{1}{2}$.

(2) If Eq. (4.8) is summed over M, the left-hand side gives $\Sigma_M 1 = [J^2]$ and we can use the usual j-summation rule while the right-hand side leads directly to the expected result.

(3) If the right-hand side of (4.8) is coupled over an intermediate momentum, $\mathbf{k} = \mathbf{j}_1 + \mathbf{j}_2$, one gets

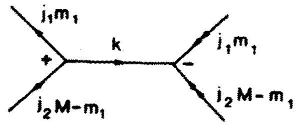
$$G = \sum_{m_1}$$ 

$$= \sum_{m_1 k} [k^2]$$ 



$$, \quad (4.9)$$

and a comparison with (4.8) gives immediately

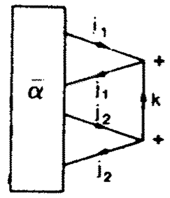
$$\sum_{m_1}$$ 

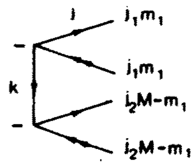
$$= [k^{-2}] = \frac{1}{2k+1}, \quad (4.10)$$

which reads analytically

$$\sum_{m_1} \begin{pmatrix} j_1 & j_2 & k \\ m_1 & M - m_1 & -M \end{pmatrix}^2 = \frac{1}{2k+1} \{ j_1 j_2 k \}. \quad (4.11)$$

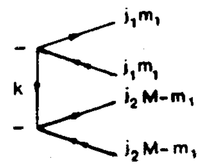
(4) If the coupling scheme defines the intermediate momentum $\mathbf{k}$ as $\mathbf{k} = \mathbf{j}_1 + \mathbf{j}_1 = \mathbf{j}_2 + \mathbf{j}_2$, so that $0 \leqslant k \leqslant 2j_1$, $0 \leqslant k \leqslant 2j_2$, one gets

$$G = \sum_{k, m_1} [k^2]$$ 



$$. \quad (4.12)$$

We set

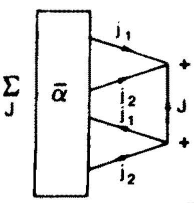
$$D(j_1 j_2; k) = \sum_{m_1}$$ 

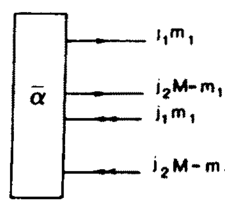
$$= \sum_{m_1} \begin{pmatrix} m_1 & k & j_1 \\ j_1 & 0 & m_1 \end{pmatrix} \times \begin{pmatrix} M - m_1 & 0 & j_2 \\ j_2 & k & M - m_1 \end{pmatrix}. \quad (4.13)$$

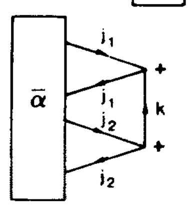
This coefficient has already been evaluated analytically by Dunlap and Judd[3]:

$$D(j_1 j_2; k) = \frac{1}{2k+1} \left[\frac{(2j_1 - k)!(2j_2 + k + 1)!}{(2j_2 - k)!(2j_1 + k + 1)!} \right]^{1/2}. \quad (4.14)$$

With this, one obtains the unusual j-summation rule

$$G = \sum_J$$ 

$$= \sum_{m_1}$$ 

$$= \sum_k [k^2] D(j_1 j_2; k)$$ 

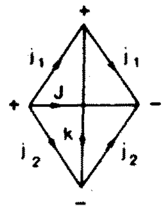
$$, \quad (4.15)$$

with $M = 0$ or $\frac{1}{2}$, $0 \leqslant k \leqslant 2j_1$, and $0 \leqslant k \leqslant 2j_2$.

Let us now take some examples.

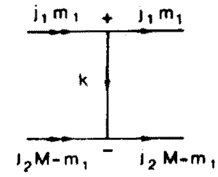
First example:

$$C_1 = \sum_J (-)^{j_1 + j_2 + k + J} \begin{Bmatrix} j_1 & j_2 & J \\ j_2 & j_1 & k \end{Bmatrix}$$

$$= \sum_J$$ 

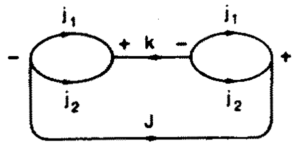
$\cdot$ (4.16)

The unusual j-summation rule allows us to erase the J-summed line and to set the m_2 magnetic moment to its $M - m_1$ values. It leads to the $D(j_1 j_2;k)$ coefficient as defined in (4.13):

$$C_1 = \sum_{m_1}$$ 

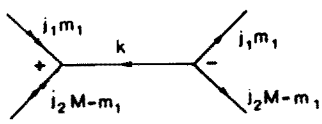
$= D(j_1 j_2;k).$ (4.17)

Second example:

$$C_2 = \sum_J$$ 

$$= \sum_J [J^{-2}]\delta_{kJ}\{j_1 j_2 J\}. \qquad (4.18)$$

The unusual j-summation rule gives

$$C_2 = \sum_{m_1}$$ 

$\cdot$ (4.19)

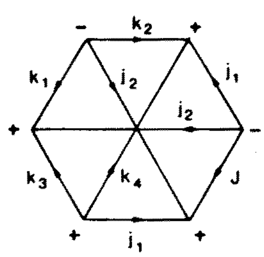
A comparison of these results gives then

$$\sum_{m_1} \begin{pmatrix} j_1 & j_2 & k \\ m_1 & M - m_1 & -M \end{pmatrix}^2 = \frac{1}{2k+1}\{j_1 j_2 k\}, \qquad (4.20)$$

a result already obtained in (4.10) and (4.11).

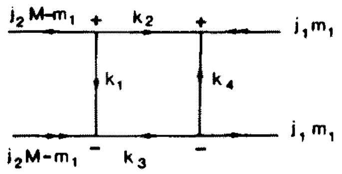
Third example:

$$C_3 = \sum_J \begin{Bmatrix} k_1 & k_2 & j_2 \\ k_3 & k_4 & j_1 \\ j_2 & j_1 & J \end{Bmatrix}$$

$$= \sum_J$$ 

$\cdot$ (4.21)

The unusual j-summation rule (4.15) gives immediately

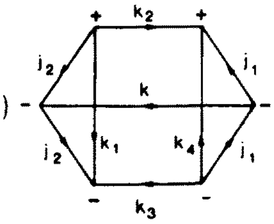
$$C_3 = \sum_{m_1}$$ 

, (4.22)

which reads analytically

$$C_3 = \sum_{M,q} \begin{pmatrix} j_1 & k_3 & k_4 \\ m_1 & q & -m_1 - q \end{pmatrix} \times \begin{pmatrix} m_1 & q & -m_1 - q \\ j_1 & k_2 & k_4 \end{pmatrix} \times \begin{pmatrix} j_2 & k_1 & k_2 \\ M - m_1 & -M + m_1 - q & q \end{pmatrix} \times \begin{pmatrix} M - m_1 & -M + m_1 - q & q \\ j_2 & k_1 & k_3 \end{pmatrix}. \qquad (4.23)$$

If we introduce the $D(j_1 j_2;k)$ coefficient, we get, with (4.15),

$$C_3 = \sum_k [k^2] D(j_1 j_2;k)$$ 

, (4.24)

which reads analytically

$$C_3 = \sum_k [k^2] D(j_1 j_2;k)(-)^{\varphi} \times \begin{Bmatrix} k_2 & k_3 & k \\ j_2 & j_2 & k_1 \end{Bmatrix} \begin{Bmatrix} k_2 & k_3 & k \\ j_1 & j_1 & k_4 \end{Bmatrix}, \qquad (4.25)$$

with $\varphi = j_1 + j_2 + k - k_1 - k_2 - k_3 + k_4$.

To summarize, when a purely geometrical expression is summed over a j-angular momentum three cases may appear.

(1) The j summation holds on normal $3jm$ poles and the $[j^2]$ weighting factor exists. The usual j-summation rule (2.3) may be used.

(2) The j summation holds on marked $3jm$ poles. The rule (3.7) can be used.

(3) The j summation holds on normal $3jm$ poles but the $[J^2]$ weighting factor is missing. The unusual j-summation rule (4.15) can be used.

ACKNOWLEDGMENTS

I am indebted to Professor M. Klapisch from the Hebrew University of Jerusalem for pointing out the unusual summation rule problem during our visit to H.B.J. and to my collaborators and especially J. Meyer for helpful discussions.

[1] A. P. Jucys, I. B. Levinson, and V. V. Vanagas, *The Theory of Angular Momentum* (Israel Program for Scientific Translation, Jerusalem, 1962); J. N. Massot, E. Elbaz, and J. Lafoucrière, Rev. Mod. Phys. **39**, 288 (1967); E. Elbaz and B. Castel, Am. J. Phys. **39**, 868 (1971); K.-N. Huan, Rev. Mod. Phys. **51**, 215 (1979).

[2] E. Elbaz and B. Castel, *Graphical Methods of Spin Algebras in Atomic, Nuclear and Particle Physics* (Dekker, New York, 1972).

[3] B. I. Dunlap and B. R. Judd, J. Math. Phys. **16**, 318 (1975).

Extension of Wick's theorem by means of the state operator formalism

Masao Nomura
Institute of Physics, College of Arts and Sciences, University of Tokyo, Tokyo, 153, Japan

(Received 20 July 1983; accepted for publication 26 October 1984)

Reordering of operators in the second quantization formalism is described by using state operators as basic operators, where the state operator is either a cluster of creation operators or that of annihilation operators specified by a set of quantum numbers of any representation. The commutation relation between state operators is simply expressed in terms of the coefficient of fractional parentage (cfp), the factor specific to state operator (de) composition. Wick's theorem, prevalently applied to a string of creation and annihilation operators, is extended to a string of state operators. Similar extension is done for the contraction theorem.

I. INTRODUCTION

In the second quantization formalism, Wick's theorem[1] has provided a powerful method for reordering operators in the form of normal products. This theorem is a direct consequence of commutation of fermions (or bosons) and relies on combinatorial analysis of an operator string $aa\cdots a$, where the operator a is either the creation operator C^+ or the annihilation operator C.

A drawback of Wick's theorem is sometimes conspicuous. An example, seen in the study of particle-pair motion in the spherical field, is reordering the operator product $(CC)_{J'M'}(C^+C^+)_{JM}$ in normal products, where the operator $(C^+C^+)_{JM}$ indicates a pair of C^+'s being coupled into the definite angular momentum J with its z component M. Reordering the operators by virtue of Wick's theorem requires three steps. First, the operator product is expanded into a linear combination of operator strings CCC^+C^+ with expansion coefficient being a pair of 3-J symbols. Next, Wick's theorem is applied to each of the operator strings. Finally, the resultant kinematical factors as well as operators are arranged, inversely to the first step, in the framework of the angular momentum representation. The complexity of the prescription originates from the fact that Wick's theorem is not applied directly to such an operator string as $(CC)_{J'M'}(C^+C^+)_{JM}$, but to each string made of uncoupled C^+'s and C's. The contraction theorem,[2] which is applied to the (anti)commutator of a pair of operator strings $aa\cdots a$, has the same drawback as Wick's theorem has.

In the present work, a general operator is reordered in normal products without recourse to its expansion into uncoupled forms of C^+'s and C's. The manipulation is presented as an extension of both Wick's theorem and the contraction theorem. We introduce state operators $A^+(n\alpha)$ and their conjugates $A(n\alpha)$, regarding them as types of basic operators for composition and commutation of involved operators. The state operator $A^+(n\alpha)$ here is defined to create an orthonormalized state $|n\alpha\rangle$, in case it acts on vacuum, and is expressed as a linear combination of uncoupled C^+'s of order n. The role of A^+ and A is schematically represented as

$$\text{General operators} —— \{A^+(n\alpha), A(n\alpha)\} —— \{C^+, C\}. \quad (1)$$

The discussion consists of the following three parts: (i) composition of A^+'s or A's, (ii) commutation of A^+'s and A's, and (iii) unification of (i) and (ii) to yield an extension of Wick's theorem and the contraction theorem. Each of (i)–(iii) is treated under the minimum destruction of pre-existing coupling and of the Pauli principle among C^+'s and/or C's. This present formalism does not rely on the way to specify quantum numbers α. For practical purposes, we keep in mind the case when quantum numbers α involve the total angular (or linear) momentum.

The kinematical factor appearing in the manipulation (i) is given by a matrix element of A^+ (or A). It is reduced, by using the Wigner–Eckart theorem, to the coefficient of fractional parentage (cfp)[3–5] in the shell model. Combining cfp with tensor algebra makes it possible to treat the many-body wave function without recourse to expansion into Slater determinants.[3] The kinematical factor appearing in (ii) is shown to be expressed simply by a pair of matrix elements of A^+ (or A), i.e., a pair of cfp's, indicating that the manipulations (i) and (ii) are closely related with each other. The commutation relation of A^+ and A was derived in a previous work.[4] It relied, however, on the first quantization formalism and was restricted to the angular momentum representation. In the present work, the lack of generality in Ref. 4 is removed and the key to commuting coupled C^+'s and C's is extracted.

The result is employed in the subsequent paper[6] where trace of a general operator product is evaluated in the truncated space of n fermions or n bosons. Application of the present manipulation to the Lanczos method[7] is also promissing since it is easy to rewrite $H^pA^+(n\alpha)|0\rangle$ as a sum of states, where H^p indicates the pth power of Hamiltonian.

In Sec. II, definition of the state operator is given. Section III concerns the state operator (de) composition. In Sec. IV, the operator product AA^+ is reordered in normal products. Section V deals with other relations involving operators A^+A and AA^+. Section VI is a summary of results in Secs. IV and V in the form of commutation relations. In Sec. VII, Wick's theorem is extended to the operator string consisting of coupled C^+'s and C's. In Sec. VII, the contraction theorem is treated in a similar way.

II. THE STATE OPERATOR

Let $\{|n\alpha\rangle, |n\beta\rangle, \cdots\}$ be an orthonormal set of n-fermion (or n-boson) wave functions

$$\langle n\alpha|n\beta\rangle = \delta(\alpha,\beta), \quad (2)$$

 0022-2488/85/040732-06$02.50

where $\alpha(\beta)$ indicates a set of quantum numbers. The wave functions are required to form the complete set spanning an n-body model space

$$\sum_\alpha |n\alpha\rangle\langle n\alpha| = P_n, \tag{3}$$

where P_n indicates the projection operator onto the n-body model space. There are various representations defining α satisfying both (2) and (3). Any representation can be chosen in the present work and explicit form of α is not necessary.

The n-body state operator $A^+(n\alpha)$ is defined by

$$|n\alpha\rangle = A^+(n\alpha)|0\rangle, \tag{4}$$

with no inclusion of annihilation operator C in $A^+(n\alpha)$. We postulate

$$A^+(0) = 1. \tag{5}$$

A well-known state operator is $A^+(n=2\,\alpha)$ with $\alpha = jj'JM$ [(2.23) of Ref. 5)]

$$A^+(jj'JM) = \sum_m (jmj'm';JM)C^+_{jm}C^+_{j'm'}/\sqrt{1+\delta(j,j')}. \tag{6}$$

A consequence of (4) is the following interrelation between the state and the state operator:

$$A^+(n\alpha)|m\beta\rangle = A^+(m\beta)|n\alpha\rangle(-1)^{mn}. \tag{7}$$

In a boson system, every sign factor of the form $(-1)^x$ is to be deleted throughout the work except (22), (23), (28), (53), and (54) to which instructions are given separately. The ensuing relation has well been used:

$$\sum_\alpha A^+(n\alpha)A(n\alpha) = \binom{\vec{n}}{n}, \tag{8}$$

where $\vec{n}$ indicates the number operator and $\binom{a}{b}$ is the binomial coefficient. The relation (8) is checked by the fact that both hand sides are n-body operators independent of specification of α.

III. STATE OPERATOR (DE) COMPOSITION

The product of state operators is rewritten in terms of a single operator as

$$A^+(n\alpha)A^+(n'\alpha') = \sum_\beta \langle n+n'\beta|A^+(n\alpha)|n'\alpha'\rangle A^+(n+n'\beta). \tag{9}$$

It is verified in the following way. The right-hand side (rhs) of (9) operating on an arbitrary state $|m\gamma\rangle$ is transformed, by using (3) and (7) to sum over β, as

$$A^+(m\gamma)A^+(n\alpha)|n'\alpha'\rangle(-1)^{m(n+n')}. \tag{10}$$

It is transformed, by using (7) again, into the same expression as the left-hand side (lhs) of (9) operating on the state $|m\gamma\rangle$.

The alternative to (9) is expressed as

$$A^+(n+n'\beta) = \sum_{\alpha\alpha'} \langle n'\alpha'|A(n\alpha)|n+n'\beta\rangle \times A^+(n\alpha)A^+(n'\alpha')/\binom{n+n'}{n}. \tag{11}$$

We verify it by showing that the rhs is transformed into the lhs. Let us substitute (9) in the operator product on the rhs of (11) and sum over quantum numbers by using (3) and (8). Then, we obtain the lhs of (11).

The matrix element of the state operator is seen in (11), as well as (9). In the angular momentum representation, it is transformed by virtue of the Wigner–Eckart theorem as

$$\langle n'J'M'|A(nJM)|n+n'J_0M_0\rangle^* = \langle n+n'J_0\|A^+(nJ)\|n'J'\rangle \times (J'M'JM;J_0M_0)/\sqrt{2J_0+1}. \tag{12}$$

The reduced matrix element on the rhs is proportional to cfp[4,5] and is called spectroscopic amplitude.

IV. REORDERING THE OPERATOR PRODUCT AA^+ IN NORMAL PRODUCTS

According to the prevalent prescription, reordering $A(n\alpha)A^+(m\beta)$ in normal products requires first the expansion of the operator product in the uncoupled form $C_1C_2\cdots C_nC^+_{n+1}\cdots C^+_{m+n}$. Wick's theorem is then applied and, finally, the resultant expression is arranged again in the form of state operators.

In fact, reordering $A(n\alpha)A^+(m\beta)$ results in a very simple form as summarized in the following theorem.

Theorem 1:

$$A(n\alpha)A^+(m\beta) = \sum_{k\gamma\delta} \langle n\alpha|A^+(n-k\gamma)A(m-k\delta)|m\beta\rangle \times A^+(m-k\delta)A(n-k\gamma)(-1)^{mn+k}, \tag{13}$$

where k runs over 0, 1, $\cdots$, min (m,n). It is rewritten as

$$A(n\alpha)A^+(m\beta) = \sum_{k\gamma\delta\epsilon} \langle n\alpha|A^+(k\epsilon)|n-k\gamma\rangle\langle m-k\delta|A(k\epsilon)|m\beta\rangle \times A^+(m-k\delta)A(n-k\gamma)(-1)^{(m-k)(n-k)}. \tag{14}$$

Proof: We use induction on both m and n. First, we prove (13) by induction on n with m being fixed to 1. The relation (13) in this case reads

$$A(n\alpha)C^+_\beta = C^+_\beta A(n\alpha)(-1)^n + (-1)^{n+1} \times \sum_\gamma \langle n\alpha|A^+(n-1\gamma)|1\beta\rangle A(n-1\gamma). \tag{15}$$

For $n=1$, it is reduced to the commutation of C_α and C^+_β. Assuming (15) for $n=N$, let us show (15) for $n=N+1$. The operator on the lhs of (15) for $n=N+1$ is rewritten, by virtue of the conjugate of (11), as

$$A(N+1\alpha)C^+_\beta = \sum_{\alpha'\epsilon} \langle N+1\alpha|A(N\alpha')|1\epsilon\rangle C_\epsilon A(N\alpha')C^+_\beta/(N+1). \tag{16}$$

The relation (15) for $n=N$ is substituted in $A(N\alpha')C^+_\beta$ on the rhs. After rearranging the resultant expression by virtue of (9) and (11), we get from (16) the relation (15) for $n=N+1$.

Next, we prove (13) for arbitrary n by induction on m. Assuming (13) for $m=M$, let us prove (13) for $m=M+1$. Decomposing $A^+(M+1\beta)$ by virtue of (11), we rewrite the operator product $A(n\alpha)A^+(M+1\beta)$, seen on the lhs of (13) for $m=M+1$, as a sum of products $A(n\alpha)A^+(M\beta')C^+_\epsilon$.

Substituting (13) for $m = M$ in the product $A(n\alpha)A^+(M\beta')$ yields

$$A(n\alpha)A^+(M+1\beta)$$
$$= \sum \langle n\alpha|A^+(n-k\gamma)A(M-k\beta_1)C_\epsilon|M+1\beta\rangle$$
$$\times A^+(M-k\beta_1)A(n-k\gamma)C_\epsilon^+(-1)^{M(n+1)+k}/(M+1), \tag{17}$$

where the sum is taken over k, β_1, γ, and ϵ. Substituting (13) for $m = 1$, i.e., (15) in the operator product $A(n-k\gamma)C_\epsilon^+$ on the rhs, we get

$$A(n\alpha)A^+(M+1\beta)$$
$$= \sum \langle n\alpha|A^+(n-k\gamma)A(M-k\beta_1)C_\epsilon|M+1\beta\rangle$$
$$\times \langle n-k\gamma|A^+(n-k-l\gamma')A(1-l\epsilon')|1\epsilon\rangle$$
$$\times A^+(M-k\beta_1)A^+(1-l\epsilon')A(n-k-l\gamma')$$
$$\times(-1)^{nM+n+M+l}/(M+1), \tag{18}$$

where the sum is taken over $k, l, \beta_1, \gamma, \gamma', \epsilon, \epsilon'$. The value of l is restricted to be either 0 or 1. Let us rearrange the rhs separately for each value of l and subsequently sum over l. Then, we obtain from (18) the relation (13) for $m = M + 1$. The relation (14) is easily derived from (13) by using (3) and (7).

V. RELATIONS AMONG OPERATOR PRODUCTS A^+A AND AA^+

The relation (13) in the last section can be regarded as a relation among operator products A^+A and AA^+. A few relations of this kind are shown here.

Theorem 2:

$$\sum_{\alpha\beta} \langle m\beta|A^+(k\gamma)A(l\epsilon)|n\alpha\rangle A^+(m\beta)A(n\alpha)$$
$$= \binom{\vec{n}-k}{m-k} A^+(k\gamma)A(l\epsilon)\delta(m-k, n-l). \tag{19}$$

Proof: In case $m - k \neq n - l$, it is clearly satisfied. Let us consider the case $m - k = n - l$. By virtue of (3) and (7), we have

$$A^+(m\beta)A(n\alpha) = \sum_{N\delta} P_N A^+(N-m\delta)|m\beta\rangle$$
$$\times \langle n\alpha|A(N-m\delta)(-1)^{(N-m)(m-n)}. \tag{20}$$

The lhs of (19) is transformed, by using (20) to sum over α and β, as

$$\text{The lhs of (19)} = \sum_{N\delta} P_N A^+(N-m\delta)A^+(k\gamma)$$
$$\times A(l\epsilon)A(N-m\delta)(-1)^{(N-m)(m-n)}. \tag{21}$$

Let us sum over δ using (8). Then, we obtain (19).

Previously, we derived (19) in the angular momentum representation using the first quantization formalism [(2.3) of Ref. 4].

It is easy to get from (19)

$$\sum_{k\gamma\delta} \langle n-k\gamma|A^+(n-l\alpha')A(m-l\beta')|m-k\delta\rangle$$
$$\times \langle n\alpha|A^+(n-k\gamma)A(m-k\delta)|m\beta\rangle(-1)^{l-k}$$
$$= \delta(l,0)\delta(\alpha,\alpha')\delta(\beta,\beta'), \tag{22}$$

where the sign on the rhs survives also in a boson system.

The following theorem provides antinormal ordering of A^+A.

Theorem 3:

$$A^+(m\beta)A(n\alpha)$$
$$= \sum_{k\gamma\delta} \langle n\alpha|A^+(n-k\gamma)A(m-k\delta)|m\beta\rangle$$
$$\times A(n-k\gamma)A^+(m-k\delta)(-1)^{(m-k)(n-k)}, \tag{23}$$

where k runs over 0, 1, ⋯, min(m, n). The sign on the rhs is replaced by $(-)^k$ in a boson system.

Proof: We show that the rhs will be transformed into the lhs. Let us substitute (13) in the operator product on the rhs, and sum over k, γ, and δ by virtue of (22). Then we obtain the lhs of (23).

VI. COMMUTATION RELATIONS AMONG STATE OPERATORS

The commutator of operator strings B and D is defined[2] by

$$[B,D]_\pm = BD - (-1)^{bd}DB, \tag{24}$$

where b (d) is the operator rank of B (D). It is shown that[2]

$$[BD,F]_\pm = B[D,F]_\pm + (-1)^{df}[B,F]_\pm D, \tag{25}$$

where f is the operator rank of F. Similarly,

$$[B,DF]_\pm = [B,D]_\pm F + (-1)^{bd}D[B,F]_\pm. \tag{26}$$

It is easy to rewrite (13) and (23) in the forms

$$[A^+(m\beta), A(n\alpha)]_\pm$$
$$= -\sum_{k\gamma\delta}{}' \langle n\alpha|A^+(n-k\gamma)A(m-k\delta)|m\beta\rangle$$
$$\times A^+(m-k\delta)A(n-k\gamma)(-1)^k \tag{27}$$

and

$$[A(n\alpha), A^+(m\beta)]_\pm$$
$$= -\sum_{k\gamma\delta}{}' \langle n\alpha|A^+(n-k\gamma)A(m-k\delta)|m\beta\rangle$$
$$\times A(n-k\gamma)A^+(m-k\delta)(-1)^{k(m+n+1)}, \tag{28}$$

respectively. The symbol Σ' indicates the sum excluding $k = 0$. The last sign factor on the rhs of (28) is replaced by $(-1)^k$ in a boson system.

Let us rewrite (27) using the contraction operator $\mathbb{C}$ defined by

$$\mathbb{C}(A^+(m\beta)A(n\alpha))$$
$$= \sum_k [C_k^+, [C_k, A^+(m\beta)A(n\alpha)]_\pm]_\pm. \tag{29}$$

It is transformed as

$$\mathbb{C}(A^+(m\beta)A(n\alpha))$$
$$= \sum_k [A^+(m\beta), C_k]_\pm [C_k^+, A(n\alpha)]_\pm$$
$$= \sum_{\gamma\delta} \langle n\alpha | A^+(n-1\gamma) A(m-1\delta) | m\beta \rangle$$
$$\times A^+(m-1\gamma) A(n-1\delta). \tag{30}$$

In the last step, we used (27). The rhs of (30) is just the $k=1$ term on the rhs of (27). The relation (27) compared with multiple applications of the contraction operator on $A^+(m\beta)A(n\alpha)$ yields

$$[A^+(m\beta), A(n\alpha)]_\pm = -\sum_k (-\mathbb{C})^k (A^+(m\beta)A(n\alpha))/k!. \tag{31}$$

VII. AN EXTENDED WICK'S THEOREM

Any operator in the second quantization formalism can be expressed in terms of the operator string

$$O_p = A(n_1\alpha_1)A^+(m_1\beta_1)A(n_2\alpha_2)$$
$$\times A^+(m_2\beta_2)\cdots A(n_p\alpha_p)A^+(m_p\beta_p) \tag{32}$$

under the convention (5). In case any of $n_1, m_1, ..., m_p$ is either 0 or 1, the operator O_p reads the string made of C^+ and C. We extend Wick's theorem so that it could be applied to O_p without any expansion into uncoupled C^+'s and C's.

In any expression of the present work, each Greek index subject to summation appears only in its bilinear form. See, e.g., (14). It reflects the fact that the expression does not rely on the way to specify the quantum number α. Further, the bilinear form in α such as $\Sigma_\alpha A^{(+)}(n\alpha)A^{(+)}(m\alpha)$ always implies $m=n$, which allows us to regard α of $(n\alpha)$ as being different from β of $(m\beta)$ in case $m \neq n$. Under these situations, we safely adopt the convention to abbreviate the set of $(n\alpha)$ as n. Each of n, m, etc., seen in the sign factor stands uniquely for the number of particles.

Lemma:

$$A(n_1)A(n_2)\cdots A(n_p)A^+(m)$$
$$= (-1)^Y \sum \langle n_1 | A^+(s_1) | l_1 \rangle \langle n_2 | A^+(s_2) | l_2 \rangle \cdots$$
$$\times \langle n_p | A^+(s_p) | l_p \rangle \langle q | A(s_1)A(s_2)...A(s_p) | m \rangle$$
$$\times A^+(q)A(l_1)A(l_2)...A(l_p), \tag{33}$$

where

$$l_i = n_i - s_i, \quad \text{for } i = 1, 2, ..., p \tag{34}$$

and

$$q = m - \sum_{i=1}^p s_i. \tag{35}$$

The sum in (33) is taken over $\{s_i\}$. The sign factor Y, characteristic of fermions, is given by

$$Y = \sum_{i=1}^p l_i (m - \sum_{j=i}^p s_j). \tag{36}$$

Proof: We prove it by induction on p. The expression (33) together with (36) for $p=1$ reads (14). Supposing (33) for $p=t$, let us derive it for $p=t+1$. We replace n_i, s_i, l_i (for any possible i) and q in (33)–(36) for $p=t$ by new notations $n_{i+1}, s_{i+1}, l_{i+1}$, and q', respectively. Both hand sides of (33) are then multiplied by $A(n_1)$ from the left. Using (14), we reorder $A(n_1)A^+(q')$ involved in the rhs as

$$A(n_1)A^+(q') = \sum_{s_1} \langle n_1 | A^+(s_1) | l_1 \rangle \langle q | A(s_1) | q' \rangle$$
$$\times A^+(q)A(l_1)(-1)^{ql_1}. \tag{37}$$

After summing over $\alpha_{q'}$ of $(q'\alpha_{q'})$, we obtain (33) together with (36) for $p=t+1$.

We present the extended Wick's theorem which is applied to the operator string (32) without recourse to decomposition into C^+'s and C's.

Theorem 4:

$$A(n_1)A^+(m_1)A(n_2)A^+(m_2)\cdots A(n_p)A^+(m_p)$$
$$= (-1)^F \sum \langle n_1 | A^+(r_{11})A^+(r_{21})\cdots A^+(r_{p1}) | k_1 \rangle \langle n_2 | A^+(r_{22})A^+(r_{32})\cdots A^+(r_{p2}) | k_2 \rangle \cdots$$
$$\times \langle n_p | A^+(r_{pp}) | k_p \rangle \langle q_1 | A(r_{11}) | m_1 \rangle \langle q_2 | A(r_{21})A(r_{22}) | m_2 \rangle \cdots$$
$$\times \langle q_p | A(r_{p1})A(r_{p2})\cdots A(r_{pp}) | m_p \rangle A^+(q_1)A^+(q_2)\cdots A^+(q_p)A(k_1)A(k_2)\cdots A(k_p), \tag{38}$$

where

$$q_i = m_i - \sum_{j=1}^i r_{ij} \tag{39}$$

and

$$k_i = n_i - \sum_{j=i}^p r_{ji}. \tag{40}$$

The sum in (38) is taken over $\{r_{ij}\}$. The sign factor F, characteristic of fermions, is given by

$$F = \sum_{i=1}^p \sum_{j=1}^i \left(m_i - \sum_{u=j}^i r_{iu}\right)\left(n_j - \sum_{v=j}^i r_{vj}\right). \tag{41}$$

Proof: We use induction on p. The expression (38) together with (41) for $p=1$ reads (14). Supposing (38) for $p=t$, let us derive it for $p=t+1$. On both hand sides of (38) for $p=t$, we multiply $A(n_{t+1})A^+(m_{t+1})$ from the right. Subsequently, we reorder the following operator which is a part of the operator string on the rhs:

$$A(k_1)A(k_2)\cdots A(k_p)A(n_{t+1})A^+(m_{t+1}). \tag{42}$$

To this end, the relation (33) for $p=t+1$ is used with s_i ($i=1,2,...t+1$) and n_i ($i=1,2,...t$) in it being replaced by new notations $r_{t+1,i}$ and k_i, respectively. The sign factor that arises from reordering (42) is shown to be $(F)_{p=t+1} - (F)_{p=t}$, by using (36). The expression (38) for $p=t$, multiplied by $A(n_{t+1})A^+(m_{t+1})$, is then transformed into (38) for $p=t+1$ with the sign factor (41) for $p=t+1$.

By using (39) and (40), the sign factor (41) is rewritten in terms of $\{q_i, k_i\}$ as

$$F=\sum_{j<i}^{p} r_{ij}\sum_{u=j}^{i-1}(q_u+k_{u+1})+\sum_{j<i}^{p} q_i k_j+\sum r_{ij} r_{uv}, \tag{43}$$

where the last sum on the rhs is taken over i,j, u, and v under the condition

$$i\leqslant v<j\leqslant u<i\leqslant p. \tag{44}$$

We interpret (38) together with (43) in the following way. In each term on the rhs of (38), r_{ij} times contractions between $A(n_j)$ and $A^+(m_i)$ are done for any pair of i and j with $j<i$. The contractions require decomposition of the operators as

$$A(n_j)\to A(k_j)A(r_{pj})A(r_{p-1,j})\cdots A(r_{jj}) \times\langle n_j|A^+(r_{jj})\cdots A^+(r_{p-1,j})A^+(r_{pj})|k_j\rangle \tag{45}$$

and

$$A^+(m_i)\to A^+(r_{ii})A^+(r_{i,i-1})\cdots A^+(r_{i1})A^+(q_i) \times\langle q_i|A(r_{i1})\cdots A(r_{i,i-1})A(r_{ii})|m_i\rangle, \tag{46}$$

respectively. Ordering of A^+'s and A's on each rhs is along (11). The matrix elements on the rhs of (38) are attributable to those in (45) and (46). The operator string on the lhs of (38), after the decompositions (45) and (46), is reordered, with contraction being neglected, as

$$\{A(k_1)A^+(q_1)\cdots A(k_p)A^+(q_p)\}\{A(r_{p1}) \times A(r_{p-1,1})\cdots A(r_{11})\cdots A^+(r_{pp})A^+(r_{p,p-1})\cdots A^+(r_{p1})\}. \tag{47}$$

It implies that the operators $A(k_1)$, etc., which are free from contraction, are gathered to the lfs. Reordering into (47) yields the sign factor, which is equal to the first term on the rhs of (43). The operator string $A(k_1)A^+(q_1)\cdots A^+(q_p)$ of (47) is rearranged in the same order as seen on the rhs of (38), yielding the sign factor equal to the second term of (43). Subsequently, we bring the pairs $\{A(r_{ij})$ and $A^+(r_{ij})\}$ to be connected next to each other without affecting relative ordering between them. The pairwise arrangement for full contractions gives rise to the sign factor given by the last term of (43).

VIII. AN EXTENDED CONTRACTION THEOREM

The contraction theorem is applied[2] to the commutator of a pair of strings B and D made of C^+'s and C's. The theorem is summarized as[2]

$$[B,D]_\pm=-\underbrace{BD}-\underbrace{BD}-\cdots, \tag{48}$$

where the symbol $\sqcup$ indicates a possible single contraction, one from the string B and the other from D.

Let us extend (48) so that operators appearing in the theorem could be written in terms of state operators. Corresponding to B in (48), we consider the operator string

$$B_p=A^+(m_1)A(n_1)A^+(m_2)A(n_2)\cdots A^+(m_p)A(n_p), \tag{49}$$

which is a general form of the string made of state operators as O_p of (32) is. To simplify the discussion, the operator corresponding to D in (48) is restricted to the form A^+A. This restriction, which is easily removed, does not matter for practical purposes cited in Ref. 2.

Theorem 5:

$$\begin{aligned}
&[B_p,A^+(M)A(N)]_\pm\\
&=-\sum{}'\langle 0|A(k_1)A(k_2)\cdots A(k_p)A(M-K_p)|M\rangle\langle N|A^+(N-L_p)A^+(l_p)\cdots A^+(l_2)A^+(l_1)|0\rangle\\
&\quad\times\langle m_1-l_1|A(l_1)|m_1\rangle\langle n_1|A^+(k_1)|n_1-k_1\rangle\langle m_2-l_2|A(l_2)|m_2\rangle\langle n_2|A^+(k_2)|n_2-k_2\rangle\times\cdots\\
&\quad\times\langle m_p-l_p|A(l_p)|m_p\rangle\langle n_p|A^+(k_p)|n_p-k_p\rangle A^+(m_1-l_1)A(n_1-k_1)A^+(m_2-l_2)A(n_2-k_2)\cdots A^+(m_p-l_p)\\
&\quad\times A(n_p-k_p)A^+(M-K_p)A(N-L_p)(-1)^{G_p},
\end{aligned} \tag{50}$$

where

$$K_p=\sum_{i=1}^{p}k_i, \tag{51}$$

$$L_p=\sum_{i=1}^{p}l_i. \tag{52}$$

The sign factor G_p in fermion and boson systems are given by

$$G_p=M(K_p+L_p)+\sum_{i=2}^{p}(m_i-l_i+n_i-k_i) \times\sum_{j=2}^{i}k_{j-1}+\sum_{i=1}^{p}(m_i+n_i)\sum_{j=1}^{i}l_j, \tag{53}$$

and $G_p=(-1)^k$, respectively. The sum Σ' on the rhs of (50) is taken over $k_1,l_1,k_2,l_2,\ldots,k_p,l_p$ except vanishing of these arguments at the same time.

Proof: We use induction on p. The relation (50) for $p=1$ reads

$$\begin{aligned}
&[A^+(m_1)A(n_1),A^+(M)A(N)]_\pm\\
&=\sum_{kl}{}'\langle k|A(M-k)|M\rangle\langle N|A^+(N-l)|l\rangle\\
&\quad\times\langle m_1-l|A(l)|m_1\rangle\langle n_1|A^+(k)|n_1-k\rangle\\
&\quad\times A^+(m_1-l)A(n_1-k)A^+(M-k)A(N-l)\\
&\quad\times(-1)^{M(k+l)+(m_1+n_1)l},
\end{aligned} \tag{54}$$

which is easily proved by using (25)–(28): In a boson system, the sign on the rhs is replaced by $(-1)^k$. Let us assume (50) for $p=q$ and prove the same relation for $p=q+1$. Using (25), we transform the expression on the lhs of (50) for $p=q+1$ as

$$\begin{aligned}
&[B_{q+1},A^+(M)A(N)]_\pm\\
&=B_q[A^+(m_{q+1})A(n_{q+1}),A^+(M)A(N)]_\pm\\
&\quad+[B_q,A^+(M)A(N)]_\pm A^+(m_{q+1})A(n_{q+1})\\
&\quad\times(-1)^{(m_{q+1}+n_{q+1})(M+N)}.
\end{aligned} \tag{55}$$

Let us apply (54) and (50) for $p=q$ to the first and the second terms on the rhs, respectively. Rearranging the second term yields the operating string

$$A^+(m_1 - l_1)A(n_1 - k_1)\cdots A(n_q - k_q)A^+(M - K_q) \times A(N - L_q)A^+(m_{q+1})A(n_{q+1}). \tag{56}$$

Its last four A^+'s and A's are rearranged as

$$[A^+(m_{q+1})A(n_{q+1}),\ A^+(M-K_q)A(N-L_q)]_\pm - A^+(m_{q+1})A(n_{q+1})A^+(M-K_q)A(N-L_q) \times(-1)^{(m_{q+1}+n_{q+1})(M-K_q+N-L_q)+1}, \tag{57}$$

which is rewritten again by using (54). After these modifications of the rhs of (55), we obtain (50) for $p = q + 1$.

Note that operators A^+'s and A's on the rhs of (50) are arranged in the same order as those of B_p given by (49). In case normal ordering of them is required, the extended Wick's theorem (38) is subsequently applied.

[1]G. C. Wick, Phys. Rev. **80**, 268 (1950).
[2]F. E. Harris, B. Jeziorski, and H. J. Monkhorst, Phys. Rev. A **23**, 1632 (1981).
[3]G. Racah, Phys. Rev. **63**, 367 (1943).
[4]M. Nomura, Prog. Theor. Phys. **59**, 451 (1978).
[5]M. H. Macfarlane, in *Lectures in Theoretical Physics*, edited by P. D. Kunz *et al.* (Univ. Colorado, Boulder, 1966).
[6]M. Nomura, submitted to J. Math. Phys.
[7]J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon, Oxford, 1965).

A reduction relation for an n-body trace of an operator product consisting of coupled C^+'s and C's

Masao Nomura
Institute of Physics, College of Arts and Sciences, University of Tokyo, Tokyo, 153, Japan

(Received 11 October 1984; accepted for publication 7 December 1984)

A reduction relation for the trace of an operator product consisting of C^+'s and C's is derived in the truncated space of n fermions or n bosons being distributed over definite orbits. The state operator formalism (derived in a preceding paper) is fully used so as to perform normal ordering (or contraction) of coupled C^+'s and C's in the operator string with the least destruction of the preexisting coupling.

I. INTRODUCTION

In the analysis of nuclear spectra, many works[1–7] have been devoted to the trace of an operator in such a truncated space as n particles being distributed over definite orbits.

The n-body trace of an operator is related to few- or several-body traces of the same kind.[1] However, this property, called propagation of the operator average, is not sufficient to reduce the n-body trace in the form of the possibly simplest matrix elements. In Refs. 2 and 3, the n-body trace of the fourth power of a two-body interaction was expressed as a sum of quartic products of interaction matrix elements by using new relations among cfp's (coefficients of fractional parentage),[2] while propagation of the average leaves three- and four-body traces of the operator as parameters.

Ginocchio[4,5] extended the above consideration to an n-body trace of a product operator. An advantage to his prescription is that it can be performed without reordering the operator in normal products. However, his prescription is insufficient at least in two points, which still subsist in its reformulation by Chang and Wong.[6] First, the trace of any boson operator cannot be treated. The fully occupied state, characteristic of a fermion system, plays a crucial role in Ref. 6, which cannot be applied to a boson system. Second, contraction of operators in Refs. 4–6 forces the decomposition of any coupled form of creation operators C^+'s and of annihilation operators C's into uncoupled forms. After contraction of operators, resultant uncoupled C^+'s and C's are to be coupled again. These are much entangled because of n-j symbols, Pauli principle among particles, etc.

In this paper, we present a reduction relation for the many-body trace of an operator product that consists of coupled C^+'s and C's, removing the above-stated defects of Ginocchio's prescription. To derive it, we make use of the state operator formalism in a preceding paper[8] where composition among coupled C^+'s (or C's) is incorporated into normal ordering among coupled C^+'s and C's. Since this preceding work is based merely on the usual (anti) commutation rules of C^+ and C, the resultant reduction relation is applied to fermions and bosons completely alike. Normal ordering (or contraction) of coupled C^+'s and C's in the operator string costs the least destruction of the preexisting coupling, and is described in terms of matrix elements.

A diagrammatic representation is available for the reduction relation A line in the present diagram implies not a single C^+ or C but a cluster of coupled C^+'s or C's.

The reduction relation makes it easy to evaluate moments of energy spectra in a many-body system. Its application to the analysis of nuclear spectra involving bosonlike excitation is in progress.

In Sec. II, the state operator formalism is summarized together with notation conventions. Section III is devoted to a theorem presenting a reduction relation for an n-body trace. In Sec. IV, the reduction relation is illustrated by means of a diagrammatic representation. In Sec. V, the reduction relation is transformed into other forms. In Sec. VI, a few remarks on trace evaluation are given.

II. THE STATE OPERATOR FORMALISM

Here is a summary of the state operator formalism[8] together with notation conventions.

Let $\{|n\alpha\rangle\}$ be an orthonormal and complete set of the n-fermion or n-boson wave functions spanning a truncated space. As an operator consisting of C^+'s of the order n, the state operator $A^+(n\alpha)$ is introduced that, acting on the vacuum, creates the wave function $|n\alpha\rangle$. We postulate that $A^+(n=0)=1$. The conjugate of $A^+(n\alpha)$ is denoted as $A(n\alpha)$.

Any expression in the present work does not rely on the way to specify the n-body orthonormal bases $\{|n\alpha\rangle\}$. As a consequence, each Greek index subject to summation appears only in its bilinear form. Further, a form such as $\Sigma_\alpha A^+(m\alpha)A^+(n\alpha)$ never appears if $m\neq n$. The quantum number α of $(n\alpha)$ is then treated as being different from β of $(m\beta)$ in the case $m\neq n$. Under these situations, we safely abbreviate the set $(n\alpha)$ as n, and do without any explicit Greek index in many of the expressions. It remarkably condenses notations. Each of n, m, etc., appearing in sign factors and binomial coefficients, stands uniquely for the number of particles.

State operators satisfy the composition rule[8]

$$A^+(n)A^+(m)=\sum\langle n'|A^+(n)|m\rangle A^+(n'), \tag{1}$$

where $n'=m+n$ due to conservation of the number of particles in the matrix element. It is short for

$$A^+(n\alpha)A^+(m\beta)=\sum_\gamma\langle n'\gamma|A^+(n\alpha)|m\beta\rangle A^+(n'\gamma). \tag{1'}$$

A pair of boson operators $A^+(n)$ and $A^+(m)$ commute with

 0022-2488/85/040738-05$02.50

each other, while commuting the fermion operators $A^+(n)$ and $A^+(m)$ yields the sign $(-1)^{mn}$. Reordering a pair of fermion operators $A(n)$ and $A^+(m)$ in normal products is expressed as[8]

$$A(n)A^+(m)=\sum_k\langle n|A^+(k)|n'\rangle\langle m'|A(k)|m\rangle \times A^+(m')A(n')(-1)^{m'n'}, \quad (2)$$

where $n'=n-k$ and $m'=m-k$ in the same sense as seen in (1). So far as the present work is concerned, every expression for a fermion system is formally translated into a boson system if any sign factor involved in the former is deleted in the latter. (This is not valid in case the expression concerns antinormal ordering which is not involved in the present work.) For example, the relation (2) is applied to a boson system if the sign $(-1)^{m'n'}$ in it is deleted, though allowed values of n and α are very different between fermion and boson systems.

We adopt the conventions

$$\sum_{i=a}^{b}=0 \text{ and } \prod_{i=a}^{b}=1, \quad \text{for } a>b. \quad (3)$$

Wick's theorem is extended to a string of state operators as[8]

$$\prod_{i=1}^{p}A(n_i)A^+(m_i)$$
$$\equiv A(n_1)A^+(m_1)\cdots A(n_p)A^+(m_p) \quad (4)$$
$$=(-1)^F\sum\left\{\prod_{i=1}^{p}\left\langle q_i\left|\prod_{s=1}^{i}A(r_{is})\right|m_i\right\rangle A^+(q_i)\right\}$$
$$\times\left\{\prod_{j=1}^{p}\left\langle n_j\left|\prod_{t=j}^{p}A^+(r_{tj})\right|k_j\right\rangle A(k_j)\right\}, \quad (5)$$

where the sum is taken over $\{r_{ij}\}$. The sign factor F, characteristic of a fermion system, is given by

$$F=\sum_{i=1}^{p}\sum_{j=1}^{i}\left(m_i-\sum_{s=j}^{i}r_{is}\right)\left(n_j-\sum_{t=j}^{i}r_{tj}\right). \quad (6)$$

The relation (5) for $p=1$ reads (2). In case any of $\{m_i\}$ and $\{n_i\}$ is either 0 or 1, the relation (5) reduces to the prevalent Wick's theorem of time-independent form. Matrix elements of $A(r_{ij})$ and of $A^+(r_{ij})$ imply r_{ij} times contractions between $A^+(m_i)$ and $A(n_j)$.

A basic relation for state operators is

$$\sum_\alpha A^+(n\alpha)A(n\alpha)=\binom{\vec{n}}{n}, \quad (7)$$

where $\vec{n}$ indicates the number operator. By using (1) and (7), it is shown that

$$\sum_{\alpha\beta}A^+(n\alpha)A^+(m\beta)\text{---}A(m\beta)A(n\alpha)$$
$$=\binom{m+n}{n}\sum_\gamma A^+(m+n\,\gamma)\text{---}A(m+n\,\gamma), \quad (8)$$

where there can be any operator at ---.

The number of single particle orbits is denoted as N: In the case of a spherical orbit j of identical fermions, $N=2j+1$. We define M by

$$M=N \text{ for fermions and } M=-N \text{ for bosons.} \quad (9)$$

The binomial coefficient in M for bosons implies, for example,

$$\binom{M-a}{b}=\binom{-N-a}{b}=\binom{N+a+b-1}{b}(-1)^b.$$

Use of M in place of N unifies, as suggested by (4.12) of Ref. 7, an expression for fermions and the corresponding one for bosons.

III. A REDUCTION RELATION FOR MANY-BODY TRACE

An operator of interest in the trace evaluation is a power of the Hamiltonian which may contain, in general, many-body interactions among different kinds of fermions and bosons. We rearrange the power of the Hamiltonian as a sum of operator products separable for each kind of particles. The operator string of our interest is then of the following form of identical particles:

$$(C^+C^+\text{---}C^+)(CC\text{---}C)\text{---}(C^+C^+\text{---}C^+)\text{---}. \quad (10)$$

We further rewrite it in terms of the string of state operators

$$O_p=\prod_{i=1}^{p}A^+(m_i)A(n_i)=A^+(m_1)A(n_1)\cdots A^+(m_p)A(n_p), \quad (11)$$

where any of $\{m_i\}$ and $\{n_i\}$ can be zero. We consider only the number conserving part of O_p because the other part does not affect the trace. The maximum particle rank of O_p is denoted as u_p:

$$u_p=\sum_{i=1}^{p}n_i=\sum_{i=1}^{p}m_i. \quad (12)$$

The uncoupled representation (e.g., the m scheme) is used to arrange state operators involved in O_p, while each of the state operators is written in the framework of a coupled representation (e.g., the J scheme).

The purpose in this section is to get a reduction relation for the n-body trace of O_p. For $p=1$, the result is simply written as

$$\sum_\alpha\langle n\alpha|A^+(m\beta)A(m\gamma)|n\alpha\rangle=\delta(\beta,\gamma)\binom{M-m}{n-m}. \quad (13)$$

The results for fermion and boson systems were discussed in Refs. 1 and 7, respectively. Extension of (13) to a general O_p is summarized in the following theorem.

Theorem:

$$\sum_\alpha\langle n\alpha|O_p|n\alpha\rangle=\sum\left|\binom{M-u_p}{n-u_p+R_p}\right.$$
$$\left.\times\left\{\prod_{i=1}^{p}\langle n_i|I_p(i,rs)|m_i\rangle\right\}(-1)^{Y(p,rs)}, \right. \quad (14)$$

where I_p is the operator string defined by

$$I_p(i,rs)=\prod_{j=i+1}^{p}A^+(r_{ji})\prod_{k=1}^{i-1}A^+(s_{ik})$$
$$\times\prod_{l=i+1}^{p}A(s_{li})\prod_{t=1}^{i-1}A(r_{it}). \quad (15)$$

Indices i and j of $r_{ij}(s_{ij})$ are defined under $1\leqslant j<i\leqslant p$. The sum of $\{r_{ij}\}$ is denoted as R_p:

$$R_p=\sum_{i>j}r_{ij}. \quad (16)$$

The sign factor $Y(p,rs)$, characteristic of fermions, is given by

$$Y(p,rs)=\sum_{q=3}^{p}\sum_{i=2}^{q-1}\sum_{m=2}^{i}\sum_{k=1}^{m-1}\{s_{qm}(s_{ik}+r_{i+1,k})$$
$$+r_{q,m-1}(s_{ik}+r_{i,k-1})\}. \qquad (17)$$

The sum on the right-hand side (rhs) of (14) is taken over $\{r_{ij},s_{ij};i>j\}$ under the restriction

$$e_i=n_i-\sum_{j=i+1}^{p}r_{ji}-\sum_{k=1}^{i-1}s_{ik} \qquad (18)$$

$$=m_i-\sum_{l=i+1}^{p}s_{li}-\sum_{t=1}^{i-1}r_{it}\geqslant 0, \qquad (18')$$

which expresses conservation of the number of particles in the matrix element of I_p between $\langle n_i|$ and $|m_i\rangle$.

An example: The expressions (14) and (17) for $p=3$ read

$$\sum_{\alpha}\left\langle n\alpha\left|\prod_{i=1}^{3}A^{+}(m_i)A(n_i)\right|n\alpha\right\rangle$$
$$=\sum\left|\binom{M-u_3}{n-u_3+R_3}\right|$$
$$\times\langle n_1|A^{+}(r_{21})A^{+}(r_{31})A(s_{21})A(s_{31})|m_1\rangle$$
$$\times\langle n_2|A^{+}(r_{32})A^{+}(s_{21})A(s_{32})A(r_{21})|m_2\rangle$$
$$\times\langle n_3|A^{+}(s_{31})A^{+}(s_{32})A(r_{31})A(r_{32})|m_3\rangle(-1)^{Y(3,rs)} \qquad (19)$$

and

$$Y(3,rs)=s_{21}s_{32}+s_{32}r_{31}+r_{31}s_{21}, \qquad (20)$$

respectively. Notice that the expression (19) involves (13).

Proof: We use induction on p. The case $p=1$ reads (13). Supposing (14) for $p=q$, let us prove it for $p=q+1$. By using (1) and (2), the operator O_{p+1} is rearranged as

$$O_{p+1}=O_{q-1}A^{+}(m_q)A(n_q)A^{+}(m_{q+1})A(n_{q+1})$$
$$=O_{q-1}\sum_{k}A^{+}(m_q')A(n_q')$$
$$\times\langle m_q'|A^{+}(m_{q+1}-k)|m_q\rangle$$
$$\times\langle n_q|A^{+}(k)A(n_{q+1})|n_q'\rangle$$
$$\times\langle m_{q+1}-k|A(k)|m_{q+1}\rangle(-1)^{X_1}, \qquad (21)$$

where the sign factor X_1, characteristic of fermions, is given by

$$X_1=(m_{q+1}-k)(m_q+n_q-k). \qquad (22)$$

The operator O_{q-1} for $q=1$ reads 1. We notice that $m_q'=m_q+m_{q+1}-k$, etc. The rhs is a sum of operators, each of which has the same form as O_q with the maximum particle rank being $u_{q+1}-k$. We apply (14) for $p=q$ to the trace of O_{q+1} written in the form of (21). It then follows that

$$\sum_{\alpha}\langle n\alpha|O_{q+1}|n\alpha\rangle$$
$$=\sum\left\{\sum_{k}\left|\binom{M-u_{q+1}+k}{n-u_{q+1}+k+R_q}\right|G_k\right\}(-1)^{Y(q,rs)}$$
$$\times\prod_{i=1}^{q-1}\langle n_i|I_q(i,rs)|m_i\rangle, \qquad (23)$$

where

$$G_k=(-1)^{X_1}\sum_{\beta}\langle m_{q+1}-k|A(k\beta)|m_{q+1}\rangle$$
$$\times\langle n_q|A^{+}(k\beta)I'|m_q\rangle \qquad (24)$$

and

$$I'=A(n_{q+1})\left\{\prod_{i=1}^{q-1}A^{+}(s_{qi})\prod_{j=1}^{q-1}A(r_{qj})\right\}A^{+}(m_{q+1}-k). \qquad (25)$$

We rearrange I' in normal products using the extended Wick's theorem (5), and put the resultant expression into (24). Then,

$$G_k=\sum\left\{\prod_{i=1}^{q-1}\langle s_{qi}'|A(w_i)|s_{qi}\rangle\langle r_{qi}|A^{+}(r_{qi}')|t_i\rangle\right\}$$
$$\times\sum_{\beta\gamma}\left\langle n_q\left|A^{+}(k\beta)A^{+}(t_0-k\gamma)\right.\right.$$
$$\times\left\{\prod_{j=1}^{q-1}A^{+}(s_{qj}')\right\}A(w_0)\left\{\prod_{l=1}^{q-1}A(r_{ql}')\right\}\left.\left|m_q\right.\right\rangle$$
$$\times\left\langle n_{q+1}\left|\left\{\prod_{j=1}^{q-1}A^{+}(w_j)\right\}A^{+}(w_0)\right.\right.$$
$$\times\left\{\prod_{l=1}^{q-1}A(t_l)\right\}A(t_0-k\gamma)A(k\beta)\left.\left|m_{q+1}\right.\right\rangle$$
$$\times(-1)^{t_0-k+X_2}, \qquad (26)$$

where conservations of the number of particles in the last two matrix elements give the same restriction expressed as

$$m_{q+1}-n_{q+1}+w_0-t_0+\sum_{i=1}^{q-1}(w_i-t_i)=0. \qquad (27)$$

Latin indices that are independently summed over on the rhs of (26) are t_0 and $\{s_{qi}',r_{qi}',t_i,w_i;\ i=1,2,\ldots,q-1\}$. The sign on the rhs of (26) is to be deleted for a boson system. The factor X_2 in it is given by

$$X_2=\sum_{i=2}^{q-1}w_i\sum_{j=1}^{i-1}s_{qj}'+\sum_{i=2}^{q-1}t_i\sum_{j=1}^{i-1}r_{qj}'+w_0\sum_{i=1}^{q-1}t_i$$
$$+\left(w_0+\sum_{i=1}^{q-1}t_i\right)\sum_{j=1}^{q-1}(w_j+s_{qj}). \qquad (28)$$

Summing over β and γ in (26) by virtue of (8), we put the expression of G_k into (23). Subsequently, Greek indices attached to $\{s_{qi},r_{qi};i=1,2,\ldots,q-1\}$ are summed over by using (1) such that

$$\sum\langle s_{qi}'|A(w_i)|s_{qi}\rangle A(s_{qi})=A(s_{qi}')A(w_i), \qquad (29)$$

where the matrix element and $A(s_{qi})$ are contained in G_k and $I_q(i,rs)$, respectively. Let us change notations in (23) such that

$$r_{qi}'\to r_{qi},\qquad s_{qi}'\to s_{qi},\qquad t_i\to r_{q+1,i},$$
$$w_i\to s_{q+1,i},\qquad t_0\to r_{q+1,q},\qquad w_0\to s_{q+1,q}. \qquad (30)$$

According to (30), arguments $r_{qi}(=r_{qi}'+t_i)$ and s_{qi} appearing in $Y(q,rs)$ are replaced by $r_{qi}+r_{q+1,i}$ and $s_{qi}+s_{q+1,i}$, respectively. Similarly, the argument R_q of the binomial coefficient is replaced by $R_{q+1}-r_{q+1,q}$. The expression (23) is now of the form

$$\sum_{\alpha} \langle n\alpha | O_{q+1} | n\alpha \rangle$$

$$= \sum \left\{ \sum_k \left| \begin{pmatrix} M - u_{q+1} + k \\ n - u_{q+1} + k + R_{q+1} - r_{q+1,q} \end{pmatrix} \right| \times \begin{pmatrix} r_{q+1,q} \\ k \end{pmatrix} (-1)^{k + r_{q+1,q}} \right\} \times \left\{ \prod_{i=1}^{q+1} \langle n_i | I_{q+1}(i,rs) | m_i \rangle \right\} (-1)^{X_3}, \tag{31}$$

where the signs are characteristic of a fermion system. The sign factor X_3 is given by $X_2 + Y(q,rs)$ being rewritten in terms of new notations (30) and is shown to agree with $Y(q+1,rs)$. The sum over k in (31) is done by using, for a boson system,

$$\sum_k \begin{pmatrix} a \\ b+k \end{pmatrix} \begin{pmatrix} c \\ k \end{pmatrix} = \begin{pmatrix} a+c \\ b+c \end{pmatrix}, \tag{32}$$

and, for a fermion system,

$$\sum_k \begin{pmatrix} a+k \\ b+k \end{pmatrix} \begin{pmatrix} c \\ k \end{pmatrix} (-1)^{c+k} = \begin{pmatrix} a \\ b+c \end{pmatrix}, \tag{33}$$

which follows from a Vandermonde convolution (32) with a in it being replaced by $-a+b-1$. The expression (31) is then written as (14) for $p = q+1$. Q.E.D.

Another proof of (14) is by means of a deductive one. It is summarized as follows. We start from reordering O_p in normal products, using Wick's theorem (5). Next, the relation (13) is applied to the normal products so that the trace of O_p could be expressed in terms of the k-body trace of a k-body operator, where $k = 0,1,\ldots,u_p$. We transform each of the traces as

$$\sum_{\alpha} \left\langle k\alpha \left| \prod_i A^+(r_i) \prod_j A(s_j) \right| k\alpha \right\rangle = \left\langle 0 \left| \prod_j A(s_j) \prod_i A^+(r_i) \right| 0 \right\rangle. \tag{34}$$

Wick's theorem (5) is used again for normal ordering of the operator string in the rhs matrix element. After somewhat lengthy rearrangement of the resultant expression, we obtain (14).

IV. A DIAGRAMMATIC REPRESENTATION OF THEOREM (14)

Here, implication of Theorem (14) is stated by means of a diagrammatic representation.

In Fig. 1 is shown the diagram for (19), i.e., (14) with $p = 3$, as an example. On each vertex of a hexagon, we set up the index of each state operator involved in $O_{p=3}$. We then draw, with the same index, an external line directed inward to or outward from the vertex according to whether the operator is A^+ or A. The polygon reflects invariance of the trace under any cyclic permutation of state operators. An internal line is drawn from the vertex m_i to n_j for any possible pair of i and j. It implies contraction between $A^+(m_i)$ and $A(n_j)$. Indices s_{ji}, e_i, and r_{ij} are assigned to the lines in the cases $i<j$, $i=j$, and $i>j$, respectively. The number of particles is conserved at each vertex, as given by (18). The arrangement of

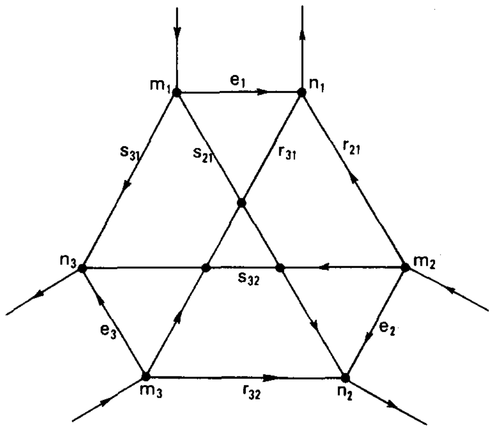


FIG. 1. The diagrammatic representation of (19), the reduction relation for the trace of an operator product $O_{p=3}$.

state operators of $I_p(i,rs)$, defined by (15), is along a clockwise ordering of line indices starting from the external line n_i and ending at the line m_i. The set of nodes inside the polygon represents the sign factor $Y(p,rs)$. There are three nodes in the hexagon of Fig. 1. These arise from the intersection of lines (s_{21},s_{32}), (s_{32},r_{31}), and (r_{31},s_{21}), which leads to (20). For the case of $O_{p=4}$, there will appear 16 nodes inside an octagon. These reproduce correctly the sign factor (17) for $p = 4$.

The present diagram is incorporated into the graphical method for angular momentum.[9] Let us consider, for example, the trace of the third power of an interaction. It is illustrated in the diagram of Fig. 1 with the external lines m_i and n_i being joined for each i, as the interaction is scalar. The diagram in the case of $e_i = 0$ for all i's has the same form as the graph for a 9-j symbol. We have checked that the diagram actually represents the term[2,3] $(\bar{V})^3$ which is a 9-j symbol multiplied by cubic products of interaction matrix elements.

There is another interpretation of the sign factor (17). Let D be the following operator defined in (15):

$$D = \prod_{k=1}^{p} I_p(p-k+1,rs). \tag{35}$$

In the diagrammatic representation, the operator D is expressed as a counterclockwise ordering of $\{r_{ij}\}$ and $\{s_{ij}\}$ around the center of the polygon. We bring the pairs $\{A^+(r_{ij}), A(r_{ij})\}$ and $\{A^+(s_{ij}), A(s_{ij})\}$ so as to be connected next to each other without affecting relative ordering between pairs. This is the same manipulation demanded in the prevalent Wick's theorem to obtain the proper sign. We obtain (17) by expressing full contraction of D in terms of the pairwise arrangement.

Summing over i on both sides of (18) yields

$$E_p \equiv \sum_i e_i = u_p - R_p - S_p, \tag{36}$$

where

$$S_p = \sum_{i>j} s_{ij}. \tag{37}$$

Comparing (14) with the result by Ginocchio,[5] we see that R_p, E_p, and S_p in the former correspond to the numbers of right, self, and left contractions in the latter, respectively. The argument t in Ref. 5 implies $u_p - R_p$ here.

V. MODIFICATION OF THEOREM (14)

We can rewrite (14) as

$$\sum_\alpha \left\langle n\alpha \left| \prod_{i=1}^{p} A(n_i) A^+(m_i) \right| n\alpha \right\rangle = \sum \left| \binom{M-u_p}{n-S_p} \right| \left\{ \prod_{i=1}^{p} \langle n_i | I_p(i,rs) | m_i \rangle \right\} \times (-1)^{Y(p,sr) + \Sigma_i e_i (m_i - n_i)}. \tag{38}$$

The sign factor $Y(p,sr)$ is given by (17) with r_{ij} and s_{ij} in it being interchanged. It is easily shown that

$$Y(p,rs) + Y(p,sr) = \sum_{i=1}^{p} (n_i - e_i)(m_i - e_i) + \text{an even number}. \tag{39}$$

The relation (38) is deduced from (14) for O_{p+1} with $A^+(m_1 = 0)$ and $A(n_{p+1} = 0)$. The product operators $\{I_p(i,rs)\}$, defined by (15), have a symmetry under interchange of r_{ij} and s_{ij}, as easily seen in (19), such that

$$\prod_{i=1}^{p} \langle n_i | I_p(i,rs) | m_i \rangle (-1)^{Y(p,rs)} = \prod_{i=1}^{p} \langle m_i | I_p(i,sr) | n_i \rangle^* (-1)^{Y(p,sr) + Z}, \tag{40}$$

where the sign factor Z, characteristic of fermions, is given by

$$Z = R_p + S_p + \sum_i \frac{(m_i - n_i)^2}{2}. \tag{41}$$

To get (41), we have used (18) and (39). Let us apply (40) to each term on the rhs of (14) in case R_p in it is larger than $[u_p/2]$, the largest integer contained in $u_p/2$. Then, we obtain

$$\sum_\alpha \langle n\alpha | O_p | n\alpha \rangle = \sum \left\{ \left| \binom{M-u_p}{n-u_p+R_p} \right| \prod_{i=1}^{p} \langle n_i | I_p(i,rs) | m_i \rangle + (-1)^Z \left| \binom{M-u_p}{n-u_p+S_p} \right| \prod_{i=1}^{p} \langle m_i | I_p(i,rs) | n_i \rangle^* \right\} \times (-1)^{Y(p,rs)} [1 + \delta(R_p, [u_p/2])]^{-1}. \tag{42}$$

The sum over R_p on the rhs is restricted to $R_p \leqslant [u_p/2]$, while in (14) $R_p < u_p$. This result is easily checked for the case of $O_{p=3}$ by using (19).

Ginocchio[5] got a relation similar to (42). However, it is applicable only to a fermion system and is restricted to the operator given by a product of number-conserving operators.

VI. REMARKS

A few remarks are given on using the reduction relation (14) or (38).

(I) The trace for $n = 0$ gives the vacuum expectation value of the operator:

$$\left\langle 0 \left| \prod_{i=1}^{p} A(n_i) A^+(m_i) \right| 0 \right\rangle = (-1)^{F'} \prod_{i=0}^{p} \left\langle 0 \left| \prod_{j=1}^{i} A(r_{ij}) \right| m_i \right\rangle \times \prod_{k=1}^{p} \left\langle n_k \left| \prod_{l=k}^{p} A^+(r_{lk}) \right| 0 \right\rangle, \tag{43}$$

where the sign factor characteristic of fermions is given by

$$F' = \sum_{abcd} r_{ab} r_{cd} \quad \text{under} \quad 1 \leqslant b < d \leqslant a < c \leqslant p.$$

The same result is obtained directly from (5).

(II) In the prescription of Refs. 4 and 6, unitary scalar decomposition should be done prior to the trace evaluation. The decomposition implies expressing of (11) as a product of commutable pairs of $A^+(m_i)$ and $A(n_i)$. The present prescription (14), as well as that of Ref. 5, does not require such a decomposition prior to the trace evaluation. The unitary scalar decomposition implies the lack of self contraction, i.e., $e_i = 0$.

(III) In order to evaluate the trace of H^2, where H is the sum of a one-body term T and a two-body interaction V, it is usually required to evaluate separately traces of T^2, V^2, and TV, which are different from each other in $\{m_i\}$ and $\{n_i\}$ of (11). In fact, we do this without decomposing H^2 into T^2, etc. We have only to replace T by its equivalent two-body operator $(\hat{n} - 1)T/(n-1)$, where n is that of the n-body trace. The Hamiltonian is then cast into the form of two-body interactions only. This manipulation is contrary to the unitary scalar decomposition. In the latter, the number operator $\hat{n}$ involved in T and V is extracted, while in the former it is attached to T.

(IV) The operator $A^+(n\,\alpha = v)$ specified by the seniority v of identical particles can be expressed in terms of the operator product[10] $A^+(n-v,0)A^+(v,v)$, each of which is much easier to handle in comparison with $A^+(n,v \neq n)$. Incorporating this and (14) makes it possible to evaluate the seniority-fixed trace.

[1] F. S. Chang, J. B. French, and T. H. Thio, Ann. Phys. **66**, 137 (1971).
[2] M. Nomura, Prog. Theor. Phys. **48**, 110 (1972).
[3] M. Nomura, Prog. Theor. Phys. **48**, 442 (1972).
[4] J. N. Ginocchio, Phys. Rev. C **8**, 135 (1973).
[5] J. N. Ginocchio, *Moment Methods in Many-Fermion Systems*, edited by B. J. Dalton, S. M. Grimes, J. P. Vary, and S. A. Williams (Plenum, New York, 1980), p. 109.
[6] B. C. Chang and S. S. M. Wong, Nucl. Phys. A **294**, 19 (1978).
[7] M. Nomura, Ann. Phys. **91**, 83 (1975).
[8] M. Nomura, J. Math. Phys. **26**, 732 (1985).
[9] H. A. Rabitz and R. Gordon, J. Math. Phys. **11**, 3339 (1970), and references cited therein.
[10] R. D. Lawson, *Theory of the Nuclear Shell Model* (Clarendon, Oxford, 1980).

Two-dimensional inverse scattering: Compactness of the generalized Marchenko operator [a)]

Margaret Cheney[b)]
Department of Mathematics, Stanford University, Stanford, California 94305

(Received 22 May 1984; accepted for publication 31 August 1984)

Recent work on the two-dimensional inverse scattering problem for the Schrödinger equation has resulted in a generalized Marchenko integral equation. The main result of this paper is that the integral operator appearing in this Marchenko equation is of Hilbert–Schmidt type. A result of Newton's shows that its spectrum therefore consists of point eigenvalues whose moduli are at most 1.

I. INTRODUCTION

The inverse scattering theory for the one-dimensional Schrödinger equation is now well understood. It is known that the potential can be recovered from the solution to the Marchenko integral equation. Moreover, it is known that this Marchenko equation has a unique solution.[1]

With the success of the one-dimensional inverse scattering theory came a renewed interest in generalizing the one-dimensional methods to higher dimensions. Perhaps the most successful of the three-dimensional methods is the theory worked out in Newton's series of papers.[2] He found that a three-dimensional (nonsymmetric) potential could also be recovered from the solution of a generalized Marchenko equation. Moreover, he showed that the operator appearing in this Marchenko equation is the square root of a Hilbert–Schmidt operator and is therefore compact. Compactness, in turn, implies that the spectrum is bounded in modulus by 1.

Much of the two-dimensional theory has been worked out.[3] Again there is a generalized Marchenko equation whose solution leads to the potential. The present paper, which is a continuation of earlier work,[3,4] contains the proof of the compactness result for the two-dimensional Marchenko operator. We shall see (Theorem 5.8) that this operator is actually of the Hilbert–Schmidt type. As in the three-dimensional case, this implies that the spectrum is bounded in modulus by 1.

The plan of the paper is as follows. The first section summarizes basic information about two-dimensional scattering and inverse scattering. Here the main result of the paper is precisely stated, and conclusions are drawn from it. The remainder of the paper is devoted to proving the compactness result. Compactness basically depends on the behavior of the wave function as a function of the energy variable k. Section II contains results showing that this dependence is smooth and well-behaved for large energies. The behavior near zero energy, however, is complicated by the divergence of the relevant fundamental solution. This small-energy behavior is investigated in Sec. III. Section IV is chiefly composed of a number of lemmas assembling information from Secs. II and III into preliminary estimates that are used in the proof of Theorem 5.8.

[a)] This is based on the author's Ph.D. thesis, "Quantum Mechanical Scattering and Inverse Scattering in Two Dimensions," Indiana University, 1982.

[b)] Permanent address: Department of Mathematics, Duke University, Durham, North Carolina 27706.

II. PRELIMINARIES

This paper is a continuation of earlier work (see Refs. 3 and 4). We will use the same notation, and we will refer to equations and results of Refs. 3 and 4 by means of the prefixes I and II, respectively. The following facts are recorded here for quick reference.

Scattering solutions of the Schrödinger equation are defined by the Lippmann–Schwinger equation

$$\psi^{\pm}(k,\theta,x)=\exp(ik\theta\cdot x)+\int G(\pm k,|x-y|)V(y)\psi^{\pm}(k,\theta,y)d^2y, \tag{2.1}$$

where θ denotes a unit vector in R^2, and G is

$$G(k,r)=-(i/4)H_0^{(1)}(kr).$$

The function $\xi^{\pm}(k,\theta,x)=|V(x)|^{1/2}\psi^{\pm}(k,\theta,x)$ satisfies

$$\xi^{\pm}(k,\theta,x)=\xi^{0}(k,\theta,x)+K^{\pm}(k)\xi^{\pm}(k,\theta,x), \tag{2.2}$$

where

$$\xi^0(k,\theta,x)=|V(x)|^{1/2}\exp(ik\theta\cdot x),$$
$$K^{\pm}(k)f(x)=\int |V(x)|^{1/2}G(\pm k,|x-y|)\times V_{1/2}(y)f(y)d^2y,$$
$$V_{1/2}(y)=V(y)|V(y)|^{-1/2}. \tag{2.3}$$

Henceforth we will usually assume that V belongs to $L^1 \cap L^2$. Under this hypothesis, for each $k\neq 0$, the function $\xi(k,\theta,x)$ is an L^2 function of x, and the $L^2(x)$ norm is uniformly bounded in θ. We will also assume throughout that the operator K never has the eigenvalue 1. The following hypothesis, which depends on the integer i, will frequently be useful.

Hypothesis (F_i): Let V belong to L^2, and suppose that for some x_0, $|V(x)|$, $|\nabla V(x)|$, and $|\Delta V(x)|$ are all bounded by $F(|x-x_0|)$, where F is a positive decreasing function satisfying

(a) $\int_0^\infty F(r)r^i\,dr\leqslant \nu<\infty$,

and

(b) for some $c>0$ and some ϵ with $0<\epsilon<\frac{1}{2}$,

$F(r)\leqslant cr^{-1+\epsilon}$ near $r=0$.

Such hypotheses, which were also used in Ref. 3, prevent certain pathological behavior of V; in particular they require the decay of V at infinity to be uniform in all direc-

 0022-2488/85/040743-10$02.50

tions. Moreover, V must be bounded and continuous at x_0. For example, near x_0, V may behave like $|x-x_0|^{1+\epsilon}+\text{const}$.

The data for the inverse scattering problem is the scattering amplitude

$$A(k,\theta,\theta')=\int \exp(-ik\theta\cdot x)V(x)\psi(k,\theta',x)d^2x. \quad (2.4)$$

Solution of the inverse problem proceeds via the Marchenko equation

$$\eta(\alpha,\theta,x)=\int_0^\infty \int_{S^1} M_x(\alpha+\beta,\theta,\theta')\eta(\beta,\theta',x)d\theta'\,d\beta + \int_{S^1} M_x(\alpha,\theta,\theta')d\theta', \quad (2.5)$$

where

$$M_x(\gamma,\theta,\theta') = i(8\pi^2)^{-1}\int_{-\infty}^{\infty} \exp[ik\{\gamma+(\theta+\theta')\cdot x\}] \times(\operatorname{sgn} k)A(k,-\theta',\theta)dk, \quad (2.6)$$

and

$$\eta(\gamma,\theta,x)=(2\pi)^{-1/2}\int_{-\infty}^{\infty}\exp(-ik\alpha)[\psi(k,\theta,x) \times\exp(-ik\cdot x)-1]dk. \quad (2.7)$$

We note that x appears only as a parameter in (2.5); henceforth we shall drop the subscript x.

Definition: Let the operator $\mathscr{M}$: $L^2(R^+\times S^1)\to L^2(R^+\times S^1)$ be defined by

$$\mathscr{M}f(\alpha,\theta)=\int_0^\infty\int_{S^1} M(\alpha+\beta,\theta,\theta')f(\beta,\theta')d\theta'\,d\beta, \quad (2.8)$$

where M is defined by (2.6). Here, $\mathscr{M}$ is the operator of the Marchenko equation (2.5).

The purpose of this paper is to present the proof of the following result, which is finally proved in Sec. V.

Theorem 5.8 (compactness): Let $V\in W^{3,1}$ with $\int|x|^4|V(x)|d^2x<\infty$. Suppose also that V satisfies hypothesis (F_2), and that V has no bound or half-bound states. Then the integral operator $\mathscr{M}$ occurring in the Marchenko equation is a Hilbert–Schmidt operator on $L^2(R^+\times S^1)$.

Newton[2] has shown that if $\mathscr{M}$ is compact, then its spectrum is in fact contained in the interval $[-1,1]$. Thus if $\mathscr{M}$ has neither the eigenvalue 1 nor -1, then $\mathscr{M}$ is a contraction and the Marchenko equation can be solved by iteration.

The above theorem allows us to apply Fredholm theory to the Marchenko equation, and, if the spectrum of $\mathscr{M}$ does not contain the point 1, to obtain a solution $\eta(\alpha,\theta,x)$ belonging to $L^2(R^+\times S^1)$ for each x. The potential can then be recovered by means of the formula

$$V(x)=-2\theta\cdot\nabla_x\eta(0,\theta,x).$$

III. SMOOTHNESS

In this section we investigate the k differentiability of the wave function.

We first estimate the Hilbert–Schmidt norms of the first and second derivatives of the integral operator $K(k)$. Close examination of the proof of the analytic Fredholm theorem[5] then shows that differentiability of $K(k)$ implies differentiability of $(I-K(k))^{-1}$. These facts not only show us that the wave function is indeed differentiable, but also give us integral equations that the derivatives satisfy, equations which in turn give us estimates on the derivatives.

Proposition 3.1: Let V belong to L^2, with $\int|x||V(x)|d^2x=M<\infty$. Then the L^2-operator-valued function $K(k)$ with kernel

$$K(k,x,y)=-(i/4)|V(x)|^{1/2}H_0^{(1)}(k|x-y|)V_{1/2}(y),$$

is differentiable for $k\neq0$, the derivative K' having the kernel

$$\frac{\partial}{\partial k}K(k,x,y) =(-i/4)|V(x)|^{1/2}|x-y|H_1^{(1)}(k|x-y|)V_{1/2}(y). \quad (3.1)$$

The Hilbert-Schmidt norm of the derivative satisfies

$$\|K'(k)\|_{\rm HS}\leqslant ck^{-1}\|V\|_1+c(M\|V\|_1k)^{1/2}. \quad (3.2)$$

Proof: Differentiability follows from the dominated convergence theorem once (3.2) has been proved. The estimate (3.2), however, follows easily from the asymptotic behavior of H_1. Q.E.D.

Proposition 3.2: Let V belong to L^2 with $\int|V(x)||x|^4d^2x$ finite. Then the L^2-operator-valued function $K'(k)$ with kernel (3.1) is continuously differentiable for $k\neq0$, the derivative having kernel

$$K''(k,x,y)=-(i/4)|V(x)|^{1/2}|x-y|^2[H_0(k|x-y|) -H_2(k|x-y|)]V_{1/2}(y).$$

The Hilbert-Schmidt norm of the derivative satisfies

$$\|K''(k)\|_{\rm HS}\leqslant ck^{-2}, \quad (3.3)$$

where c is a constant depending on V.

Proof: The proof is similar to that of Proposition 3.1. Q.E.D.

Corollary 3.3: Suppose V satisfies the hypotheses of Proposition 3.1. Then for $k>0$, $(I-K(k))^{-1}$ is a continuously differentiable operator-valued function of k. If V satisfies the hypotheses of Proposition 3.2, then $(d^2/dk^2)(I-K(k))^{-1}$ is continuous for $k>0$.

Proof: A straightforward generalization of the proof in Ref. 5, p. 201, shows that differentiability of K implies that of $(I-K)^{-1}$. Q.E.D.

Corollary 3.4: Let V belong to L^2 with $\int|x|^2|V(x)|d^2x$ finite. Then for all θ the function $\xi(k,\theta,x)$ is a continuously differentiable L^2-valued function of k. If in addition $\int|x|^4|V(x)|d^2x$ is finite, then ξ has a continuous second derivative in $L^2(x)$.

Proof: Derivatives of ξ can be computed from the equation $\xi(k,\theta,x)=[I-K(k)]^{-1}\xi^0$. Q.E.D.

Corollary 3.5: Let V belong to L^2 with $\int|x|^2|V(x)|d^2x$ finite. Then $\xi'=(d/dk)\xi$ satisfies

$$\xi'=\theta\cdot x\xi^0+K'(k)\xi+K(k)\xi'. \quad (3.4)$$

If, in addition $\int|x|^4|V(x)|d^2x$ is finite, then ξ'' satisfies

$$\xi''=(\theta\cdot x)^2\xi^0+K''(k)\xi+2K'(k)\xi'+K(k)\xi''. \quad (3.5)$$

Proof: We differentiate Eq. (2.2). Q.E.D.

Corollary 3.6: Let V belong to L^2 with $\int|x|^2|V(x)|d^2x$ finite. Then for $k \geqslant k_0 > 0$, $\|\xi'(k)\|_2$ is uniformly bounded.

Proof: From Eq. (3.4) we obtain

$$\|\xi'(k)\|_2 \leqslant \|(I-K)^{-1}\| \times\left(\int |x|^2|V(x)|d^2x + \|K'\|\,\|\xi(k)\|_2\right).$$

We then use Propositions I.1.1 and 3.1. Q.E.D.

Corollary 3.7: Let V belong to L^2, with $\int|x|^2|V(x)|d^2x$ finite. Then for $k \geqslant k_0 > 0$, $\|\xi'(k,\theta,x) - \theta\cdot x\xi^0(k,\theta,x)\|_2 \leqslant ck^{-1/2}$.

Proof: From (3.4) we obtain

$$\|\xi' - \theta\cdot x\xi^0\| \leqslant \|K'(k)\|\,\|\xi\| + \|K(k)\|\,\|\xi'\|.$$

Again we use Propositions I.1.1. and 3.1. Q.E.D.

IV. BEHAVIOR NEAR $k=0$

For later estimates we will need detailed knowledge of the small-k behavior of the scattering amplitude. In particular, we will need small-k estimates of the scattering amplitude and its first two derivatives. Naturally, to obtain this information we must study the small-k behavior of the wave function ψ and its first two derivatives. And to study the wave function, we must study the small-k behavior of the operator $K(k)$. For this we use the small-argument behavior of the Hankel function to rewrite K in terms of the operators L and P defined as follows[7]:

$$L(k)f(x) = -\left(\frac{i}{4}\right)\int |V(x)|^{1/2}[H_0^{(1)}(k|x-y|) - (2i/\pi)\log k\,]V_{1/2}(y)f(y)d^2y, \tag{4.1a}$$

$$Pf(x) = (2\pi)^{-1}|V(x)|^{1/2}(V_{1/2},f), \tag{4.1b}$$

where $(\cdot,\cdot)$ denotes the $L^2(R^2)$ inner product. With this notation, we can write $K(k)$ as

$$K(k)f = \log kPf + L(k)f.$$

We note that P is of rank 1 and $L(k)$ is well behaved at $k=0$. From Refs. (3) and (4) we recall the following facts.

Proposition 4.1: Let $V\in L^2$, with $\int|x||V(x)|d^2x < \infty$, and suppose $(I-L(0))^{-1}$ exists. Then near $k=0$ we have

$$\|(I-L(k))^{-1}\xi_0(k) - (I-L(k))^{-1}|V|^{1/2}\|_2 \leqslant ck^{1/2}, \tag{4.2}$$

and

$$(V_{1/2},[I-L(k)]^{-1}|V|^{1/2}) = a_0 + a_1k, \tag{4.3}$$

where $a_0 = (V_{1/2},[I-L(0)]^{-1}|V|^{1/2})$ and a_1 is a bounded function of k. Moreover, $\xi(k)$ satisfies

$$\xi(k) = (I-L(k))^{-1}\xi^0(k) + \frac{(V_{1/2},(I-L(k))^{-1}\xi^0(k))\log k}{2\pi - (V_{1/2},(I-L(k))^{-1}|V|^{1/2})\log k} \times (I-L(k))^{-1}|V|^{1/2}. \tag{4.4}$$

The L^2 norm is bounded by

$$\|\xi(k)\|_2 \leqslant \begin{cases} c|\log k|^{-1}, & \text{if } a_0\neq 0\\ c, & \text{if } a_0 = 0.\end{cases} \tag{4.5}$$

Use of these facts allows us to find the small-k behavior of the scattering amplitude.

Theorem 4.2: Let V belong to L^2 with $\int|x||V(x)|d^2x < \infty$, and suppose $[I-L(0)]^{-1}$ exists. Then for small k, the scattering amplitude satisfies

$$|A(,k,\theta,\theta')| \leqslant \begin{cases} c/|\log k|, & \text{if } a_0\neq 0,\\ ck, & \text{if } a_0 = 0.\end{cases} \tag{4.6}$$

Proof: In (2.4) we split the exponential into two pieces:

$$|A(k,\theta,\theta')| \leqslant |(V_{1/2},\xi(k))| + \left|\int \psi(k,\theta',x)V(x) \times[1-\exp(-ik\theta\cdot x)]d^2x\right|. \tag{4.7}$$

We use (4.4) to calculate $(V_{1/2},\xi)$ and discover that after we find a common denominator, the logarithms in the numerator cancel. Use of this fact together with Eqs. (4.2) and (4.3) allow us to write (4.7) as

$$|A(k,\theta,\theta')| \leqslant \frac{2\pi(a_0 + a_1k + ck^{1/2})}{2\pi - (a_0 + a_1k)\log k} + k\int|\psi(k,\theta',x)V(x)||x|d^2x \leqslant \begin{cases} c/|\log k|, & \text{if } a_0\neq 0,\\ ck, & \text{if } a_0 = 0.\end{cases} \tag{4.8}$$

Q.E.D.

Remark: In contrast to the three-dimensional theory (see Ref. 6, p. 16), in two dimensions the Born series for a nonzero potential is never convergent at $k=0$. This is due to the logarithmic singularity of G.

In order to investigate derivatives of the scattering amplitude, we will need the following result concerning differentiability of $L(k)$.

Proposition 4.3: Let $V\in L^2$ with $\int|x|^2|V(x)|d^2x = M < \infty$. Then the L^2-operator-valued function $L(k)$ is differentiable for $k\neq 0$. The Hilbert–Schmidt norm of the derivative is bounded, the bound depending only on V.

If, in addition, $\int|x|^4|V(x)|d^2x$ is finite, then $L(k)$ has two derivatives whose Hilbert-Schmidt norms are bounded.

Proof: We must check that

$$\lim_{l\to k}\|[L(k)-L(l)](k-l)^{-1} - L'(k)\|_{\mathrm{HS}} = 0.$$

This, however, follows from the dominated convergence theorem and the boundedness of $\|L'(k)\|_{\mathrm{HS}}$. This latter estimate is straightforward and requires only knowledge of the asymptotic forms of $H_1^{(1)}$. Q.E.D.

The proof for the second derivative is similar.

Proposition 4.4: Let V belong to L^2 with $\int|x|^2|V(x)|d^2x$ finite. Assume $[I-L(0)]^{-1}$ exists. Then for k near zero, the L^2 norm of ξ' satisfies

$$\|\xi'(k)\|_2 \leqslant ck^{-1}. \tag{4.9}$$

Moreover, we have the small-k estimates

$$\left|\int V(x)\psi'(k,\theta',x)\exp(ik\theta\cdot x)d^2x\right| \leqslant \frac{c}{k|\log k|^2}. \tag{4.10}$$

Proof: We note that Eq. (3.4) for ξ' is the same as Eq. (2.2) for ξ except that the inhomogeneity ξ^0 of (2.2) is replaced by the quantity η, which we define to be

$$\eta(k,\theta,x) = (\theta\cdot x)\xi^0 + K'(k)\xi. \tag{4.11}$$

Equation (4.4) therefore holds with ξ replaced by ξ' and ξ^0 replaced by η. Explicitly, we have

$$\xi'(k)=[I-L(k)]^{-1}\eta(k) + \frac{(V_{1/2},[I-L(k)]^{-1}\eta(k))\log k}{2\pi-(V_{1/2},[I-L(k)]^{-1}|V|^{1/2})\log k} \times [I-L(k)]^{-1}|V|^{1/2}. \tag{4.12}$$

We use (4.1) to compute the L^2-norm of η:

$$\|\eta(k)\|_2 \leqslant \left(\int |x|^2|V(x)|d^2x\right)^{1/2} + \|L'\xi\|_2 + k^{-1}\|P\xi\|_2 \leqslant c + \|L'\|\,\|\xi\|_2 + 2\pi k^{-1}\|V\|_1^{1/2}|(V_{1/2},\xi)|. \tag{4.13}$$

We use the estimate (4.5) in (4.13), obtaining

$$\|\eta(k)\|_2 \leqslant \begin{cases} c|k\log k|^{-1}, & \text{if } a_0\neq 0, \\ c, & \text{if } a_0=0 \end{cases}. \tag{4.14}$$

Next we use (4.3) in estimating (4.12):

$$\|\xi'(k)\|_2 \leqslant \|(I-L(k))^{-1}\|\,\|\eta\|_2 + \frac{\|V\|_1^{1/2}\|(I-L(k))^{-1}\|\|\eta\|_2|\log k|}{|2\pi-(a_0+a_1k)\log k|}. \tag{4.15}$$

We then apply (4.14) to (4.15), obtaining

$$\|\xi'(k)\|_2 \leqslant \frac{c}{|k\log k|} + \frac{c/k}{|2\pi-(a_0+a_1k)\log k|}.$$

If a_0 is zero, we obtain (4.9); if a_0 is nonzero, we obtain the slightly better estimate $\|\xi'(k)\|_2 \leqslant c|k\log k|^{-1}$.

In preparation for proving (4.10), we prove the estimate

$$\left|\int V(x)\psi'(k,\theta,x)d^2x\right| \leqslant \frac{c}{k\,|\log k|^2}. \tag{4.16}$$

This follows from taking the inner product of both sides of (4.12) with $V_{1/2}$. When we find a common denominator for the terms on the right side of the resulting expression, we discover that the logarithmic terms in the numerator cancel. We then have

$$(V_{1/2},\xi') = (V_{1/2},(I-L)^{-1}\eta) \times \frac{2\pi}{2\pi-(V_{1/2},(I-L)^{-1}|V|^{1/2})\log k}. \tag{4.17}$$

The Schwarz inequality, Eq. (4.3) and (4.14) applied to (4.17) give us (4.16).

We now write the left side of (4.10) as

$$\left|\int V(x)\psi'(k,\theta',x)\exp(ik\theta\cdot x)d^2x\right| \leqslant |(V_{1/2},\xi)| + \left|\int V(x)\psi'(k,\theta',x)(1-\exp(ik\theta\cdot x))d^2x\right|. \tag{4.18}$$

To the first term of (4.18), we apply the estimate (4.16). To the second term we apply the Schwarz inequality. We thus have

$$\left|\int V(x)\psi'(k,\theta',x)\exp(ik\theta\cdot x)d^2x\right| \leqslant \frac{c}{k\,|\log k|^2} + k\|\xi'\|_2\int |x|^2|V(x)|d^2x.$$ Q.E.D.

Lemma 4.5 (differentiability of the scattering amplitude): Suppose V belongs to L^2 with $\int|x|^2|V(x)|d^2x<\infty$. Then for all θ and θ' and for all $k>0$, the function $A(k,-\theta,\theta')=\int\psi(k,\theta',x)V(x)\exp(ik\theta\cdot x)d^2x$ is differentiable with respect to k, the derivative being

$$\int V(x)[\psi'(k,\theta,x)\exp(ik\theta\cdot x) + (i\theta\cdot x)\psi(k,\theta',x)\exp(ik\theta\cdot x)]d^2x. \tag{4.19}$$

If in addition, V satisfies $\int|x|^4|V(x)|d^2x<\infty$, then the second derivative of $A(k,-\theta,\theta')$ exists and is given by

$$\int V(x)[\psi''(k,\theta,x)\exp(ik\theta\cdot x) + 2(i\theta\cdot x)\psi'(k,\theta',x)\exp(ik\theta\cdot x) - (\theta\cdot x)^2\psi(k,\theta',x)\exp(ik\theta\cdot x)]d^2x. \tag{4.20}$$

Proof: Corollary 3.6 allows us to apply the dominated convergence theorem to the difference quotient approximation to A'. We conclude that differentiation under the integral sign is legitimate; this gives us (4.19). The proof of (4.20) is similar. Q.E.D.

Corollary 4.6 (small-k estimate for A'): Suppose V belongs to L^2 with $\int|x|^2V(x)|d^2x$ finite. Assume $[I-L(0)]^{-1}$ exists. Then for small k, A' satisfies

$$|A'(k,\theta,\theta')| \leqslant ck^{-1}|\log k|^{-2}. \tag{4.21}$$

Proof: The first term of (4.19) is estimated by (4.10); to the second term we apply the Schwarz inequality and use (4.5). Q.E.D.

We now turn to estimates for the second derivatives of ψ and A.

Proposition 4.7: Let V belong to L^2 with $\int|x|^4|V(x)|d^2x$ finite. Assume $[I-L(0)]^{-1}$ exists. Then for small k, $\xi''(k)$ satisfies

$$\|\xi''(k)\|_2 \leqslant ck^{-2}. \tag{4.22}$$

Moreover, for small k we also have

$$\left|\int V(x)\psi''(k,\theta',x)\exp(ik\theta\cdot x)d^2x\right| \leqslant \frac{c}{k^2|\log k|}. \tag{4.23}$$

Proof: This proof is similar to that of Proposition 4.6. Q.E.D.

Corollary 4.8 (small-k estimate for A"): Suppose V belongs to L^2 with $\int|x|^4|V(x)|d^2x$ finite. Assume $[I-L(0)]^{-1}$ exists. Then for small k, A'' satisfies

$$|A''(k,\theta,\theta')| \leqslant ck^{-2}|\log k|^{-1}. \tag{4.24}$$

Proof: We use (4.23) to estimate the first term of (4.20). To the second term of (4.20) we apply the Schwarz inequality and (4.9). Similarly, we use the Schwarz inequality and (4.5) to estimate the last term. Q.E.D.

V. COMPACTNESS

Proposition 5.1 (large-k behavior of the Born remainder term): Let V belong to L^1 and satisfy hypothesis (F_0). Then the Born remainder term R, defined by

$$R=\int V(y)\exp(-ik\theta\cdot y)(\psi(k,\theta',y) - \exp(ik\theta'\cdot y))d^2y, \tag{5.1}$$

is uniformly bounded and decays as $k^{-1-\epsilon/2}$ for large k. In particular, $R(k,\theta,\theta')$ is uniformly bounded in L^1 as a function of k.

Proof: First we note that R is uniformly bounded:

$$|R| \leqslant \|V\|_1^{1/2}\|\xi\|_2 + \|V\|_1^{1/2}.$$

Next we investigate the large k decay. Without loss of generality we consider only positive k, because the relation $\psi(-k,\theta,x) = \overline{\psi(k,\theta,x)}$ [which follows from Eq. (2.1)] implies $R(-k,\theta,\theta') = \overline{R(k,\theta,\theta')}$.[3] We use Eq. (2.1) to obtain for R an expression containing H_0. We then split R into four pieces corresponding to large and small argument behavior of the Hankel function. We write

$$4iR = B_1 + B_2 + B_3 + B_4,$$

where

$$B_1 = \int V(y)\exp(-ik\theta\cdot y)(2i/\pi) \times \int_{|y-z|<k^{-1}} \log(k|y-z|)V(z)\psi(k,\theta',z)d^2z\,d^2y, \tag{5.2}$$

$$B_2 = \int V(y)\exp(-ik\theta\cdot y)\int_{|y-z|<k^{-1}} [H_0^{(1)}(k|y-z|) - (2i/\pi)\log(k|y-z|)]V(z)\psi(k,\theta',z)d^2z\,d^2y, \tag{5.3}$$

$$B_3 = \int V(y)\exp(ik\theta\cdot y)\int_{|y-z|>k^{-1}}\left(\frac{2}{\pi k|z-y|}\right)^{1/2} \times \exp(ik|y-z| - i\pi/4)V(z)\psi(k,\theta',z)d^2z\,d^2y, \tag{5.4}$$

$$B_4 = \int V(y)\exp(-ik\theta\cdot y)\int_{|y-z|>k^{-1}}\left[H_0^{(1)}(k|y-z|) - \left(\frac{2}{\pi k|z-y|}\right)^{1/2}\exp\left(ik|y-z| - \frac{i\pi}{4}\right)\right] \times V(z)\psi(k,\theta',z)d^2z\,d^2y. \tag{5.5}$$

First we consider B_1. Since B_1 converges absolutely, it is permissible to do the y integral of (5.2) first. In the y integral (which we label J below), we write $z - y = r\hat{\phi}$, the hats denoting unit vectors:

$$J = \int_{|y-z|<k^{-1}} V(y)\exp(-ik\hat{\theta}\cdot y)\log(k|y-z|)d^2y$$
$$= \exp(-ik\hat{\theta}\cdot z)\int_0^{k^{-1}} r\log kr \times \int_{S^1} V(z + r\hat{\phi})\exp(-ikr\cos\phi)d\phi\,dr.$$

We then apply the stationary phase approximation (See Ref. 3, Appendix D) to the angular integral:

$$J = \exp(-ik\theta\cdot z)\int_0^{k^{-1}} r\log(kr)[(kr)^{-1/2}(aV(z+r\theta) + bV(z-r\theta)) + \widetilde{R}]dr, \tag{5.6}$$

where $\widetilde{R}$ is of order $(kr)^{-1}$; a and b are constants; and we have once again dropped the hats on unit vectors. In modulus, (5.6) is bounded by

$$|J| \leqslant ck^{-1/2}\int_0^{k^{-1}} \sqrt{r}\log(kr)F(||z+x_0| - r|)dr$$
$$\leqslant ck^{-1/2-\epsilon}\int_0^{k^{-1}} r^{1/2-\epsilon}F(||z+x_0| - r|)dr$$
$$\leqslant ck^{-1}\int_0^{k^{-1}} r^{-1+\epsilon}dr$$
$$\leqslant ck^{-1-\epsilon}.$$

With this, an application of the Schwarz inequality to B_1 gives the desired estimate:

$$|B_1| \leqslant ck^{-1-\epsilon}\int |V(z)\psi(k,\theta',z)|d^2z$$
$$\leqslant ck^{-1-\epsilon}\|V\|_1^{1/2}\|\xi\|_2.$$

Precisely the same argument also works for B_2.

Next we consider B_3. Again B_3 converges absolutely, so the y integral of (5.4) can be done first. As before, in the y integral (which we label I below), we make the substitution $z - y = r\hat{\phi}$, the hats distinguishing the unit vectors:

$$I = \int_{|z-y|>k^{-1}} V(y)\exp(-ik\hat{\theta}\cdot y) \times \exp(ik|z-y|)|z-y|^{-1/2}d^2y$$
$$= \exp(-ik\hat{\theta}\cdot z)\int_{k^{-1}}^{\infty}\exp(ikr)r^{-1/2} \times \int V(z + r\hat{\phi})\exp(ikr\cos\phi)d\phi\,r\,dr.$$

We apply the stationary phase approximation to the ϕ integral, dropping hats on unit vectors again:

$$I = \exp(-k\theta\cdot z)\int_{k^{-1}}^{\infty}\sqrt{r}\exp(ikr) \times [(kr)^{-1/2}(aV(z+r\theta) + bV(z-r\theta)) + R]dr, \tag{5.7}$$

where a,b, and R are as in (5.6). The remainder term we leave alone for the moment; the leading term we write as $\exp(-ik\theta\cdot z)(\int\cdots dr)^{1/4}(\int\cdots dr)^{3/4}$. We integrate by parts in the $(\cdot)^{1/4}$ term, obtaining

$$I = k^{-3/4}\exp(-ik\theta\cdot z)\Big\{\exp(ikr)[aV(z+r\theta) + bV(z-r\theta)]\big|_{k^{-1}}^{\infty} - \int_{k^{-1}}^{\infty}\exp(ikr)[a\theta\cdot\nabla V(z+r\theta) - b\theta\cdot\nabla V(z-r\theta)]\,dr\Big\}^{1/4}$$
$$\times\left[\int_{k^{-1}}^{\infty}\exp(ikr)(aV(z+r\theta) + bV(z-r\theta))dr\right]^{3/4}$$
$$+ \exp(-ik\theta\cdot z)\int_{k^{-1}}^{\infty}\sqrt{r}\exp(ikr)R\,dr. \tag{5.8}$$

Next we obtain a bound for I by replacing the V's by F's in (5.8):

$$|I| \leqslant ck^{-3/4}\left[F(||z|-k^{-1}|)+\int_{k^{-1}}^{\infty}F(|r-|z||)dr\right]^{1/4}$$
$$\times\left[\int_{k^{-1}}^{\infty}F(|r-|z||)dr\right]^{3/4}$$
$$+c\int_{k^{-1}}^{\infty}(kr)^{-3/2}F(|r-|z||)r^{1/2}\,dr. \tag{5.9}$$

We then carry out the integrations in the first term of (5.9) and use the fact that $kr>1$ in the second term:

$$|I|<ck^{-3/4}(F(||z|-k^{-1}|)+v)^{1/4}v^{3/4}$$
$$+ck^{-(1+\epsilon)/2}\int_{k^{-1}}^{\infty}r^{-\epsilon/2}F(|r-|z||)dr. \tag{5.10}$$

The remaining integral in (5.10) is bounded for large k by

$$c\int_{k^{-1}}^{1}r^{-1+\epsilon/2}\,dr+v\leqslant c. \tag{5.11}$$

We now recall that I is the y integral of B_3; we use (5.10) and (5.11) in (5.4), obtaining

$$|B_3|\leqslant ck^{-1-\epsilon/2}\int|V(z)\psi(k,\theta,z)|$$
$$\times[(F(|z|-k^{-1})+v)^{1/4}v^{3/4}+c]dz. \tag{5.12}$$

Two applications of the Schwarz inequality to (5.12) give us

$$|B_3|\leqslant ck^{-1-\epsilon/2}\|\xi\|_2$$
$$\times\left(\|V\|_2\int F(|z|-k^{-1})d^2z+\|V\|_1^{1/2}\right)$$
$$\leqslant ck^{-1-\epsilon/2}.$$

Next we consider B_4. Information on the asymptotic behavior of the Hankel function allows us to estimate (5.5):

$$|B_4|\leqslant c\int\int_{|y-z|<k^{-1}}\frac{|V(y)||V(z)||\psi(k,\theta,z)|}{|k(z-y)|^{3/2}}\,d^2z\,d^2y. \tag{5.13}$$

In (5.13) we make the substitution $z-y=r\phi$, where ϕ is a unit vector; we also use the fact that $kr>1$. This gives

$$|B_4|\leqslant ck^{-1-\epsilon/2}\int\int_{S_1}\int_{k^{-1}}^{\infty}r^{-1-\epsilon/2}\,V(z-r\phi)r\,dr\,d\phi$$
$$\times V(z)\psi(k,\theta,z)d^2z\leqslant ck^{-1-\epsilon/2}. \qquad \text{Q.E.D.}$$

Lemma 5.2 (Uniform estimate for A): Let V belong to $W^{1,1}$ and satisfy hypothesis (F_0). Then $A(k,\theta,\theta')=I+R$, where

$$I\leqslant c\|V\|_{1,1}/(1+k|\theta'-\theta|), \tag{5.14}$$

and $\int|R|^2\,dk$ is uniformly bounded. For fixed $|\theta'-\theta|\neq 0$,

$$\lim_{k\to\infty}A(k,\theta,\theta')=0.$$

Proof: We split A into the Born term and the remainder

$$A(k,\theta,\theta')=I+R,$$

where

$$I=\int\exp(ik(\theta'-\theta)\cdot x)V(x)d^2x,$$

$$R=\int\exp(-ik\theta\cdot x)V(x)(\psi(k,\theta',x)-\exp(ik\theta'\cdot x))d^2x.$$

To estimate I, we integrate by parts:

$$ik(\theta'-\theta)I=\int[\nabla\exp(ik(\theta'-\theta)\cdot x)]V(x)d^2x$$
$$=-\int\exp(ik(\theta'-\theta)\cdot x)]\nabla V(x)d^2x.$$

This shows that

$$I\leqslant\|V\|_{1,1}/(k|\theta'-\theta|).$$

However, since V is in L^1, its Fourier transform is uniformly bounded; this shows that

$$I\leqslant c\|V\|_{1,1}/(1+k|\theta'-\theta|).$$

To the term R, we apply Proposition 5.1. Q.E.D.

Lemma 5.3: Let V satisfy hypothesis (F_1). Then for $k\geqslant k_0>0$, the following holds:

$$\left|\int V(x)\psi'(k,\theta',x)\exp(ik\theta\cdot x)d^2x\right|$$
$$\leqslant c/(1+k|\theta+\theta'|), \tag{5.15}$$

where c is a constant depending only on V, and where ψ' denotes the k derivative of ψ.

Proof: Equation (3.4) allows us to write

$$\int V(x)\psi'(k,\theta',x)\exp(ik\theta\cdot x)d^2x=I_1+I_2+I_3,$$

where

$$I_1=\int V(x)(\theta'\cdot x)\exp(ik(\theta+\theta')\cdot x)d^2x,$$

$$I_2=\int V(x)\int|x-y|H_1^{(1)}(k|x-y|)V(y)$$
$$\times\psi(k,\theta',y)d^2y\exp(ik\theta\cdot x)d^2x,$$

$$I_3=\int V(x)\int H_0(k|x-y|)V(y)$$
$$\times\psi'(k,\theta',y)d^2y\exp(ik\theta\cdot x)d^2x.$$

First we consider I_1. Our hypotheses imply that $(\theta'\cdot x)V(x)$ satisfies the hypotheses of Lemma 5.2. We therefore have

$$|I_1|\leqslant c\|xV\|_{1,1}/(1+k|\theta+\theta'|).$$

Next we consider I_2. We split I_2 into two pieces: $I_2=I_4+I_5$, where

$$I_4=\int V(x)\int|x-y|H_1^{(1)}(k|x-y|)V(y)$$
$$\times\exp(ik\theta'\cdot y)d^2y\exp(ik\theta\cdot x)d^2x, \tag{5.16}$$

$$I_5=\int V(x)\int|x-y|H_1(k|x-y|)V(y)$$
$$\times(\psi(k,\theta',y)-\exp(ik\theta'\cdot y))d^2y$$
$$\times\exp(ik\theta\cdot x)d^2x. \tag{5.17}$$

First we attack I_4 by letting $z=x-y$ and writing z in polar coordinates as $z=r\hat{\phi}$:

$$I_4 = \int V(x)\exp(ik(\hat\theta + \hat\theta')\cdot x)\int_0^\infty r^2 H_1(kr)$$
$$\times \int_{S^1} V(x + r\hat\phi)\exp(ikr\cos\phi)d\hat\phi\, dr\, d^2x. \quad (5.18)$$

We split (5.18) into large-r and small-r pieces: $I_4 = I_6 + I_7$, where

$$I_6 = \int V(x)\exp(ik(\hat\theta + \hat\theta')\cdot x)\int_0^{k^{-1}} r^2 H_1(kr)$$
$$\times \int_{S^1} V(x + r\hat\phi)\exp(ikr\cos\phi)d\hat\phi\, dr\, d^2x,$$
$$I_7 = \int V(x)\exp(ik(\hat\theta + \hat\theta')\cdot x)\int_{k^{-1}}^\infty r^2 H_1(kr)$$
$$\times \int_{S^1} V(x + r\hat\phi)\exp(ikr\cos\phi)d\hat\phi\, dr\, d^2x.$$

We use the small-argument asymptotic form of the Hankel function and the assumption on V to estimate I_6:

$$|I_6| \leqslant c \int |V(x)| \int_0^{k^{-1}} r^2 (kr)^{-1}$$
$$\times 2\pi F(|r - |x + x_0||)dr\, d^2x$$
$$\leqslant ck^{-2}.$$

To I_7 we apply the stationary phase approximation (see Ref. 3):

$$I_7 = \int V(x)\exp(ik(\theta + \theta')\cdot x)\int_{k^{-1}}^\infty r^2 H_1(kr)$$
$$\times [(kr)^{-1/2}(aV(x + r\theta') + bV(x - r\theta')) + R\,]dr\, d^2x, \quad (5.19)$$

where a and b are constants and R is of order $(kr)^{-1}$. Because of this behavior of R and the $(kr)^{-1/2}$ behavior of the Hankel function, the final term of (5.19) is easily seen to decay as $k^{-3/2}$. Similarly, the leading term of (5.19) decays as k^{-1}. We have thus shown that I_7, and therefore also I_4, is bounded by ck^{-1}.

Next we consider I_5. We use Lemma I.1.2 to obtain

$$|I_5| \leqslant \int |V(x)| \int |x - y||H_1(k|x - y|)|$$
$$\times |V(y)|ck^{-1/2-\epsilon/4}\, d^2y\, d^2x;$$

estimates similar to those above give us decay of $k^{-1-\epsilon/4}$. We have now shown that I_2 is bounded by ck^{-1}.

Finally, we consider I_3. We use the asymptotic form of ψ' (Corollary 3.7) to split I_3 into $I_3 = I_8 + I_9$, where

$$I_8 = \int V(x)\int H_0(k|x - y|)V(y)(\theta'\cdot y)$$
$$\times \exp(ik\theta'\cdot y)d^2y\, \exp(ik\theta\cdot x)d^2x, \quad (5.20)$$
$$I_9 = \int V(x)\int H_0(k|x - y|)V(y)[\psi'(k, \theta', y)$$
$$\times -(\theta'\cdot y)\exp(ik\theta'\cdot y)]d^2y\, \exp(ik\theta\cdot x)d^2x. \quad (5.21)$$

Let us first consider I_8. Since I_8 converges absolutely, we may do the x integral of (5.20) first. In the x integral of (5.20), we make the substitution $x - y = r\hat\phi$:

$$I_8 = \int V(y)(\hat\theta'\cdot y)\exp(ik(\hat\theta + \hat\theta)\cdot y)$$
$$\times \int_0^\infty H_0(kr)\int_{S^1} V(y + r\hat\phi)$$
$$\times \exp(ikr\cos\phi)d\hat\phi\, r\, dr\, d^2y. \quad (5.22)$$

We split (5.22) into large-r and small-r pieces: $I_8 = I_8' + I_8''$, where

$$I_8' = \int V(y)(\hat\theta'\cdot y)\exp(ik(\hat\theta' + \hat\theta)\cdot y)$$
$$\times \int_0^{k^{-1}} H_0(kr)\int V(y + r\hat\phi)$$
$$\times \exp(ikr\cos\phi)d\phi\, r\, dr\, d^2y,$$
$$I_8'' = \int V(y)(\hat\theta'\cdot y)\exp(ik(\hat\theta' + \hat\theta)\cdot y)$$
$$\times \int_{k^{-1}}^\infty H_0(kr)\int (\text{same integrand}).$$

The term I_8' is easily handled; we obtain decay of k^{-1} because the domain of integration shrinks as k increases. In I_8'', we use the large-argument asymptotic form of the Hankel function and apply the stationary phase approximation to the angular integral, obtaining

$$|I_8''| \leqslant c \int |V(y)||y|$$
$$\times \int_{k^{-1}}^\infty (kr)^{-1}F(|r - |y + x_0||)r\, dr\, d^2y$$
$$\leqslant ck^{-1}\int |V(y)||y|d^2y < ck^{-1}.$$

This shows that $|I_8| \leqslant ck^{-1}$.

Next we consider I_9. We apply the Schwarz inequality to the y integral of (5.21)

$$|I_9| \leqslant \int |V(x)|\left(\int |H_0(k|x - y|)|^2 |V(y)|d^2y\right)^{1/2}$$
$$\times \|\xi'(k) - (\theta'\cdot y)\xi^0\| d^2x. \quad (5.23)$$

We then use Corollary 3.7 and write the remaining y integral of (5.23) in polar coordinates:

$$|I_9| \leqslant \int |V(x)|\left(\int_0^\infty |H_0(kr)|^2 \int_{S^1} V(x - r\phi)d\phi\, r\, dr\right)^{1/2}$$
$$\times ck^{-1/2}\, d^2x. \quad (5.24)$$

Finally, we split the r integral of (5.24) into terms corresponding to integration over $r < k^{-1}$ and $r > k^{-1}$, respectively. In each term we use the appropriate asymptotic form of the Hankel function:

$$|I_9| \leqslant ck^{-1/2}\int |V(x)|\left[\int_0^{k^{-1}} |\log kr|^2 2\pi\right.$$
$$\times F(|r - |x + x_0||)r\, dr$$
$$\left. + \int_{k^{-1}}^\infty (kr)^{-1}\, 2\pi F(|r - |x + x_0||)r\, dr\right]^{1/2} d^2x$$
$$< ck^{-1}.$$

Q.E.D.

Corollary 5.4 (large-k estimate for A′): Suppose V satisfies hypothesis (F_1) and $\int|x|^2|V(x)|d^2x<\infty$. Let k_0 be positive. Then for $k\geqslant k_0>0$, A' satisfies

$$|A'(k,-\theta,\theta')|\leqslant c/(1+k|\theta+\theta'|), \tag{5.25}$$

where c depends only on V and on k_0.

Proof: We apply Lemma 5.3 to the firm term of (4.19). To the second term we apply Lemma 5.2. Q.E.D.

Notation: We will write

$$B(k,\theta,\theta',x)=(8\pi^2)^{-1}i \times\exp(ik(\theta+\theta')\cdot x)A(k,-\theta',\theta). \tag{5.26}$$

Corollary 5.5 (estimates for B and its derivatives): Suppose V belongs to L^2 with $\int|x|^2|V(x)|d^2x$ finite. Assume $(I-L(0))^{-1}$ exists. Then for k sufficiently small, B' satisfies

$$|B'(k,\theta,\theta',x)|\leqslant ck^{-1}|\log k|^{-2}, \tag{5.27}$$

where c depends on x and on V. If V also satisfies $\int|x|^4|V(x)|d^2x<\infty$, then the second derivative satisfies

$$|B''(k,\theta,\theta',x)|\leqslant ck^{-2}|\log k|^{-1}, \tag{5.28}$$

where c depends on x and on V. If V satisfies hypothesis F_1 and $\int|x|^2|V(x)|d^2x<\infty$, then for $k_0>0$, the derivative $B'=(d/dk)B$ satisfies

$$\|\chi_{|k|>k_0}(k)B'(k,\theta,\theta',x)\|_2^{(k)}<c|\theta+\theta'|^{-1/2}, \tag{5.29}$$

where χ denotes the characteristic function that is one on the set $|k|>k_0$, and zero elsewhere.

Proof: Equations (5.27) and (5.28) follow directly from Corollaries 4.6 and 4.8, respectively. Equation (5.29) follows from integration of (5.25). Q.E.D.

Notation: We recall from (2.6) the definition

$$M(\gamma,\theta,\theta')=\int_{-\infty}^{\infty}\exp(ik\gamma)(\operatorname{sgn}k)B(k,\theta,\theta',x)dk, \tag{5.30}$$

where B is defined by (5.26). The function M is the kernel of the integral operator appearing in the Marchenko equation (2.5).

Lemma 5.6 (integration by parts for M): Let V belong to $W^{1,1}$ and satisfy hypothesis (F_0). Suppose also that $\int|x|^2|V(x)|d^2x$ is finite, and that $(I\text{-}L(0))^{-1}$ exists. Then for $\theta+\theta'\neq0$,

$$M(\gamma,\theta,\theta')=-i\gamma^{-1}\int_{-\infty}^{\infty}\exp(ik\gamma)(\operatorname{sgn}k)B'(k,\theta,\theta',x)dk. \tag{5.31}$$

Proof: We split the right side of (5.30) into two pieces, and integrate by parts in each. The boundary terms disappear because of Lemma 5.2 and Theorem 4.2. Q.E.D.

Remark: The vanishing of the boundary terms on the above integration-by-parts lemma is crucial. It allows us to prove that M decays faster than γ^{-1} at infinity, which in turn is needed in the proof of the main theorem [see Eq. (5.48)].

Lemma 5.7 (large-γ estimate for small-k part of M): Suppose V belongs to L^2 with $\int|x|^4|V(x)|d^2<\infty$. Suppose also that $(I\text{-}L(0))^{-1}$ exists. Then for k_0 sufficiently small and $\gamma>k_0^{-2}$, we have the estimate

$$\int_0^{k_0}\exp(ik\gamma)B'(k,\theta,\theta',x)dk<c|\log\gamma|^{-1}. \tag{5.32}$$

Proof: We denote the left side of (5.32) by I. We write $I=I_1+I_2$, where

$$I_1=\int_0^{\gamma^{-1/2}}\exp(ik\gamma)B'(k)dk, \tag{5.33}$$

$$I_2=\int_{\gamma^{-1/2}}^{k_0}(\text{same integrand}). \tag{5.34}$$

First we consider I_1. By Corollary 5.5, we can bound the right side of (5.33) by

$$|I_1|\leqslant c\int_0^{\gamma^{-1/2}}\frac{dk}{k(\log k)^2}. \tag{5.35}$$

The substitution $u=\log k$ gives us (for $\gamma>1$)

$$|I_1|\leqslant c\int_{-\infty}^{-(1/2)\log\gamma}u^{-2}\,du=c|\log\gamma|^{-1}.$$

Now attacking I_2, we integrate by parts in (5.34):

$$I_2=(i\gamma)^{-1}\exp(ik\gamma)B'(k)|_{\gamma^{-1/2}}^{k_0} -\int_{\gamma^{-1/2}}^{k_0}(i\gamma)^{-1}\exp(ik\gamma)B''(k)dk. \tag{5.36}$$

We then use Corollary 5.5 to estimate the right side of (5.36):

$$|I_2|\leqslant c\gamma^{-1}+c\gamma^{-1/2}|\log\gamma|^{-2}+\gamma^{-1}\int_{\gamma^{-1/2}}^{k_0}\frac{c}{k^2\log k}dk. \tag{5.37}$$

In the third term of (5.37), we use $k^{-1}<\gamma^{1/2}$. The remaining integrand is $(k\log k)^{-1}$, which can be integrated exactly with the help of the substitution $u=\log k$. The third term of (5.37) is therefore bounded by $c\gamma^{-1/2}\log\log\gamma$. Q.E.D.

Theorem 5.8: Let V belong to $W^{3,1}$ with $\int|x|^4|V(x)|d^2$ finite. Suppose also that V satisfies hypothesis (F_2), and that $(I\text{-}L(0))^{-1}$ exists. Then the operator $\mathscr{M}$ defined in (2.8) is of Hilbert–Schmidt type; in particular $\mathscr{M}$ is a compact operator.

Proof: We split the operator into two pieces:

$$\mathscr{M}f(\alpha,\theta)=\mathscr{M}_1f+\mathscr{M}_2f,\quad\text{where for some } b>0,$$

$$\mathscr{M}_1f(\alpha,\theta)=\int_0^{\max(b-\alpha,0)}\int_{S^1}Mf\,d\theta'\,d\beta, \tag{5.38}$$

$$\mathscr{M}_2f(\alpha,\theta)=\int_{\max(b-\alpha,0)}^{\infty}\int_{S^1}Mf\,d\theta'\,d\beta. \tag{5.39}$$

The claim is that both $\mathscr{M}_1$ and $\mathscr{M}_2$ are Hilbert–Schmidt operators.

To show that $\mathscr{M}_2$ is of Hilbert–Schmidt type, we will compute

$$\|\mathscr{M}_2\|_{HS}^2=\int_{S^1}\int_{S^1}J(\theta,\theta')d\theta\,d\theta', \tag{5.40}$$

where

$$J(\theta,\theta')=\int_0^{\infty}\int_0^{\infty}|\chi_{[\max(b-\alpha,0),\infty]}(\beta) \times M(\alpha+\beta,\theta,\theta')|^2\,d\alpha\,d\beta, \tag{5.41}$$

χ being the characteristic function of the interval $[\max(b-\alpha,0),\infty]$. We shall prove the estimate

$$J(\theta,\theta')\leqslant c|\theta+\theta'|^{-1/2}, \tag{5.42}$$

which clearly establishes the boundedness of (5.40).

To prove (5.42), we let $\gamma=\alpha+\beta$ in (5.41), and then interchange the α and γ integrals. This gives us

$$J(\theta,\theta')=\int_b^\infty \gamma|M(\gamma,\theta,\theta')|^2\,d\gamma. \tag{5.43}$$

Next we use (5.31) in (5.43). In doing so, we split the k integral of (5.31) into large- and small-k pieces, and then use the triangle inequality. We obtain the following inequality, which holds for $\theta+\theta'\neq 0$:

$$\left(\int_b^\infty \gamma|M(\gamma,\theta,\theta')|^2\,d\gamma\right)^{1/2}\leqslant J_1+J_2, \tag{5.44}$$

where

$$J_1^2=\int_b^\infty \gamma^{-1}\left|\int_{|k|>k_0}\exp(ik\gamma)(\operatorname{sgn}k)B'(k,\theta,\theta',x)dk\right|^2 d\gamma, \tag{5.45}$$

$$J_2^2=\int_b^\infty \gamma^{-1}\left|\int_{|k|<k_0}\exp(ik\gamma)(\operatorname{sgn}k)B'(k,\theta,\theta',x)dk\right|^2 d\gamma. \tag{5.46}$$

Fortunately, we have already done most of the work involved in estimating J_1 and J_2. In (5.45), we note that $1/\gamma$ is bounded by $1/b$. We can therefore use the Plancherel theorem to obtain

$$J_1^2\leqslant b^{-1}\|\chi_{|k|>k_0}(k)B'(k,\theta,\theta',x)\|^2_{L^2(k)}, \tag{5.47}$$

which, by (5.29) is bounded by $c|\theta+\theta'|^{-1}$. The k integral occurring in J_2 has already been estimated in Lemma 5.7. This gives

$$|J_2|^2\leqslant c\int_b^\infty \frac{d\gamma}{\gamma(\log\gamma)^2}<\infty, \tag{5.48}$$

which establishes (5.42).. We have thus shown that $\mathscr{M}_2$ is a Hilbert–Schmidt operator.

Next we shall show that $\mathscr{M}_1$ is of Hilbert–Schmidt type. To do this we must estimate $M(\gamma,\theta,\theta')$ for small γ. We recall that M is given by (2.6). We split the scattering amplitude appearing in (2.6) into the Born approximation, plus remainder. This corresponds to writing M as $M=E_1+E_2$, where

$$E_1=i(8\pi^2)^{-1}\int_{-\infty}^\infty \exp(ik\gamma)(\operatorname{sgn}k)\int V(y)\times\exp(ik(\theta+\theta')\cdot(x+y))d^2y\,dk, \tag{5.49}$$

$$E_2=i(8\pi^2)^{-1}\int_{-\infty}^\infty \exp[ik(\gamma+(\theta+\theta')\cdot x)](\operatorname{sgn}k)\times\int V(y)[\psi(k,\theta',y)-\exp(ik\theta'\cdot y)]\exp(ik\theta\cdot y)d^2y\,dk. \tag{5.50}$$

Proposition 5.1 shows that E_2 is bounded and that the corresponding integral operator is of Hilbert–Schmidt type.

To prove that $\mathscr{M}_2$ is a Hilbert–Schmidt operator, we thus need only show that the integral operator $\mathscr{E}_1$ corresponding to E_1 is of Hilbert–Schmidt type. To do this, we shall establish the estimate

$$|E_1|\leqslant c|\theta+\theta'|^{-1/4}\gamma^{-3/4}. \tag{5.51}$$

Estimate (5.51) suffices to prove the theorem; we see this as follows. The Hilbert–Schmidt norm of $\mathscr{E}_1$ is

$$\|\mathscr{E}_1\|_{\rm HS}=\int_{S_1}\int_{S_1}\int_0^b\int_0^{b-\alpha}|E_1(\alpha+\beta,\theta,\theta')|^2\,d\beta\,d\alpha\,d\theta\,d\theta'. \tag{5.52}$$

In (5.52), we make the substitution $\gamma=\alpha+\beta$ and interchange the α and γ integrals. We obtain

$$\|\mathscr{E}_1\|_{\rm HS}=\int_{S_1}\int_{S_1}\int_0^b \gamma|E_1|^2 d\gamma\,d\theta\,d\theta'. \tag{5.53}$$

In (5.53), we use (5.51) and easily conclude that $\mathscr{E}_1$ is a Hilbert–Schmidt operator.

It remains to prove (5.51). To do this, we first split E_1 into

$$E_1=\int_{-\infty}^0+\int_0^\infty=D_2+D_1. \tag{5.54}$$

We shall consider only D_1. (The estimate for D_2 is similar.) In (5.49) we make the substitution $z=x+y$ and write the Fourier transform as $\widetilde{V}_x(\omega)=\int V(z-x)\exp(i\omega\cdot z)d^2z$. With this notation, D_1 can be written

$$D_1=i(8\pi^2)^{-1}\int_0^\infty \exp(ik\gamma)\widetilde{V}_x(k(\theta+\theta'))dk. \tag{5.55}$$

In (5.55) we make the changes of variables $t=k|\theta+\theta'|$, $\hat{n}=(\theta+\theta')/|\theta+\theta'|$, and $\eta=\gamma/|\theta+\theta'|$:

$$D_1=i(8\pi^2|\theta+\theta'|)^{-1}\int_0^\infty \exp(it\eta)\widetilde{V}_x(t\hat{n})dt. \tag{5.56}$$

We denote the integral appearing in (5.56) by J:

$$J=\int_0^\infty \exp(it\eta)\widetilde{V}_x(t\hat{n})dt. \tag{5.57}$$

We note that J is uniformly bounded because $\widetilde{V}\in L^1$. This is due to the following two facts: first, since $V\in L^1$, $\widetilde{V}_x(\omega)$ is uniformly bounded in x and ω; second, if $V\in W^{3,1}$ then three integrations by parts show the large t decay to be t^{-3}.

Next we perform an integration by parts in (5.57) to obtain η^{-1} decay of J for large η:

$$J=-i\eta^{-1}\widetilde{V}_x(t\hat{n})|_0^\infty+i\eta^{-1}\int_0^\infty \exp(it\eta)\hat{n}\cdot\nabla\widetilde{V}_x(t\hat{n})dt. \tag{5.58}$$

We now need to show that $\hat{n}\cdot\nabla\widetilde{V}_x(t\hat{n})\in L^1(t)$.

First we note that

$$\hat{n}\cdot\nabla\widetilde{V}(t\hat{n})=i\int \hat{n}\cdot zV(z-x)\exp(it\hat{n}\cdot z)d^2z \tag{5.59}$$

is uniformly bounded in $t\hat{n}$ because the first moment of the potential is finite. Next we integrate (5.59) by parts twice, obtaining $|\hat{n}\cdot\nabla\widetilde{V}(t\hat{n})|\leqslant c/t^2$. This shows that $\hat{n}\cdot\nabla\widetilde{V}_x(tn)\in L^1(t)$.

We use this fact in (5.58), finding that $|J|<c\eta^{-1}$. Since J is also uniformly bounded, we have $|J|<c\eta^{-3/4}$. We use this estimate in (5.57), obtaining

$$|D_1|\leqslant c|\theta+\theta'|^{-1/4}\gamma^{-3/4},$$

which in turn leads to (5.51). Q.E.D.

ACKNOWLEDGMENTS

I am grateful to my thesis advisor, Roger G. Newton, for suggesting the problem and discussing it with me frequently. In particular, he caught one of my early mistakes, and later, suggested Eq. (5.48).

This work was supported in part by the National Science Foundation, the Office of Naval Research, the Army Research Office, and the Air Force Office of Scientific Research.

[1]Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, New York, 1956).
[2]R. G. Newton, J. Math. Phys. **21**, 1698 (1980); **22**, 2191 (1981); **23**, 594 (1982).
[3]M. Cheney, J. Math. Phys. **25**, 94 (1984).
[4]M. Cheney, J. Math. Phys. **25**, 1449 (1984).
[5]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1972).
[6]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton, U. P., Princeton, NJ, 1971).
[7]M. Klaus and B. Simon, Ann. Phys. **130**, 251 (1980).

Time decay and spectral kernel asymptotics

T. A. Osborn[a)]
Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742
R. Wong[b)]
Department of Mathematics, University of Manitoba, Winnipeg, Manitoba R3T 2N2 Canada

(Received 8 May 1984; accepted for publication 24 August 1984)

For quantum systems in $\mathbb{R}^3$ defined by a Hamiltonian H given as the sum of the negative Laplacian perturbed by a real-valued potential $v(x)$, the large time behavior of the fundamental solution of the time-dependent Schrödinger equation is investigated. For a suitably restricted class of potentials that have algebraic decay as $|x|\to\infty$, the continuous spectrum portion of the fundamental solution is characterized by an asymptotic expansion as $t\to\pm\infty$, which is uniform in compact subsets of $\mathbb{R}^3\times\mathbb{R}^3$. These results are then applied to derive the large energy asymptotic expansions of the spectral kernel associated with H.

I. INTRODUCTION

One-body quantum mechanics in $\mathbb{R}^3$ assumes the following form. Let H be the operator generating time evolution on the Hilbert space $L^2(\mathbb{R}^3)$. Take x to be the generic point in $\mathbb{R}^3$. Then H is defined as the self-adjoint extension of the elliptic differential operator

$$H_{(x)} = -(\hbar^2/2m)\Delta_x + v(x), \tag{1.1}$$

where Δ_x is the Laplacian in $\mathbb{R}^3$ and $v(x)$ is a real-valued potential. The constants $\hbar$ and m denote the rationalized value of Planck's constant and the particle mass, respectively.

If t is the real-valued parameter representing time displacement then the time evolution group may be represented, in a generalized function sense, by an associated family of kernels $U(x,y;it/\hbar)$. For sufficiently smooth potentials, $U(x,y;it/\hbar)$ is the fundamental solution of the time-dependent Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}U(x,y;it/\hbar) = H_{(x)}U(x,y;it/\hbar), \quad t\neq 0, \tag{1.2}$$

that satisfies the delta-function initial condition

$$U(x,y;it/\hbar)\to\delta(x-y), \quad \text{as } t\to 0. \tag{1.3}$$

The objective of this paper is to characterize—by an appropriate asymptotic expansion—the large-t behavior of the fundamental solution of Eq. (1.2). Hereafter we simplify our notation by setting $\hbar = 1$ and $m=\frac{1}{2}$. The inner product on $L^2(\mathbb{R}^3)$ is denoted by $\langle\cdot,\cdot\rangle$ and the related L^2-norm by $\|f\| = \langle f,f\rangle^{1/2}$. The inner product is defined to be linear in the right element. For potentials that decay as $|x|\to\infty$ [e.g., $v(x)\in L^2(\mathbb{R}^3)$], it is known[1] that H has a countable number of independent eigenfunctions

$$(H\psi_j)(x) = \lambda_j\psi_j(x), \quad \|\psi_j\| = 1, \quad j=1,2,\dots, \tag{1.4}$$

and that the point spectrum $\sigma_p(H) = \{\lambda_j : j=1,2,\dots\}$ has zero as the only possible cluster point. Since the manifold $\mathbb{R}^3$ on which $H_{(x)}$ is defined is noncompact, this causes H to have a continuous spectrum in addition to the point spectrum. This is typically true in the case of decaying potentials. Each eigenfunction ψ_j contributes a term $e^{-it\lambda_j}\psi_j(x)\psi_j(y)^*$ to the spectral decomposition of $U(x,y;it)$. Thus the portion of the fundamental solution associated with the continuous spectrum may be defined by

$$S(x,y;t) = U(x,y;it) - \sum_j e^{-i\lambda_j t}\psi_j(x)\psi_j(y)^*. \tag{1.5}$$

Our interest in the large-t decay of $S(x,y;t)$ stems from the following problem. Suppose $\{e(x,y;\lambda):\lambda\in\mathbb{R}\}$ is the family of spectral kernels defined by H. Take $z\in\mathbb{C}$ and let D be the open right-half plane in $\mathbb{C}$. The family of bounded operators $\{e^{-zH}: z\in D\}$ constitutes the analytic semigroup induced by H. Each operator of this semigroup is represented by a Carleman kernel. These integral kernels admit a small-$|z|$ asymptotic expansion of the form[2]

$$U(x,y;z) = U_0(x,y;z)\left[\sum_{n=0}^{M-1}\frac{(-z)^n}{n!}P_n(x,y) + E_M(x,y;z)\right], \tag{1.6}$$

where $U_0(x,y;z)$ is the semigroup kernel associated with the case $v(x)=0$, i.e.,

$$U_0(x,y;z) = [1/(4\pi z)^{3/2}]e^{(-|x-y|^2)/4z}. \tag{1.7}$$

The coefficient functions $P_n(x,y)$ are functions of $v(x)$ and its first $2(n-1)$ partial derivatives, and are given by explicit formulas[3] for all $n\geqslant 0$. The order M of expansion (1.6) is proportional to the number of bounded derivatives that $v(x)$ possesses. The error term $E_M(x,y;z)$ has a bound which is of order $|z|^M$ and is uniform for all $(x,y)\in\mathbb{R}^3\times\mathbb{R}^3$. If $\operatorname{Im} z = 0$ and $x=y$, then (1.6) is the well-known heat kernel expansion for $H_{(x)}$ (see Ref. 4).

Under what circumstances does the asymptotic expansion (1.6) imply the existence of an asymptotic expansion of $e(x,y;\lambda)$ as $\lambda\to\infty$? In a previous work (Ref. 5: hereafter OW), a preliminary answer to this question has been obtained. The method used in OW is based on the formula

$$e(x,y;\lambda) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{e^{z\lambda}}{z}U(x,y;z)dz, \quad c>0, \tag{1.8}$$

for $\lambda\neq\sigma_p(H)$. Equation (1.8) transforms semigroup kernels into spectral kernels. This formula is the inverse Laplace transform companion to the spectral kernel representation

$$U(x,y;z) = \int_{-\infty}^{\infty} e^{-z\lambda}\,de(x,y;\lambda), \quad z\in D. \tag{1.9}$$

[a)]On leave from the University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada.
[b)]I. W. Killam Research Fellow.

 0022-2488/85/040753-16$02.50

For $\gamma \in \mathbb{R}$ and integer $N \geqslant 0$ we say that *H belongs to class* $T_p(\gamma;N)$ if

$$\int_{|t|>1} dt \frac{1}{|t|^{\gamma+1}} \left| \left(\frac{d}{dt}\right)^j S(x,y;t) \right| < \infty \tag{1.10}$$

for each $(x,y) \in \mathbb{R}^3 \times \mathbb{R}^3$ and for $j = 0,1,...,N$. In OW the following was established: If $v(x)$ supports a sufficient number of uniformly bounded derivatives, if $\sigma_p(H)$ is contained in a finite interval, and $H \in T_p(\gamma;N)$ for some $\gamma < 0$, then one may combine asymptotic expansion (1.6) and transform (1.8) to obtain the spectral kernel expansion

$$e(x,y;\lambda) = \sum_{n=0}^{M-1} \frac{(-1)^n}{n!} P_n(x,y) \left(\frac{d}{d\lambda}\right)^n e_0(x,y;\lambda) + R_M(x,y;\lambda). \tag{1.11}$$

In this formula, $e_0(x,y;\lambda)$ is the spectral kernel for the Hamiltonian H_0 (the self-adjoint extension of $-\Delta_x$). This free spectral kernel has the following analytic definition:

$$e_0(x,y;\lambda) = (\lambda^{1/2}/2\pi|x-y|)^{3/2} J_{3/2}(\lambda^{1/2}|x-y|), \tag{1.12}$$

where $J_{3/2}$ is the Bessel function of the first kind. Here, M is the largest integer less than $2N + \frac{3}{2}$. Although it is not apparent from the form of (1.11), the small parameter in expansion (1.11) is $\lambda^{-1/2}$. The error term $R_M(x,y;\lambda)$ is $o(\lambda^{-N})$ uniformly for x and y in compact subsets of $\mathbb{R}^3$. This summary makes it apparent that if we can obtain an appropriate decay estimate of $S(x,y;t)$, then we have established $H \in T_p(\gamma;N)$ and, hence, verified the expansion (1.11).

We base our investigation of the large-t behavior of $S(x,y;t)$ on the formula

$$S(x,y;t) = \lim_{\beta \to 0^+} \frac{1}{\pi} \int_0^\infty e^{-(\beta+it)\lambda} \operatorname{Im} r(x,y;\lambda + i0) d\lambda. \tag{1.13}$$

Here $r(x,y;z)$ is the resolvent kernel for the operator $r(z) = (H-z)^{-1}, z \in \mathbb{C}, \operatorname{Im} z \neq 0$. The notation $r(x,y;\lambda + i0)$, $\lambda > 0$, is used to denote the limiting value of $r(x,y;\lambda + i\nu)$ as $\nu \to 0^+$. By obtaining small-λ asymptotic expansions and large-λ estimates for $\operatorname{Im} r(x,y;\lambda + i0)$, one may transform the Abel-limit integral (1.13) into an asymptotic expansion for $S(x,y;t)$. We provide sufficient restrictions on the potentials $v(x)$ which enable us to prove that

$$S(x,y;t) = \Phi(x,y) t^{-3/2} + \delta(x,y;t). \tag{1.14}$$

The remainder is characterized by

$$\delta(x,y;t) = o(t^{-2}), \quad \text{as } |t| \to \infty, \tag{1.15}$$

uniformly for x and y in compact subsets of $\mathbb{R}^3$. Furthermore, we establish sufficient conditions on the potential that allow us to differentiate (1.14) with respect to t any number of times.

Consider for the moment the nature of the several restrictions which we shall impose on the potentials $v(x)$. Basically we have three types of restrictions. First, a smoothness requirement that ensures the validity of expression (1.6); second, an algebraic decay as $|x| \to \infty$ is needed to control the pointwise estimates for the kernel $r(x,y;\lambda + i0)$ for both large and small values of λ; and, finally, a condition that prohibits appearance of a zero energy resonance.[6]

The smoothness criteria is met by supposing that the potential $v(x)$ is a Fourier image of a complex bounded measure μ

$$v(x) = \int_{\mathbb{R}^3} e^{ik\cdot x} d\mu(k), \tag{1.16}$$

where $k \cdot x$ denotes the scalar product in $\mathbb{R}^3$. Of course, the potential $v(x)$ must be real if H is to be self-adjoint. The measure μ is said to obey the *reflection property* if $\mu(-e) = \mu(e)^*$ for all measurable sets e. Let $\mathscr{M}^*(\mathbb{R}^3)$ be the set of all complex bounded measures satisfying the reflection property defined on the Borel field on $\mathbb{R}^3$. Transform (1.16) then defines a natural class by

$$\mathscr{F}^* = \left\{ v(x) = \int_{\mathbb{R}^3} e^{ik\cdot x} d\mu(k); \mu \in \mathscr{M}^*(\mathbb{R}^3) \right\}. \tag{1.17}$$

Potentials in this class are uniformly bounded and uniformly continuous in $\mathbb{R}^3$. A convenient way to ensure the existence of partial derivatives of order M or less is to define the class of functions

$$\mathscr{F}_M^* = \left\{ v \in \mathscr{F}^*: \int_{\mathbb{R}^3} |k|^n d|\mu|(k) < \infty ; n = 0,1,...,M \right\}, \tag{1.18}$$

where $|\mu|(e)$ denotes the total variation of $\mu(e)$. For $v \in \mathscr{F}_M^*$ there exists a smallest finite positive constant K (depending on μ and M) such that

$$\int_{\mathbb{R}^3} |k|^n d|\mu|(k) \leqslant K^n \|\mu\|, \tag{1.19}$$

where

$$\|\mu\| \equiv |\mu|(\mathbb{R}^3). \tag{1.20}$$

With norm $\|v\| \equiv \|\mu\|$, it may be shown that $\mathscr{F}^*$ is a Banach space. Suppose D_x^L is a partial derivative with multi-index $L = (l_1,l_2,l_3)$ and length $|L| = l_1 + l_2 + l_3$. If $v \in \mathscr{F}_M^*$, then

$$|D_x^L v(x)| \leqslant K^L \|\mu\|, \quad \text{for all } |L| \leqslant M. \tag{1.21}$$

For more details on the class $\mathscr{F}_M^*$, see Ref. 2. The validity of expansion (1.6) requires only that $v \in \mathscr{F}_{2M}^*$.

We characterize the algebraic decay of $v(x)$ and its gradient by two classes, $\mathscr{A}_n$ and $\mathscr{B}_n$, where n is a positive integer. Potential $v(x)$ is said to be in class $\mathscr{A}_n$ if there is an $\epsilon > 0$ such that

$$|v(x)| < C/(1+|x|)^{n+\epsilon}, \quad \text{all } x, \tag{1.22}$$

for some $C < \infty$. Similarly, $v(x)$ is said to be in class $\mathscr{B}_n$ if there is an $\epsilon > 0$ such that

$$|\nabla v(x)| < C/(1+|x|)^{n+1+\epsilon}, \quad \text{all } x, \tag{1.23}$$

for some $C < \infty$. The symbol ∇ is the gradient in $\mathbb{R}^3$.

Finally we say $v(x)$ is in class $\mathscr{C}_0$ if the eigenfunction problem

$$\int \frac{[\operatorname{sgn} v(x)] |v(x)v(y)|^{1/2}}{4\pi|x-y|} \phi(y) dy = -\phi(x) \tag{1.24}$$

has only the trivial solution $\phi = 0$ in $L^2(\mathbb{R}^3)$. Condition $\mathscr{C}_0$ suffices to show that $\operatorname{Im} r(x,y;\lambda)$ behaves like $\sqrt{\lambda}$ as $\lambda \to 0$. If (1.24) has nontrivial solutions then, as Jensen and Kato have shown,[6] $r(\lambda)$ has a $\lambda^{-1/2}$ singularity as $\lambda \to 0$. This latter case permits the existence of zero energy resonances. However, it is only for exceptional potentials that (1.24) will have a nontrivial solution. Our method would permit us to omit condition $\mathscr{C}_0$ in our analysis of $S(x, y;t)$, but this would substantially lengthen our discussion.

The plan of this paper is as follows. Section II gathers together the basic facts about the resolvent kernel $r(x, y;\lambda + i\nu)$ and characterizes the boundary behavior as $\nu \to 0^{\pm}$. Section III proves the validity of representation (1.13). Large-λ estimates are found in Sec. IV and small-λ asymptotic expansions of the resolvent kernel are given in Sec. V. In the final section, we establish the asymptotic expansion (1.14) and apply it to obtain the spectral kernel expansion (1.11).

II. RESOLVENT KERNELS

In order to use (1.13) as the basis for the study of the large-t behavior of $S(x, y;t)$, it is evident that one requires detailed estimates of $\operatorname{Im} r(x, y;\lambda + i0)$. First we introduce operators w and u on $L^2(\mathbb{R}^3)$ that are defined by multiplication with the functions

$$w(x) = |v(x)|^{1/2} \tag{2.1}$$

and

$$u(x) = w(x)\operatorname{sgn} v(x) . \tag{2.2}$$

In terms of $r_0(z) = (H_0 - z)^{-1}$, the free resolvent, we also define the family of operators $\{A(z): z \in \mathbb{C} \setminus [0, \infty)\}$ by

$$A(z) = u r_0(z) w \tag{2.3}$$

with the boundary values of $A(z)$ being defined by replacing z by $\lambda \pm i0$. A convenient way[7] to investigate the resolvent kernel $r(x, y;z)$ is to employ the operator identity

$$r(z) = r_0(z) - r_0(z) w (1 + A(z))^{-1} u r_0(z) \tag{2.4}$$

[see Ref. 7, (3.1)]. Formula (2.4) implies that the $z = \lambda \pm i0$, $\lambda > 0$ boundary value behavior of $(1 + A(z))^{-1}$ should suffice to control the behavior of the kernels $r(x, y;\lambda \pm i0)$. This technique of studying $r(x, y;\lambda \pm i0)$ is particularly straightforward since it is easily seen that $A(z)$ is a Schmidt-class operator on the boundary of the resolvent set.

This section focuses on characterizing the linkage between the following two eigenvalue problems. Let C_+ and C_- represent the open upper- (lower-) half complex plane and $\overline{C}_\pm$ the respective closures of these half-planes. The two eigenvalue problems are

$$A(z)\phi = -\phi , \quad \phi \in L^2(\mathbb{R}^3) , \tag{2.5}$$

$$H\psi = z\psi , \quad \psi \in L^2(\mathbb{R}^3) , \tag{2.6}$$

for z in either $\overline{C}_+$ or $\overline{C}_-$. For convenience, let us also introduce the symbol Π for the plane cut along the positive real axis. (Hence, Π is an open set.) With Π considered as a sheet in the Riemann surface for $\sqrt{z}$, we denote its closure by Π_c. From the literature[1,7–10] we select and modify a number of results that will serve as an appropriate foundation for deriving $\lambda \to \infty$ and $\lambda \to 0$ asymptotic expansions of $r(x, y;\lambda + i0)$.

Suppose $K: L^2(\mathbb{R}^3) \to L^2(\mathbb{R}^3)$ is a bounded operator. We say K is an *integral operator* if there exists a complex-valued measurable function $k(x, y)$ on $\mathbb{R}^3 \times \mathbb{R}^3$ such that

$$(Kf)(x) = \int_{\mathbb{R}^3} k(x, y) f(y) dy , \quad \text{a.a. } x \in \mathbb{R}^3 , \tag{2.7}$$

for f chosen from some dense subset of $L^2(\mathbb{R}^3)$. Furthermore the kernel is said to be *Carleman type* if

$$\int_{\mathbb{R}^3} |k(x, y)|^2 dy < \infty , \quad \text{a.a. } x \in \mathbb{R}^3 . \tag{2.8}$$

For example, the free resolvent $r_0(z)$ is an integral operator and represented by

$$(r_0(z)f)(x) = \int_{\mathbb{R}^3} \frac{e^{i\sqrt{z}|x-y|}}{4\pi|x-y|} f(y) dy . \tag{2.9}$$

If $z \in \Pi$, then the free resolvent kernel is of Carleman type and (2.9) is valid for all $f \in L^2(\mathbb{R}^3)$. With this terminology one has the following proposition.

Proposition 1: Suppose $v \in L^2(\mathbb{R}^3)$ and assume $\operatorname{Im} z \neq 0$. Then we have the following.

(i) $r(z)$ is an integral operator of Carleman type and its kernel $r(x, y;z)$ (the resolvent kernel) satisfies the integral equation

$$r(x, y;z) = \frac{e^{i\sqrt{z}|x-y|}}{4\pi|x-y|} - \int \frac{e^{i\sqrt{z}|x-x'|}}{4\pi|x-x'|} v(x') r(x', y;z) dx' \tag{2.10}$$

as a function of x a.e. in $\mathbb{R}^3$ for a.a. fixed $y \in \mathbb{R}^3$.

(ii) Uniqueness. Suppose $\tilde{r}(x, y;z)$ is a solution of (2.10) such that $\tilde{r}(\cdot, y;z) \in L^2(\mathbb{R}^3)$ for each fixed y then $\tilde{r}(x, y;z)$ is the resolvent kernel of $r(z)$, i.e.,

$$\tilde{r}(x, y;z) = r(x, y;z) , \quad \text{a.a. } (x, y) \in \mathbb{R}^3 \times \mathbb{R}^3 . \tag{2.11}$$

Proof: See Ikebe, Ref. 1. □

Let us record the basic features of the family of operators $A(z)$. Recall that a potential is said to be of *Rollnik class* ($v \in \mathscr{R}$) if

$$B_r \equiv \int \int \frac{|v(x)||v(y)|}{(4\pi)^2 |x-y|^2} dx\, dy < \infty . \tag{2.12}$$

A sufficient condition[8] for v to belong to $\mathscr{R}$ is $v \in L^1(\mathbb{R}^3) \cap L^2(\mathbb{R}^3)$. We denote the derivatives in the strong operator topology of $A(z)$ with respect to $z \in \Pi_c$ (when they exist) by

$$A^{(n)}(z) = \left(\frac{d}{dz}\right)^n A(z) . \tag{2.13}$$

Proposition 2: Suppose that $v \in L^1(\mathbb{R}^3) \cap L^2(\mathbb{R}^3)$.

(i) For all $z \in \Pi_c$, $A(z)$ is an integral operator with kernel

$$A(x, y;z) = u(x) e^{i\sqrt{z}|x-y|} w(y) / 4\pi|x-y| . \tag{2.14}$$

(ii) For all $z \in \Pi_c$ the operator $A(z)$ is Schmidt class. The Schmidt norm has the uniform bound in Π_c

$$\|A(z)\|_2 \leqslant \sqrt{B_r} . \tag{2.15}$$

(iii) The family of operators $\{A(z): z \in \Pi_c\}$ is Schmidt-norm continuous. For all z, z' in Π_c the continuity condition

$$\|A(z) - A(z')\|_2 \leqslant (\|v\|_1 / 4\pi) |\sqrt{z} - \sqrt{z'}| \tag{2.16}$$

is valid.

(iv) There exists a $\Lambda_1 < \infty$ such that if $|z| > \Lambda_1$ and $z \in \Pi_c$ then

$$\|A^2(z)\|_2 < \tfrac{1}{2} . \tag{2.17}$$

Furthermore,

$$\lim_{|z| \to \infty} \|A^2(z)\|_2 = 0 . \tag{2.18}$$

(v) Let n be a positive integer. In addition to the above hypothesis, suppose that $(1+|x|)^{2(n-1)}v(x) \in L^1(\mathbb{R}^3)$. Then $A^{(n)}(z)$ is a Schmidt-class operator for all $z \in \Pi_c \setminus \{0\}$. If $z \in \Pi_c$ and $|z| > 1$, then

$$\|A^{(n)}(z)\|_2 < C_1 |z|^{-n/2} \|(1+|x|)^{2(n-1)}v(x)\|_1 , \tag{2.19}$$

where C_1 is finite and independent of z.

Proof: (i) Formula (2.14) follows from the fact that $r_0(z)$ is an integral operator with the kernel given in (2.9) and the fact that u and w are operators of multiplication. Statements (ii), (iii), and (v) are immediate consequences of formula (2.14). (iv) (for $z = \lambda \pm i0$) is proved by Simon[8] (Theorem I.23). The extension of Simon's proof to $z \in \Pi_c$ is trivial. □

Next we summarize a number of results associated with the nature of the operator $(1 + A(\lambda \pm i0))^{-1}$.

Lemma 1: Assume $(1+|x|^2)v(x) \in L^1(\mathbb{R}^3)$ and $v(x) \in L^\infty(\mathbb{R}^3)$. If $\lambda > 0$ is such that there exists a nontrivial $L^2(\mathbb{R}^3)$ solution of

$$A(\lambda \pm i0)\phi = -\phi , \tag{2.20}$$

then the Fourier transform $\widehat{w\phi}$ of $w\phi$ has the representation

$$\widehat{w\phi}(k) = (|k| - \sqrt{\lambda})g(k) , \tag{2.21}$$

where $g(k)$ is a continuous function of $k \in \mathbb{R}^3$ having a uniform bound

$$|g(k)| < C(\lambda) \tag{2.22}$$

for some finite k-independent $C(\lambda)$.

Proof: The function $\widehat{w\phi}(k)$ is defined by

$$\widehat{w\phi}(k) = \frac{1}{(2\pi)^{3/2}} \int e^{ik\cdot x} w(x)\phi(x)dx . \tag{2.23}$$

The Schwartz inequality applied to (2.23) gives the k-independent bound

$$|\widehat{w\phi}(k)| \leqslant [1/(2\pi)^{3/2}] \|v\|_1^{1/2} \|\phi\| . \tag{2.24}$$

Furthermore, since $|w(x)\phi(x)|$ is an $L^1(\mathbb{R}^3)$ function, $\widehat{w\phi}(k)$ is uniformly continuous in k. Represent $k \in \mathbb{R}^3$ by the spherical coordinate system $(\eta, \hat{k})$, where $\hat{k}$ is the two-dimensional unit vector and $\eta = |k|$. The radial derivative of $\widehat{w\phi}(\eta, \hat{k})$ is given by the integral

$$\frac{\partial}{\partial \eta} \widehat{w\phi}(\eta, \hat{k}) = \frac{1}{(2\pi)^{3/2}} \int i|x|\hat{k}\cdot\hat{x} e^{i\eta|x|\hat{k}\cdot\hat{x}} w(x)\phi(x)dx . \tag{2.25}$$

Here $\hat{k}\cdot\hat{x}$ is the scalar product on the unit sphere in $\mathbb{R}^3$. Equation (2.25) yields the η-independent bound

$$\left| \frac{\partial}{\partial \eta} \widehat{w\phi}(\eta, \hat{k}) \right| \leqslant \frac{1}{(2\pi)^{3/2}} \|x^2 v\|_1^{1/2} \|\phi\| . \tag{2.26}$$

The radial derivative in (2.25) is continuous in both η and $\hat{k}$. So, for $\eta > \sqrt{\lambda} > 0$ one may write

$$\widehat{w\phi}(\eta, \hat{k}) - \widehat{w\phi}(\sqrt{\lambda}, \hat{k}) = (\eta - \sqrt{\lambda})g(\eta, \hat{k}) , \tag{2.27}$$

where g is defined by

$$g(\eta, \hat{k}) = \frac{1}{\eta - \sqrt{\lambda}} \int_{\sqrt{\lambda}}^{\eta} \frac{\partial}{\partial \eta'} \widehat{w\phi}(\eta', \hat{k})d\eta' . \tag{2.28}$$

Bound (2.26) for the radial derivative implies that $|g(\eta, \hat{k})|$ is also bounded by the right-hand side of (2.26).

It remains to show that $\widehat{w\phi}(\sqrt{\lambda}, \hat{k}) = 0$ for all $\hat{k}$. This conclusion results from the following argument. The potential $v(x)$ is real-valued and $\phi \in L^2(\mathbb{R}^3)$, so

$$0 = -\operatorname{Im}\langle \phi, (v/|v|)\phi \rangle = \operatorname{Im}\langle \phi, (v/|v|)A(\lambda + i0)\phi \rangle . \tag{2.29}$$

Since $v \in L^1(\mathbb{R}^3) \cap L^2(\mathbb{R}^3)$, by the continuity property [Proposition 2(iii)] Eq. (2.29) takes the limiting form

$$0 = \lim_{\nu \to 0^+} \operatorname{Im}\langle \phi, w r_0(\lambda + i\nu) w\phi \rangle . \tag{2.30}$$

Employing the Parseval theorem for Fourier integrals and the fact that $r_0(\lambda + i\nu)$ is given by multiplication with $(k^2 - \lambda - i\nu)^{-1}$ in the Fourier transform space, the inner product above is

$$\langle w\phi, r_0(\lambda + i\nu) w\phi \rangle = \int \frac{1}{k^2 - \lambda - i\nu} |\widehat{w\phi}(k)|^2 dk . \tag{2.31}$$

Finally, since $\widehat{w\phi}(k)$ is uniformly continuous in k, the limiting relation (2.30) becomes

$$0 = \pi \int \delta(|k|^2 - \lambda) \left\{ \int |\widehat{w\phi}(k)|^2 d\hat{k} \right\} |k|^2 d|k| , \tag{2.32}$$

where δ is the one-dimensional delta function. Thus $\widehat{w\phi}(\sqrt{\lambda}, \hat{k}) = 0$ for all $\hat{k}$. □

Let $\mathscr{D}(H)$ and $\mathscr{D}(H_0)$ be the domains in $L^2(\mathbb{R}^3)$ of the operators H and H_0, respectively. For potentials that are bounded ($\|v\|_\infty < \infty$) then $\mathscr{D}(H) = \mathscr{D}(H_0)$.

Lemma 2: Assume $(1+|x|^2)v(x) \in L^1(\mathbb{R}^3)$ and $v(x) \in L^\infty(\mathbb{R}^3)$. Suppose $\lambda > 0$ is such that (2.20) has a nontrivial $L^2(\mathbb{R}^3)$ solution and let $g(k)$ be the function defined by (2.21). If $\hat{\psi}_\lambda : \mathbb{R}^3 \to \mathbb{C}$ is defined by the formula

$$\hat{\psi}_\lambda(k) \equiv g(k)/(|k| + \sqrt{\lambda}) , \tag{2.33}$$

then

(i) $\psi_\lambda \in \mathscr{D}(H_0) = \mathscr{D}(H)$, (2.34)

(ii) $\psi_\lambda = \operatorname{s-lim}_{\nu \to 0^+} r_0(\lambda + i\nu) w\phi \equiv r_0(\lambda + i0) w\phi$, (2.35)

(iii) $(H_0 - \lambda) r_0(\lambda + i0) w\phi = w\phi$. (2.36)

Proof: (i) It suffices to show

$$\|H_0 \psi_\lambda\|^2 = \|\widehat{H_0 \psi_\lambda}\|^2 = \int |k^2 \hat{\psi}_\lambda(k)|^2 dk < \infty . \tag{2.37}$$

The contribution to integral (2.37) for $|k| \leqslant \sqrt{2\lambda}$ is seen to be finite from estimate (2.22), whereas the contribution for $|k| > \sqrt{2\lambda}$ is shown finite by using (2.21) in combination with $\|\widehat{w\phi}\|^2 \leqslant \|v\|_\infty \|\phi\|^2$. Similar reasoning shows $\|\psi_\lambda\| < \infty$.

(ii) Equation (2.35) is the statement

$$\lim_{\nu \to 0^+} \|\psi_\lambda - r_0(\lambda + i\nu) w\phi\| = 0 . \tag{2.38}$$

Observe that the Fourier transform of $r_0(\lambda + i\nu) w\phi$ may be written as $\hat{r}_0(\lambda + i\nu)\widehat{w\phi}$, where $\hat{r}_0(\lambda + i\nu)$ indicates multiplication by $(k^2 - \lambda - i\nu)^{-1}$. By Parseval's theorem

$$\|\psi_\lambda - r_0(\lambda + i\nu) w\phi\|^2 = \int \left| \frac{g(k)}{|k| + \sqrt{\lambda}} - \frac{\widehat{w\phi}(k)}{k^2 - \lambda - i\nu} \right|^2 dk . \tag{2.39}$$

Let $I(\nu)$ denote the k-space integral on the right-hand side of (2.39). Recalling (2.21), $I(\nu)$ is

$$I(\nu)=\int\left|1-\frac{k^2-\lambda}{k^2-\lambda-i\nu}\right|^2\left|\frac{g(k)}{|k|+\sqrt{\lambda}}\right|^2 dk . \quad (2.40)$$

The first absolute value factor is smaller than 1 for all arguments. Using (2.33) and $\|\hat{\psi}_\lambda\|=\|\psi_\lambda\|<\infty$, then dominated convergence shows $I(\nu)\to 0$ as $\nu\to 0^+$.

(iii) As in (2.39), evaluate the relevant norm in k-space. Using (2.33) and (2.35) to represent $r_0(\lambda+i0)w\phi$, one has

$$\|(H_0-\lambda)r_0(\lambda+i0)w\phi-w\phi\|^2$$
$$=\int\left|(k^2-\lambda)\frac{g(k)}{|k|+\sqrt{\lambda}}-\widehat{w\phi}(k)\right|^2 dk . \quad (2.41)$$

However, identity (2.21) shows that the integrand of (2.41) is strictly zero. Thus (iii) is proved. □

Lemma 3: Let $(1+|x|^2)v(x)\in L^1(\mathbb{R}^3)$ and $v(x)\in L^\infty(\mathbb{R}^3)$. Suppose $z\in\Pi_c\setminus\{0\}$ is such that there exists a nontrivial $L^2(\mathbb{R}^3)$ solution of

$$A(z)\phi=-\phi . \quad (2.42)$$

Then

(i) $\psi_z\equiv r_0(z)w\phi\in\mathscr{D}(H)$, (2.43)

(ii) $H\psi_z=z\psi_z$. (2.44)

Proof: (i) Define ρ as the distance from z to the spectrum of H_0, i.e.,

$$\rho=\inf_{\lambda\in[0,\infty)}|z-\lambda| . \quad (2.45)$$

Suppose $\rho>0$, then one has the operator identity

$$(H_0-z)r_0(z)=1 , \quad (2.46)$$

and it follows that

$$H_0\psi_z=w\phi+zr_0(z)w\phi . \quad (2.47)$$

Since $\|r_0(z)\|=(1/\rho)<\infty$, the two vectors on the right of (2.47) are both in $L^2(\mathbb{R}^3)$. Thus $\psi_z\in\mathscr{D}(H_0)=\mathscr{D}(H)$ so (i) is demonstrated for $\rho>0$. If $\rho=0$ and $z=\lambda+i0$, then (i) is implied by statements (i) and (ii) of Lemma 2. If $z=\lambda-i0$, the argument of Lemma 2 is easily modified to include this case.

(ii) If $\rho>0$, (2.44) is a trivial consequence of (2.42). Assume $\rho=0$ and $z=\lambda+i0$, $\lambda>0$. Apply $H-\lambda$ to $\psi_{\lambda+i0}$. Multiplication by $v(x)$ is a bounded operator and $\psi_{\lambda+i0}$ is in $\mathscr{D}(H_0)$, so one has

$$(H-\lambda)\psi_{\lambda+i0}$$
$$=(H_0-\lambda)r_0(\lambda+i0)w\phi+w[ur_0(\lambda+i0)w]\phi$$
$$=w\phi-w\phi=0 . \quad (2.48)$$

In the second equality we have used (2.36) and hypothesis (2.42). A similar argument applies to $z=\lambda-i0$. □

The following lemma provides us with a simple criteria that guarantees the absence of positive energy eigenvalues of H.

Lemma 4: If $v\in\mathscr{A}_1$ then $\sigma_p(H)\subseteq(-\infty,0]$.

Proof: See Kato.[11] □

The criteria given in Lemma 4 is only one of several[1,12,13] available in the literature to rule out positive eigenvalues. Our study of large-t asymptotic expansions eventually requires $v\in\mathscr{A}_n\cap\mathscr{B}_n$ for $n>1$, thus the Kato criteria is adequate for our purposes.

To proceed further we invoke restriction $\mathscr{C}_0$ on $v(x)$ in order to rule out any possible singularity of $(1+A(z))^{-1}$ as $z\to 0$. As noted in the Introduction both zero energy resonances and zero energy eigenfunctions are prohibited by this condition.

Let S_1 be the semi-infinite strip in $\mathbb{C}$ given by

$$S_1=\{z:z\in\mathbb{C},\quad 0\leqslant\operatorname{Re}z,\quad 0\leqslant\operatorname{Im}z\leqslant 1\} . \quad (2.49)$$

In the proof of Proposition 3 we use the fact that the product AB of a Schmidt-class operator A and a bounded operator B is a Schmidt-class operator and has bound $\|AB\|_2\leqslant\|A\|_2\|B\|$. This fact will be used again without further mention.

Proposition 3: Assume $(1+|x|^2)v(x)\in L^1(\mathbb{R}^3)$ and $v\in\mathscr{A}_1\cap\mathscr{C}_0$.

(i) For each $z\in\Pi_c\setminus\sigma_p(H)$ the operator $(1+A(z))^{-1}$ exists as a bounded linear transformation from $L^2(\mathbb{R}^3)$ to $L^2(\mathbb{R}^3)$.

(ii) The operator norm $\|(1+A(z))^{-1}\|$ is uniformly continuous for z in any compact subset of $\Pi_c\setminus\sigma_p(H)$.

(iii) There is an $M<\infty$ such that

$$\|(1+A(z))^{-1}\|<M , \quad (2.50)$$

for all $z\in S_1$.

Proof: (i) For $z\in\Pi_c$, $A(z)$ is a compact operator on $L^2(\mathbb{R}^3)$. Recall, for compact operators, that $(1+A(z))^{-1}$ is a bounded operator if -1 is not an eigenvalue of $A(z)$. Assume there is some z in $\Pi_c\setminus\sigma_p(H)$ such that -1 is an eigenvalue of $A(z)$ and ϕ is the associated eigenfunction of Eq. (2.42). Then Lemma 3 states that $r_0(z)w\phi$ is an eigenfunction of H with eigenvalue z. If $\operatorname{Im}z\neq 0$, this is a contradiction since H is self-adjoint. If $\operatorname{Im}z=0$ then (2.44) states that $z\in\sigma_p(H)$; this is also impossible since z is chosen from the complement of $\sigma_p(H)$. So (i) is proved.

(ii) Let z_2 be an arbitrary fixed point in $\Pi_c\setminus\sigma_p(H)$. For $z_1\in\Pi_c\setminus\sigma_p(H)$, define the operator $\widetilde{A}$ by

$$\widetilde{A}=1+(A(z_1)-A(z_2))(1+A(z_2))^{-1} . \quad (2.51)$$

The Schmidt-norm continuity of $A(z)$ as given in Proposition 2(iii) means that there is a $\delta>0$ such that for $|\sqrt{z_2}-\sqrt{z_1}|<\delta$, $\|A(z_1)-A(z_2)\|_2\|(1+A(z_2))^{-1}\|<\frac{1}{2}$. Thus $\widetilde{A}^{-1}$ exists and $\|\widetilde{A}^{-1}\|<2$ for z_1 sufficiently close to z_2. In this circumstance the operator identity

$$(1+A(z_1))^{-1}-(1+A(z_2))^{-1}$$
$$=(1+A(z_2))^{-1}(A(z_2)-A(z_1))(1+A(z_2))^{-1}\widetilde{A}^{-1} \quad (2.52)$$

is valid. The $\|\cdot\|_2$ norm of the right-hand side vanishes as $\sqrt{z_1}\to\sqrt{z_2}$. This follows, since by Proposition 2(iii) $\|A(z_1)-A(z_2)\|_2\to 0$ in this limit while the other three factors on the right have bounded operator norms.

(iii) Divide the strip S_1 into two disjoint parts L and R. Let the set L contain $z\in S_1$ with $\operatorname{Re}z\leqslant A_1$ and $R=S_1\setminus L$, where A_1 is the constant occurring in statement (iv) of Proposition 2. Assume first that $z\in R$. Then (2.17) is valid and we have

$$(1+A(z))^{-1}=(1-A(z))(1-A^2(z))^{-1} . \quad (2.53)$$

Applying Proposition 2(ii) and (iv) to (2.53) gives

$$\|(1+A(z))^{-1}\|<2(1+B_r) ,\quad z\in R . \quad (2.54)$$

Now let $z \in L$. Lemma 4, together with the condition $\mathscr{C}_0$, asserts that the intersection of $\sigma_p(H)$ and L is empty. Thus $\|(1 + A(z))^{-1}\|: L \to \mathbb{R}$ is a continuous function on the compact set L, so there exists some point $z_m \in L$, where $\|(1 + A(z))^{-1}\|$ assumes its maximum value. But $\|(1 + A(z_m))^{-1}\| < \infty$. So M in Eq. (2.50) is any bound larger than $\|(1 + A(z_m))^{-1}\|$ or $2(1 + B_r)$. □

The results of Propositions 2 and 3 are similar to the findings of Agmon[10] and Ginibre and Moulin.[7] However, these authors analyze $r_0(z)$ as an operator acting between a pair of Sobolev spaces. For a somewhat narrower class of potentials we have obtained Propositions 2 and 3 by utilizing only Hilbert space methods.

III. THE ABEL LIMIT REPRESENTATION OF $S(x, y; t)$

This section brings together the necessary analysis and estimates needed to establish the validity of

$$\left(\frac{d}{dt}\right)^j S(x, y; t) = \lim_{\beta \to 0^+} \frac{1}{\pi} \int_0^\infty (-i\lambda)^j e^{-(\beta + it)\lambda} \times \operatorname{Im} r(x, y; \lambda + i0) d\lambda, \tag{3.1}$$

for $j = 0, 1, 2, \ldots$. The Abel limit $\beta \to 0^+$ is an essential feature of this formula, since if $\beta = 0$ the integral is in general divergent for $j \geqslant 0$. The justification of (3.1) rests on two basic results. The first is that the spectral kernel $e(x, y; \lambda)$ is absolutely continuous for $\lambda \geqslant 0$. The second is that the analytic semigroup kernels $U(x, y; \beta + it)$ have a well-defined limit as $\beta \to 0^+$, if $t \neq 0$.

Consider the behavior of the kernels $r_0(x, y; z) w(y)$ and $r_0(x, y; z) u(y)$. We have the following lemma.

Lemma 5: Assume $v \in L^1(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)$. For all $z \in \Pi_c$ and each $x \in \mathbb{R}^3$, $r_0(x, y; z) w(y)$ is an $L^2(\mathbb{R}^3)$ function of y satisfying the following.

(i) $$\|r_0(x, \cdot\, ; z) w(\cdot)\| < C_2 < \infty, \tag{3.2}$$

where C_2 is independent of both x and z.

(ii) The function $r_0(x, \cdot\, ; z) w(\cdot)$ is norm continuous in Π_c. For each $z_2 \in \Pi_c$,

$$\lim_{z_1 \to z_2} \|r_0(x, \cdot\, ; z_1) w(\cdot) - r_0(x, \cdot\, ; z_2) w(\cdot)\| = 0. \tag{3.3}$$

Formulas (3.2) and (3.3) hold with $w \leftrightarrow u$.

Proof: Fix $x \in \mathbb{R}^3$ and $z \in \Pi_c$. By definition,

$$\|r_0(x, \cdot\, ; z) w(\cdot)\|^2 = \int |e^{i\sqrt{z}|x-y|}|^2 \frac{|v(y)|}{(4\pi|x-y|)^2} dy. \tag{3.4}$$

Note for $z \in \Pi_c$ the exponential factor in the integrand is bounded by 1. Take K_x to be the unit sphere in $\mathbb{R}^3$ with center at x. For $y \notin K_x$, $|x - y| \geqslant 1$, and so one has

$$\int \frac{|v(y)|}{|x-y|^2} dy = \int_{K_x} \frac{|v(y)|}{|x-y|^2} dy + \int_{\mathbb{R}^3 \setminus K_x} \frac{|v(y)|}{|x-y|^2} dy \leqslant 4\pi \|v\|_\infty + \|v\|_1. \tag{3.5}$$

Thus

$$\|r_0(x, \cdot\, ; z) w(\cdot)\|^2 \leqslant [1/(4\pi)^2](4\pi\|v\|_\infty + \|v\|_1). \tag{3.6}$$

Similar considerations combined with the dominated convergence theorem demonstrate (3.3). □

In the following we investigate the real-axis boundary behavior of the resolvent kernel $r(x, y; z)$ with the aid of the operator

$$T(z) = u r(z) w, \quad \operatorname{Im} z \neq 0. \tag{3.7}$$

Since $r(z)$ is bounded for $|\operatorname{Im} z| > 0$ and u and w are bounded operators if $\|v\|_\infty < \infty$, it follows that $T(z)$ is bounded.

Proposition 4: Assume $(1 + |x|^2) v(x) \in L^1(\mathbb{R}^3)$ and $v \in \mathscr{A}_1 \cap \mathscr{C}_0$.

(i) $$T(z) = (1 + A(z))^{-1} A(z), \quad \operatorname{Im} z \neq 0. \tag{3.8}$$

(ii) Defining $T(\lambda + i0)$ by (3.8) with $z = \lambda + i0$,

$$\|T(z)\|_2 < M\sqrt{B_r} < \infty, \quad z \in S_1, \tag{3.9}$$

where M is the constant defined in (2.50) and B_r is the Rollnik constant in (2.12).

(iii) For $x \neq y$ and all $z \in S_1$,

$$r(x, y; z) = r_0(x, y; z) - [r_0(z) v r_0(z)](x, y) + \langle r_0(z^*) w(x, \cdot), T(z) u r_0(z)(\cdot, y)\rangle, \tag{3.10}$$

where the second term on the right denotes the kernel of the operator $r_0(z) v r_0(z)$.

(iv) For all $z \in S_1$, all $x, y \in \mathbb{R}^3$,

$$|r(x, y; z) - r_0(x, y; z)| \leqslant C_3(1 + M\sqrt{B_r}), \tag{3.11}$$

where C_3 is finite and independent of x, y, and z.

Proof: (i) Start with the operator identity

$$r(z) = r_0(z) - r_0(z) v r(z), \quad \operatorname{Im} z \neq 0. \tag{3.12}$$

If we multiply by u from the right and w from the left and use definition (3.7) then

$$T(z) = A(z) - A(z) T(z). \tag{3.13}$$

Since $(1 + A(z))^{-1}$ exists for $\operatorname{Im} z \neq 0$, (3.8) follows from (3.13). Statement (ii) is an immediate consequence of Proposition 2(ii) and Proposition 3(iii).

(iii) If we combine the identity

$$r(z) = r_0(z) - r(z) v r_0(z) \tag{3.14}$$

with (3.12), then we have for $\operatorname{Im} z \neq 0$

$$r(z) = r_0(z) - r_0(z) v r_0(z) + r_0(z) v r(z) v r_0(z). \tag{3.15}$$

Observe that the last factor on the right may be written as

$$r_0(z) v r(z) v r_0(z) = r_0(z) w T(z) u r_0(z). \tag{3.16}$$

From Proposition 1 it is known that $r(z)$ has a unique kernel for $\operatorname{Im} z \neq 0$. Furthermore the three terms on the right of (3.15) are all integral operators. Consider specifically the operator expression on the right of (3.16). Here, $T(z)$ is Schmidt class and $r_0(z) w$ and $u r_0(z)$ are bounded operators. Thus the product of these three operators is Schmidt class. The kernel of (3.16) is easily shown to be

$$[r_0(z) w T(z) u r_0(z)](x, y) = \int\int \frac{e^{i\sqrt{z}|x-y'|}}{4\pi|x-y'|} w(y') T(y', x'; z) u(x') \times \frac{e^{i\sqrt{z}|x'-y|}}{4\pi|x'-y|} dy'\, dx' = \langle r_0(x, \cdot\, ; z^*) w(\cdot), T(z) u(\cdot) r_0(\cdot, y; z)\rangle. \tag{3.17}$$

Here $T(y', x'; z)$ is the Hilbert–Schmidt kernel for $T(z)$.

Lemma 5 justifies the interpretation of the dy' integral as an inner product. Similar remarks apply to the term $r_0(z)vr_0(z)$, provided that $T(z)$ in (3.17) is replaced by the identity. Thus for $x\neq y$, (3.10) is proved. Note that the inner product form in (3.17) is simple to bound. Again from Lemma 5, we have the obvious estimate

$$
\begin{aligned}
&|\langle r_0(x,\cdot\,;z^*)w(\cdot),T(z)u(\cdot)r_0(\cdot\,,y;z)\rangle| \\
&\quad<\|r_0(x,\cdot\,;z^*)w(\cdot)\|\|u(\cdot)r_0(\cdot\,,y;z)\|\|T(z)\| \\
&\quad<(C_2)^2M\sqrt{B_r}\,, \qquad (3.18)
\end{aligned}
$$

valid for all $x,y\in\mathbb{R}^3$ and all $z\in S_1$. Formula (3.18) and its companion formula for $r_0(z)vr_0(z)$ give one the statement (3.11). □

Let us establish the absolute continuity for positive λ of the spectral kernel $e(x,y;\lambda)$. We first modify a result of Titchmarsh.[14]

Proposition 5: Let H be a self-adjoint operator on $L^2(\mathbb{R}^3)$ and $\{E_\lambda:\lambda\in\mathbb{R}\}$ the associated family of spectral projectors. For all $\lambda<\infty$ assume that E_λ are integral operators represented by

$$(E_\lambda f)(x)=\int e(x,y;\lambda)f(y)dy\,,\quad f\in L^2(\mathbb{R}^3)\,, \qquad (3.19)$$

where the spectral kernel $e(x,y;\lambda):\mathbb{R}^3\times\mathbb{R}^3\to\mathbb{C}$ is (for each fixed λ) an L^2 function on every compact subset of $\mathbb{R}^3\times\mathbb{R}^3$. Furthermore suppose that the resolvent $r(z)$ ($\operatorname{Im} z\neq0$) defined by H is an integral operator with a symmetric Carleman kernel, that is,

$$(r(z)f)(x)=\int r(x,y;z)f(y)dy\,, \qquad (3.20)$$

where

$$r(x,y;z)=r(y,x;z)\,,\quad x\neq y\,,\quad \operatorname{Im} z\neq0\,. \qquad (3.21)$$

Suppose further that λ_2,λ_1, ($\lambda_2>\lambda_1$) are a pair of real numbers not in the point spectrum of H. If (a) there exists a positive function $C(x,y)$ such that for a.a. x, $C(x,y)$ is $L^2(dy)$ on every compact subset of $\mathbb{R}^3$ and

$$\int_{\lambda_1}^{\lambda_2}|\operatorname{Im} r(x,y;\lambda+i\nu)|d\lambda<C(x,y)\,, \qquad (3.22)$$

where $C(x,y)$ is independent of $\nu\in[0,1]$; and if (b) almost everywhere in $\mathbb{R}^3\times\mathbb{R}^3$ the limit

$$\lim_{\nu\to0^+}\int_{\lambda_1}^{\lambda_2}\operatorname{Im} r(x,y;\lambda+i\nu)d\lambda \qquad (3.23)$$

exists and defines an L^2 function on compact subsets of $\mathbb{R}^3\times\mathbb{R}^3$; then

$$e(x,y;\lambda_2)-e(x,y;\lambda_1)=\lim_{\nu\to0^+}\frac{1}{\pi}\int_{\lambda_1}^{\lambda_2}\operatorname{Im} r(x,y;\lambda+i\nu)d\lambda \qquad (3.24)$$

a.e. in $\mathbb{R}^3\times\mathbb{R}^3$.

Proof: Equation (3.24) is the kernel analog of the well-known[15] operator identity $[\lambda_2,\lambda_1\notin\sigma_p(H)]$

$$E_{\lambda_2}-E_{\lambda_1}=\operatorname*{s-lim}_{\nu\to0^+}\frac{1}{2\pi i}\int_{\lambda_1}^{\lambda_2}[r(\lambda+i\nu)-r(\lambda-i\nu)]d\lambda\,. \qquad (3.25)$$

If we take the kernel form of (3.25) and use (a) and (b) to justify changing orders of integration and passing the $\nu\to0^+$ limiting process through the multiple integral arising from the right side of (3.25), then (3.24) follows without difficulty. □

Lemma 6: If $v\in\mathscr{F}^*$ then H has a unique (after adjustment on a null set of $\mathbb{R}^3\times\mathbb{R}^3$) family of spectral kernels $\{e(x,y;\lambda):\lambda\in\mathbb{R}\}$ satisfying (3.19). If in addition, v satisfies the hypothesis of Proposition 4, then for each $(x,y)\in\mathbb{R}^3\times\mathbb{R}^3$ the spectral kernel $e(x,y;\lambda)$ is absolutely continuous with respect to $d\lambda$ in the interval $\lambda\in[0,\infty)$. The derivative of $e(x,y;\lambda)$ is given by

$$\frac{d}{d\lambda}e(x,y;\lambda)=\frac{1}{\pi}\operatorname{Im} r(x,y;\lambda+i0)\,,\quad \lambda>0\,. \qquad (3.26)$$

Proof: The existence of $e(x,y;\lambda)$ for $v\in\mathscr{F}^*$ and the L^2-character for compact subsets of $\mathbb{R}^3\times\mathbb{R}^3$ of the spectral kernels are contained in Theorem 2 of OW. For alternate existence statements see Povzner[16] and Gårding.[17] To proceed we employ Proposition 5 and verify that the resolvent kernel satisfies hypotheses (a) and (b). First consider requirement (a). The absolute value of the imaginary part of representation (3.10) for $r(x,y;z)$ gives

$$
\begin{aligned}
|\operatorname{Im} r(x,y;z)|&\leqslant|\operatorname{Im} r_0(x,y;z)| \\
&\quad+|\langle r_0(z^*)w(x,\cdot),[T(z)-1]ur_0(z)(\cdot\,,y)\rangle|\,. \qquad (3.27)
\end{aligned}
$$

The inner product term is bounded by estimate (3.2) and the bound (3.9) for $T(z)$. With $z=\lambda+i\nu$, $\lambda\geqslant0$, we have the bound

$$
\begin{aligned}
&|\operatorname{Im} r(x,y;\lambda+i\nu)| \\
&\quad\leqslant(\lambda^2+1)^{1/4}/4\pi+(C_2)^2[1+M\sqrt{B_r}]\,,\quad \nu\in[0,1]\,, \qquad (3.28)
\end{aligned}
$$

uniform in x, y, and ν. Clearly, for $\lambda_2>\lambda_1\geqslant0$,

$$
\begin{aligned}
&\int_{\lambda_1}^{\lambda_2}|\operatorname{Im} r(x,y;\lambda+i\nu)|d\lambda \\
&\quad\leqslant\left\{\frac{(\lambda_2^2+1)^{1/4}}{4\pi}+(C_2)^2[1+M\sqrt{B_r}]\right\}(\lambda_2-\lambda_1)\,. \qquad (3.29)
\end{aligned}
$$

Since the right-hand side of (3.29) is finite and x,y,ν independent, hypothesis (a) is verified.

Next examine hypothesis (b). Fix $x,y,\lambda\geqslant0$ in the representation (3.10) for $r(x,y;\lambda+i\nu)$. The operator norm continuity of $T(\lambda+i\nu)-1=-[1+A(\lambda+i\nu)]^{-1}$ as given in Proposition 3(ii) combined with the vector norm continuity (3.3) of $r_0(\lambda+i\nu)w(x,\cdot)$ and $ur_0(\lambda+i\nu)(\cdot\,,y)$ imply $\operatorname{Im} r(x,y;\lambda+i\nu)$ is continuous for $\nu\in[0,1]$,

$$\lim_{\nu\to0}\operatorname{Im} r(x,y;\lambda+i\nu)=\operatorname{Im} r(x,y;\lambda+i0)\,, \qquad (3.30)$$

for $(x,y;\lambda)\in\mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}^+$. The pointwise limit (3.30) and the ν-independent integral estimate (3.29) allows application of the dominated convergence theorem to show

$$\lim_{\nu\to0^+}\int_{\lambda_1}^{\lambda_2}\operatorname{Im} r(x,y;\lambda+i\nu)d\lambda=\int_{\lambda_1}^{\lambda_2}\operatorname{Im} r(x,y;\lambda+i0)d\lambda\,. \qquad (3.31)$$

Hypothesis (b) is thus satisfied.

Combining Eqs. (3.24) and (3.31) leads to

$$e(x,y;\lambda_2) - e(x,y;\lambda_1) = \frac{1}{\pi}\int_{\lambda_1}^{\lambda_2} \mathrm{Im}\, r(x,y;\lambda + i0)d\lambda\,, \quad (3.32)$$

for all x, y. Equation (3.26) follows as an immediate consequence. □

The following theorem summarizes the existence and properties of the analytic semigroup kernels $U(x,y;z)$. Here, as in the Introduction, D denotes the open right-half complex plane and $\overline{D}$ its closure. An nth order derivative with respect to z will be indicated by the notation in Eq. (2.13).

Theorem 1: Suppose $v \in \mathscr{F}^*$. For every $z \in \overline{D}\setminus\{0\}$ the operator e^{-zH} is an integral operator with a kernel $U(x,y;z)$ given by the following series representation:

$$U(x,y;z) = U_0(x,y;z)F(x,y;z)\,, \quad (3.33)$$

where $U_0(x,y;z)$ is the free semigroup kernel in Eq. (1.7). The function F is defined by (for all $z \in \overline{D}$)

$$F(x,y;z) = 1 + \sum_{n=1}^{\infty} B_n(x,y;z)\,, \quad (3.34)$$

where

$$B_n(x,y;z) = \frac{(-z)^n}{n!}\int_0^1 \cdots \int_0^1 d\xi_1 \cdots d\xi_n \int d\mu(k_1) \cdots \int d\mu(k_n) e^{-za_n + ib_n} \quad (3.35)$$

and

$$a_n(\xi_1\cdots\xi_n;k_1\cdots k_n) = \sum_{i,j=1}^{n} k_i\cdot k_j \min\{\xi_i(1-\xi_j),\xi_j(1-\xi_i)\}\,, \quad (3.36)$$

$$b_n(\xi_1\cdots\xi_n;k_1\cdots k_n;x,y) = \sum_{i=1}^{n}((1+\xi_i)x + \xi_i y)\cdot k_i\,. \quad (3.37)$$

For each x, y the series (3.34) defines a holomorphic function in D. The series (3.34) is uniformly convergent for every compact subset of $\overline{D}$.

Proof: See Ref. 2, Theorem 1. □

Corollary 1: Let M be a non-negative integer. Suppose $v \in \mathscr{F}^*_{2M}$, then for all x, y, and $|t| \neq 0$

$$\lim_{\beta\to 0^+} U^{(j)}(x,y;\beta + it) = U^{(j)}(x,y;it)\,, \quad j = 0,1,\ldots,M\,. \quad (3.38)$$

Proof: Fix x, y and $|t| \neq 0$. The function $U_0^{(j)}(x,y;\beta+it)$ is continuous for $\beta \in [0,1]$ and for any $j \geq 0$, thus it suffices to show that $F^{(j)}(x,y;\beta+it)$ is continuous for $\beta \in [0,1]$. This in turn is established by differentiating series (3.34) term by term and showing that a convergent majorizing series exists. The jth derivative of $B_n(x,y;z)$ is

$$B_n^{(j)}(x,y;z) = \frac{(-1)^n}{n!}\int_0^1 d^n\xi \int d^n\mu \left(\frac{\partial}{\partial z}\right)^j \{z^n e^{-za_n}\} e^{ib_n}\,. \quad (3.39)$$

Here passing the derivative through the multiple integral is permissible since $v \in \mathscr{F}^*_{2M}$ and $j \leq M$. Note that

$$\left|\left(\frac{\partial}{\partial z}\right)^j \{z^n e^{-za_n}\}\right| \leq n^j |z|^n \sum_{r=0}^{j}\binom{j}{r}|z|^{r-j}|a_n|^r\,, \quad (3.40)$$

where $|e^{-za_n}| \leq 1$ since $a_n \geq 0$ (Ref. 2, Lemma 5). The symbol $\binom{j}{r}$ is the binomial coefficient. For $v \in \mathscr{F}^*_{2M}$ one has the estimate [Ref. 2, Eq. (4.13)]

$$\int d|\mu|(k_1) \cdots \int d|\mu|(k_n)|a_n|^r \leq \|\mu\|^n\left(\frac{n^2K^2}{4}\right)^r, \quad r = 0,1,\ldots,M\,. \quad (3.41)$$

If this estimate is combined with (3.40) one finds the bound

$$|B_n^{(j)}(x,y;z)| \leq \frac{n^{3j}(|z|\,\|\mu\|)^n}{n!}\left(\frac{1}{|z|} + \frac{K^2}{4}\right)^j. \quad (3.42)$$

So the majorizing sum is

$$\sum_{n=0}^{\infty}|B_n^{(j)}(x,y;z)| \leq \left(\frac{1}{|z|} + \frac{K^2}{4}\right)^j \sum_{n=0}^{\infty}\frac{n^{3j}(|z|\,\|\mu\|)^n}{n!} < \infty\,. \quad (3.43)$$

Each term $B_n^{(j)}(x,y;\beta+it)$ is continuous for $\beta \in [0,1]$. The uniform convergence of

$$F^{(j)}(x,y;\beta+it) = \sum_{n=0}^{\infty} B_n^{(j)}(x,y;\beta+it)\,, \quad j = 0,1,\ldots,M \quad (3.44)$$

shows that $F^{(j)}(x,y;\beta+it)$ is continuous for $\beta \in [0,1]$. □

Proposition 6: Assume $v \in \mathscr{F}^*_{2M}$.

(i) If $t \neq 0$, then

$$U^{(j)}(x,y;it) = \lim_{\beta\to 0^+}\int_{-\|\mu\|}^{\infty}(-\lambda)^j e^{-(\beta+it)\lambda}\, de(x,y;\lambda) \quad (3.45)$$

for all x, y and $j = 0,1,\ldots,M$.

(ii) If in addition v satisfies the hypothesis of Proposition 4, then Eq. (3.1) is valid for all x, y and $j = 0,1,\ldots,M$.

Proof: (i) First suppose $v \in \mathscr{F}^*$. In this circumstance it is proved in OW (Corollary 1) that the semigroup kernel has the representation

$$U(x,y;z) = \int_{-\|\mu\|}^{\infty} e^{-z\lambda}\, de(x,y;\lambda)\,, \quad (3.46)$$

for $z \in D$. The right side is analytic for $\mathrm{Re}\, z > 0$ and satisfies

$$U^{(j)}(x,y;z) = \int_{-\|\mu\|}^{\infty}(-\lambda)^j e^{-z\lambda}\, de(x,y;\lambda) \quad (3.47)$$

(see Widder,[18] Theorem 5a). Corollary 1 implies identity (3.45).

(ii) From the relation (3.26) for the spectral kernel, (3.45) may be written

$$U^{(j)}(x,y;it) = \lim_{\beta\to 0^+}\int_{-\|\mu\|}^{0}(-\lambda)^j e^{-(\beta+it)\lambda}\, de(x,y;\lambda) + \lim_{\beta\to 0^+}\frac{1}{\pi}\int_0^{\infty}(-\lambda)^j e^{-(\beta+it)\lambda}\,\mathrm{Im}\, r(x,y;\lambda+i0)d\lambda\,. \quad (3.48)$$

From Lemma 4 we know the first integral contains only point spectrum. Since $v \in \mathscr{F}^*$, it is also known that $de(x,y;\lambda)$ has finite total variation on the interval $[-\|\mu\|,0]$ (see Theorem 2, OW). Thus the dominated convergence theorem gives

$$\lim_{\beta\to 0^+}\int_{-\|\mu\|}^{0}(-\lambda)^j e^{-(\beta+it)\lambda}\, de(x,y;\lambda) = \sum_l (-\lambda_l)^j e^{-it\lambda_l}\psi_l(x)\psi_l(y)^*\,. \quad (3.49)$$

Lemma 4 in OW shows that the right side of (3.49) is finite for all x, y. Thus (3.1) is established for all x, y and $j = 0,1,...M$. □

In order to complete the discussion of $U(x, y;z)$, we recall the connection of this kernel to the fundamental solutions of the time-dependent Schrödinger equation.

Corollary 2: If $v \in \mathscr{F}_2^*$ then $U(x, y;it/\hbar)$ is the fundamental solution of partial differential equation (1.2) that satisfies initial condition (1.3).

Proof: See Ref. 19, Proposition 1. □

IV. A TECHNIQUE OF BUSLAEV

In this section we are concerned with the large-λ behavior of the operators $A(\lambda)A^{(n)}(\lambda)$ and $A^{(n)}(\lambda)A(\lambda)$, $n = 0,1,2,...$, $\lambda > 0$. For notational convenience, we have simply written λ for $\lambda + i0$. Thus $A^{(n)}(\lambda)$ is meant for $A^{(n)}(\lambda + i0)$. In Lemma 7 below, we shall establish the following decay estimates:

$$\|A(\lambda)A^{(n)}(\lambda)\| = O(\lambda^{-(n+1)/2}) \tag{4.1}$$

and

$$\|A^{(n)}(\lambda)A(\lambda)\| = O(\lambda^{-(n+1)/2}), \tag{4.2}$$

as $\lambda \to +\infty$, for $n = 0,1,2,...$. To prove these, we adopt a technique of Buslaev.[9]

Let $k = 0,1,2,...$, and consider the integral

$$B_k(x, x';\lambda) = \int_{\mathbb{R}^3} dy \frac{e^{i\sqrt{\lambda}|x-y|}}{4\pi|x-y|} \frac{e^{i\sqrt{\lambda}|y-x'|}}{4\pi|y-x'|} |x'-y|^k v(y). \tag{4.3}$$

Following Buslaev, we introduce a new Cartesian coordinate system $y = (y_1, y_2, y_3)$ with its center at the point $X = (x + x')/2$, and y_3 directed toward the point x. Now, in the integral (4.3), we consider x, x', and y in the new coordinate system and replace v by $\tilde{v}$, where $\tilde{v}(y) = v(X + y)$. Furthermore, let us represent y in the ellipsoidal coordinates

$$\xi = \tfrac{1}{2}(|x-y| + |y-x'|),$$
$$\eta = (|x'-y| - |y-x|)/|x-x'|. \tag{4.4}$$

Let $d = \frac{1}{2}|x - x'|$ and ϕ be the angle between the y_1-axis and the projection of y onto the (y_1, y_2) plane. In terms of the new variables, the integral (4.3) becomes

$$B_k(x,x';\lambda) = \frac{1}{(4\pi)^2} \int_d^\infty d\xi\, e^{i2\sqrt{\lambda}\xi}\, \Psi_k(\xi,d), \tag{4.5}$$

where

$$\Psi_k(\xi,d) = \int_{-1}^1 d\eta \int_0^{2\pi} d\phi\, (\xi + d\eta)^k V(\xi, \eta,\phi,d) \tag{4.6}$$

and

$$V(\xi, \eta,\phi,d) = \tilde{v}(\cos\phi \sqrt{(1-\eta^2)(\xi^2 - d^2)},$$
$$\sin\phi \sqrt{(1-\eta^2)(\xi^2-d^2)}, \xi\eta). \tag{4.7}$$

The function $\tilde{v}$ in (4.7) depends explicitly on the new center X as well as the Euler angles that characterize the rotation involved in the change of the original coordinate system to the one with center at X. Let us now assume $v \in \mathscr{A}_k$. Integration by parts then gives

$$(4\pi)^2 B_k(x,x';\lambda) = -\frac{\Psi_k(d,d)}{i2\sqrt{\lambda}} e^{i2\sqrt{\lambda}d} - \frac{1}{i2\sqrt{\lambda}} \int_d^\infty d\xi\, e^{i2\sqrt{\lambda}\xi} \frac{\partial}{\partial\xi} \Psi_k(\xi,d). \tag{4.8}$$

Note that, in view of (1.21) with $L = 0$, we have $|\tilde{v}| \leqslant \|\mu\|$. Thus it is easily seen that

$$|\Psi_k(d,d)| \leqslant [4\pi/(k+1)]\,|x-x'|^k \|\mu\|, \tag{4.9}$$

and that the first term on the right-hand side of (4.8) is dominated by $2\pi|x-x'|^k \|\mu\|/(k+1)\sqrt{\lambda}$. Next observe that

$$\frac{\partial}{\partial\xi} V(\xi, \eta,\phi,d) = \frac{\partial\tilde{v}}{\partial y_1} \frac{\sqrt{1-\eta^2}}{\sqrt{\xi^2-d^2}} \xi\cos\phi + \frac{\partial\tilde{v}}{\partial y_2} \frac{\sqrt{1-\eta^2}}{\sqrt{\xi^2-d^2}} \xi\sin\phi + \frac{\partial\tilde{v}}{\partial y_3} \eta \tag{4.10}$$

and

$$\frac{\partial}{\partial\xi} \Psi_k(\xi,d) = k\Psi_{k-1}(\xi,d) + \int_{-1}^1 d\eta \times \int_0^{2\pi} d\phi\, (\xi + d\eta)^k \frac{\partial}{\partial\xi} V(\xi, \eta,\phi,d). \tag{4.11}$$

Hence

$$\int_d^\infty d\xi \left|\frac{\partial}{\partial\xi} \Psi_k(\xi,d)\right| \leqslant k \int_d^\infty d\xi\, |\Psi_{k-1}(\xi,d)| + 2^k \int_d^\infty d\xi \int_{-1}^1 d\eta \int_0^{2\pi} d\phi\, \xi^k \left\{\frac{\xi}{\sqrt{\xi^2-d^2}}\right\} |\nabla\tilde{v}|. \tag{4.12}$$

Since $\Psi_{k-1}(\xi,d) = O(\xi^{-1-\epsilon})$ in view of (1.22), the first term on the right-hand side of (4.12) is finite. We now impose our second condition: $v \in \mathscr{B}_k$. This condition ensures that the second term on the right of (4.12) is also finite. Let us now rewrite Eq. (4.8) as

$$B_k(x,x';\lambda) \equiv B_k^{(1)}(x,x';\lambda) + B_k^{(2)}(x,x';\lambda). \tag{4.13}$$

If $(1 + |x|)^{2(k+1)} v(x) \in L^1(\mathbb{R}^3)$, then the estimates (4.9) and (4.12) show that for $i = 1,2$, $u(x)B_k^{(i)}(x,x';\lambda)w(x')$ is a Schmidt kernel with norm being $O(\lambda^{-1/2})$.

Lemma 7: Let $v(x) \in L^2(\mathbb{R}^3)$ and $(1 + |x|)^{2(n+1)} v(x) \in L^1(\mathbb{R}^3)$. If, in addition, $v(x) \in \mathscr{A}_n \cap \mathscr{B}_n$, then the operators $A(\lambda)A^{(n)}(\lambda)$ and $A^{(n)}(\lambda)A(\lambda)$ are of Schmidt class and satisfy the norm estimates in (4.1) and (4.2), respectively.

Proof: We first recall the identity

$$\frac{d^n}{dx^n} e^{c\sqrt{x}} = \sum_{j=0}^n d_{jn} \frac{c^{n-j}}{(2\sqrt{x})^{n+j}} e^{c\sqrt{x}}, \tag{4.14}$$

where c is any complex constant and the coefficients d_{jn} are also constants with $d_{00} = 1$ and $d_{nn} = 0$ for $n = 1,2,...$; see Ref. 20, p. 20. From (2.14), it then follows that there are constants c_{jn} with $c_{00}\neq 0$, for $n = 1,2,\ldots$, such that the kernel for $A^{(n)}(\lambda)$ is given by

$$A^{(n)}(\lambda)(x,y) = A^{(n)}(x,y;\lambda) = \frac{1}{4\pi}\sum_{j=0}^{n}\frac{c_{jn}}{\lambda^{(n+j)/2}}b_{n-j}(x,y;\lambda), \tag{4.15}$$

where

$$b_k(x,y;\lambda) = u(x)|x-y|^{k-1}e^{i\sqrt{\lambda}|x-y|}w(y) \tag{4.16}$$

and the derivatives are taken with respect to λ. This in turn implies that the kernel for $A(\lambda)A^{(n)}(\lambda)$ is given by

$$[A(\lambda)A^{(n)}(\lambda)](x,y) = \int_{\mathbf{R}^3} A(x,x';\lambda)A^{(n)}(x',y;\lambda)dx' = \sum_{j=0}^{n}\frac{c_{jn}}{\lambda^{(n+j)/2}}u(x)B_{n-j}(x,y;\lambda)w(y), \tag{4.17}$$

where $B_k(x,y;\lambda)$ is as defined in Eq. (4.3). Since each term under the sum is of Schmidt class with norm being $O(\lambda^{-1/2})$, we have the estimate in (4.1). A similar argument establishes the result in (4.2). This completes the proof of the lemma.

□

Lemma 8: Let $v\in\mathscr{A}_{s+2}\cap\mathscr{B}_{s+2}$ and $(1+|x|)^{2s+2}\times v(x)\in L^1(\mathbf{R}^3)$. Define

$$\epsilon_L(x,y;\lambda) = \langle r_0(\lambda^*)w(x,\cdot),\ [1+A(\lambda)]^{-1}A^{L+1}(\lambda)ur_0(\lambda)(\cdot,y)\rangle. \tag{4.18}$$

Then, as $\lambda\to+\infty$,

$$\left(\frac{d}{d\lambda}\right)^{s+2}\epsilon_{3s+3}(x,y;\lambda) = O(\lambda^{-s-1-1/4}). \tag{4.19}$$

Proof: Since $(1+|x|)^{2s+2}v(x)\in L^1(R^3)$, we have

$$\left(\frac{d}{d\lambda}\right)^{s+2}\epsilon_{3s+3}(x,y;\lambda) = \sum_{l=0}^{s+2}\binom{s+2}{l}\left\langle\left(\frac{d}{d\lambda}\right)^{s+2-l}r_0(\lambda^*)w(x,\cdot),\ \left(\frac{d}{d\lambda}\right)^{l}(1+A(\lambda))^{-1}A^{3s+4}(\lambda)ur_0(\lambda)(\cdot,y)\right\rangle. \tag{4.20}$$

The second element in the inner product in (4.20) can be written in the form

$$\sum_{i+m+n=l}c_{imn}\left(\frac{d}{d\lambda}\right)^{i}(1+A(\lambda))^{-1}\left(\frac{d}{d\lambda}\right)^{m}A^{3s+4}(\lambda)\times\left(\frac{d}{d\lambda}\right)^{n}ur_0(\lambda)(\cdot,y). \tag{4.21}$$

Simple calculation then shows that

$$\left\|\left(\frac{d}{d\lambda}\right)^{s+2-l}r_0(\lambda)w(x,\cdot)\right\| = O(\lambda^{-(s+2-l)/2}) \tag{4.22}$$

and

$$\left\|\left(\frac{d}{d\lambda}\right)^{n}ur_0(\lambda)(\cdot,y)\right\| = O(\lambda^{-n/2}). \tag{4.23}$$

By Proposition 2(v), we also have

$$\left\|\left(\frac{d}{d\lambda}\right)^{i}(1+A(\lambda))^{-1}\right\| = O(\lambda^{-i/2}). \tag{4.24}$$

Thus, to prove (4.19), it suffices to show that

$$\left\|\left(\frac{d}{d\lambda}\right)^{m}A^{3s+4}(\lambda)\right\| = O(\lambda^{-(m+s+1/2)/2}). \tag{4.25}$$

The operator inside the norm is equal to a finite sum of products of the derivatives of $A(\lambda)$. Note that each product has $3s+4$ factors, and at most m factors involve derivatives. The constraints on the summations in (4.20) and (4.21) imply $m\leqslant s+2$. Thus there are at least $2s+2$ factors of $A(\lambda)$ not involving derivatives. If the undifferentiated $A(\lambda)$'s can be grouped into products of $A^2(\lambda)$, then by Proposition 2(v) and Lemma 7 (with $n=0$) we have $\|(d/d\lambda)^m A^{3s+4}(\lambda)\| = O(\lambda^{-(m+s+1)/2})$, which implies (4.25). The worst possible case is when the undifferentiated $A(\lambda)$'s are separated by differentiated ones as frequently as possible. Let us consider a term in the finite sum, which is of the form

$$A(\lambda)^{\mu_1+1}A^{(m_1)}(\lambda)A^{(n_1)}(\lambda)A^{(m_2)}(\lambda)A(\lambda)^{\mu_2+1}\times A^{(m_3)}(\lambda)A(\lambda)\cdots A(\lambda)A^{(m_i)}(\lambda)A^{(n_j)}(\lambda)\times A^{(m_{i+1})}(\lambda)A(\lambda)\cdots A(\lambda)A^{(m_u)}(\lambda)A(\lambda)^{\mu_w+1}, \tag{4.26}$$

where $1\leqslant m_i\leqslant m$ for $1\leqslant i\leqslant u$, $0\leqslant n_j\leqslant m$ for $1\leqslant j\leqslant v$, and $m_1+\cdots+m_u+n_1+\cdots+n_v=m$. Note that $u\leqslant m$. The worst possible case is included in (4.26) by taking $m_1=\cdots=m_u=1$ and $n_1=\cdots=n_v=0$. Terms not of the form in (4.26) can be handled in a similar but simpler manner. We shall always group $A^{(m_i)}(\lambda)$ with an $A(\lambda)$ so that the results in Lemma 7 can be applied. For instance, in (4.26), we have

$$\begin{gathered}\|A(\lambda)A^{(m_1)}(\lambda)\| = O(\lambda^{-(m_1+1)/2})\\ \|A(\lambda)^{(m_2)}A(\lambda)\| = O(\lambda^{-(m_2+1)/2})\\ \vdots\end{gathered} \tag{4.27}$$

To $A^{(n_j)}(\lambda)$, we apply Proposition 2(v). Thus, $\|A^{(n_j)}(\lambda)\| = O(\lambda^{-n_j/2})$. Let μ_k denote the number of factors in a block $A(\lambda)\cdots A(\lambda)$ between two consecutive derivatives of $A(\lambda)$, as indicated in (4.26). There are totally $w+1$ such blocks, where $w=u-v$. Thus we have the relationship

$$\mu_1+\cdots+\mu_{w+1}+v+2u = 3s+4, \tag{4.28}$$

from which we have

$$\sum_{i=1}^{w+1}\left[\frac{1}{2}\mu_i\right]\geqslant s+\frac{1}{2}\{s+4-v-2u-(w+1)\}, \tag{4.29}$$

where $[\frac{1}{2}\mu_i]$ stands for the largest integer less than or equal to $\frac{1}{2}\mu_i$. The product in (4.26) can now be shown to be of the order

$$O(\lambda^{-(m+u+\Sigma_{i=1}^{w+1}[\mu_i/2])/2}).$$

[Note that $\|A^2(\lambda)\| = O(\lambda^{-1/2})$ by Lemma 7.] Since, by the inequality in (4.29),

$$u+\sum_{i=1}^{w+1}\left[\frac{1}{2}\mu_i\right]\geqslant s+\frac{1}{2}, \tag{4.30}$$

this product is, in fact, of the order $O(\lambda^{-(m+s+1/2)/2})$. This establishes the estimate in (4.25) and hence the lemma. □

V. THE SMALL-λ EXPANSION OF $\operatorname{Im} r(x,y;\lambda+i0)$

As in Sec. IV, we shall again write λ for $\lambda+i0$. Thus, from Proposition 4 we have

$$\begin{aligned} r(x,y;\lambda) &= r_0(x,y;\lambda) - [r_0(\lambda)vr_0(\lambda)](x,y) \\ &\quad + \langle r_0(\lambda^*)w(x,\cdot),(1+A(\lambda))^{-1}A(\lambda)ur_0(\lambda)(\cdot,y)\rangle, \end{aligned} \tag{5.1}$$

if $(1+|x|^2)v(x)\in L^1(\mathbb{R}^3)$ and $v\in\mathscr{A}_1\cap\mathscr{C}_0$. In order to obtain the small-λ expansion of $\operatorname{Im} r(x,y;\lambda+i0)$, we first need the following lemmas.

Lemma 9: Assume $(1+|x|^2)v(x)\in L^1(\mathbb{R}^3)$ and $v\in\mathscr{A}_1\cap\mathscr{C}_0$. Then, for any $L\geqslant 1$,

$$\begin{aligned} r(x,y;\lambda) &= r_0(x,y;\lambda) - [r_0(\lambda)vr_0(\lambda)](x,y) \\ &\quad + \sum_{l=0}^{L-1}(-1)^l[r_0(\lambda)(vr_0(\lambda))^{l+2}](x,y) \\ &\quad + (-1)^L\langle r_0(\lambda^*)w(x,\cdot), \\ &\quad (1+A(\lambda))^{-1}A^{L+1}(\lambda)ur_0(\lambda)(\cdot,y)\rangle, \end{aligned} \tag{5.2}$$

where the lth term in the finite sum on the right-hand side denotes the kernel of the operator $r_0(\lambda)(vr_0(\lambda))^{l+2}$.

Proof: For any $k\geqslant 2$, we have the operator identity

$$r_0(\lambda)(vr_0(\lambda))^k = (r_0(\lambda)w)(A(\lambda))^{k-1}(ur_0(\lambda)). \tag{5.3}$$

The kernel form of this identity is given by

$$[r_0(\lambda)(vr_0(\lambda))^k](x,y) = [r_0(\lambda)wA^{k-1}(\lambda)ur_0(\lambda)](x,y). \tag{5.4}$$

The right-hand side of (5.4) is a multiple integral and can be reinterpreted as the inner product

$$\langle r_0(\lambda^*)w(x,\cdot),[A^{k-1}(\lambda)ur_0(\lambda)](\cdot,y)\rangle, \tag{5.5}$$

since both $r_0(\lambda^*)w(x,\cdot)$ and $ur_0(\lambda)(\cdot,y)$ are square integrable. Thus

$$\begin{aligned} &[r_0(\lambda)(vr_0(\lambda))^k](x,y) \\ &\quad = \langle r_0(\lambda^*)w(x,\cdot),[A^{k-1}(\lambda)ur_0(\lambda)](\cdot,y)\rangle. \end{aligned} \tag{5.6}$$

The desired result in (5.2) now follows from (5.1) and the expansion

$$\begin{aligned} (1+A(\lambda))^{-1} &= 1 - A(\lambda) + \cdots + (-1)^{L-1}A^{L-1}(\lambda) \\ &\quad + (-1)^L(1+A(\lambda))^{-1}A^L(\lambda). \end{aligned} \tag{5.7}$$

□

Lemma 10: Let $v\in\mathscr{F}^*\cap L^1(\mathbb{R}^3)$. Then, for $n\geqslant 1$,

$$\begin{aligned} &\lim_{\beta\to 0^+}\frac{1}{\pi}\int_0^\infty \operatorname{Im}\{[r_0(\lambda)(vr_0(\lambda))^n](x,y)\}e^{-\beta\lambda-it\lambda}\,d\lambda \\ &= \frac{1}{(4\pi)^{n+3/2}}\frac{1}{(it)^{3/2}}\int dy_1\cdots dy_n \\ &\quad\times\left\{\frac{v(y_1)\cdots v(y_n)[\,|x-y_1|+\cdots+|y_n-y|\,]}{|x-y_1||y_1-y_2|\cdots|y_n-y|}\right\} \\ &\quad\times\exp\left[\frac{i}{4t}(|x-y_1|+\cdots+|y_n-y|)^2\right]. \end{aligned} \tag{5.8}$$

Formula (5.8) also holds for $n=0$ if the multiple integral in (5.8) is replaced by $\exp((i/4t)|x-y|^2)$.

Proof: For $k\geqslant 0$ and $\operatorname{Re} z>0$, we have

$$\frac{e^{-k/z}}{z^{3/2}} = \frac{1}{\sqrt{\pi k}}\int_0^\infty e^{-z\lambda}\sin(2\sqrt{k\lambda})\,d\lambda; \tag{5.9}$$

see Ref. 20, p. 1146. Let

$$I \equiv \frac{1}{\pi}\int_0^\infty \operatorname{Im}\{[r_0(\lambda)(vr_0(\lambda))^n](x,y)\}e^{-z\lambda}\,d\lambda, \tag{5.10}$$

and note that

$$\begin{aligned} &[r_0(\lambda)(vr_0(\lambda))^n](x,y) \\ &\quad = \left(\frac{1}{4\pi}\right)^{n+1}\int dy_1\cdots dy_n\left\{\frac{v(y_1)\cdots v(y_n)}{|x-y_1|\cdots|y_n-y|}\right\} \\ &\qquad\times\exp[i\sqrt{\lambda}(|x-y_1|+\cdots+|y_n-y|)]. \end{aligned} \tag{5.11}$$

Since $v\in\mathscr{F}^*\cap L^1(\mathbb{R}^3)\subset L^2(\mathbb{R}^3)\cap L^1(\mathbb{R}^3)$, it follows that $v\in\mathscr{R}$ (see Ref. 8, Corollary I.14). Hence the above multiple integral is absolutely convergent. By Fubini's theorem

$$\begin{aligned} I &= \frac{1}{\pi}\left(\frac{1}{4\pi}\right)^{n+1}\int dy_1\cdots dy_n\left\{\frac{v(y_1)\cdots v(y_n)}{|x-y_1|\cdots|y_n-y|}\right\} \\ &\quad\times\int_0^\infty \sin[\sqrt{\lambda}(|x-y_1|+\cdots+|y_n-y|)]e^{-z\lambda}\,d\lambda. \end{aligned} \tag{5.12}$$

Now we use (5.9) with $k=\frac{1}{4}(|x-y_1|+\cdots+|y_n-y|)^2$. The result is

$$\begin{aligned} I &= \frac{1}{(4\pi)^{n+3/2}}\frac{1}{z^{3/2}}\int dy_1\cdots dy_n \\ &\quad\times\left\{\frac{v(y_1)\cdots v(y_n)[\,|x-y_1|+\cdots+|y_n-y|\,]}{|x-y_1||y_1-y_2|\cdots|y_n-y|}\right\} \\ &\quad\times\exp\left[-\frac{1}{4z}(|x-y_1|+\cdots+|y_n-y|)^2\right]. \end{aligned} \tag{5.13}$$

Write $z=\beta+it$ and let β tend to zero. The result in (5.8) follows immediately since it can be shown that

$$\begin{aligned} &\int dy_1\cdots dy_n\left|\left\{\frac{v(y_1)\cdots v(y_n)[\,|x-y_1|+\cdots+|y_n-y|\,]}{|x-y_1||y_1-y_2|\cdots|y_n-y|}\right\}\right| \\ &\quad\leqslant L<\infty, \end{aligned} \tag{5.14}$$

where L is independent of x and y. □

Lemma 11: Let $v\in\mathscr{F}^*$ and $(1+|x|)^{2(m+2)}\times v(x)\in L^1(\mathbb{R}^3)$. Then

$$A(\lambda) = A_0 + i\sqrt{\lambda}A_1 + \lambda A_2 + R_3(\lambda), \tag{5.15}$$

where each A_i, $i=0,1,2$, is a Schmidt class operator with a real-valued kernel, and $R_3(\lambda)$ satisfies

$$\|R_3(\lambda)\|_2 = O(\lambda^{3/2}),\quad\text{as }\lambda\to 0^+. \tag{5.16}$$

Furthermore, the expansion (5.15) can be termwise differentiated m times, and, for $\nu=0,1,\ldots,m$,

$$\|R_3^{(\nu)}(\lambda)\|_2 = O(\lambda^{3/2-\nu}),\quad\text{as }\lambda\to 0^+. \tag{5.17}$$

Proof: We first recall the Taylor formula

$$e^{icx} = 1 + icx - \frac{(cx)^2}{2!} - i\frac{(cx)^3}{2!}\int_0^1(1-t)^2e^{icxt}\,dt. \tag{5.18}$$

With $c = |x-y|$ and $x = \sqrt{\lambda}$, we have, from the definition of $A(x,y;\lambda)$,

$$A(x,y;\lambda) = A_0(x,y) + i\lambda^{1/2}A_1(x,y) + \lambda A_2(x,y) + R_3(x,y;\lambda), \tag{5.19}$$

where

$$A_0(x,y) = u(x)w(y)/4\pi|x-y|, \tag{5.20}$$

$$A_1(x,y) = u(x)w(y)/4\pi, \tag{5.21}$$

and

$$A_2(x,y) = -(|x-y|/8\pi)u(x)w(y). \tag{5.22}$$

The remainder has the explicit form

$$R_3(x,y;\lambda) = -i\frac{\lambda^{3/2}}{8\pi}|x-y|^2u(x)w(y) \times \int_0^1 (1-t)^2 e^{i\sqrt{\lambda}|x-y|t}\,dt. \tag{5.23}$$

Expansion (5.19) is the kernel form of the expansion in (5.15). It is easy to see that if $(1+|x|)^4v(x) \in L^1(\mathbb{R}^3)$ then for each $i = 0,1,2$, $\|A_i\|_2 < \infty$. Furthermore,

$$\|R_3(\lambda)\|_2 = O(\lambda^{3/2}), \quad \text{as } \lambda \to 0^+. \tag{5.24}$$

The second statement of the lemma now follows from differentiating expansion (5.19) and applying Eq. (4.14) to Eq. (5.23). [Observe that in (4.14), $d_{00} \neq 0$.] This completes the proof of the lemma. □

Lemma 12: Let v be as in Lemma 11, and suppose further that $v \in \mathscr{A}_1 \cap \mathscr{C}_0$. Then we have

$$(1+A(\lambda))^{-1} = B_0 + i\sqrt{\lambda}B_1 + \lambda B_2 + E_3(\lambda), \tag{5.25}$$

where B_0 is a bounded operator which maps a real-valued L^2-function into a real-valued function, B_1 and B_2 are Schmidt class operators with real-valued kernels, and $E_3(\lambda)$ satisfies

$$\|E_3(\lambda)\|_2 = O(\lambda^{3/2}), \quad \text{as } \lambda \to 0^+. \tag{5.26}$$

Furthermore, the expansion in (5.25) can be termwise differentiated m times, and

$$\|E_3^{(\nu)}(\lambda)\|_2 = O(\lambda^{3/2-\nu}), \quad \text{as } \lambda \to 0^+. \tag{5.27}$$

for $\nu = 0,1,\dots,m$.

Proof: Since $(1+A(\lambda))^{-1}$ is a function of $\sqrt{\lambda}$, it is convenient to put $(1+A(\lambda))^{-1} = f(\sqrt{\lambda})$ and introduce the notation

$$A^{[n]}(\lambda) = \left(\frac{d}{d\sqrt{\lambda}}\right)^n A(\lambda). \tag{5.28}$$

Simple calculations then show that $(d/d\sqrt{\lambda})^l f(\sqrt{\lambda})$ is equal to a finite sum of the form

$$(1+A(\lambda))^{-1}A^{[m_1]}(\lambda)(1+A(\lambda))^{-1} \times A^{[m_2]}(\lambda)\cdots(1+A(\lambda))^{-1}A^{[m_\mu]}(\lambda)(1+A(\lambda))^{-1}, \tag{5.29}$$

where $m_1 + \cdots + m_\mu = l$. In particular, we have

$$f(0) = (1+A_0)^{-1}, \tag{5.30}$$

$$\frac{d}{d\sqrt{\lambda}}f(\sqrt{\lambda})\bigg|_{\lambda=0} = -i(1+A_0)^{-1}A_1(1+A_0)^{-1}, \tag{5.31}$$

and

$$\left(\frac{d}{d\sqrt{\lambda}}\right)^2 f(\sqrt{\lambda})\bigg|_{\lambda=0} = -2(1+A_0)^{-1}\{A_1(1+A_0)^{-1}A_1 + A_2\}(1+A_0)^{-1}, \tag{5.32}$$

where A_0, A_1, and A_2 are the operators appearing as coefficients in the expansion (5.15). An application of Taylor's theorem with remainder then gives

$$(1+A(\lambda))^{-1} = B_0 + i\sqrt{\lambda}B_1 + \lambda B_2 + E_3(\lambda), \tag{5.33}$$

with $B_0 = (1+A_0)^{-1}$, $B_1 = -B_0A_1B_0$, $B_2 = B_1A_1B_0 - B_0A_2B_0$, and

$$E_3(\lambda) = \frac{\lambda^{3/2}}{2!}\int_0^1 (1-t)^2 f^{(3)}(\sqrt{\lambda}\,t)\,dt. \tag{5.34}$$

In (5.34), it is understood that

$$f^{(3)}(\sqrt{\lambda}\,t) = \left(\frac{d}{d\sqrt{u}}\right)^3 f(\sqrt{u})\bigg|_{\sqrt{u}=\sqrt{\lambda}\,t}. \tag{5.35}$$

To obtain norm bounds for B_i, $i = 0,1,2$, we apply Proposition 3 and Lemma 11. The results are $\|B_0\| \leqslant M$, $\|B_1\|_2 \leqslant M^2\|A_1\|_2$, and $\|B_2\|_2 \leqslant M^3\|A_1\|_2^2 + M^2\|A_2\|_2$, where M is the constant appearing in (2.50).

By the mean-value theorem for integrals, the remainder in (5.34) can be written as

$$E_3(\lambda) = \frac{\lambda^{3/2}}{3!}f^{(3)}(\sqrt{\lambda}\,t_1) = \frac{\lambda^{3/2}}{3!}\left(\frac{d}{d\sqrt{u}}\right)^3(1+A(u))^{-1}\bigg|_{\sqrt{u}=\sqrt{\lambda}\,t_1}, \tag{5.36}$$

where $0 < t_1 < 1$. The third-order derivative on the right-hand side is equal to a finite sum of terms of the form given in (5.29). By Proposition 3, it is easily seen that this derivative is bounded by

$$6M^4\|A^{[1]}(\lambda\xi)\|_2^3 + 6M^3\|A^{[2]}(\lambda\xi)\|_2\|A^{[1]}(\lambda\xi)\|_2 + M^2\|A^{[3]}(\lambda\xi)\|_2, \tag{5.37}$$

where $\xi = t_1^2 \in (0,1)$. Since $A(\lambda)$ is an analytic function in $\sqrt{\lambda}$, for each $i = 0,1,2,\dots$, $\|A^{[i]}(\lambda\xi)\|_2$ is bounded for bounded λ. This establishes the result in (5.26). To prove the O-relation in (5.27), we first recall the formula of Faa di Bruno

$$\left(\frac{d}{dx}\right)^n F(\sqrt{x}) = \frac{F^{(n)}(\sqrt{x})}{(2\sqrt{x})^n} - \frac{n(n-1)}{1!}\frac{F^{(n-1)}(\sqrt{x})}{(2\sqrt{x})^{n+1}} + \frac{(n+1)n(n-1)(n-2)}{2!}\frac{F^{(n-2)}(\sqrt{x})}{(2\sqrt{x})^{n+2}} + \cdots. \tag{5.38}$$

(See Ref. 20, p. 20.) By this formula, together with essentially the same argument as given for (5.26), it is easily shown that

$$\left\|\left(\frac{d}{d\lambda}\right)^k f^{(3)}(\sqrt{\lambda}\,t_1)\right\|_2 = O(\lambda^{-k+1/2}), \quad k = 1,2,\dots,m. \tag{5.39}$$

(The exponent 1/2 is absent when $k = 0$.) The estimate in (5.27) now follows immediately, after a simple application of Leibniz's rule to the first equality in (5.36).

We finally come to show the reality properties of the operators B_i, $i=0,1,2$. The fact that $B_0=(1+A_0)^{-1}$ maps real-valued functions to real-valued functions is obvious, since A_0 has a real-valued kernel (Lemma 11) and $v\in\mathscr{E}_0$. This also implies the operators B_1 and B_2 have the same property. Since B_1 and B_2 are Schmidt class operators, it can now be shown that they must have real-valued kernels. □

Lemma 13: Assume $v\in\mathscr{F}^*$ and $(1+|x|)^{2(m+2)}v(x)\in L^1(\mathbb{R}^3)$. Then

$$\frac{u(x)}{4\pi|x-x'|}e^{i\sqrt{\lambda}|x-x'|}=a_0(x,x')+i\sqrt{\lambda}\,a_1(x,x')+\lambda a_2(x,x')+\Phi_3(x,x',\lambda), \quad (5.40)$$

and similarly

$$\frac{w(y)}{4\pi|x'-y|}e^{i\sqrt{\lambda}|x'-y|}=b_0(x',y)+i\sqrt{\lambda}\,b_1(x',y)+\lambda b_2(x',y)+\Psi_3(x',y;\lambda), \quad (5.41)$$

where the coefficients $a_i(x,x')$ and $b_i(x',y)$, $i=0,1,2$, are all real-valued square integrable functions of x' in $\mathbb{R}^3$, and the remainders satisfy

$$\|\Phi_3(x,\cdot;\lambda)\|_2\leqslant\lambda^{3/2}M_x \quad (5.42)$$

and

$$\|\Psi_3(\cdot,y;\lambda)\|_2\leqslant\lambda^{3/2}N_y \quad (5.43)$$

for all sufficiently small λ, M_x and N_y being constants depending on x and y, respectively. Furthermore, the expansions (5.40) and (5.41) can be termwise differentiated m times with respect to λ, and for all sufficiently small λ,

$$\|\Phi_3^{(\nu)}(x,\cdot;\lambda)\|_2\leqslant\lambda^{3/2-\nu}M_{x,\nu} \quad (5.44)$$

and

$$\|\Psi_3^{(\nu)}(\cdot,y;\lambda)\|_2\leqslant\lambda^{3/2-\nu}N_{y,\nu}, \quad (5.45)$$

$\nu=1,2,...,m$, where $M_{x,\nu}$ and $N_{y,\nu}$ are again constants independent of λ.

Proof: Since the arguments here are similar to those given in Lemma 11, we omit them altogether. □

We are now ready to present our main result of the section. First we return to Eq. (5.1) and observe that to obtain the small-λ behavior of Im $r(x,y;\lambda)$, it suffices to consider the function

$$h(x,y;\lambda)\equiv\mathrm{Im}\langle r_0(\lambda^*)w(x,\cdot),\; [1+A(\lambda)]^{-1}A(\lambda)ur_0(\lambda)(\cdot,y)\rangle. \quad (5.46)$$

Proposition 7: Let $v\in\mathscr{F}^*$ and $(1+|x|)^{2(m+2)}v(x)\in L^1(\mathbb{R}^3)$, and suppose that $v\in\mathscr{A}_1\cap\mathscr{E}_0$. Then

$$h(x,y;\lambda)=\lambda^{1/2}h_1(x,y)+\phi_2(x,y;\lambda), \quad (5.47)$$

where $h_1(x,y)$ is uniformly bounded in x and y, and

$$\phi_2(x,y;\lambda)=O(\lambda^{3/2}),\quad \text{as } \lambda\to0^+. \quad (5.48)$$

Furthermore, Eq. (5.47) can be differentiated m times with respect to λ and

$$\phi_2^{(\nu)}(x,y;\lambda)=O(\lambda^{3/2-\nu}),\quad \text{as } \lambda\to0^+, \quad (5.49)$$

for $\nu=1,...,m$. The O-symbols in (5.48) and (5.49) are uniform with respect to x and y for (x,y) in any compact subset of $\mathbb{R}^3\times\mathbb{R}^3$.

Proof: By Lemmas 11 and 12, we have

$$(1+A(\lambda))^{-1}A(\lambda)=D_0+i\sqrt{\lambda}D_1+\lambda D_2+\widetilde{D}_3(\lambda), \quad (5.50)$$

where D_i, $i=0,1,2$, and $\widetilde{D}_3(\lambda)$ are Schmidt class operators. The kernels of D_i, denoted by $D_i(x,y)$, are real valued, and the remainder $\widetilde{D}_3(\lambda)$ satisfies

$$\|\widetilde{D}_3(\lambda)\|_2=O(\lambda^{3/2}),\quad \text{as } \lambda\to0^+. \quad (5.51)$$

The coefficients D_i are explicitly given by $D_0=B_0A_0$, $D_1=B_1A_0+B_0A_1$, and $D_2=B_2A_0-B_1A_1+B_0A_2$. A combination of (5.46), (5.50), and Lemma 13 gives the desired approximation (5.47) with

$$h_1(x,y)=\int\int dy_1\,dy_2\{b_0(x,y_1)D_0(y_1,y_2)a_1(y_2,y)+b_1(x,y_1)D_0(y_1,y_2)a_0(y_2,y)+b_0(x,y_1)D_1(y_1,y_2)a_0(y_2,y)\}. \quad (5.52)$$

A similar formula exists for the remainder ϕ_2 in (5.47). We next observe that the coefficients $a_i(x,x')$ and $b_i(x',y)$, $i=0,1,2$, in (5.40) and (5.41) have the norm estimates $\|a_0(\cdot,y)\|_2<C$, $\|b_0(x,\cdot)\|_2<C$, $\|a_i(\cdot,y)\|_2\leqslant C(1+|y|)^{i-1}$, and $\|b_i(x,\cdot)\|_2\leqslant C(1+|x|)^{i-1}$, $i=1,2$, C being independent of x and y. These estimates, together with (5.52), immediately give

$$|h_1(x,y)|\leqslant C_1, \quad (5.53)$$

for some constant C_1. The remainders $\Phi_3(x,y;\lambda)$ and $\Psi_3(x,y;\lambda)$ in (5.40) and (5.41) also satisfy the norm estimates

$$\|\Phi_3^{(\nu)}(x,\cdot;\lambda)\|_2\leqslant M_\nu\lambda^{3/2-\nu}(1+|x|)^{2+\nu}, \quad (5.54)$$

and

$$\|\Psi_3^{(\nu)}(\cdot,y;\lambda)\|_2\leqslant N_\nu\lambda^{3/2-\nu}(1+|y|)^{2+\nu}, \quad (5.55)$$

for $\nu=0,1,...,m$. From these estimates, it is now easy to see that we have

$$\phi_2^{(\nu)}(x,y;\lambda)=O(\lambda^{3/2-\nu}),\quad \text{as } \lambda\to0^+, \quad (5.56)$$

$\nu=0,1,...,m$, uniformly valid for x and y restricted to any fixed compact subset of $\mathbb{R}^3$. □

VI. THE LARGE-t EXPANSION OF $S(x,y;t)$

Our starting point is the representation

$$S(x,y;t)=\lim_{\beta\to0^+}\frac{1}{\pi}\int_0^\infty e^{-(\beta+it)\lambda}\,\mathrm{Im}\,r(x,y;\lambda)d\lambda \quad (6.1)$$

given in (3.1). In view of (5.1) we can rewrite this as

$$S(x,y;t)=(1/\pi)[H_0(x,y;t)-H_1(x,y;t)+H_2(x,y;t)], \quad (6.2)$$

where

$$H_0(x,y;t)=\lim_{\beta\to0^+}\int_0^\infty e^{-(\beta+it)\lambda}\,\mathrm{Im}\,r_0(x,y;\lambda)d\lambda, \quad (6.3)$$

$$H_1(x,y;t)=\lim_{\beta\to0^+}\int_0^\infty e^{-(\beta+it)\lambda}\,\mathrm{Im}\{[r_0(\lambda)vr_0(\lambda)](x,y)\}d\lambda, \quad (6.4)$$

and

$$H_2(x,y;t)=\lim_{\beta\to0^+}\int_0^\infty e^{-(\beta+it)\lambda}h(x,y;\lambda)d\lambda. \quad (6.5)$$

The function $h(x, y;\lambda)$ is defined in (5.46), and denotes the last term in expansion (5.1). In Lemma 10, both $H_0(x, y;t)$ and $H_1(x, y;t)$ have been evaluated in closed forms. Hence it is easily seen that they have asymptotic expansions of the forms

$$H_0(x, y;t) = \alpha t^{-3/2} + O(t^{-5/2}) \tag{6.6}$$

and

$$H_1(x, y;t) = \beta(x, y) t^{-3/2} + O(t^{-5/2}), \tag{6.7}$$

as $t \to +\infty$, where α is a constant and $\beta(x, y)$ is uniformly bounded in x and y. The order estimates in (6.6) and (6.7) hold also uniformly for x and y restricted to any compact subset of $\mathbb{R}^3$.

To derive a similar result for $H_2(x, y;t)$, we insert (5.47) into (6.5) and obtain

$$H_2(x, y;t) = e^{-i3\pi/4}\Gamma(\tfrac{3}{2}) h_1(x, y) t^{-3/2} + \widetilde{H}_2(x, y;t), \tag{6.8}$$

where

$$\widetilde{H}_2(x, y;t) = \lim_{\beta\to 0^+} \int_0^\infty e^{-(\beta+it)\lambda}\phi_2(x, y;\lambda)\, d\lambda. \tag{6.9}$$

To the last expression, we apply integration by parts twice. The integrated terms vanish, in view of (5.48) and (5.49) and the fact that $\phi_2(x, y;\lambda)$ and $\phi_2'(x, y;\lambda)$ have at most polynomial growth at $\lambda = +\infty$. Thus

$$\widetilde{H}_2(x, y;t) = \left(-\frac{1}{t^2}\right) \lim_{\beta\to 0^+} \int_0^\infty \phi_2''(x, y;\lambda) e^{-(\beta+it)\lambda}\, d\lambda. \tag{6.10}$$

If we can show that the integral

$$\int_0^\infty \phi_2''(x, y;\lambda) e^{it\lambda}\, d\lambda \tag{6.11}$$

converges uniformly for all large values of t and for x and y in compact subsets of $\mathbb{R}^3$, then by the modified Riemann–Lebesgue lemma (OW, Lemmas 6 and 9), we have

$$\widetilde{H}_2(x, y;t) = o(t^{-2}), \quad \text{as } t\to\infty, \tag{6.12}$$

uniformly for x and y in a compact subset of $\mathbb{R}^3$.

To prove the uniform convergence of the integral in (6.11), we first note that the estimate in (5.49) implies that this integral is absolutely convergent at the lower limit. Hence, in view of (5.47), we need be concerned with only the integral

$$\int^\infty h''(x, y;\lambda) e^{i\lambda t}\, d\lambda. \tag{6.13}$$

Now we recall the identity in (5.2), and, for convenience, put

$$A_s(x, y;\lambda) \equiv [r_0(\lambda)(vr_0(\lambda))^s](x, y), \quad s = 2,3,\dots. \tag{6.14}$$

A combination of (5.1), (5.2), and (5.46) gives

$$h(x, y;\lambda) = \sum_{l=0}^{L-1} (-1)^l \operatorname{Im} A_{l+2}(x, y;\lambda) + (-1)^L \operatorname{Im} \epsilon_L(x, y;\lambda), \tag{6.15}$$

where $\epsilon_L(x, y;\lambda)$ is as defined in (4.18). For the moment we need consider only the case $L = 3$, and Lemma 8 immediately gives

$$\epsilon_3''(x, y;\lambda) = O(\lambda^{-1-1/4}), \quad \text{as } \lambda\to\infty. \tag{6.16}$$

By definition,

$$(4\pi)^{s+1} A_s(x, y;\lambda) = \int dy_1\cdots dy_s \left\{\frac{v(y_1)\cdots v(y_s)}{|x-y_1|\cdots|y_s-y|}\right\} \times \exp(i\sqrt{\lambda}(|x-y_1| + \cdots + |y_s-y|)), \tag{6.17}$$

see (6.14) and (5.11). Upon differentiating (6.17) with respect to λ twice, we have

$$(4\pi)^{s+1} A_s''(x, y;\lambda) = \lambda^{-3/2} F_s(x, y;\lambda) - 1/(4\lambda) G_s(x, y;\lambda), \tag{6.18}$$

where $F_s(x, y;\lambda)$ is uniformly bounded in λ and in (x, y) for (x, y) belonging to any compact subset of $\mathbb{R}^3\times\mathbb{R}^3$. The function $G_s(x, y;\lambda)$ is given by

$$G_s(x, y;\lambda) = \int dy_1\cdots dy_s \frac{v(y_1)\cdots v(y_s)}{|x-y_1|\cdots|y_s-y|} \times (|x-y_1| + \cdots + |y_s-y|)^2 \times \exp(i\sqrt{\lambda}(|x-y_1| + \cdots + |y_s-y|)). \tag{6.19}$$

By applying the Buslaev technique (Sec. IV) to the dy_1 integral, it can be shown that

$$G_s(x, y;\lambda) = O(\lambda^{-1/2}), \quad \text{as } \lambda\to\infty, \tag{6.20}$$

uniformly for (x, y) in compact subsets of $\mathbb{R}^3\times\mathbb{R}^3$. This, together with (6.18), infers that the integrals

$$\int^\infty A_s''(x, y;\lambda) e^{i\lambda t}\, d\lambda, \quad s = 2,3,\dots, \tag{6.21}$$

and also the integral in (6.13) are uniformly convergent for all large t and for all x and y in compact subsets of $\mathbb{R}^3$. As remarked previously, this proves the result in (6.12) and hence the following theorem.

Theorem 2: Assume that (i) $v\in\mathscr{F}^*\cap\mathscr{C}_0$, (ii) $v\in\mathscr{A}_2\cap\mathscr{B}_2$, and (iii) $(1+|x|)^8 v(x)\in L^1(\mathbb{R}^3)$. Then Eq. (1.14) holds, i.e.,

$$S(x, y;t) = \Phi(x, y) t^{-3/2} + \delta(x, y;t), \tag{6.22}$$

where $\Phi(x, y)$ is uniformly bounded in x and y and the remainder satisfies

$$\delta(x, y;t) = o(t^{-2}), \quad \text{as } |t|\to\infty, \tag{6.23}$$

uniformly for x and y in compact subsets of $\mathbb{R}^3$.

This is a kernel analog of Theorem 10.3 given in Jensen and Kato.[6] As mentioned in Sec. I, to investigate the asymptotic expansion of the spectral kernel $e(x, y;\lambda)$, we require the hypotheses $T_p(\gamma;N)$. For this reason we need the following extension of Theorem 2.

Corollary 3: Assume that (i) $v\in\mathscr{F}^*_{2N}\cap\mathscr{C}_0$, (ii) $v\in\mathscr{A}_{N+2}\cap\mathscr{B}_{N+2}$, and (iii) $(1+|x|)^{2(N+4)} v(x)\in L^1(\mathbb{R}^3)$. Then the asymptotic expansion (6.22) can be termwise differentiated N times, and for $j = 1,\dots,N$,

$$\delta^{(j)}(x, y;t) = o(t^{-2-j}), \quad \text{as } |t|\to\infty, \tag{6.24}$$

uniformly for x and y in compact subsets of $\mathbb{R}^3$.

Proof: The argument here is similar to that given for Theorem 2, and will hence be kept to a minimum. We begin with (3.1).

Replacing $r(x, y;\lambda)$ by its representation in (5.1) gives

$$\left(\frac{d}{dt}\right)^j S(x,y;t) = (1/\pi)[H_0^{(j)}(x,y;t) - H_1^{(j)}(x,y;t) + H_2^{(j)}(x,y;t)], \tag{6.25}$$

where $H_i^{(j)}(x,y;t)$, $i = 0,1$, is easily shown to be the jth derivative, with respect to t, of the function $H_i(x,y;t)$ given in (6.3) and (6.4). Since these functions, as previously mentioned, have already been explicitly evaluated, their asymptotic expansions can be easily obtained by differentiating the expansions in (6.6) and (6.7). The last term in (6.25) is given by

$$H_2^{(j)}(x,y;t) = \lim_{\beta\to 0^+}\int_0^\infty (-i\lambda)^j e^{-(\beta+it)\lambda} h(x,y;\lambda)d\lambda. \tag{6.26}$$

Repeating the argument leading to (6.8), we obtain

$$H_2^{(j)}(x,y;t) = (-1)^j\Gamma(j+\tfrac{3}{2})e^{-i3\pi/4}h_1(x,y)t^{-j-3/2} + \widetilde{H}_2^{(j)}(x,y;t), \tag{6.27}$$

where

$$\widetilde{H}_2^{(j)}(x,y;t) = \lim_{\beta\to 0^+}(-1)^j\int_0^\infty e^{-(\beta+it)\lambda}\lambda^j\phi_2(x,y;\lambda)d\lambda. \tag{6.28}$$

Put

$$\psi_2(x,y;\lambda) = \lambda^j\phi_2(x,y;\lambda). \tag{6.29}$$

Proposition 7 then gives

$$\left(\frac{d}{d\lambda}\right)^i\psi_2(x,y;\lambda) = O(\lambda^{3/2+j-i}), \quad \text{as } \lambda\to 0^+, \tag{6.30}$$

for $i = 0,1,\ldots,j+2$. We now integrate (6.28) by parts $j+2$ times. The results in (6.30) ensure that the integrated terms all vanish at the lower limit $\lambda = 0$. Since ψ_2 and all its derivatives (with respect to λ) have at most polynomial growth, the integrated terms also vanish at the upper limit $\lambda = +\infty$. Thus

$$\widetilde{H}_2^{(j)}(x,y;t) = \lim_{\beta\to 0^+}(-1)^j t^{-j-2} \times\int_0^\infty e^{-(\beta+it)\lambda}\left(\frac{d}{d\lambda}\right)^{j+2}\psi_2(x,y;\lambda)d\lambda. \tag{6.31}$$

We wish to show that as $t\to\infty$,

$$\widetilde{H}_2^{(j)}(x,y;t) = o(t^{-j-2}), \tag{6.32}$$

uniformly for x and y in compact subsets of $\mathbb{R}^3$. This is equivalent to showing that for $s = 0,1,\ldots,j$, we have

$$\lim_{\beta\to 0^+}\int_0^\infty e^{-(\beta+it)\lambda}\lambda^s\left(\frac{d}{d\lambda}\right)^{s+2}\phi_2(x,y;\lambda)d\lambda\to 0, \tag{6.33}$$

as $t\to\infty$, uniformly for x and y in compact subsets of $\mathbb{R}^3$. (The cases $s = -2$ and $s = -1$ are simpler.) At the lower limit, the integrals in (6.33) are absolutely convergent on account of (5.49). (Take $m = N+2$ in Proposition 7.) Thus, we need be concerned with only the upper limit. In view of (5.47), it suffices to show that for $s = 0,1,\ldots,j$,

$$\lim_{\beta\to 0^+}\int_\rho^\infty e^{-(\beta+it)\lambda}\lambda^s\left(\frac{d}{d\lambda}\right)^{s+2}h(x,y;\lambda)d\lambda\to 0, \tag{6.34}$$

as $t\to\infty$, where ρ is some fixed positive number. With $L = 3s+3$, Eq. (6.15) becomes

$$h(x,y;\lambda) = \sum_{l=0}^{3s+2}(-1)^l\,\mathrm{Im}\,A_{l+2}(x,y;\lambda) + (-1)^{3s+3}\,\mathrm{Im}\,\epsilon_{3s+3}(x,y;\lambda). \tag{6.35}$$

From Lemma 8, it follows that the integral

$$\int_\rho^\infty \lambda^s\left(\frac{d}{d\lambda}\right)^{s+2}\epsilon_{3s+3}(x,y;\lambda)d\lambda \tag{6.36}$$

is absolutely convergent. By the Riemann–Lebesgue lemma, the result in (6.34) holds provided that as $t\to\infty$,

$$\lim_{\beta\to 0^+}\int_\rho^\infty e^{-(\beta+it)\lambda}\lambda^s\left(\frac{d}{d\lambda}\right)^{s+2}A_{l+2}(x,y;\lambda)d\lambda\to 0, \tag{6.37}$$

for $l = 0,1,\ldots,3s+2$. Since (6.37) follows from integrations by parts s times [cf. (6.17) and (4.14)], the proof of the theorem is now complete. □

The above derivations were carried out for large positive values of t. However, the case in which t is a negative value can be included by changing the sign of i throughout. The validity of the expansion (1.11) was established in OW under the condition that $H\in T_p(\gamma;N)$ for some $\gamma<0$; see OW, Theorem 5 and the concluding remark. From Corollary 3, it is easy to see that $H\in T_p(\gamma;N)$ for all $\nu\in(-\frac{3}{2},0)$. Thus we have the following conclusion.

Theorem 3: Let N be a positive integer and M be the integer satisfying

$$2N+\tfrac{5}{2}<M<2N+\tfrac{7}{2}. \tag{6.38}$$

If (i) $v\in\mathscr{F}^*_{2(M+N)}\cap\mathscr{C}_0$, (ii) $v\in\mathscr{A}_{N+2}\cap\mathscr{B}_{N+2}$, and (iii) $(1+|x|)^{2(N+4)}v(x)\in L^1(\mathbb{R}^3)$, then

$$e(x,y;\lambda) = \sum_{n=0}^{M-1}\frac{(-1)^n}{n!}P_n(x,y)\left(\frac{d}{d\lambda}\right)^n e_0(x,y;\lambda) + o(\lambda^{-N}), \tag{6.39}$$

as $\lambda\to+\infty$, where the o-symbol is uniform with respect to x and y for (x,y) in any compact subset of $\mathbb{R}^3\times\mathbb{R}^3$.

ACKNOWLEDGMENTS

This research has been supported in part by the U. S. Department of Energy and the Natural Sciences and Engineering Research Council of Canada.

[1] T. Ikebe, Arch. Ration. Mech. Anal. **5**, 1 (1960).
[2] T. A. Osborn and Y. Fujiwara, J. Math. Phys. **24**, 1093 (1983).
[3] S. F. J. Wilk, Y. Fujiwara, and T. A. Osborn, Phys. Rev. A **24**, 2187 (1981).
[4] P. Greiner, Arch. Ration. Mech. Anal. **41**, 163 (1971).
[5] T. A. Osborn and R. Wong, J. Math. Phys. **24**, 1487 (1983).
[6] A. Jensen and T. Kato, Duke Math. J. **46**, 583 (1979).
[7] J. Ginibre and M. Moulin, Ann. Inst. H. Poincaré A **21**, 97 (1974).
[8] B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, 1971).
[9] V. S. Buslaev, in *Spectral Theory and Wave Processes*, edited by M. Sh. Birman (Consultants Bureau, London, 1967), Vol. 1, p. 69.
[10] S. Agmon, Ann. Scuola Norm. Sup. Pisa. IV **2**, 151 (1975).
[11] T. Kato, Commun. Pure Appl. Math. **12**, 403 (1959).
[12] R. Froese, I. Herbst, M. Hoffmann-Ostenhof, and T. Hoffmann-Ostenhof, J. Anal. Math. **41**, 272 (1982).

[13]B. Simon, Commun. Pure Appl. Math. **22**, 531 (1969).
[14]E. C. Titchmarsh, *Eigenfunction Expansions* (Oxford U. P., London, 1958), Part II.
[15]T. Kato, *Perturbation Theory for Linear Operator* (Springer-Verlag, Berlin, 1976), Chap. VI.
[16]A. Povzner, Mat. Sb. **32**, 109 (1953) [Am. Math. Soc. Transl. Ser. 2 **60**, 1 (1967)].
[17]L. Gårding, K. Fysiogr. Sallsk. Lund. Förh. **24**(21), 1 (1954); Math. Scand. **1**, 237 (1953).
[18]D. Widder, *The Laplace Transform* (Princeton U. P., Princeton, 1941).
[19]T. A. Osborn, R. A. Corns, and Y. Fujiwara, J. Math. Phys. **26**, 453 (1985).
[20]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1980).

A remarkable property of spherical harmonics

R. Beig
Institut für Theoretische Physik, Universität Wien, A-1090 Wien, Austria

(Received 26 July 1984; accepted for publication 19 October 1984)

It is shown that a certain bilinear expression on the space of spherical harmonics of equal angular momentum is constant on the two-sphere.

I. INTRODUCTION

In this paper I want to present a new identity for spherical harmonics. The proof of this relation is a good example of the geometrical viewpoint introduced by Newman and Penrose[1] in the study of spherical harmonics. However, to make the discussion self-contained, I shall avoid their formalism and, essentially, rederive their results concerning spin-weighted functions on S^2 in standard tensorial language.[2]

Let f, g be spherical harmonics of angular momentum l, i.e., functions on S^2 satisfying

$$[D^2 + l(l+1)]f = [D^2 + l(l+1)]g = 0. \tag{1}$$

Here $D^2 = g^{ab}D_aD_b$, g_{ab} is the standard metric on S^2, and D_a its covariant derivative. Define

$$F_k := \mathrm{TS}[D_{a_1}\cdots D_{a_k}f]\,\mathrm{TS}[D^{a_1}\cdots D^{a_k}g], \tag{2}$$

where TS denotes the operation of taking the trace-free symmetric part. Then, I claim,

$$\sum_{k=0}^{l} 2^k\frac{(l-k)!}{(l+k)!}F_k = \frac{2l+1}{4\pi}\int_{S^2} F_0\, d^2\Omega. \tag{3}$$

Note that Eq. (3) says, first of all, that the left side is constant and, secondly, that this constant is computable in terms of the integral of $f\cdot g$ over S^2.

II. THE PROOF

We proceed by a series of lemmas. Let $f_{a_1\cdots a_k}$ $(k\geqslant 3)$ be trace-free and symmetric (in short: TS) in the indices $a_2,\ldots, a_k$. Then, on a two-dimensional manifold, one easily checks the identity with Lemma 1.

Lemma 1:

$$f_{a_1\cdots a_k} = \mathrm{TS}[f_{a_1a_2\cdots a_k}] + \tfrac{1}{2}g_{a_1a_2}f^a{}_{a_3\cdots a_k a} + g_{a_3[a_1}f^a{}_{a_2]a_4\cdots a_k a}. \tag{4}$$

Now specialize to S^2 with curvature

$$R_{abcd} = 2g_{c[a}g_{b]d}. \tag{5}$$

Let f be an arbitrary function on S^2. Then, for $k\geqslant 2$, we have Lemma 2.

Lemma 2:

$$D^{a_k}\mathrm{TS}[D_{a_1}\cdots D_{a_k}f] = \tfrac{1}{2}\mathrm{TS}[D_{a_1}\cdots D_{a_{k-1}}\times(D^2 + k(k-1))f]. \tag{6}$$

Proof: Equation (6) is easily seen to be valid for $k=2$. For $k\geqslant 3$ we apply Eq. (4) to the expression $\mathrm{TS}[D_{a_1}\cdots D_{a_k}f]$ to find that the left side of Eq. (6) is given by

$$D^{a_k}D_{a_1}\mathrm{TS}[D_{a_2}\cdots D_{a_k}f] + \text{trace terms}.$$

Commuting derivatives, there results

$$D^{a_k}\mathrm{TS}[D_{a_1}\cdots D_{a_k}f] = D_{a_1}D^{a_k}\mathrm{TS}[D_{a_2}\cdots D_{a_k}f] + (k-1)\mathrm{TS}[D_{a_1}\cdots D_{a_{k-1}}f] + \text{trace terms}, \tag{7}$$

and Eq. (6) follows by operating with TS on Eq. (7) and using the induction hypothesis.

Lemmas 1 and 2 are purely local statements. A crucial piece of global information is contained in Lemma 3.

Lemma 3: Let $f_{a_1\cdots a_k}$ $(k\geqslant 2)$ be a TS tensor which is also divergence-free. Then $f_{a_1\cdots a_k}$ is identically zero.

Proof: Let ξ be in the three-parameter set of functions satisfying

$$D_aD_b\xi = -\xi g_{ab}. \tag{8}$$

(Essentially, ξ is a Euclidean component of the unit normal of S^2 in $\mathbb{R}^3$). Define $(\xi_a := D_a\xi)$

$$t_a = f_{a a_2\cdots a_k}\xi^{a_2}\cdots\xi^{a_k}. \tag{9}$$

One verifies that t_a is both divergence- and curl-free whence, as a vector field on S^2, it has to vanish identically for all ξ's. But, since $f_{a_1\cdots a_k}$ has only two independent components, this implies that $f_{a_1\cdots a_k}$ itself is zero.

For completeness sake we mention an interesting corollary of Lemma 3 which we do not need.

Lemma 4: Let $f_{a_1\cdots a_k}$ $(k\geqslant 2)$ be a TS tensor. Then there exists a vector field f_a such that

$$f_{a_1\cdots a_k} = \mathrm{TS}[D_{a_1}\cdots D_{a_{k-1}}f_{a_k}]. \tag{10}$$

Proof: One first shows

$$f_{a_1\cdots a_k} = \mathrm{TS}[D_{a_1}f_{a_2\cdots a_k}], \tag{11}$$

for suitable $f_{a_1\cdots a_{k-1}}$. Look at the right side of Eq. (11) as defining an operator L from TS tensors of valence k-1 $(k\geqslant 2)$ to TS tensors of valence k. Taking the obvious inner product on these spaces, we find that the adjoint operator L^* is given by "minus divergence." From Lemma 3, L^* has trivial kernel. Since L is elliptic, Eq. (11) is soluble because of the Fredholm alternative. Iterating, we obtain Eq. (10).

Remark: Since every vector field on S^2 is the sum of a curl and a divergence, Lemma 4 enables one to reduce any tensor equation on S^2 to a system of scalar equations. It is essentially this fact which makes the formalism of Ref. 1 or, equivalently, of the present paper, so useful in practical computations.

Lemma 5: Let f be a spherical harmonic of angular momentum k. Then $(k=0,1,2,\ldots)$

$$\mathrm{TS}[D_{a_1}\ldots D_{a_{k+1}}f] = 0. \tag{12}$$

Note that Eq. (12) means that f is completely characterized by the values of f, D_af, $\mathrm{TS}[D_{a_1}D_{a_2}f]$,..., $\mathrm{TS}[D_{a_1}\cdots D_{a_k}f]$ at

 0022-2488/85/040769-02$02.50

some point of S^2. This makes $2k+1$ independent spherical harmonics, as it should be. In the language of Newman and Penrose,[2] Eq. (12) says that spherical harmonics whose spin weight exceeds their angular momentum value have to vanish.

Proof of Lemma 5: Eq. (12) is obviously true for $k=0$. For higher k compute the divergence of the left side of Eq. (12). Using Eq. (6) we obtain

$$D^{a_{k+1}}\mathrm{TS}[D_{a_1}\cdots D_{a_{k+1}}f] = \tfrac{1}{2}\mathrm{TS}[D_{a_1}\cdots D_{a_k}(D^2+k(k+1))f]=0. \tag{13}$$

Using Lemma 3, it follows that $\mathrm{TS}[D_{a_1}\cdots D_{a_{k+1}}f]$ has to vanish.

Let F_k now be defined as in Eq. (2) and let

$$F_{a|k} := \mathrm{TS}[D_aD_{a_1}\cdots D_{a_k}f]\ \mathrm{TS}[D^{a_1}\cdots D^{a_k}g] + TS\,[D_{a_1}\cdots D_{a_k}f]\ \mathrm{TS}[D_aD^{a_1}\cdots D^{a_k}g]. \tag{14}$$

We have

$$D_aF_0 = F_{a|0}. \tag{15}$$

To compute D_aF_k we note that, for $k\geqslant 2$,

$$D_a\mathrm{TS}[D_{a_1}\cdots D_{a_k}f] = \mathrm{TS}[D_aD_{a_1}\cdots D_{a_k}f] + \tfrac{1}{4}g_{aa_1}\mathrm{TS}[D_{a_2}\cdots D_{a_k}(D^2+k(k-1))f] + \tfrac{1}{2}g_{a_2[a}\mathrm{TS}[D_{a_1]}D_{a_3}\cdots D_{a_k}(D^2+k(k-1))f], \tag{16}$$

where we have used Lemma 1 and Lemma 2. Equation (16) and the fact that f and g have angular momentum l imply

$$D_aF_k = F_{a|k} - \tfrac{1}{2}[l(l+1)-k(k-1)]F_{a|k-1}, \tag{17}$$

and this relation is also valid for $k=1$. Note, finally, because of Lemma 5 we have

$$F_{a|l}=0. \tag{18}$$

Armed with this information it is an easy matter to check that the gradient of the left side of Eq. (3) vanishes. Now we integrate over both sides of

$$\sum_{k=0}^{l} 2^k \frac{(l-k)!}{(l+k)!}F_k = \mathrm{const.} \tag{19}$$

Using ($k\geqslant 1$)

$$\int_{S^2} F_k\, d^2\Omega = \tfrac{1}{2}[l(l+1)-k(k-1)]\int_{S^2}F_{k-1}\,d^2\Omega, \tag{20}$$

we find that the constant in Eq. (19) is given by

$$\frac{2l+1}{4\pi}\int_{S^2}F_0\,d^2\Omega.$$

This completes the proof of our theorem.

Let us sum up. Let f be a spherical harmonic of angular momentum l. Then the expression

$$[\mathrm{TS}\,D_{a_1}\cdots D_{a_l}f]$$

is a conformal Killing tensor on S^2 [i.e., it obeys Eq. (12)]. It follows that f is characterized by its $2l+1$ "Killing data" at some point of S^2. Remarkably, there exists a bilinear form on the space of such Killing data which remains constant throughout S^2, namely the left side of Eq. (3).

ACKNOWLEDGMENTS

I thank H. Urbantke for discussions.

This research has been supported by Fonds zur Förderung der Wissenschaftlichen Forschung, Projekt Nr. P4961.

[1]E. T. Newman and R. Penrose, J. Math. Phys. 7, 863 (1966), see also J. N. Goldberg *et al.*, J. Math. Phys. 8, 15 (1967).

[2]For a formulation in the framework of complex analysis, see M. Eastwood and P. Tod, Math. Proc. Cambridge Philos. Soc. 92, 317 (1982).

On spherically symmetric perfect fluid solutions

J. Hajj-Boutros
Physics Department, Lebanese University, Mansourieh el-Metn, P.O. Box 72, Lebanon

(Received 20 July 1984; accepted for publication 28 September 1984)

Many exact solutions for the spherically symmetric perfect fluid distribution of matter with shear, acceleration, and expansion are obtained. One of them is expressed in terms of Painlevé's third transcendent.

I. INTRODUCTION

Although, spherically symmetric stars are a very important class of objects for the application of the general relativity theory, only very few perfect fluid solutions of Einstein's equations, with shear, expansion, and acceleration have been found so far.

Starting from the metric $ds^2 = Y^2(r,t)d\Omega^2 + e^{2\lambda(r)}dr^2 - e^{2\nu(r)}dt^2$ and writing $Y(r,t)$ as a product $R(r)\cdot T(t)$, we are able to obtain the field equations as a set of two differential equations.

For $T(t)$ we get a second-order linear differential equation; for a particular choice of $\lambda(r)$, we get for $R(r)$ a nonlinear second-order differential equation which is exactly the third transcendent equation of Painlevé.

II. LINE ELEMENT AND FIELD EQUATIONS

In a comoving frame of reference, and for a spherically symmetric distribution of matter, a coordinate system can be introduced such as

$$ds^2 = Y(r,t)d\Omega^2 + e^{2\lambda(r,t)}dr^2 - e^{2\nu(r,t)}dt^2, \tag{2.1}$$

where

$$d\Omega^2 = d\theta^2 + \sin^2\theta\, d\varphi^2.$$

Like all perfect fluid solutions the spherically symmetric solutions can be classified according to their kinematical properties; i.e., the four-velocity's rotation, acceleration, expansion, and shear.

For a timelike vector field u, like the four-velocity of a fluid, the acceleration, rotation, shear, and expansion can be defined as follows:

$$\dot u_a = u_{a;b}u^b, \quad \dot u_a u^a = 0, \tag{2.2}$$

$$\omega_{ab} = u_{[a;b]} + \dot u_{[a}u_{b]}, \quad \omega_{ab}u^b = 0, \tag{2.3}$$

$$\sigma_{ab} = u_{(a;b)} + \dot u_{(a}u_{b)} - \Theta h_{ab}/3, \quad \sigma_{ab}u^b = 0, \tag{2.4}$$

$$\Theta = u^a_{;a}. \tag{2.5}$$

Here $h_{ab} = g_{ab} + u_a u_b$ and ; means covariant derivative. An explicit list of acceleration, expansion, and shear in the case of (2.1), is given in Ref. 1.

For nonvanishing acceleration, expansion, and shear we must have, respectively,

$$\nu' \neq 0, \tag{2.6}$$

$$2\dot Y/Y \neq -\dot\lambda, \tag{2.7}$$

$$\dot Y/Y \neq \dot\lambda. \tag{2.8}$$

Here prime and dot denote differentiation with respect to the coordinates r and t, respectively.

Spherical symmetry implies that $\omega_{ab} = 0$.

Following Gutman and Bespal'ko,[2] we can restrict ourselves to the metric of the form

$$ds^2 = R^2(r)T^2(t)d\Omega^2 + e^{2\lambda(r)}dr^2 - e^{2\nu(r)}dt^2. \tag{2.9}$$

The conditions (2.6), (2.7), and (2.9) imply

$$\nu' \neq 0, \quad \dot T/T \neq 0. \tag{2.10}$$

The field equations in general relativity read

$$R_{ab} - \tfrac{1}{2}Rg_{ab} = \kappa_0 T_{ab}. \tag{2.11}$$

For a perfect fluid distribution of matter we have

$$T_{ab} = (\mu + p)u_a u_b + pg_{ab}, \quad u_a u^a = -1, \tag{2.12}$$

μ being the energy density and p the pressure.

If we choose a line element of the form (2.9) the field equations become

$$\kappa_0\mu = \frac{1}{R^2T^2} - \frac{2}{R}e^{-2\lambda(r)}\left[R'' - R'\lambda' + \frac{R'^2}{2R}\right] + \frac{2}{T}e^{-2\nu(r)}\left(\frac{\dot T^2}{2T}\right), \tag{2.13}$$

$$\kappa_0 p = \frac{1}{RT}\left\{-\frac{1}{RT} + T\left(2R'\nu' + \frac{R'^2}{R}\right)e^{-2\lambda(r)}\right\} - 2R\left(\ddot T + \frac{\dot T^2}{2T}\right)e^{-2\nu(r)}, \tag{2.14}$$

$$\kappa_0 p = \left\{e^{-2\lambda(r)}\left[(\nu'' + \nu'^2 - \nu'\lambda') + \frac{1}{R}(R'' + R'\nu' - R'\lambda')\right]\right\} - \frac{\ddot T}{T}e^{-2\nu(r)}, \tag{2.15}$$

$$\dot TR' - R\dot T\nu' = 0. \tag{2.16}$$

Furthermore, the Bianchi identities imply the relations

$$p' = -(\mu + p)\nu', \quad \dot\mu = -(\mu + p)(2\dot T/T). \tag{2.17}$$

III. SOLUTIONS

From (2.14) and (2.15) we obtain

$$-1/RT + Te^{-2\lambda}[-R'' + R'^2/R + R'(\nu' + \lambda') + R(\lambda'\nu' - \nu'^2 - \nu'')] - R(\ddot T + \dot T^2/T)e^{-2\nu} = 0. \tag{3.1}$$

By suitable arrangement (3.1) becomes

$$Re^{-2\lambda}[-R'' + R'^2/R + R'(\lambda' + \nu') + R(\lambda'\nu' - \nu'^2 - \nu''] = \frac{1}{T^2} + + \frac{R^2}{e^{2\nu}}\left(\frac{\ddot T}{T} + \frac{\dot T^2}{T^2}\right). \tag{3.2}$$

 0022-2488/85/040771-03$02.50

Performing the substitution

$$T(t)=u^{1/2}(t),$$

we obtain

$$\frac{1}{u}+\frac{R^2}{e^{2\nu}}\left(\frac{1}{2}\frac{\ddot u}{u}\right)$$
$$=Re^{-2\lambda}\left[-R''+R'^2/R+R'(\nu'+\lambda')+R(\nu'\lambda'-\nu'^2-\nu'')\right]. \tag{3.3}$$

From (2.16) we get

$$\nu'=R'/R,\quad 2\nu=\ln(R^2/c^2), \tag{3.4}$$

c being a constant.

The relations (3.3) and (3.4) give

$$\frac{1}{u}+\frac{c^2}{2}\cdot\frac{\ddot u}{u}=Re^{-2\lambda}\left(-2R''+\frac{2R'^2}{R}+2R'\lambda'\right). \tag{3.5}$$

We can set both sides of (3.5) equal to an arbitrary constant α, and we obtain

$$\frac{1}{u}+\frac{c^2}{2}\frac{\ddot u}{u}=\alpha, \tag{3.6}$$

$$2R(-R''+R'^2/R+R'\lambda')e^{-2\lambda}=\alpha. \tag{3.7}$$

We start by studying (3.6). There are three cases.

Case A.1: $\alpha<0$. We set $\alpha=-\omega^2c^2/2$ and we obtain

$$u(t)=A_1\cos(\omega t)+B_1\sin(\omega t)-2/c^2\omega^2, \tag{3.8}$$

so

$$T(t)\equiv[A_1\cos(\omega t)+B_1\sin(\omega t)-2/c^2\omega^2]^{1/2}, \tag{3.9}$$

A_1 and B_1 being two constants.

Case A.2: $\alpha=0$. From (2.6) we obtain

$$u(t)=-t^2/c^2+A_2t+B_2, \tag{3.10}$$

so

$$T(t)=[-t^2/c^2+A_2t+B_2]^{1/2}, \tag{3.11}$$

where A_2 and B_2 are two constants.

Case A.3: $\alpha>0$. We set $\alpha=\gamma^2c^2/2$ and we obtain

$$u(t)=A_3e^{-\gamma t}+B_3e^{\gamma t}+2/c^2\gamma^2, \tag{3.12}$$

so

$$T(t)=[A_3e^{-\gamma t}+B_3e^{\gamma t}+2/c^2\gamma^2]^{1/2}, \tag{3.13}$$

A_3 and B_3 being two constants.

In order to obtain a metric we must find the solutions corresponding to (3.7), which contains two unknowns, $\lambda(r)$ and $R(r)$; we are restricted to use an *ad hoc* choice for $\lambda(r)$, and then deduce $R(r)$.

Case B.1: $\lambda=c_0$, c_0 being a constant. Equation (3.7) becomes

$$R''=R'^2/R-\alpha_0e^{(2c_0)}/R, \tag{3.14}$$

where $\alpha_0=\alpha/2$.

If α_0 is positive, a trivial solution is

$$R(r)=r\alpha_0^{1/2}e^{c_0}. \tag{3.15}$$

Equations (3.15) and (3.13) thus give us the Gutman solution.

A more general solution of (3.14) can be obtained by the change of function

$$R(r)=e^{\sigma(r)}. \tag{3.16}$$

From (3.14) and (3.16) we get

$$\sigma''=-\alpha_0e^{2c_0}\cdot e^{-2\sigma}. \tag{3.17}$$

Consequently we know a first integral

$$\sigma'^2=\alpha_0e^{2c_0}\cdot e^{-2\sigma}+c_1, \tag{3.18}$$

c_1 being a constant.

For α_0 positive, and by quadrature we obtain[3]

$$r+c_2=\frac{1}{-2c_1^{1/2}}\ln\left\{\frac{(c_1+\alpha_0e^{2c_0}e^{-2\sigma})^{1/2}-c_1^{1/2}}{(c_1+\alpha_0e^{2c_0}e^{-2\sigma})^{1/2}+c_1^{1/2}}\right\},\quad c_1>0, \tag{3.19}$$

$$r+c_3=-\frac{1}{(-c_1)^{1/2}}\arctan\left\{\frac{(c_1+\alpha_0e^{2c_0}e^{-2\sigma})^{1/2}}{(-c_1)^{1/2}}\right\},\quad c_1<0, \tag{3.20}$$

where c_2,c_3 are two constants. For $c_1=0$ and from (3.18) we obtain (3.15)

$$R(r)=\alpha_0^{1/2}e^{c_0}\cdot r+c_4, \tag{3.21}$$

c_4 being a constant.

By suitable arrangement (3.19) and (3.20) reduce, respectively, to

$$R^2=\frac{\alpha_0e^{2c_0}}{4c_1}\frac{[e^{-2\sqrt{c_1}(r+c_2)}-1]^2}{e^{-2\sqrt{c_1}(r+c_2)}}, \tag{3.22}$$

$$R^2=-(\alpha_0/c_1)e^{2c_0}\cos^2[\sqrt{-c_1}(r+c_3)]. \tag{3.23}$$

So from (3.4), (3.13), (3.22), and the choice $\lambda=c_0$ we obtain the following metric:

$$ds^2=\frac{\alpha_0e^{2c_0}(e^{-2\sqrt{c_1}(r+c_2)}-1)^2}{4c_1e^{-2\sqrt{c_1}(r+c_2)}}\times\left[A_3e^{-\gamma t}+B_3e^{\gamma t}+\frac{2}{c^2\gamma^2}\right]d\Omega^2+e^{2c_0}dr^2-\frac{\alpha_0e^{2c_0}}{4c_1c^2}\frac{(e^{-2\sqrt{c_1}(r+c_2)}-1)^2}{e^{-2\sqrt{c_1}(r+c_2)}}dt^2, \tag{3.24}$$

where

$$c_1>0.$$

Another metric is obtained from (3.4), (3.13), (3.23), and the choice $\lambda=c_0$ so we have

$$ds^2=-(\alpha_0/c_1)e^{2c_0}\cos^2[\sqrt{-c_1}(r+c_3)]d\Omega^2+e^{2c_0}dr^2-\left[-\frac{\alpha_0}{c_1c_2}e^{2c_0}\cos^2(\sqrt{-c_1}(r+c_3)\right]dt^2, \tag{3.25}$$

where $c_1<0$.

Case B.2: $\lambda=c_0$, $\alpha=0$. Equation (3.7) becomes

$$R''=R'^2/R. \tag{3.26}$$

By quadrature we obtain

$$R(r)=c_5e^{c_6r}, \tag{3.27}$$

c_5,c_6 being two constants.

The line element reads

$$ds^2 = c_5^2 e^{(2c_6 r)}(-t^2/c^2 + A_2 t + B^2)d\Omega^2 + e^{2c_0} dr^2 - (c_5^2 e^{2c_6 r}/c^2)dt^2. \quad (3.28)$$

Case B.3: $\lambda = \ln(a/r)$, $\alpha \neq 0$ (a being a constant). Equation (3.7) becomes

$$R'' = \frac{R'^2}{R} - \frac{R'}{r} + \frac{a^2\alpha}{r^2 R}. \quad (3.29)$$

Setting $\delta = a^2\alpha$ and using the substitution $R(r) = P(r)/r$, we obtain

$$P'' = \frac{P'^2}{P} - \frac{P'}{r} + \frac{\delta}{P}. \quad (3.30)$$

So we obtain a particular case of the canonical equations[4–6] of type III of Painlevé (see Appendix).

From (3.4), (3.13), (3.30), and the choice $\lambda = \ln(a/r)$, we get the metric

$$ds^2 = (P_{\rm III}^2(r)/r^2)(A_3 e^{-\gamma t} + B_3 e^{\gamma t} + 2/c^2\gamma^2)d\Omega^2 + (a^2/r^2)dr^2 - [P_{\rm III}^2(r)/c^2 r^2]dt^2. \quad (3.31)$$

Another one is obtained from (3.4), (3.9), (3.30), and the choice $\lambda = \ln(a/r)$.

So we have

$$ds^2 = \frac{P_{\rm III}^2(r)}{r^2}\left[A_1 \cos \omega t + B_1 \sin \omega t - \frac{2}{c^2\omega^2}\right] d\Omega^2 + \frac{a^2}{r^2} dr^2 - [P_{\rm III}^2(r)/r^2 c^2]dt^2, \quad (3.32)$$

where $P_{\rm III}(r)$ is the third Painlevé transcendent.

Case B.4: $\lambda = \ln(a/r)$, $\alpha = 0$. Equation (3.7) becomes

$$R'' = R'^2/R - R'/r. \quad (3.33)$$

By the change of function $R(r) = e^{\sigma(r)}$ we get

$$\sigma'' = -\sigma'/r. \quad (3.34)$$

By quadrature we obtain

$$R(r) = (r/c_8)^{c_7}, \quad (3.35)$$

c_7, c_8 being two constants. So from (3.4), (3.11), (3.35), and the choice $\lambda = \ln(a/r)$, we get another metric given in the following form:

$$ds^2 = (r/c_8)^{2c_7}(-t^2/c^2 + A_2 t + B_2)d\Omega^2 + (a^2/r^2)dr^2 - (1/c^2)(r/c_8)^{2c_7} dt^2 \quad (3.36)$$

Case B.5: $\lambda = br$, b being a constant. Equation (3.7) becomes

$$R'' = R'^2/R + bR' - \alpha_0 e^{2br}/R, \quad (3.37)$$

where $\alpha_0 = \alpha/2$.

By the change of variable $Z = e^{br}$ we get

$$\frac{d^2R}{dZ^2} = \frac{1}{R}\left(\frac{dR}{dZ}\right)^2 - \frac{\alpha_0}{b^2 R}. \quad (3.38)$$

Equation (3.38) is of the same type as (3.14). So the solutions for (3.38) are identical to those of (3.19) and (3.20) with $e^{2c_0} = 1/b^2$.

IV. CONCLUDING REMARKS

We have proved in this work that the differential equation for $R(r)$ reduces essentially to a nonlinear differential equation of the form

$$R'' = (1/R)R'^2 + [A(r)R + B(r) + c(r)/R]R' + D(r)R^3 + E(r)R^2 + F(r)R + G(r) + H(r)/R. \quad (4.1)$$

The latter takes one of the three following canonical[4] forms:

$$R'' = (1/R)R'^2, \quad (4.2)$$

$$R'' = (1/R)R'^2 + \alpha R^3 + \beta R^2 + \gamma + \delta/R, \quad (4.3)$$

$$R'' = \frac{1}{R}R'^2 - \frac{1}{r}R' + \frac{1}{r}(\alpha R^2 + \beta) + \gamma R^3 + \frac{\delta}{R}. \quad (4.4)$$

The first two equations may be solved by classical transcendents; the third requires the Painlevé transcendent.

APPENDIX: CANONICAL EQUATIONS

The canonical equations of the Painlevé and Gambier classification are the following:

$$R'' = 6R^2 + r, \quad (A1)$$

$$R'' = 2R^3 + rR + \alpha, \quad (A2)$$

$$R'' = \frac{1}{R}R'^2 - \frac{1}{r}R' + \frac{1}{r}(\alpha R^2 + \beta) + \gamma R^3 + \frac{\delta}{R}, \quad (A3)$$

$$R'' = \frac{1}{2R}R'^2 + \frac{3R^3}{2} + 4rR^2 + 2(r^2 - \alpha)R + \frac{\beta}{R}, \quad (A4)$$

$$R'' = \left(\frac{1}{2R} + \frac{1}{R-1}\right)R'^2 - \frac{1}{r}R' + \frac{(R-1)^2}{r^2} \times \left(\alpha R + \frac{\beta}{R}\right) + \gamma \frac{R}{r} + \frac{\delta R(R+1)}{R-1}, \quad (A5)$$

$$R'' = \frac{1}{2}\left[\frac{1}{R} + \frac{1}{R-1} + \frac{1}{R-r}\right]R'^2 - \left[\frac{1}{r} + \frac{1}{r-1} + \frac{1}{R-r}\right]R' + \frac{R(R-1)(R-r)}{2r^2(r-1)^2}\left[\alpha - \frac{\beta r}{R^2} + \frac{\gamma(r-1)}{(R-1)^2} - \frac{(\delta-1)r(r-1)}{(R-r)^2}\right]. \quad (A6)$$

[1] D. Kramer, H. Stephani, M. A. H. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Deutscher Verlag der Wissenschaften, Berlin, 1980), pp. 165–173.
[2] I. I. Gutman and R. M. Bespal'ko Sb. Sov. Probl. Grav. Tbilissi, 201 (1967) (in Russian).
[3] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1980).
[4] P. Painlevé, Acta Math. **25**, 1 (1902).
[5] B. Gambier, Acta Math. **33**, 1 (1909).
[6] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).

Minimal acceleration requirements for "time travel" in Gödel space-time

David B. Malament
Department of Philosophy, The University of Chicago, Chicago, Illinois 60637

(Received 8 June 1984; accepted for publication 23 November 1984)

It is demonstrated that the total integrated acceleration of any closed timelike curve in Gödel space-time must be at least $\ln(2+\sqrt{5})$. This answers a question posed by Geroch.

I. INTRODUCTION

In Gödel space-time, even though there exist closed timelike curves, there do not exist any closed timelike geodesics.[1] Thus any "time traveler" who would return to an "earlier" point on his own world line must undergo some acceleration, sometime during the trip. The question arises whether there is some minimal amount that is needed.

Let γ be a closed timelike curve.[2] We take its *total (integrated) acceleration* to be

$$\mathrm{TA}(\gamma) = \int_\gamma a\, ds,$$

where s is elapsed proper time along γ, and a is the magnitude of its acceleration. [Thus, if we let ξ^n be the unit tangent to γ, and let $\alpha^n = \xi^m \nabla_m \xi^n$ be its acceleration, then $a = (-\alpha^n \alpha_n)^{1/2}$.] Our question is this. Does there exist some number $k>0$ such that $\mathrm{TA}(\gamma) \geqslant k$ for all closed timelike curves γ in Gödel space-time? [Notice that $\mathrm{TA}(\gamma)$ is invariant under rescaling of the space-time metric.[3] It does not depend on our choice of units for space-time length.]

The simplest closed timelike curves in Gödel space-time ("Gödel circles") exhibit enormous total acceleration. (See Sec. II below.) But it is just possible that a would-be economical "time traveler" can make do with arbitrarily small quantities of total acceleration by properly choosing his navigational strategy. (For example, he might try using large bursts of acceleration for ultrashort periods of proper time, rather than sustaining acceleration over the entire trip. And he might try wandering over large regions of the space-time manifold, rather than staying close to home.)

Chakrabarti, Geroch, and Liang[4] have shown that this possibility can be ruled out *if* the "time traveler" is required to start out at rest relative to the major, field producing, mass points of the Gödel universe. In effect they show that "time travel" is not possible *at all* unless, during at least part of the trip, high relative speed is achieved. If one is starting from a state of relative rest, this is impossible without the accumulation of considerable total acceleration. Their argument establishes that $\mathrm{TA}(\gamma) \geqslant \frac{1}{2}\ln 2$ for all closed timelike curves γ in the restricted class.

We show below in Sec. IV that this bound holds (and can be raised) even if the "time traveler" is allowed arbitrarily large initial relative speed. Thus, $\mathrm{TA}(\gamma) \geqslant \ln(2+\sqrt{5})$ for *all* closed timelike curves in Gödel space-time [$\ln(2+\sqrt{5})$ is approximately 1.44]. This bound can probably be raised still further, but it is not clear by how much. In any case, the answer to our question above is certainly positive.

II. PRELIMINARIES

In this section we list several basic features of Gödel space-time that will be needed later, and then compute the total acceleration for a special class of closed timelike curves.

We start with the following characterization of Gödel space-time (M, g_{mn}). Here, M is the manifold R^4, and g_{mn} is such that for some point (and hence, by homogeneity, any point) p in M, there is a global adapted (cylindrical) coordinate system t, r, φ, y in which $t(p) = r(p) = y(p) = 0$ and

$$\begin{aligned} g_{mn} = {} & (\nabla_m t)(\nabla_n t) - (\nabla_m r)(\nabla_n r) - (\nabla_m y)(\nabla_n y) \\ & + (\mathrm{sh}^4 r - \mathrm{sh}^2 r)(\nabla_m \varphi)(\nabla_n \varphi) \\ & + 2\sqrt{2}\, \mathrm{sh}^2 r (\nabla_{(m} \varphi)(\nabla_{n)} t). \end{aligned}$$

(We shall use sh r and ch r, respectively, to abbreviate sinh r and cosh r.) Here $-\infty < t < \infty$, $-\infty < y < \infty$, $0 \leqslant r < \infty$, and $0 \leqslant \varphi \leqslant 2\pi$ with $\varphi = 0$ identified with $\varphi = 2\pi$.

Clearly $(\partial/\partial t)^n$ is a timelike Killing field of unit length. It represents the four-velocity of the major, field-producing, mass points of the universe, and determines a temporal orientation. The integral curves of $(\partial/\partial t)^n$, characterized by constant values for r, φ, and y, will be called *matter lines*.

Here, $(\partial/\partial\varphi)^n$ is a rotational Killing field with squared norm $(\mathrm{sh}^4 r - \mathrm{sh}^2 r)$. It will play an essential role in our argument. The (closed) integral curves of $(\partial/\partial\varphi)^n$, characterized by constant values for t, r, and y, will be called *Gödel circles*.

Given any two points p and q in M, we take the *(radial) distance* from p to q to be the r-coordinate value of q in any cylindrical coordinate system (of the sort above) adapted to p. This distance function is symmetric, and induces a natural geometric structure on all $t = \text{const}$, $y = \text{const}$ submanifolds of Gödel space-time. Indeed, the following is true.

(1) Under the radial distance function every $t = \text{const}$, $y = \text{const}$ submanifold is a model for the axioms of hyperbolic (i.e., Lobatchevskian) plane geometry.[5]

Given a point p, we take the *critical cylinder* associated with p to be the set of all points whose radial distance from p is less than $r_c = \ln(1+\sqrt{2})$. Since sh $r_c = 1$, and the squared norm of $(\partial/\partial\varphi)^n$ is $(\mathrm{sh}^4 r - \mathrm{sh}^2 r)$, (2) follows immediately.

(2) Gödel circles with radius $r = r_c$ are closed null curves. Those with radius $r > r_c$ are closed timelike curves. Thus there exist closed timelike curves fully contained in any arbitrarily small radial expansion of a critical cylinder. But the expansion is essential.

(3) There are no closed timelike curves contained within any critical cylinder. Indeed, within a critical cylinder t is a universal time function (i.e., it increases along all future-di-

 0022-2488/85/040774-04$02.50

rected timelike curves),[6] and so the cylinder considered as a space-time in its own right is stably causal.

(4) All timelike geodesics through a point p are confined to the critical cylinder associated with p. [Hence, by (3), there are no closed timelike geodesics.]

There is an easy proof of this statement which does not require a prior characterization of all geodesics in Gödel space-time. Since the argument will help to motivate our own proof in Sec. III, we present it here in detail.

Consider any timelike geodesic γ passing through p. Let ξ^n be its unit tangent, and let the function E_φ be defined by

$$\begin{aligned} E_\varphi &= \xi^n\left(\frac{\partial}{\partial\varphi}\right)_n \\ &= (\mathrm{sh}^4 r - \mathrm{sh}^2 r)(\xi^n\nabla_n\varphi) + \sqrt{2}\,\mathrm{sh}^2 r(\xi^n\nabla_n t). \end{aligned}$$

Here, E_φ must be constant along γ since

$$\xi^n\nabla_n E_\varphi = \xi^n\xi^m\nabla_n\left(\frac{\partial}{\partial\varphi}\right)_m + \xi^n\left(\frac{\partial}{\partial\varphi}\right)_m\nabla_n\xi^m = 0.$$

[The first term vanishes because $(\partial/\partial\varphi)^n$ is a Killing field; the second because γ is a geodesic.] Its constant value must be 0 since γ passes through p.[7]

Now suppose that γ escapes from the critical cylinder associated with p. Let q be the point where it reaches the critical radius r_c. Then at q we have $\mathrm{sh}\, r = 1$ and $E_\varphi = 0$. So $\xi^n\nabla_n t = 0$. But ξ^n is of unit length. So at all points

$$\begin{aligned} 1 = \xi^n\xi_n &= (\xi^n\nabla_n t)^2 - (\xi^n\nabla_n r)^2 - (\xi^n\nabla_n y)^2 \\ &\quad + (\mathrm{sh}^4 r - \mathrm{sh}^2 r)(\xi^n\nabla_n\varphi)^2 \\ &\quad + 2\sqrt{2}\,\mathrm{sh}^2 r(\xi^n\nabla_n\varphi)(\xi^m\nabla_m t). \end{aligned}$$

Hence at q

$$1 = -(\xi^n\nabla_n r)^2 - (\xi^n\nabla_n y)^2,$$

which is impossible. ∎

Now we do a simple calculation so as to have a numerical value for total acceleration in at least one case.

Lemma 1: A Gödel circle γ with radius $r > r_c$ has total acceleration

$$\pi\,\mathrm{sh}\, 2r(2\,\mathrm{sh}^2 r - 1)/(\mathrm{sh}^4 r - \mathrm{sh}^2 r)^{1/2}.$$

Proof: The unit tangent to the circle is $\xi^n = f(\partial/\partial\varphi)^n$, where $f = (\mathrm{sh}^4 r - \mathrm{sh}^2 r)^{-1/2}$. Clearly $\xi^n\nabla_n f = 0$. The acceleration vector α_n is given by

$$\begin{aligned} \alpha_n &= f^2\left(\frac{\partial}{\partial\varphi}\right)^m\nabla_m\left(\frac{\partial}{\partial\varphi}\right)_n = -f^2\left(\frac{\partial}{\partial\varphi}\right)^m\nabla_n\left(\frac{\partial}{\partial\varphi}\right)_m \\ &= \frac{-f^2}{2}\nabla_n(\mathrm{sh}^4 r - \mathrm{sh}^2 r) \\ &= \frac{-f^2}{2}\mathrm{sh}\, 2r(2\,\mathrm{sh}^2 r - 1)\nabla_n r. \end{aligned}$$

(For the second equality we have used the fact that $(\partial/\partial\varphi)^n$ is a Killing field.) Hence

$$a = (-\alpha^n\alpha_n)^{1/2} = (f^2/2)\mathrm{sh}\, 2r(2\,\mathrm{sh}^2 r - 1).$$

Therefore

$$\mathrm{TA}(\gamma) = \int_\gamma a\, ds = \int_0^{2\pi} a f^{-1}\, d\varphi = \frac{2\pi a}{f},$$

and our claim follows. (In the second equality we have used $d\phi/ds = \xi^n\nabla_n\phi = f$.) ∎

Notice that $\mathrm{TA}(\gamma)$ blows up as $r \to \infty$ and $r \to r_c$. A minimal value for total acceleration is reached when r satisfies $\mathrm{sh}^2 r = (1+\sqrt{3})/2$. It comes out to $2\pi(9+6\sqrt{3})^{1/2}$, which is approximately 27.67.

III. AN INEQUALITY

We know from our proof of statement (4) above that the function E_φ (as determined relative to any cylindrical coordinate system) cannot increase along a timelike curve if the curve is a geodesic. A key idea in our proof is that when E_φ *does* increase, its *magnitude* of increase can be used to monitor the accumulation of total acceleration along the curve.

We start with a quite general inequality.[4]

Lemma 2: Let λ^m be a Killing field, not necessarily timelike, in a space-time (M, g_{mn}). Let γ be an arbitrary timelike curve in (M, g_{mn}) with tangent ξ^n, and let $E = \xi^m\lambda_m$. Then

$$|\xi^n\nabla_n E| \leqslant a[E^2 - \lambda^m\lambda_m]^{1/2}.$$

Proof: Direct computation shows

$$\begin{aligned} \xi^n\nabla_n E &= \xi^n\xi^m\nabla_n\lambda_m + \xi^n\lambda_m\nabla_n\xi^m \\ &= \lambda_m\alpha^m = h_{mn}\lambda^m\alpha^n, \end{aligned}$$

where $h_{mn} = g_{mn} - \xi_m\xi_n$ is the (negative semidefinite) "spatial metric" which projects g_{mn} orthogonal to ξ^m. Hence, by the Schwarz inequality (applied to $-h_{mn}$),

$$\begin{aligned} |\xi^n\nabla_n E| &= |-h_{mn}\lambda^m\alpha^n| \\ &\leqslant (-h_{mn}\alpha^m\alpha^n)^{1/2}(-h_{mn}\lambda^m\lambda^n)^{1/2} \\ &= a[E^2 - \lambda^m\lambda_m]^{1/2}. \end{aligned}$$

∎

We are interested in the case where (M, g_{mn}) is Gödel space-time, λ^n is $(\partial/\partial\varphi)^n$, and E is E_φ. So the inequality comes to

$$|\xi^n\nabla_n E_\varphi| \leqslant a[E_\varphi^2 - (\mathrm{sh}^4 r - \mathrm{sh}^2 r)]^{1/2}.$$

There are two subcases to consider. If $r > r_c$, then $(\mathrm{sh}^4 r - \mathrm{sh}^2 r) > 0$ and the square root term is dominated by $\sqrt{E_\varphi^2}$. If $0 \leqslant r \leqslant r_c$, then $(\mathrm{sh}^4 r - \mathrm{sh}^2 r) \leqslant 0$ and the term assumes a minimal (negative) value of $-\frac{1}{4}$ when $\mathrm{sh}\, r = 1/\sqrt{2}$. So in both cases we have

$$|\xi^n\nabla_n E_\varphi| \leqslant a[E_\varphi^2 + \tfrac{1}{4}]^{1/2}.$$

This is the inequality we shall exploit.

Lemma 3: Let γ be an arbitrary timelike curve in Gödel space-time passing through the point p. Let the rotational Killing field $(\partial/\partial\varphi)^n$ be centered at p, and let q be any point on γ. Then

$$\mathrm{TA}(\gamma) \geqslant \ln[2E_\varphi(q) + (4E_\varphi^2(q) + 1)^{1/2}].$$

Proof: Just integrate

$$\begin{aligned} \mathrm{TA}(\gamma) &\geqslant \int_{\substack{\gamma \\ p \text{ to } q}} \frac{|\xi^n\nabla_n E_\varphi|}{[E_\varphi^2 + \frac{1}{4}]^{1/2}}\, ds \geqslant \int_{\substack{\gamma \\ p \text{ to } q}} \frac{dE_\varphi}{[E_\varphi^2 + \frac{1}{4}]^{1/2}} \\ &= \ln[2E_\varphi(q) + (4E_\varphi^2(q) + 1)^{1/2}]. \end{aligned}$$

IV. THE THEOREM

Now we concentrate attention on *closed* timelike curves in Gödel space-time. Given any one such γ, and any two points p and q on γ, there is a well-defined (radial) distance between p and q. Let the *diameter* of γ be the maximal value of this distance as p and q range over γ. The second key idea in our proof is the demonstration that this diameter cannot be arbitrarily small. Here we invoke statement (3) from Sec. II. We show that if the diameter were less than some minimal value, then the entire curve would have to fall within some critical cylinder; and that is impossible. The only slightly delicate point is that in computing that minimal value we cannot fall back on Euclidean plane geometry. The radial distance function is hyperbolic, not Euclidean.

Lemma 4: Let γ be a closed timelike curve in Gödel space-time with diameter D. Then $\operatorname{ch} D \geqslant (1+\sqrt{5})/2$.

Proof: We are really interested not so much in γ itself, but rather its (possibly self-intersecting) projection in some $t=\text{const}$, $y=\text{const}$ submanifold of Gödel space-time. Let γ' be this projection and let p and q be points on γ' which are maximally distant from one another. [So $d(p,q)=D$, where d is our distance function.] Further let s be the midpoint of the line segment connecting p and q. We show that if $\operatorname{ch} D < (1+\sqrt{5})/2$, then γ' is fully contained in the (open) disk of radius r_c centered at s. It will follow that γ itself is contained in the critical cylinder which has this disk as its projection, and we shall be done. (See Fig. 1.)

Let u be any point on γ'. Then $d(p,u) \leqslant D$ and $d(q,u) \leqslant D$. The angles $\measuredangle psu$ and $\measuredangle qsu$ cannot both be acute. Without loss of generality assume the former is not. By the counterpart to the "law of cosines" which holds in hyperbolic plane geometry[8] we have

$$\begin{aligned}\operatorname{ch}[d(p,u)] &= \operatorname{ch}[d(p,s)]\operatorname{ch}[d(s,u)]\\ &\quad - \operatorname{sh}[d(p,s)]\operatorname{sh}[d(s,u)]\cos\measuredangle psu.\end{aligned}$$

Since $\cos\measuredangle psu \leqslant 0$ it follows that

$$\operatorname{ch}[d(s,u)] \leqslant \frac{\operatorname{ch}[d(p,u)]}{\operatorname{ch}[d(p,s)]} \leqslant \frac{\operatorname{ch} D}{\operatorname{ch} D/2} = \frac{\sqrt{2}\operatorname{ch} D}{[\operatorname{ch} D + 1]^{1/2}}.$$

But now if $\operatorname{ch} D < (1+\sqrt{5})/2$, then $\operatorname{ch}[d(s,u)] < \sqrt{2}$ and we may conclude that $d(s,u) < r_c$.[9] ∎

Our proposition is a simple consequence of Lemmas 3 and 4. All we need is the fact that given any two timelike vectors λ^m, μ^m at a point (in any space-time), $\gamma^m\mu_m \geqslant (\lambda^m\lambda_m)^{1/2}(\mu^m\mu_m)^{1/2}$.

Proposition: Let γ be a closed timelike curve in Gödel space-time. Then

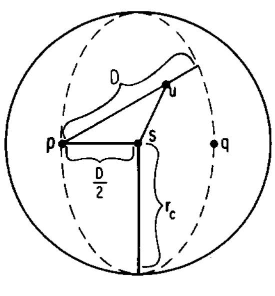


FIG. 1. Figure for Lemma 4.

$$\mathrm{TA}(\gamma) \geqslant \ln(2+\sqrt{5}).$$

Proof: Let p and q be any two points on γ which are maximally distant from one another. Consider a cylindrical coordinate system adapted to p. By Lemma 4 the r coordinate of q satisfies $\operatorname{ch} r \geqslant (1+\sqrt{5})/2$. Hence $(\operatorname{sh}^4 r - \operatorname{sh}^2 r)^{1/2} \geqslant 1$, and q falls outside the critical cylinder centered at p. Since $(\partial/\partial\varphi)^n$ is timelike at q, it must be the case at that point that

$$\begin{aligned}E_\varphi &= \xi^n\left(\frac{\partial}{\partial\varphi}\right)_n \geqslant \left[\left(\frac{\partial}{\partial\varphi}\right)^n\left(\frac{\partial}{\partial\varphi}\right)_n\right]^{1/2}\\ &= (\operatorname{sh}^4 r - \operatorname{sh}^2 r)^{1/2} \geqslant 1.\end{aligned}$$

Our claim now follows from Lemma 3. ∎

It is important that the point p in our proof need not be the one point on γ where there is a kink (if there is one at all).[2] Even if it is not, at least one of the connecting segments of γ between p and q must be smooth, and Lemma 3 can be applied to that one. If γ is smooth everywhere, then the argument can be applied *twice*, once on each segment, and the lower bound on $\mathrm{TA}(\gamma)$ can be raised by another factor of 2.

We can get some sense for magnitudes of total acceleration by considering another inequality[4] involving "fuel consumption." Suppose a point particle "rocket ship" traverses a timelike curve γ. Suppose its (rest) mass at any point is m. Then $\xi^n\nabla_n m \leqslant 0$ (since the rocket uses up fuel during the trip). Let J^n be the energy momentum of the rocket's exhaust. Assuming that the rocket is suitably isolated, J^n must balance precisely the rate at which the rocket itself loses energy momentum. So

$$J^n = -\xi^p\nabla_p(m\xi^n) = -[\xi^n(\xi^p\nabla_p m) + m\alpha^n].$$

Now J^n must be causal, i.e., $J^nJ_n \geqslant 0$. Therefore,

$$(\xi^p\nabla_p m)^2 - m^2a^2 \geqslant 0.$$

Since $(\xi^p\nabla_p m) \leqslant 0$ it follows that

$$a \leqslant -\xi^p\nabla_p(\ln m).$$

If m_i and m_f are, respectively, the initial and final mass of the rocket, then (by integration)

$$\mathrm{TA}(\gamma) \leqslant \ln(m_i/m_f).$$

This is the inequality we were looking for. It gives us a lower bound on that percent of the rocket's initial mass which must be in the form of fuel. Since $m_f - m_i$ is the fuel expended during the trip, we have

$$\text{Percent of initial mass as fuel} \geqslant \frac{m_i - m_f}{m_i} \geqslant 1 - \frac{1}{e^{\mathrm{TA}(\gamma)}}.$$

If $\mathrm{TA}(\gamma) = \ln(2+\sqrt{5})$ then the percent must already be greater than 76%. If $\mathrm{TA}(\gamma) = 2\pi(9+6\sqrt{3})^{1/2}$ (recall our calculation for Gödel circles), then the percent cannot differ from 100% by more than 2×10^{-12}.

We close by mentioning explicitly several questions which our discussion leaves open.

(a) Is there any closed timelike curve in Gödel space-time with total acceleration less than $2\pi(9+6\sqrt{3})^{1/2}$?

If the answer is yes, then we have the following questions.

(b) What is the greatest lower bound of $\mathrm{TA}(\gamma)$ as γ ranges over all closed timelike curves?

(c) Is that lower bound realized?

(d) What do the curves look like which realize or approach the bound?

ACKNOWLEDGMENTS

The problem here discussed was posed to me by Robert Geroch several years ago. He aroused my interest by taking very seriously the possibility that "time travel" *is* possible in Gödel space-time using arbitrarily small quantities of total acceleration. I wish to thank him, and also David Garfinkle, Lee Lindblom, and Robert Wald, for numerous helpful discussions.

[1]See, for example, K. Gödel, Rev. Mod. Phys. **21**, 447 (1949); W. Kundt, Z. Phys. **145**, 611 (1956); S. Chandrasekhar and J. P. Wright, Proc. Natl. Acad. Sci. **47**, 341 (1961); H. Stein, Philos. Sci. **37**, 589 (1970); J. Pfarr, Gen. Relativ. Gravit. **13**, 1073 (1981).

[2]In what follows, "timelike curves" will be taken to be smooth everywhere unless they are closed, in which case smoothness will be allowed to fail at initial (= terminal) points.

[3]By means of such a rescaling we can always make the maximal value of a along γ as small as we like. The point is that any such "saving" is exactly balanced by a corresponding increase in elapsed proper time along γ.

[4]S. Chakrabarti, R. Geroch, and G. Liang, J. Math. Phys. **24**, 597 (1983).

[5]Let $h_{mn} = g_{mn} - (\partial/\partial t)_m(\partial/\partial t)_n$ be the (negative semidefinite) metric which results from projecting g_{mn} orthogonal to $(\partial/\partial t)^m$. Since $(\partial/\partial t)_m = \nabla_m t + \sqrt{2}\,\mathrm{sh}^2 r \nabla_m \varphi$, we have

$$-h_{mn} = (\nabla_m r)(\nabla_n r) + \tfrac{1}{4}\mathrm{sh}^2 2r(\nabla_m \varphi)(\nabla_n \varphi).$$

Let S be any $t = \text{const}$, $y = \text{const}$ submanifold, and construe $-h_{mn}$ as a (positive definite) metric on S. It suffices for us to show that $(S, -h_{mn})$ is a complete Riemannian manifold of constant negative curvature. (The value of curvature is $-\frac{1}{4}$.) There are various ways to do this. One is the following. Consider new coordinates on S defined by

$$x_1 = \tfrac{1}{2}\mathrm{ch}\, 2r, \quad x_2 = \tfrac{1}{2}\mathrm{sh}\, 2r \cos \varphi, \quad x_3 = \tfrac{1}{2}\mathrm{sh}\, 2r \sin \varphi.$$

For all r and φ we have $x_1 > 0$ and $x_1^2 - x_2^2 - x_3^2 = \frac{1}{4}$. Furthermore, in these coordinates the metric assumes the form

$$-h_{mn} = -(\nabla_m x_1)(\nabla_n x_1) + (\nabla_m x_2)(\nabla_n x_2) + (\nabla_m x_3)(\nabla_n x_3).$$

Therefore, $(S, -h_{mn})$ is isometric to the upper half of a two-sheeted hyperboloid of radius $\frac{1}{2}$ in R^3, with respect to the metric induced on the latter by a background flat metric of Lorentz signature. It is a standard result that this hyperboloid (under the induced metric) is a complete Riemannian manifold of constant curvature $-\frac{1}{4}$. [See, for example, J. Wolf, *Spaces of Constant Curvature* (Publish or Perish, Boston, 1974), Chap. 2.]

[6]It must be shown that $\nabla_m t$ is timelike and future directed within a critical cylinder. That is easy. The inverse to g_{mn} is given by

$$\begin{aligned} g^{mn} = \frac{1}{(\mathrm{sh}^4 r + \mathrm{sh}^2 r)} \Big[& -(\mathrm{sh}^4 r - \mathrm{sh}^2 r)\left(\frac{\partial}{\partial t}\right)^m\left(\frac{\partial}{\partial t}\right)^n \\ & -(\mathrm{sh}^4 r + \mathrm{sh}^2 r)\left(\frac{\partial}{\partial r}\right)^m\left(\frac{\partial}{\partial r}\right)^n \\ & -(\mathrm{sh}^4 r + \mathrm{sh}^2 r)\left(\frac{\partial}{\partial y}\right)^m\left(\frac{\partial}{\partial y}\right)^n - \left(\frac{\partial}{\partial \varphi}\right)^m\left(\frac{\partial}{\partial \varphi}\right)^n \\ & + 2\sqrt{2}\,\mathrm{sh}^2 r\left(\frac{\partial}{\partial \varphi}\right)^{(m}\left(\frac{\partial}{\partial t}\right)^{n)}\Big], \end{aligned}$$

and hence

$$(\nabla_m t)(\nabla^m t) = (1 - \mathrm{sh}^2 r)/(1 + \mathrm{sh}^2 r).$$

So clearly $\nabla_n t$ is timelike if and only if $r < r_c$. Also, $\nabla_n t$ is future directed within the cylinder since $(\partial/\partial t)^n(\nabla_n t) = 1$.

[7]The angular coordinate ϕ is not defined at p, but that does not matter. The vector $(\partial/\partial\varphi)^n$ goes to the zero vector as p is approached, and E_φ goes to 0.

[8]See almost any book on non-Euclidean geometry; e.g., W. T. Fishback, *Projective and Euclidean Geometry* (Wiley, New York, 1969), p. 257.

[9]Of course the value $(1 + \sqrt{5})/2$ was obtained by working this computation backwards.

Perfect fluid spheres admitting a one-parameter group of conformal motions

L. Herrera[a)]
Departamento de Física, Facultad de Ciencias, Universidad Central de Venezuela, Caracas 1051, Venezuela
J. Ponce de León
Departamento de Física, Facultad de Ciencias, Universidad Central de Venezuela, Caracas 1051, Venezuela and Centro de Física, I.V.I.C., Apdo. 1827, Caracas 1010-A, Venezuela

(Received 25 May 1984; accepted for publication 3 August 1984)

Some exact analytical solutions of the Einstein equations for perfect fluids were found under the assumption of spherical symmetry and the existence of a one-parameter group of conformal motions. The first solution exhibited represents a nonstatic homogeneous spherically symmetric distribution of matter which is singular at $t = 0$. Two other solutions represent contracting and expanding fluids, respectively, whose evolution tends asymptotically to a static sphere with a surface gravitational potential equal to $\frac{1}{3}$. These two solutions possess vanishing pressure surfaces which are not the boundary of matter except in the static limit. Finally an oscillating distribution of matter is presented.

I. INTRODUCTION

The extent to which analytical solutions of the Einstein equations[1–11] could be of any use in the study of different stages of stellar evolution is conditioned by the number and character of the simplifying assumptions when integrating the field equations and also when choosing the equations of state. Unfortunately, integrating the Einstein equations, for realistic equations of state, without any further simplification is extremely difficult. It seems useful then to introduce additional restrictions, which although not fully justified from the physical point of view may lead to solutions which contain some of the essential features of a realistic situation.

We propose in this paper to integrate the Einstein equations for perfect fluids, under the assumption that the space-time admits, besides the spherical symmetry, a one-parameter group of conformal motions, i.e.,

$$\underset{\xi}{L}\, g_{\alpha\beta} = \psi g_{\alpha\beta}, \tag{1}$$

where the left-hand side is the Lie derivative of the metric tensor and ψ is an arbitrary function of the coordinates.

For the case $\psi = \text{const}$ (homothetic motions), a number of interesting results have been obtained in the past with possible applications in astrophysics and cosmology.[12–15] Also, for arbitrary choices of ψ we were able to find static solutions, and to establish a link between the group of special conformal motions ($\nabla_\alpha \nabla_\beta \psi = 0$) and the stiff equation of state (pressure equal energy density).[16]

In this work we shall see how the analytical integration of the Einstein equations (under the assumptions above) may be accomplished, leading to solutions representing different pictures of self-gravitating spheres.

The simplest solution which can be found represents an homogeneous spherically symmetric distribution of matter which is singular at the initial time $t = 0$.

Two other solutions represent expanding and contracting fluids, respectively. Both solutions possess positive energy and tend asymptotically to a static sphere with a surface gravitational potential equal to $\frac{1}{3}$. We shall see that there exist vanishing pressure surfaces, for both solutions, which are not the boundary of matter except in the static limit. Finally we shall present a solution with an oscillating behavior.

The paper is organized as follows. The field equations as well as the conventions used are included in Sec. II. In Sec. III we integrate the field equations and display some solutions. Some details of calculations are included in Appendixes A and B.

II. THE FIELD EQUATIONS AND CONVENTIONS

Let us consider a nonstatic distribution of matter represented by a perfect fluid and which is spherically symmetric.

In comoving coordinates the line element may be written as[17]

$$ds^2 = e^\nu dt^2 - e^\lambda dr^2 - e^\mu d\Omega^2, \tag{2}$$

with

$$d\Omega^2 = d\theta^2 + \sin^2\theta\, d\phi^2; \quad x^{0,1,2,3} \equiv t,r,\theta,\phi,$$

where λ, ν, and μ are functions of r and t. For the energy-momentum tensor we have the usual expression

$$T^\mu_\nu = (\rho + p) U^\mu U_\nu - p\delta^\mu_\nu, \tag{3}$$

with ρ, p, and U^μ denoting the energy density, the pressure, and the four velocity of the fluid, respectively. Also, since we are in a comoving frame,

$$U^\mu = \delta^\mu_0 e^{-\nu/2}. \tag{4}$$

Thus the Einstein equations

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = 8\pi T_{\mu\nu} \tag{5}$$

read

$$-8\pi T^1_1 = 8\pi p = \tfrac{1}{2} e^{-\lambda}(\mu'^2/2 + \mu'\nu') - e^{-\nu}(\ddot\mu - \tfrac{1}{2}\dot\mu\dot\nu + \tfrac{3}{4}\dot\mu^2) - e^{-\mu}, \tag{6}$$

$$-8\pi T^2_2 = 8\pi p = \tfrac{1}{4} e^{-\lambda}(2\nu'' + \nu'^2 + 2\mu''$$

a) Postal address: Apartado 80793, Caracas 1080 A, Venezuela.

 0022-2488/85/040778-07$02.50

$$+\mu'^2-\mu'\lambda'-\nu'\lambda'+\mu'\nu')$$
$$+\tfrac{1}{4}e^{-\nu}(\dot\lambda\dot\nu+\dot\mu\dot\nu-\dot\lambda\dot\mu-2\ddot\lambda-\dot\lambda^2-2\ddot\mu-\dot\mu^2), \tag{7}$$

$$8\pi T_0^0=8\pi\rho=-e^{-\lambda}(\mu''+\tfrac{3}{4}\mu'^2-\mu'\lambda'/2)$$
$$+\tfrac{1}{2}e^{-\nu}(\dot\lambda\dot\mu+\dot\mu^2/2)+e^{-\mu}, \tag{8}$$

$$8\pi T_0^1=0=\tfrac{1}{2}e^{-\lambda}(2\dot\mu'+\dot\mu\mu'-\dot\lambda\mu'-\nu'\dot\mu), \tag{9}$$

(dots and primes denote differentiation with respect to t and r, respectively).

Next, we shall assume that the space-time admits a one-parameter group of conformal motions, i.e.,

$$\underset{\xi}{L}\, g_{\mu\nu}\equiv\xi_{\mu;\nu}+\xi_{\nu;\mu}=\psi g_{\mu\nu}, \tag{10}$$

where ψ is an arbitrary function of t and r. We shall further restrict the vector field ξ^α, by demanding

$$\xi^\alpha U_\alpha=0. \tag{11}$$

Then as a consequence of the spherical symmetry and from Eq. (11) we have

$$\xi^0=\xi^2=\xi^3=0. \tag{12}$$

Thus, using (2) and (12) we get from Eq. (10)

$$\nu'\xi^1=\psi, \tag{13}$$

$$\xi^1{}_{,0}=0, \tag{14}$$

$$\lambda'\xi^1+2\xi^1{}_{,1}=\psi, \tag{15}$$

$$\mu'\xi^1=\psi, \tag{16}$$

(subscripted commas denote partial derivatives). It can be seen at once from (13) and (16) that

$$\nu-\mu=f_1(t), \tag{17}$$

where $f_1(t)$ is an arbitrary function of t.

Next, taking derivatives of (15) and (16) with respect to t, and using (14), we obtain the equation

$$\dot\lambda'=\dot\mu', \tag{18}$$

from which it follows that

$$\lambda-\mu=f_2(t)+g_1(r), \tag{19}$$

f_2 and g_1 being arbitrary functions of their arguments.

We still have the freedom to perform a coordinate transformation of the form[17]

$$t=t(\tilde t),\quad r=r(\tilde r).$$

Thus, without loss of generality we may choose $f_1(t)=g_1(r)=0$, and then one obtains

$$\nu-\mu=0,\quad \lambda-\mu=f(t), \tag{20}$$

Feeding back (20) in (15) and using (13) we obtain

$$\xi^1=A=\text{const},\quad \psi=A\nu'. \tag{21}$$

Expressions (20) and (21) contain all the implications derived from the existence of the conformal motion with ξ^μ orthogonal to U^μ.

Let us now turn to the field equations (6)–(9). Using (20) we get from Eq. (9)

$$2\dot\lambda'-\dot\lambda\lambda'=0. \tag{22}$$

Introducing the new variable

$$Z\equiv e^{-\lambda/2},$$

Eq. (22) becomes

$$\dot Z'=0,$$

whose solution has the form

$$Z\equiv e^{-\lambda/2}=h_1(r)+h_2(t), \tag{23}$$

where h_1 and h_2 are two unknown functions of their arguments. Again using Eq. (20) we obtain

$$e^{-\nu/2}=e^{-\mu/2}=e^{f(t)/2}[h_1(r)+h_2(t)]. \tag{24}$$

We can now write down the field equations (6)–(9) in terms of the functions $f(t)$, $h_1(r)$ and $h_2(t)$. We get

$$-8\pi T_1^1=8\pi p=[3h_1'^2(r)-3\dot h_2^2(t)e^{f(t)}]+e^{-\lambda/2}e^{f(t)}$$
$$\times[2\ddot h_2(t)-\dot h_2(t)\dot f(t)]+e^{-\lambda}e^{f(t)}$$
$$\times[\ddot f(t)-\dot f^2(t)/4-1], \tag{25}$$

$$-8\pi T_2^2=8\pi p=[3h_1'^2(r)-3\dot h_2^2(t)e^{f(t)}]-2e^{-\lambda/2}$$
$$\times[h_1''(r)-\ddot h_2(t)e^{f(t)}]+\tfrac{1}{2}\ddot f(t)e^{-\lambda}e^{f(t)}, \tag{26}$$

$$8\pi T_0^0=8\pi\rho=-[3h_1'^2(r)-3\dot h_2^2(t)e^{f(t)}]+2e^{-\lambda/2}$$
$$\times[h_1''(r)+\dot h_2(t)\dot f(t)e^{f(t)}]$$
$$+e^{-\lambda}e^{f(t)}[\dot f^2(t)/4+1], \tag{27}$$

and for the line element we have

$$ds^2=\frac{e^{-f(t)}}{[h_1(r)+h_2(t)]^2}[dt^2-e^{f(t)}dr^2-d\Omega^2]. \tag{28}$$

It will be seen in the next section that the system (25)–(27) can be integrated analytically without the introduction of additional restrictions on the metric functions or on the equation of state.

III. INTEGRATING THE FIELD EQUATIONS

Because of the local isotropy of the pressure, we obtain from (25) and (26)

$$e^{\lambda/2}[\dot h_2(t)\dot f(t)e^{f(t)}-2h_1''(r)]=\Phi(t), \tag{29}$$

with

$$4\Phi(t)=(2\ddot f(t)-\dot f^2(t)-4)e^{f(t)}. \tag{30}$$

Taking derivatives of (29) with respect to r, we get

$$(\lambda'/2)\Phi(t)-2h_1'''(r)e^{\lambda/2}=0, \tag{31}$$

and from (23)

$$-(\lambda'/2)e^{-\lambda/2}=h_1'(r). \tag{32}$$

Using (31) and (32) we are led to

$$-2h_1'''(r)/h_1'(r)=\Phi(t), \tag{33}$$

which implies at once that $\Phi(t)=C_1=\text{const}$. We can now integrate (33) to obtain

$$2h_1''(r)+C_1h_1(r)=2C_2, \tag{34}$$

where C_2 is a constant of integration. Next we can use (34) to rewrite (29) in the form

$$\dot h_2(t)\dot f(t)e^{f(t)}-C_1h_2(t)=2C_2. \tag{35}$$

Also Eq. (30) may be written as

$$[2\ddot f(t)-\dot f^2(t)-4]e^{f(t)}=4C_1. \tag{36}$$

Thus, the symmetry assumptions allowed us to reduce the Einstein equations to a system of ordinary differential equations for the three unknown functions f, h_1, and h_2 [Eqs. (34)–(36)].

Let us now turn to the integration of Eq. (34). Setting $h_1'(r) = Q(h_1)$ we have

$$2Q\frac{dQ}{dh_1} + C_1 h_1 = 2C_2, \tag{37}$$

whose solution may be written as

$$Q^2 = h_1'^2(r) = 2C_2 h_1 - (C_1/2)h_1^2 + C_4. \tag{38}$$

Introducing the new variable

$$R(r,t) = e^{-f(t)/2}/(h_1(r) + h_2(t)) = e^{-f(t)/2}e^{\lambda/2}, \tag{39}$$

we may express $h_1(r)$ in terms of $f(t)$, R, and $h_2(t)$,

$$h_1(r) = e^{-f(t)/2}/R - h_2(t), \tag{40}$$

and feeding (40) back into (38) gives

$$h_1'^2 = 2\frac{C_2 e^{-f(t)/2}}{R} - 2C_2 h_2(t) - \frac{C_1}{2}\frac{e^{-f(t)}}{R^2} + C_1\frac{e^{-f(t)/2}}{R}h_2(t) - \frac{C_1}{2}h_2^2(t) + C_4. \tag{41}$$

Next, a first integral of (36) may be obtained at once; viz.

$$\dot f^2(t) = 4C_3 e^{f(t)} - 4 - 2C_1 e^{-f(t)}, \tag{42}$$

where C_3 is a new constant of integration. Now the expressions for the pressure and density, using (41), (42), (34), and (35), read

$$8\pi p = \frac{C_3}{R^2}e^{f(t)} - \frac{4C_3 e^{f(t)/2}}{R}\frac{[2C_2 + C_1 h_2(t)]}{\dot f^2(t)} - G(t), \tag{43}$$

$$8\pi\rho = (C_3/R^2)e^{f(t)} + G(t), \tag{44}$$

where

$$G(t) \equiv 6C_2 h_2(t) + \tfrac{3}{2}C_1 h_2^2(t) - 3C_4 + 3[2C_2 + C_1 h_2(t)]^2/\dot f^2 e^{f(t)}. \tag{45}$$

Since we are going to consider solutions for which $\rho \geqslant p$, $G(t)$ must satisfy the inequality

$$4C_3 e^{f(t)/2}[2C_2 + C_1 h_1(t)]/R\dot f^2(t) + 2G(t) \geqslant 0, \tag{46}$$

we notice that C_3 should be different from zero, for otherwise the pressure would be negative.

Let us now exhibit the different solutions which may be obtained by different choices of the constants of integration.

Solution 1: Let us start by considering the case $C_1 = C_2 = 0$. From (45) and (46) it follows that $C_4 \leqslant 0$. On the other hand, from (41), $C_4 \geqslant 0$. Thus we must put $C_4 = 0$. Using (41) and (35) we have

$$h_1'^2(r) = C_4 = 0, \quad \dot h_2(t) = 0,$$

and from (42), using $x = e^{-f(t)/2}$,

$$\dot x^2 + x^2 = C_3. \tag{47}$$

Integrating, we obtain

$$x = e^{-f(t)/2} = \sqrt{C_3}\sin(t - t_0).$$

We may choose $t_0 = 0$ without loss of generality, then for the pressure, energy density, and the line element we obtain

$$\rho = p = \frac{\text{const}}{\sin^4 t}, \tag{48}$$

$$ds^2 = \text{const } C_3 \sin^2 t\left\{dt^2 - \frac{dr^2}{C_3\sin^2 t} - d\Omega^2\right\}. \tag{49}$$

Also for the kinematical quantities a_μ and θ we get

$$a_\mu = U_{\mu;\sigma}U^\sigma = h_1'(r)\delta_\mu^1/(h_1(r) + h_2(t)) = 0, \tag{50}$$

$$\theta = U^\mu_{;\mu} = -e^{f(t)/2}\{3\dot h_2(t) + \dot f(t)e^{-\lambda/2}\} = \text{const}\cos t/\sin^2 t. \tag{51}$$

Solution 2: Let us consider the case $C_1 \neq 0$. First of all, it should be noted that the choice of the sign of C_1 will have a strong influence in the behavior of the different possible solutions. In fact, if we choose C_1 as a positive or zero constant, then it is clear from Eq. (36) that any possible solution will never enter into the stationary regime. We shall analyze the case in which C_1 is negative; thus, we put

$$C_1 = -2\omega^2 \quad (\omega \neq 0).$$

Then Eq. (42) reads

$$\dot x^2 = \omega^2 x^4 - x^2 + C_3, \tag{52}$$

with $x = e^{-f(t)/2}$. The solution of the above equation, in general, is expressed in terms of elliptic functions (see Appendix A). However, there is one choice of the constant C_3 for which it is possible to find solutions expressed solely in terms of elementary functions; namely $C_3 = 1/4\omega^2$. Since we are interested in analytical solutions, we shall restrict ourselves to this choice. Thus, we obtain from (52)

$$\pm dt = \frac{dx}{(\omega x^2 - 1/2\omega)}, \tag{53}$$

from which it follows that

$$\pm(t - t_0) = \frac{1}{(2)^{1/2}}\ln\left|\frac{\omega x(2)^{1/2} - 1}{\omega x(2)^{1/2} + 1}\right|. \tag{54}$$

Without loss of generality we may choose $t_0 = 0$. Then, solving (54) for x, we found two possible solutions

$$e_a^{-f(t)/2} = (1/\sqrt{2}\omega)\coth(t/\sqrt{2}), \tag{55}$$

$$e_b^{-f(t)/2} = (1/\sqrt{2}\omega)\tanh(t/\sqrt{2}) \tag{56}$$

[a and b denoting hereafter the solutions corresponding to (55) and (56), respectively]. We can now integrate Eq. (35). After some elementary manipulations we get

$$\ln|2C_2 - 2\omega^2 h_2(t)| = -2\omega^2\int\frac{e^{-f(t)}}{\dot f(t)}dt. \tag{57}$$

Using (55) and (56) in (57) we find the following two solutions:

$$h_{2,a}(t) = \frac{B}{2\omega^2}\frac{e^{-(1/2)\sinh^2(t/2^{1/2})}}{\sinh(t/2^{1/2})} + \frac{C_2}{\omega^2}, \tag{58}$$

and

$$h_{2,b}(t) = \frac{D}{2\omega^2}\frac{e^{-(1/2)\cosh^2(t/2^{1/2})}}{\cosh(t/2^{1/2})} + \frac{C_2}{\omega^2}, \tag{59}$$

where B and D are two constants of integrations.

Finally, Eq. (34) can be easily integrated, we obtain

$$h_1(r) = C\cosh(\omega r + \alpha) - C_2/\omega^2, \tag{60}$$

where C is related with the previous constants C_2 and C_4 of Eq. (38) as follows:

$$C^2 = C_2^2/\omega^4 - C_4/\omega^2,$$

α is a new constant of integration.

Let us write the metric functions in its final form. Using (55), (58), and (60) we get for solution a

$$e_a^\lambda = \left[C\cosh(\omega r+\alpha) + \frac{B}{2\omega^2} \times \frac{\exp[-(1/2)\sinh^2(t/2^{1/2})]}{\sinh(t/2^{1/2})}\right]^{-2},$$

$$e_a^\nu = e_a^\mu = \frac{\coth^2(t/2^{1/2})}{2\omega^2}\left[C\cosh(\omega r+\alpha) + \frac{B}{2\omega^2} \times \frac{\exp[-(1/2)\sinh^2(t/2^{1/2})]}{\sinh(t/2^{1/2})}\right]^{-2}. \tag{61}$$

For solution b, using Eqs. (56), (59), and (60), we obtain

$$e_b^\lambda = \left[\cosh(\omega r+\alpha) + \frac{D}{2\omega^2}\frac{\exp[(1/2)\cosh^2(t/2^{1/2})]}{\cosh(t/2^{1/2})}\right]^{-2},$$

$$e_b^\mu = e_b^\nu = \frac{\tanh^2(t/2^{1/2})}{2\omega^2}\left[C\cosh(\omega r+\alpha) + \frac{D}{2\omega^2} \times \frac{\exp[(1/2)\cosh^2(t/2^{1/2})]}{\cosh(t/2^{1/2})}\right]^{-2}. \tag{62}$$

Now we proceed to analyze solutions a and b. For the sake of simplicity we shall restrict our attention to the specific subcases $B = D = 0$. Then we obtain for the line element, the pressure, and the energy density the following expressions:

$$ds_a^2 = \frac{\coth^2(t/\sqrt{2})}{2\omega^2 C^2\cosh^2(\omega r+\alpha)} \times [dt^2 - 2\omega^2\tanh^2(t/2^{1/2})\,dr^2 - d\Omega^2], \tag{63}$$

$$ds_b^2 = \frac{\tanh^2(t/\sqrt{2})}{2\omega^2 C^2\cosh^2(\omega r+\alpha)} \times [dt^2 - 2\omega^2\coth^2(t/2^{1/2})\,dr^2 - d\Omega^2], \tag{64}$$

$$p_{a,b} = \overline{C}\,[R_0^2(t)_{a,b}/R^2 - 1], \tag{65}$$

$$\rho_{a,b} = \overline{C}\,[R_0^2(t)_{a,b}/R^2 + 1], \tag{66}$$

with $\overline{C} = 3C^2\omega^2/8\pi$,

$$R_{0,a}(t) = (1/\sqrt{6}\omega C)\tanh(t/\sqrt{2}), \tag{67}$$

$$R_{0,b}(t) = (1/\sqrt{6}\omega C)\coth(t/\sqrt{2}). \tag{68}$$

The following remarks are in order at this point.

(i) The vanishing pressure surface $R = R_0(t)_{a,b}$ is not the boundary of the source. This can be seen from the fact that

$$U^\mu n_\mu \neq 0,$$

where n_μ is the normal vector to the vanishing pressure surface. Indeed

$$U^\mu n_\mu = e^{-\nu/2}\frac{\partial}{\partial x^0}(R - R_0(t)_{a,b}) \neq 0.$$

(ii) Solution a represents a contracting sphere (ρ is an increasing function of t) with an expanding vanishing pressure surface.

(iii) Solution b represents an expanding sphere (ρ is a decreasing function of t) with a contracting vanishing pressure surface.

(iv) The energy density is positive everywhere and bigger than the pressure for all values of t and r.

(v) Both solutions (a and b) have the same final (asymptotically) configuration. In fact, taking the limit $t \to \infty$ in (63)–(66), we get

$$ds_{a,b}^2 = (2\omega^2 C^2\cosh^2(\omega r+\alpha))^{-1}(dt^2 - 2\omega^2\,dr^2 - d\Omega^2), \tag{69}$$

$$p_{a,b} = \overline{C}\,[R_0^2(\infty)/R^2 - 1], \tag{70}$$

$$\rho_{a,b} = \overline{C}\,[R_0^2(\infty)/R^2 + 1], \tag{71}$$

where

$$R_0^2(\infty) = 1/\sqrt{6}\omega C.$$

Next, transforming the solution (70) to Schwarzschild-like coordinates (see Appendix B),

$$R = (\sqrt{2}\omega C\cosh(\omega r+\alpha))^{-1}, \quad T = t/a, \quad a = \text{const}, \tag{72}$$

we get

$$ds_{a,b}^2 = a^2R^2dT^2 - \frac{2}{[1 - R^2/3R_0^2]}dR^2 - R^2\,d\Omega^2. \tag{73}$$

This solution was previously found[16] and represents a static sphere with a surface potential $M/R_0 = \frac{1}{3}$.

Finally, we would like to discuss further the process of expansion (contraction) of the vanishing pressure surface by means of the concept of the radial velocity of the surface, as measured by a locally Minkowskian observer.[18]

In the Schwarzschild-like coordinates the line element has the form (see Appendix B)

$$ds^2 = g_{TT}(R,T)dT^2 + g_{RR}(R,T)dR^2 - R^2(d\theta^2 + \sin^2\theta\,d\theta^2). \tag{74}$$

Then introducing, purely locally, Minkowski coordinates (τ,x,y,z) by

$$d\tau = \sqrt{g_{TT}}dt, \quad dx = \sqrt{-g_{RR}}\,dR,$$
$$dy = R\,d\theta, \quad dz = R\sin\theta\,d\phi,$$

we define the radial velocity of the surface as

$$V_{a,b} = \left(\frac{dx}{d\tau}\right)_S = \left[\sqrt{-\frac{g_{RR}}{g_{TT}}}\right]_S\frac{dR_0(t)_{a,b}}{dT}, \tag{75}$$

where the subscript S indicates that the quantity is evaluated at the surface of the sphere, and $R_0(t)_{a,b}$ are given by (67) and (68).

Using (B5) we can put Eq. (75) in the form

$$V_a = \frac{R_0(\infty)}{2^{1/2}\cosh^2(t/2^{1/2})}\left[\frac{R'}{J_a}\sqrt{-\frac{g_{RR}}{g_{TT}}}\right]_S,$$

$$V_b = \frac{R_0(\infty)}{2^{1/2}\sinh^2(t/2^{1/2})}\left[\frac{R'}{J_b}\sqrt{-\frac{g_{RR}}{g_{TT}}}\right]_S, \tag{76}$$

where $J_{a,b}$ denotes the Jacobian of the transformation (B11) and (B15).

It is clear from (B11), (B15), and (B10) that V_a is always positive and V_b is always negative.

Furthermore, from the fact that

$$g_{rt} = \frac{\partial R}{\partial r}\frac{\partial R}{\partial t} g_{RR} + \frac{\partial T}{\partial r}\frac{\partial T}{\partial t} g_{TT} = 0,$$

we get

$$\frac{T'\dot{T}}{R'\dot{R}} = -\frac{g_{RR}}{g_{TT}},$$

which leads to the following expression for the square of the radial velocity of the surface:

$$V_a^2 = \frac{R_0^2(\infty)}{2}\frac{1}{\cosh^4(t/2^{1/2})}\left[\frac{\dot{T}T'R'}{J_a^2\dot{R}}\right]_S, \tag{77}$$

$$V_b^2 = \frac{R_0^2(\infty)}{2}\frac{1}{\sinh^4(t/2^{1/2})}\left[\frac{\dot{T}T'R'}{J_b^2\dot{R}}\right]_S. \tag{78}$$

Now, from the definition of the Jacobian

$$J = R'\dot{T} - \dot{R}T'$$

and

$$g_{RT} = \frac{\partial t}{\partial R}\frac{\partial t}{\partial T} g_{tt} + \frac{\partial r}{\partial T}\frac{\partial r}{\partial R} g_{rr} = 0,$$

we obtain

$$\frac{\dot{T}}{J} = \left(R'\left[\frac{\dot{R}^2}{R'^2}\frac{g_{rr}}{g_{tt}} + 1\right]\right)^{-1},$$

$$\frac{T'}{J} = -\dot{R}g_{11}\left(R'^2 g_{00}\left[\frac{\dot{R}^2}{R'^2}\frac{g_{rr}}{g_{tt}} + 1\right]\right)^{-1}. \tag{79}$$

Using metric functions of (63) and (64) we finally obtain for V_a and V_b, respectively,

$$V_a^2 = \frac{3[3-\tanh^4(t/2^{1/2})]\cosh^4(t/2^{1/2})}{[(3-\tanh^4(t/2^{1/2}))\cosh^4(t/2^{1/2}) - 3]^2}, \tag{80}$$

$$V_b^2 = \frac{3[3-\coth^4(t/2^{1/2})]\sinh^4(t/2^{1/2})}{[(3-\coth^4(t/2^{1/2}))\sinh^4(t/2^{1/2}) - 3]^2}. \tag{81}$$

From this expression it follows that $V_{a,b} \to 0$ as $t \to \infty$, as was indicated before.

Solution 3: We shall see that it is possible to construct an oscillating source, if we only choose appropriately the constants of integration. With this aim, let us take $C_2 = 0$ and $\dot{h}_2 = 0$. Then from (35) we obtain

$$C_1 h_2(t) = 0. \tag{82}$$

Choosing $C_1 \neq 0$, we have $h_2 = 0$.

Next, from Eqs. (43)–(44) we get for the pressure and the energy density

$$p = \overline{C}\,[R_0^2(t)/R^2 - 1], \tag{83}$$

$$\rho = \overline{C}\,[R_0^2(t)/R^2 + 1], \tag{84}$$

with

$$\overline{C} = -3C_4/8\pi > 0, \quad C_4 < 0,$$

and

$$\rho = p + 2\overline{C},$$

$$R_0^2(t) = (C_3/8\pi C)e^{f(t)}. \tag{85}$$

If we now substitute (85) in (42) we obtain the following equation for $R_0(t)$:

$$\dot{R}_0^2 = -3C_4R_0^4 - R_0^2 + C_1C_3/6C_4.$$

Taking derivatives with respect to t, we obtain

$$\ddot{R}_0 + R_0 = -6C_4R_0^3. \tag{86}$$

Equation (86) can be solved approximately if $|C_4| \ll 1$. In this case, one gets (see pp. 86–87, in Ref. 19), up to terms of first order in C_4

$$R_0(t) = a\cos\omega t - (3a^3/16)C_4\cos 3\omega t \tag{87}$$

with

$$\omega = 1 + \tfrac{9}{4}a^2C_4: \quad a = \text{const}.$$

Feeding back (87) into (83)–(84), we explicitly display the oscillating behavior of the matter variables.

APPENDIX A: THE GENERAL INTEGRAL OF EQ. (52)

In this appendix we shall give the most general integrals of Eq. (52). First of all, observe that Eq. (52) may be written as

$$\pm \int dt = \int \frac{dx}{((\omega x^2 - 1/2\omega)^2 + (C_3 - 1/4\omega^2))^{1/2}}, \tag{A1}$$

which leads to three different solutions depending upon the three possible choices of C_3 namely,

$$C_3 = 1/4\omega^2, \quad C_3 > 1/4\omega^2, \quad C_3 < 1/4\omega^2.$$

Thus we can write

$$\pm \int dt = I_i, \quad i = 0,1,2, \tag{A2}$$

with

$$I_0 = \int \frac{dx}{(\omega x^2 - 1/2\omega)}, \tag{A3}$$

$$I_1 = \int \frac{dx}{((\omega x^2 - 1/2\omega)^2 + H^2)^{1/2}}, \tag{A4}$$

$$I_2 = \int \frac{dx}{((\omega x^2 - 1/2\omega)^2 - H^2)^{1/2}}, \tag{A5}$$

where

$$H^2 = |C_3 - 1/4\omega^2| \neq 0.$$

The integral (A3) may be calculated easily, with the result

$$I_0 = \frac{1}{2^{1/2}}\ln\left|\frac{\omega x 2^{1/2} - 1}{\omega x 2^{1/2} + 1}\right|. \tag{A6}$$

This expression leads to the solutions (55) and (56) presented in Sec. III of the paper.

To calculate the integral (A4) it is useful to introduce the new variable β

$$H\sinh\beta = (\omega x^2 - 1/2\omega). \tag{A7}$$

In terms of β, I_1 may be written as

$$I_1 = \frac{1}{2\omega}\int \frac{d\beta}{((H/\omega)\sinh\beta + 1/2\omega^2)^{1/2}}$$

or (see Ref. 20, p. 130)

$$I_1 = [\sqrt{2}\,(1 + 4H^2\omega^2)^{1/4}]^{-1}F(\phi,K), \tag{A8}$$

where

$\phi = \arccos(\sqrt{1+4H^2\omega^2} - 1 - 2\omega H \sinh\beta)/$

$(\sqrt{1+4H^2\omega^2} + 1 + 2\omega H \sinh\beta),$

$$k = \left(\frac{1+(1+4H^2\omega^2)^{1/2}}{2(1+4H^2\omega^2)^{1/2}}\right)^{1/2},$$

$H/\omega > 0, \quad \beta > -\operatorname{arcsinh}(1/2\omega H),$

and $F(\phi,k)$ is an elliptic integral of the first kind, defined by

$$F(\phi,k) = \int_0^\phi \frac{d\beta}{(1-K^2\sin^2\beta)^{1/2}}$$

(see Ref. 20, p. 918).

To calculate I_2 we introduce the new variable γ, as

$$H\cosh\gamma = (\omega x^2 - 1/2\omega).$$

A simple calculation gives (see Ref. 20, p. 130)

$$I_2 = \frac{1}{2\omega}\int \frac{d\gamma}{((H/\omega)\cosh\gamma + 1/2\omega^2)^{1/2}} = \frac{\sqrt{2}}{(1+2\omega H)^{1/2}} F(\phi,k) \tag{A9}$$

with

$$\phi = \arcsin(\tanh\gamma/2),$$

$$k = \sqrt{\frac{1-2\omega H}{1+2\omega H}}, \quad \frac{1}{2\omega H} > 1, \quad \gamma > 0$$

and $F(\phi,k)$ is again an elliptic integral of the first kind.

APPENDIX B: THE SCHWARZSCHILD-LIKE COORDINATES

In this appendix we shall derive the formulas for the transformation from the comoving coordinates to Schwarzschild-like coordinates, for both, the expanding and the contracting solution.

Let us start with the contracting solution, whose line element is given by (63). We want to transform to a coordinate system (T,R,Θ,Φ), such that the metric takes the form

$$ds^2 = g_{TT}(R,T)dT^2 + g_{RR}(R,T)dR^2 - R^2[d\Theta^2 + \sin^2\Theta\, d\Phi^2]. \tag{B1}$$

The comparison of (B1) with (63), suggests the following transformation:

$$R = \frac{\coth(t/\sqrt{2})}{2^{1/2}\omega\, C\cosh(\omega r + \alpha)}, \tag{B2}$$

$$T = T(r,t), \quad \Theta = \theta, \quad \Phi = \phi,$$

where $T(r,t)$ is a function to be found from the condition

$$g_{RT} = \frac{\partial t}{\partial R}\frac{\partial t}{\partial T} g_{tt} + \frac{\partial r}{\partial R}\frac{\partial r}{\partial T} g_{rr} = 0. \tag{B3}$$

From the definition of the Jacobian

$$J = \begin{vmatrix} R' & \dot{R} \\ T' & \dot{T} \end{vmatrix}. \tag{B4}$$

The following relationships may be found:

$$\frac{\partial r}{\partial R} = \frac{\dot{T}}{J}, \quad \frac{\partial r}{\partial T} = -\frac{\dot{R}}{J},$$

$$\frac{\partial t}{\partial R} = -\frac{T'}{J}, \quad \frac{\partial t}{\partial T} = \frac{R'}{J}. \tag{B5}$$

Using (B5) and (63) in (B3) we get

$$\frac{\tanh(\omega r + \alpha)}{\omega} T' - \sqrt{2}\frac{\sinh(t/2^{1/2})}{\cosh^3(t/2^{1/2})}\dot{T} = 0, \tag{B6}$$

whose solution may be written as

$$T = F(y) \tag{B7}$$

with $y = \ln|\sinh(\omega r + \alpha)| + \ln|\sinh(t/\sqrt{2})| + \sinh^2(t/\sqrt{2})/2$ and $F(y)$ is an arbitrary function of the argument.

In order to fix $F(y)$, let us write the Jacobian in the form

$$J = \frac{\partial F}{\partial y}\left[R'\frac{\partial y}{\partial t} - \dot{R}\frac{\partial y}{\partial r}\right]$$

or, using (B2) and (B7)

$$J = \frac{\partial F/\partial y}{2C\sinh^2(t/\sqrt{2})\sinh(\omega r + \alpha)} \times [1 - \cosh^4(t/\sqrt{2})\tanh^2(\omega r + \alpha)]. \tag{B8}$$

In order to avoid divergences in (B8) in the limit $t \to \infty$, we take

$$F = (1/\sqrt{2}a)\ln y, \quad a > 0. \tag{B9}$$

This choice gives us in the limit $t \to \infty$

$$\lim_{t\to\infty} J = -\frac{1}{2^{1/2}Ca}\frac{\sinh(\omega r + \alpha)}{\cosh^2(\omega r + \alpha)}, \tag{B10}$$

which is the value of the Jacobian for the transformation (72).

Thus we are led to the transformation

$$R = \frac{1}{2^{1/2}\omega}\frac{\coth(t/\sqrt{2})}{C\cosh(\omega r + \alpha)},$$

$$T = (1\sqrt{2}a)\ln y, \quad \Theta = \theta, \quad \Phi = \phi, \tag{B11}$$

with

$$y \equiv \ln|\sinh(\omega r + \alpha)| + \ln|\sinh(t/\sqrt{2})| + \tfrac{1}{2}\sinh^2(t/\sqrt{2}).$$

For the contracting solution, the transformation formulas may be deduced in a very similar way. In fact, demanding the line element (64) to have, in the new coordinates (T,R,Θ,Φ), the form given by (B1), we write

$$R = \tanh(t/\sqrt{2})/\sqrt{2}\omega C\cosh(\omega r + \alpha),$$

$$T = T(t,r), \quad \Theta = \theta, \quad \Phi = \phi. \tag{B12}$$

From the condition $g_{RT} = 0$, we arrive at the equation

$$\frac{\tanh(\omega r + \alpha)}{\omega} T' + \sqrt{2}\frac{\cosh(t/2^{1/2})}{\sinh^3(t/2^{1/2})}\dot{T} = 0, \tag{B13}$$

whose solution may be written as

$$T = F(u),$$

with

$$u \equiv \cosh^2(t/\sqrt{2})/2 - \ln|\cosh(t/\sqrt{2})| - \ln|\sinh(\omega r + \alpha)|.$$

From the asymptotic condition on the Jacobian

$$\lim_{t\to\infty} J = \frac{-1}{2^{1/2}aC}\frac{\sinh(\omega r + \alpha)}{\cosh^2(\omega r + \alpha)}$$

we get

$$F = (1/\sqrt{2}a)\ln u. \tag{B14}$$

Thus we have the coordinate transformation

$$R = \tanh(t/\sqrt{2})/\sqrt{2}\omega C \cosh(\omega r + \alpha),$$

$$T = (1/\sqrt{2}a)\ln u, \quad \Theta = \theta, \quad \Phi = \phi, \tag{B15}$$

with

$$u = \cosh^2(t/\sqrt{2})/2 - \ln|\cosh(t/\sqrt{2})| - \ln|\sinh(\omega r + \alpha)|.$$

[1]M. E. Cahill and G. C. Mc Vittie, J. Math. Phys. **11**, 1392 (1970).
[2]G. C. Mc Vittie, Mon. Not. R. Astron. Soc. **93**, 325 (1933).
[3]G. C. Mc Vittie, Astrophys. J. **143**, 682 (1966).
[4]G. C. Mc Vittie, Ann. Inst. Henri Poincaré **6**, 1 (1967).
[5]H. Thompson and G. J. Whitrow, Mon. Not. R. Astron. Soc. **136**, 207 (1967).
[6]W. B. Bonnor and M. C. Faulkes, Mon. Not. R. Astron. Soc. **137**, 239 (1967).
[7]H. Bondi, Mon. Not. R. Astron. Soc. **142**, 333 (1969).
[8]H. Nariai, Prog. Theor. Phys. **40**, 3, 679 (1968).
[9]A. Banergee, J. Phys. A: Gen. Phys. **5**, 1305 (1972).
[10]E. N. Glass, J. Math. Phys. **20**, 1508 (1979).
[11]H. D. Wahlquist and F. B. Estabrook, Phys. Rev. **156**, 1359 (1967).
[12]M. E. Cahill and A. H. Taub, Commun. Math. Phys. **21**, 1 (1971).
[13]D. M. Eardley, Commun. Math. Phys. **37**, 287 (1974).
[14]G. V. Bickhell and R. N. Henriksen, Astrophys. J. **219**, 1043 (1978).
[15]R. N.Henriksen and P. S. Wesson, Astrophys. Space Sci. **53**, 429 (1978).
[16]L. Herrera, J. Jiménez, L. Leal, J. Ponce de Léon, M. Esculpi, and V. Galina, J. Math. Phys. **25**, 3274 (1984).
[17]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley, Reading, MA, 1951), p. 311.
[18]H. Bondi, Proc. R. Soc. London Ser. A **281**, 39 (1964).
[19]L. D. Landau and E.M. Lifshitz, *Mechanics* (Pergamon, Oxford, 1960).
[20]I. S. Gradshteyn and I.M.Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).

Space-times with intrinsic symmetries on the three-spaces $t=$ constant

E. Martinez[a] and J. L. Sanz
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Santander, Santander, Spain

(Received 23 March 1984; accepted for publication 7 September 1984)

We consider metrics which possess *a priori* certain "intrinsic symmetries" on the three-spaces $t=$ const. In the vacuum case, we obtain a generalization of Birkhoff's theorem and a set of solutions with a translational isometry operating on the whole space-time. In the nonvacuum case, we assume a perfect fluid matter content and a fluid flow orthogonal to the three-spaces $t=$ const, and obtain several exact analytical solutions, some of them satisfying the standard energy conditions. In particular, a stiff equation of state is obtained in some cases, and we also have found particular solutions where a plane, spherical, or hyperbolic intrinsic symmetry is manifest.

I. INTRODUCTION

In the framework of classical general relativity, the standard representations of a star or the Universe are done on the basis of idealized symmetries operating on the whole space-time (isometries). Recently, there has been some criticism[1–4] concerning what really can be inferred for the geometry of space-time from astronomical observations. Collins[5] has suggested as an alternative approach to space-time symmetries the idea of "intrinsic symmetries," i.e., symmetries operating on submanifolds. This method has been explored by some authors in order to generate inhomogeneous cosmological models[3,6–8] (we mention the invariant characterization[7] of Szekeres models[9] and an interesting generalization[3–4] of Friedmann–Robertson–Walker models) in which the spatial curvature index k is no longer constant, found by Stephani.[10]

In this paper, we have assumed this idea of intrinsic symmetry on the geometry. Specifically, we have considered certain symmetries, which we shall explicitly give in Sec. II, operating on three-spaces $t=$ const. These include the physically interesting cases: spherical, plane, and hyperbolic intrinsic symmetry, which we shall denote as SIS, PIS, and HIS, respectively.

We solve Einstein's field equations in the vacuum case (Sec. III) and for a perfect fluid matter content (Sec. V). In the last case, we introduce the new concept of intrinsic symmetry as referred to the energy-momentum tensor, and we consider perfect fluid matter content that leads to tractable field equations. Finally, the main conclusions of the paper are raised in Sec. VI.

II. INTRINSIC ISOMETRIES

A space-time possesses an intrinsic isometry if there exist certain submanifolds such that the induced metric has an isometry in the usual sense. We are interested in the study of intrinsic symmetries operating on the three-spaces $t=$ const in such a way that the induced metric, on every three-space, has the form

$$dl^2=\overline{B}(r)dr^2+\overline{C}(r)[d\theta^2+\overline{M}^2(\theta)d\varphi^2]\,, \tag{1}$$

where $\overline{B}$, $\overline{C}$, and $\overline{M}$ are arbitrary functions of their arguments. Every section $t=$ const admits the obvious isometry generated by the Killing vector $\xi=\partial/\partial\varphi$.

[a] Present address: Nordita, Kobenhavn, Denmark.

We shall deal in this paper with space-times possessing the mentioned intrinsic symmetry, Eq. (1), but of the particular form

$$ds^2=-A(t,r,\theta,\varphi)dt^2+B(t,r)dr^2+C(t,r)[d\theta^2+M^2(\theta)d\varphi^2]\,. \tag{2}$$

For simplicity, we have assumed that the t lines are orthogonal to the three-spaces $t=$ const and that the function M does not depend on the coordinate t. In addition, A, B, C, and M are arbitrary functions of their arguments.

There are three physically interesting cases covered by the metric (2): if M takes the value $\sin\theta$, θ, or $\sinh\theta$, then the space-time possesses spherical (SIS), plane (PIS), or hyperbolic (HIS) intrinsic symmetry, respectively.

III. VACUUM SOLUTIONS

For the metric given by Eq. (2), Einstein's equations in the vacuum case ($R_{ab}=0$) lead to

$$e^{-2\alpha}\dot\gamma(2\dot\beta+\dot\gamma)-e^{-2\beta}(2\gamma''+3\gamma'^2-2\beta'\gamma')-e^{-2\gamma}M^{-1}M_{\theta\theta}=0\,, \tag{3a}$$

$$e^{-2\alpha}(\ddot\beta+\dot\beta^2-\dot\alpha\dot\beta+2\dot\beta\dot\gamma)-e^{-2\beta}[\alpha''+\alpha'^2-\alpha'\beta'+2(\gamma''+\gamma'^2-\beta'\gamma')]=0\,, \tag{3b}$$

$$e^{-2\alpha}(\ddot\gamma+2\dot\gamma^2-\dot\alpha\dot\gamma+\dot\beta\dot\gamma)-e^{-2\beta}\times(\gamma''+2\gamma'^2-\beta'\gamma'+\alpha\gamma')-e^{-2\gamma}\times(\alpha_{\theta\theta}+\alpha_\theta^2+M^{-1}M_{\theta\theta})=0\,, \tag{3c}$$

$$\dot\gamma'+\dot\gamma\gamma'-\dot\gamma\alpha'-\dot\beta\gamma'=0\,, \tag{3d}$$

$$(\dot\beta+\dot\gamma)\alpha_\theta=0\,, \tag{4a}$$

$$(\dot\beta+\dot\gamma)\alpha_\varphi=0\,, \tag{4b}$$

$$\alpha_\theta'+(\alpha'-\gamma')\alpha_\theta=0\,, \tag{4c}$$

$$\alpha_\varphi'+(\alpha'-\gamma')\alpha_\varphi=0\,, \tag{4d}$$

$$\alpha_{\theta\varphi}+(\alpha_\theta-M^{-1}M_\theta)\alpha_\varphi=0\,, \tag{4e}$$

$$\alpha_{\theta\theta}+(\alpha_\theta-M^{-1}M_\theta)\alpha_\theta-M^{-2}(\alpha_{\varphi\varphi}+\alpha_\varphi^2)=0\,, \tag{4f}$$

where $\alpha(t,r,\theta,\varphi)$, $\beta(t,r)$, $\gamma(t,r)$ are defined by

$$A=e^{2\alpha},\quad B=e^{2\beta},\quad C=e^{2\gamma}, \tag{5}$$

and

$$\dot f\equiv\frac{\partial f}{\partial t},\quad f'\equiv\frac{\partial f}{\partial r},\quad f_\theta\equiv\frac{\partial f}{\partial\theta},\quad f_\varphi\equiv\frac{\partial f}{\partial\varphi}$$

for any function $f(t,r,\theta,\varphi)$.

 0022-2488/85/040785-07$02.50

Two cases can be considered.

(A) Case $\alpha_\theta = \alpha_\varphi = 0$: From Eq. (3a) we obtain

$$M^{-1}M_{\theta\theta} = \text{const}. \tag{6}$$

Then, it is obvious that M can only take three different values

$$M_{+1} = \sin\theta, \quad M_0 = \theta, \quad \text{or } M_{-1} = \sinh\theta, \tag{7}$$

after rescaling the coordinates (θ, φ). Now, the general solution for the metric given by Eq. (2), with $A(t,r)$, is the well-known generalized Schwarzschild solution[11]

$$ds^2 = (k - m/r)dt^2 + (k - m/r)^{-1} dr^2 + r^2[d\theta^2 + M_k^2(\theta) d\varphi^2], \tag{8}$$

where m is a constant and $k = +1, 0, -1$. Therefore, in this case we do not obtain a true intrinsic symmetry, but an isometry (spherical, plane, or hyperbolic symmetry) of the whole space-time.

(B) Case $\alpha_\theta \neq 0$ or $\alpha_\varphi \neq 0$: From Eqs. (4a) or (4b) we obtain

$$\dot\beta + \dot\gamma = 0, \tag{9}$$

whereas differentiating Eq. (3a) with respect to the variable φ and taking into account Eq. (9), one arrives at

$$\dot\gamma^2 \alpha_\varphi = 0. \tag{10}$$

Three subcases can be studied: (Bi) $\dot\gamma \neq 0$, (Bii) $\gamma = \text{const}$, and (Biii) $\dot\gamma = 0$, $\gamma' \neq 0$.

(Bi) $\dot\gamma \neq 0$: In this case, Eq. (10) leads to $\alpha_\varphi = 0$; then $\alpha_\theta \neq 0$. Differentiating Eq. (3d) with respect to the variable θ, one trivially obtains $\alpha'_\theta = 0$, and this result, substituted into Eq. (4c) gives

$$\alpha' - \gamma' = 0. \tag{11}$$

By differentiating Eq. (3a) with respect to the variable θ and taking into account Eq. (9),

$$e^{-2\alpha}\alpha_\theta = e^{-2\gamma}(2\dot\gamma^2)^{-1}(M^{-1}M_{\theta\theta})_\theta. \tag{12}$$

This equation can be differentiated with respect the variable r, which implies the condition

$$\dot\gamma' = 0. \tag{13}$$

Then, Eqs. (3d), (9), (11), and (13) lead to

$$\alpha' = \gamma' = 0. \tag{14}$$

Summing up, the set of equations (3) and (4) can be reduced to the equivalent set

$$e^{-2\alpha}\dot\gamma^2 + e^{-2\gamma}M^{-1}M_{\theta\theta} = 0, \tag{15}$$

$$\ddot\gamma + \dot\gamma^2 - \dot\alpha\dot\gamma = 0, \tag{16}$$

$$\alpha_{\theta\theta} + \alpha_\theta^2 + M^{-1}M_{\theta\theta} = 0, \tag{17}$$

$$\alpha_{\theta\theta} + \alpha_\theta^2 - M^{-1}M_\theta\alpha_\theta = 0, \tag{18}$$

for the functions $\alpha(t,\theta)$ and $\gamma(t)$, and $\beta(t,r)$ is given by

$$\beta = -\gamma + A(r), \tag{19}$$

where $A(r)$ is an arbitrary function of its argument.

By taking the derivative with respect to θ into Eq. (16) one obtains $\dot\alpha_\theta = 0$, which can be integrated, giving $\alpha = \alpha(\theta)$ after a trivial rescaling of the coordinate t. Then, Eq. (16) can be integrated into the form

$$e^\gamma = at + b, \tag{20}$$

where $a \neq 0$ and b are arbitrary constants. Equations (15) and (20) lead to

$$e^{2\alpha} = -a^2M(M_{\theta\theta})^{-1}. \tag{21}$$

On the other hand, integrating Eqs. (17) and (18) gives

$$e^\alpha = c(M_\theta)^{-1}, \tag{22}$$

$$M_{\theta\theta} = -dMM_\theta^2, \tag{23}$$

where c and d are arbitrary constants. By comparing Eqs. (21) and (22), and taking into account Eq. (23), we obtain the relation $a^2 = dc^2$. Therefore we can assume $d > 0$ in Eq. (23), keep Eq. (22), and drop Eq. (21). In fact, by rescaling the coordinate φ we can always assume that $d \equiv 2$ in Eq. (23), whose general solution is given by

$$l\theta = \int dM\, e^{M^2} \tag{24}$$

(after rescaling the coordinate θ), where l is an arbitrary constant.

Therefore, the metric given by Eq. (2) is

$$ds^2 = -e^{2M^2}dt^2 + t^{-2}dr^2 + A^2t^2(d\theta^2 + M^2d\varphi^2),$$
$$\theta = \int dM\, e^{M^2}, \tag{25}$$

where only a constant A survives after a rescaling of all coordinates.

On the other hand, a calculation of the components of the Riemann tensor in the orthonormal tetrad

$$\{\omega^0 = e^\alpha dt, \quad \omega^r = e^\beta dr, \quad \omega^\theta = e^\gamma d\theta, \quad \omega^\varphi = e^\gamma M\, d\varphi\}$$

leads, for example, to

$$R^0{}_{r0r} = 2t^{-2}e^{-2M^2}, \tag{26}$$

for the previous metric, Eq. (25), i.e., we have obtained a nonflat space-time.

Finally, as $\alpha = \alpha(\theta)$, the metric (25) possesses the isometry $\xi = \partial/\partial\varphi$, operating on the whole space-time and thus the translations along the coordinate φ_0 do not constitute an intrinsic symmetry. It is also obvious that this metric does not possess SIS, PIS, or HIS.

(Bii) $\gamma = \text{const}$: From Eq. (3a) we obtain $M_{\theta\theta} = 0$, and Eq. (9) implies $\dot\beta = 0$, therefore

$$M = \theta, \quad \beta = \gamma = 0, \tag{27}$$

after rescaling the coordinate (r, θ, φ).

By substituting Eq. (27) into Eqs. (3b)–(3d) and (4), we obtain

$$\alpha'' + \alpha'^2 = 0, \quad \alpha'_\theta + \alpha'\alpha_\theta = 0, \quad \alpha'_\varphi + \alpha'\alpha_\varphi = 0, \tag{28}$$

$$\alpha_{\theta\theta} + \alpha_\theta^2 = 0, \quad \alpha_{\theta\varphi} + (\alpha_\theta - \theta^{-1})\alpha_\varphi = 0,$$
$$\alpha_{\varphi\varphi} + \alpha_\varphi^2 + \theta\alpha_\theta = 0, \tag{29}$$

whose integration gives

$$e^\alpha = a_1(t) + a_2(t)r + a_3(t)\theta\cos[\varphi + a_4(t)], \tag{30}$$

where $a_1(t), \ldots, a_4(t)$ are arbitrary functions of the variable t. Now, it is very easy to prove that the Reimann tensor, corresponding to the metric given by Eqs. (27) and (30), vanishes. Therefore, there is nothing in the assumption $\gamma = \text{const}$ but Minkowski space-time.

(Biii) $\dot\gamma = 0$, $\gamma' \neq 0$: By defining a new radial variable, we can take $\gamma = \ln r$, and the set of basic equations (3) and (4) can be rewritten as

$$e^{-2\beta}(1-2r\beta')+M^{-1}M_{\theta\theta}=0\,, \tag{31}$$

$$\alpha''+\alpha'^2-\alpha'\beta'-2r^{-1}\beta'=0\,, \tag{32}$$

$$e^{-2\beta}(1+r\alpha'-r\beta')+\alpha_{\theta\theta}+\alpha_\theta^2+M^{-1}M_{\theta\theta}=0\,, \tag{33}$$

$$\alpha'_\theta+(\alpha'-r^{-1})\alpha_\theta=0,\quad \alpha'_\varphi+(\alpha'-r^{-1})\alpha_\varphi=0\,, \tag{34}$$

$$\begin{aligned}&\alpha_{\theta\varphi}+(\alpha_\theta-M^{-1}M_\theta)\alpha_\varphi=0\,,\\ &\alpha_{\theta\theta}+\alpha_\theta^2-M^{-1}M_\theta\alpha_\theta=M^{-2}(\alpha_{\varphi\varphi}+\alpha_\varphi^2)\,,\end{aligned} \tag{35}$$

where $\beta(r)$ according to Eq. (9).

From Eq. (31) one obtains $M^{-1}M_{\theta\theta}=$ const, and, as in case (A), M can only take essentially the three values given by Eq. (7). On the other hand, the integrability condition $(\alpha'')_\theta=(\alpha'_\theta)'$ applied to Eqs. (32) and (34) leads to $\beta'\alpha_\theta=0$. The condition $\alpha_\theta=0$ implies, through Eq. (35), $\alpha_\varphi=0$, which is not possible in case (B). Therefore $\beta'=0$, but Eq. (31) gives $e^{-2\beta}=k$, and this implies $\beta=0$ and $k=+1$ ($M=\sin\theta$). The remaining equations of the set (31)–(35) can be rewritten in terms of $x=e^\alpha$ as follows:

$$x''=0,\quad x'_\theta-r^{-1}x_\theta=0,\quad x'_\varphi-r^{-1}x_\varphi=0\,, \tag{36}$$

$$\begin{aligned}&x_{\theta\theta}+rx'=0,\quad x_{\theta\varphi}=(\cot\theta)x_\varphi\,,\\ &x_{\theta\theta}-(\cot\theta)x_\theta=\sin^{-2}\theta k_{\varphi\varphi}\,,\end{aligned} \tag{37}$$

whose general solution is

$$\begin{aligned}x=e^\alpha=a_1(t)+r\{a_2(t)\cos\theta+a_3(t)\\ \times\sin\theta\cos[\varphi+a_4(t)]\}\,,\end{aligned} \tag{38}$$

where $a_1(t),\ldots,a_4(t)$ are arbitrary functions of its argument. A direct calculation proves that the Riemann tensor vanishes for the metric obtained in this case [$\beta=0$, $\gamma=\ln r$, $M=\sin\theta$, and α given by Eq. (38)]. So there is nothing new in the assumption $\dot\gamma=0$, $\gamma'\neq 0$.

Summing up, the main result of this section is that the only vacuum solutions of Einstein's field equations, different from flat space-time, with the structure given by Eq. (2) are the generalized Schwarzschild solution [Eq. (8)] and the metric, expressed by Eq. (25). Then, there is no possibility of having a SIS, PIS, or HIS in vacuum if the t lines are orthogonal to the three-spaces $t=$ const, because the metric must be the generalized Schwarzschild solution. This obviously represents a generalization of Birkhoff's theorem.[12]

IV. INTRINSIC SYMMETRIES IN THE ENERGY-MOMENTUM TENSOR

The energy-momentum tensor T possesses an intrinsic symmetry if there exist certain submanifolds such that the induced tensor has a symmetry in the usual sense. We are interested in the study of intrinsic symmetries operating on the three-spaces $t=$ const, as referred to an energy-momentum tensor corresponding to a perfect fluid, i.e., $T=(\rho+p)u\otimes u+pq$ (ρ is the energy density, p is the pressure, and $u=u_t\,dt+u_r\,dr+u_\theta\,d\theta+u_\varphi\,d\varphi$ the one-form representing the velocity of matter).

Let us assume, for the sake of simplicity, that the fluid flow is orthogonal to the three-spaces $t=$ const [i.e., $u_r=u_\theta=u_\varphi=0$, and $u_t^2=-g_{tt}=A(t,r,\theta,\varphi)$, because $u^2=-1$]. Then the induced tensor is

$$T(t=\text{const})=pg(t=\text{const})\,. \tag{39}$$

It is now clear that any intrinsic symmetry on g is translated to T, provided p ($t=$ const) also possesses the same symmetry.

Finally, we shall assume that the energy density is always positive ($\rho>0$). Other additional conditions usually assumed as regards physical solutions are $p\geqslant-\rho$ (weak energy condition), $\rho\geqslant p>-\rho$ (dominant energy condition), and $p\geqslant-\frac{1}{3}\rho$ (strong energy condition).

V. NONVACUUM SOLUTIONS

For the metric given by Eq. (2) and the energy momentum assumed in the last section (i.e., perfect fluid with flow orthogonal to the three-spaces $t=$ const), Einstein's field equations lead to

$$\begin{aligned}&e^{-2\alpha}\dot\gamma(2\dot\beta+\dot\gamma)-e^{-2\beta}(2\gamma''+3\gamma'^2-2\beta'\gamma')\\ &\quad-e^{-2\gamma}M^{-1}M_{\theta\theta}=\rho\,,\end{aligned} \tag{40a}$$

$$\begin{aligned}&e^{-2\alpha}(\ddot\beta+\dot\beta^2-\dot\alpha\dot\beta+2\dot\beta\dot\gamma)-e^{-2\beta}[\alpha''+\alpha'^2-\alpha'\beta'\\ &\quad+2(\gamma''+\gamma'^2-\beta'\gamma')]=\tfrac{1}{2}(\rho-p)\,,\end{aligned} \tag{40b}$$

$$\begin{aligned}&e^{-2\alpha}(\ddot\gamma+2\dot\gamma^2-\dot\alpha\dot\gamma+\dot\beta\dot\gamma)-e^{-2\beta}(\gamma''+2\gamma'^2\\ &\quad-\beta'\gamma'+\alpha'\gamma')-e^{-2\gamma}(\alpha_{\theta\theta}+\alpha_\theta^2\\ &\quad+M^{-1}M_{\theta\theta})=\tfrac{1}{2}(\rho-p)\,,\end{aligned} \tag{40c}$$

$$\dot\gamma'+\dot\gamma\gamma'-\dot\gamma\alpha'-\dot\beta\gamma'=0\,, \tag{40d}$$

and also to the set of equations (4).

We are not interested in the case $\alpha_\theta=\alpha_\varphi=0$, which has been extensively considered in the literature[11] for $M\equiv\sin\theta$ (spherical symmetry) and $M\equiv\theta$ (plane symmetry), with the isometry operating on the whole space-time $\xi\equiv\partial/\partial\varphi$. Therefore, we shall assume $\alpha_\theta\neq 0$ or $\alpha_\varphi\neq 0$, which implies, taking into account Eqs. (4a) and (4b),

$$\dot\beta+\dot\gamma=0\,. \tag{41}$$

We shall consider three cases: (A) $\gamma=$ const, (B) $\dot\gamma=0$, $\gamma'\neq 0$, and (C) $\dot\gamma\neq 0$,

(A) Case $\gamma=$const: Equation (41) implies $\beta(r)$, and an obvious redefinition of the coordinates (r,θ,φ) leads to

$$\beta=\gamma=0\,. \tag{42}$$

Then, the set of equations (40) is reduced, after some elementary algebra, to

$$\rho=-M_{\theta\theta}/M\,, \tag{43}$$

$$p=2(\alpha_{\theta\theta}+\alpha_\theta^2)+M_{\theta\theta}/M\,, \tag{44}$$

$$\alpha_{\theta\theta}+\alpha_\theta^2+M_{\theta\theta}/M=\alpha''+\alpha'^2\,. \tag{45}$$

The Eqs. (43) and (44) can be used as definitions of ρ and p. Therefore, the problem has been simplified to finding the general solution for α satisfying Eqs. (4) and (45). Equations (4c) and (4d) can be integrated into the form

$$e^\alpha=F(t,r)+G(t,\theta,\varphi)\,, \tag{46}$$

and Eq. (45) leads to

$$F'=G_{\theta\theta}+M^{-1}M_{\theta\theta}(G+F)\,. \tag{47}$$

If we take the derivative with respect to r in the last equation, it is obvious that

$$F'''=M^{-1}M_{\theta\theta}\,F'. \tag{48}$$

At this stage we can consider the two subcases: (Ai) $F'\neq 0$ and (Aii) $F'=0$.

(Ai) $F'\neq 0$: From Eq. (48), $(F')^{-1}F''' = M^{-1}M_{\theta\theta}$ = const, and M can be reduced to one of the three forms expressed by Eq. (7), which implies

$$F'' = -kF + \phi(t), \quad k = 0, \pm 1, \tag{49}$$

where $\phi(t)$ is an arbitrary function. By substituting Eq. (49) into Eq. (47) we obtain

$$G_{\theta\theta} = kG + \phi(t). \tag{50}$$

On the other hand, Eqs. (4e) and (4f) are rewritten in terms of F and G as

$$G_{\theta\varphi} = M^{-1}M_\theta G_\varphi, \tag{51}$$

$$G_{\theta\theta} - M^{-1}M_\theta G_\theta = M^{-2}G_{\varphi\varphi}. \tag{52}$$

The integrability condition of Eqs. (50) and (51), $(G_{\theta\theta})_\varphi = (G_{\theta\varphi})_\theta$, leads to $kG_\varphi = 0$. Now, if we consider the cases $k = \pm 1$, we must have $G_\varphi = 0$, and then Eqs. (50) and (52) imply $G_\theta = 0$. But this contradicts our initial assumption ($\alpha_\theta \neq 0$ or $\alpha_\varphi \neq 0$); therefore, we conclude that $k = 0$ (i.e., $M \equiv \theta$) and Eqs. (49)–(52) can be integrated into the form

$$F = \tfrac{1}{2}\phi(t)r^2 + \psi(t)r + \eta(t), \tag{53}$$

$$G = \tfrac{1}{2}\phi(t)\theta^2 + \theta[\lambda(t)\sin\varphi + \mu(t)\cos\varphi] + \Delta(t). \tag{54}$$

Summing up, the metric we have found can be written as (after redefining t) $s\equiv 0,1$,

$$ds^2 = -\{\psi(t)r + s(r^2+\theta^2) + \theta[a_1(t)\sin\varphi + a_2(t)\cos\varphi] + a_3(t)\}^2 dt^2 + dr^2 + d\theta^2 + \theta^2 d\varphi^2, \tag{55}$$

which possesses plane intrinsic symmetry. However, a calculation of ρ and p through Eqs. (43) and (44) leads to

$$\rho = 0, \quad p = 2e^{-\alpha}. \tag{56}$$

Therefore, the solution we have found is unphysical because it does not satisfy the energy condition $\rho > 0$. Finally, as p is θ dependent we see that the PIS on the geometry is not translated to the energy-momentum tensor.

(Aii) $F' = 0$: It is obvious that the expression (46) can be written as $e^\alpha = G(t,\theta,\varphi)$. Then, we must find the general solution of the set constituted by Eqs. (51), (52), and (47), which is rewritten as

$$G_{\theta\theta} + M^{-1}M_{\theta\theta}G = 0. \tag{57}$$

The integrability condition of Eqs. (51) and (57), $(G_{\theta\theta})_\varphi = (G_{\theta\varphi})_\theta$, leads to $M_{\theta\theta}G_\varphi = 0$. Now, $G_\varphi = 0$; since otherwise $M_{\theta\theta} = 0$ and $G_{\theta\theta} = 0$ [Eq. (57)], which imply $\rho = p \equiv 0$, i.e., we are in the vacuum case considered in Sec. IV. Then, the general solution for $G(t,\theta)$ satisfying Eqs. (52) and (57) is

$$G = \int d\theta\, M + \text{const}, \tag{58}$$

where $M(\theta)$ must satisfy

$$M = u_\theta, \quad u_{\theta\theta} = bu^{-1}, \quad b = \text{const}. \tag{59}$$

Summing up, the metric is given by [see Eqs. (46) and (58)]

$$ds^2 = -\left\{\int d\theta\, M + \text{const}\right\}^2 dt^2 + dr^2 + d\theta^2 + M^2 d\varphi^2, \tag{60}$$

and $M(\theta)$ is the general solution of Eq. (59), which can be easily integrated to give an analytical expression for $\theta = \theta(M)$. Moreover, a direct calculation of ρ and p—using Eqs. (43) and (44)—gives

$$p = \rho = bu^{-2}. \tag{61}$$

Therefore, for $b > 0$ the solution corresponds to a positive density and also satisfies all the standard energy conditions because we have a stiff equation of state. However, the metric possesses the isometry $\xi = \partial/\partial\varphi$, operating on the whole space-time, and it is also obvious that SIS, PIS, or HIS are not particular solutions of this subcase.

(B) Case $\gamma = 0$, $\dot\gamma' \neq 0$: By defining a new radial coordinate, we can take $\gamma = \ln r$, and the set of Eqs. (40) can be rewritten as

$$\rho = -r^{-2}[e^{-2\beta}(1 - 2r\beta') + M^{-1}M_{\theta\theta}], \tag{62}$$

$$p - \rho = 2e^{-2\beta}[\alpha'' + \alpha'^2 - (2r^{-1} + \alpha')\beta'], \tag{63}$$

$$r^2(\alpha'' + \alpha'^2) - r(\alpha' + \beta' + r\alpha'\beta') = 1 + e^{2\beta}(\alpha_{\theta\theta} + \alpha_\theta^2 + M^{-1}M_{\theta\theta}), \tag{64}$$

whereas Eqs. (4) adopt the form expressed by Eqs. (34) and (35). The couple of Eqs. (34) admits as a general solution

$$e^\alpha = F(t,r) + rf(t,\theta,\varphi), \tag{65}$$

and this expression, when substituted into Eqs. (35), leads to

$$f = Ms(t,\varphi) + l(t,\theta), \tag{66}$$

where $M(\theta)$ is the function which appears in the metric and s, l must satisfy

$$(M^{-1}l_\theta)_\theta = M^{-2}s_{\varphi\varphi} - s(M^{-1}M_\theta)_\theta. \tag{67}$$

We shall consider two subcases: (Bi) $\alpha_\varphi \neq 0$ and (Bii) $\alpha_\varphi = 0$.

(Bi) $\alpha_\varphi \neq 0$: By taking the derivative with respect to φ in Eq. (67), we obtain

$$(s_\varphi)^{-1}s_{\varphi\varphi\varphi} = MM_{\theta\theta} - M_\theta^2 = \text{const}, \tag{68}$$

which implies

$$M_{\theta\theta} = -kM, \quad k = 0, \pm 1 \tag{69}$$

(after a simple rescaling of θ and φ), and then M takes the values given by Eq. (7). By introducing Eq. (69) into Eq. (68) we obtain $S_{\varphi\varphi\varphi} = -S_\varphi$, which can be integrated into the form

$$s = a_1(t) + a_2(t)\sin\varphi + a_3(t)\cos\varphi. \tag{70}$$

When this last expression is substituted into Eq. (67), one obtains

$$l = -a_1(t)M + a_4(t)\int d\theta\, M + a_5(t). \tag{71}$$

Therefore, e^α given by Eq. (65) can be rewritten, taking into account Eq. (69), as

$$e^\alpha = F(t,r) + r\left[M_k \sin\varphi\, a_1(t) + M_k \times \cos\varphi\, a_2(t) + a_3(t)\int d\theta\, M_k\right]. \tag{72}$$

Now, if Eqs. (62) and (63) are considered as definitions of ρ and p, the only equation to be resolved is Eq. (64). By substituting e^α—given by Eq. (72)—in the mentioned equation, one obtains the equivalent conditions

$$k=0:\quad \beta' + r^{-1} = 0, \quad F'' - r^{-1}a_3(t)e^{2\beta} = 0\,, \tag{73a}$$

$$k=\pm 1:\quad \beta' + r^{-1}(1-ke^{2\beta}) = 0,$$
$$F'' - kr^{-1}F'e^{2\beta} = 0\,, \tag{73b}$$

whose integration gives

$$k=0:\quad e^{-2\beta} = br^2, \quad b = \text{const};$$
$$F = (2br)^{-1}a_3(t) + ra_4(t) + a_5(t)\,; \tag{74a}$$
$$k=\pm 1:\quad e^{-2\beta} = k + br^2, \quad b = \text{const};$$
$$F = a_4(t)(k+br^2)^{1/2} + a_5(t)\,. \tag{74b}$$

Summing up, the metric in this subcase can be written, according to Eqs. (72) and (74), as

$$\begin{aligned} ds^2 = &-\{rM_k[a_1(t)\sin\varphi + a_2(t)\cos\varphi] + rN_k a_3(t) \\ &+ (k+br^2)^{1/2}a_4(t) + c + r^{-1}d_k(t)\}^2\,dt^2 \\ &+ (k+br^2)^{-1}\,dr^2 + r^2[d\theta^2 + M_k^2(\theta)d\varphi^2]\,, \end{aligned} \tag{75}$$

where $c = 0, +1$ (by redefining t), $d_0(t)$ is an arbitrary function, $d_{\pm 1}(t) \equiv 0$, b is a constant, M_k adopts the values expressed by Eq. (7) and $N_0 = \theta^2, N_{+1} = \cos\theta, N_{-1} = \cosh\theta$.

On the other hand, a calculation of ρ and p through Eqs. (62) and (63) leads to

$$\rho = -3b\,, \tag{76}$$
$$p = b(3 - 2ce^{-\alpha})\,. \tag{77}$$

Obviously, Eqs. (74) imply $b>0$ for $k=0$ or -1, but then $\rho<0$ and the solutions we have obtained are unphysical. In the case $k=+1$, we can consider $b<0$, which implies $\rho = \text{const} > 0$ and also that the weak energy condition is satisfied.

Anyway, the solutions we have obtained possess PIS ($k=0$), SIS ($k=+1$), and HIS ($k=-1$), whereas these intrinsic symmetries are absent in the energy-momentum tensor unless we have $c\equiv 0$ (which implies $p=-\rho = \text{const}$). The solution, Eqs. (75)–(77), for $k=+1$ was found by Stephani[10] in another context and rediscovered by Krasinski.[3]

(Bii) $\alpha_\varphi = 0$: From Eqs. (65) and (66) we obtain $s = s(t)$. Thus, integrating Eq. (67) we can write

$$e^\alpha = F(t,r) + r\int d\theta\, M\,, \tag{78}$$

after a trivial redefinition of t, where $F(t,r)$ and $M(\theta)$ are arbitrary functions.

By substituting Eq. (76) into Eq. (64) we obtain the differential equation

$$\begin{aligned} &-rF'' + \left(F' + r^{-1}F + 2\int d\theta\, M\right)(1+r\beta') \\ &+ e^{2\beta}\left[M_\theta + M^{-1}M_{\theta\theta}\left(r^{-1}F + \int d\theta\, M\right)\right] = 0\,. \end{aligned} \tag{79}$$

If we take the derivative with respect to θ, one obtains an equation which implies $F = F(r)$, and taking the derivative with respect to r in the last equation the following is obtained:

$$2M[e^{-2\beta}(1+r\beta')]' + (r^{-1}F)'(M^{-1}M_{\theta\theta})_\theta = 0\,. \tag{80}$$

Let us consider two subcases: (Biia) $(r^{-1}F)' \neq 0$ and (Biib) $(r^{-1}F)' = 0$.

(Biia) $(r^{-1}F)' \neq 0$: Equation (80) is equivalent to

$$(M^{-1}M_{\theta\theta})_\theta = CM\,, \tag{81}$$
$$2e^{-2\beta}(1+r\beta') = -r^{-1}CF + D\,, \tag{82}$$

where C and D are arbitrary constants. If we introduce Eqs. (81) and (82) into the equation obtained from Eq. (79), taking the derivative with respect to θ, one concludes that $C\equiv 0$ and $D = -2M^{-1}M_{\theta\theta}$. Thus, Eqs. (81) and (82) can be rewritten as

$$M_{\theta\theta} = -kM, \quad k = 0, \pm 1; \quad \beta' + r^{-1}(1-ke^{2\beta}) = 0\,, \tag{83}$$

and then Eq. (79) gives a differential equation on $F(r)$ similar to Eq. (73). It is clear that in this subcase we obtain the structure expressed by Eqs. (75)–(77) but with $a_1 = a_2 \equiv 0$, $a_3 \equiv 1$ for $k=-1$, $a_3 \equiv -1$ for $k=+1$, $a_3 \equiv \frac{1}{2}$ for $k=0$, and d_0, a_4 are constants.

(Biib) $(r^{-1}F)' = 0$: We write $F = Br$, with $B = \text{const}$ and, by itegrating Eq. (80),

$$e^{-2\beta} = \lambda + br^2\,, \tag{84}$$

where λ and B are arbitrary constants. By substituting these expressions for F and β into Eq. (64) we obtain a differential equation for the variable $s \equiv \int d\theta\, M + B$, which can be integrated into the form

$$ss_{\theta\theta} + \lambda s^2 = D, \quad D = \text{const}. \tag{85}$$

Summing up, we have obtained the following metric, according to Eqs. (78) and (84),

$$ds^2 = -r^2s^2(\theta)dt^2 + (\lambda + br^2)^{-1}\,dr^2 + r^2(d\theta^2 + s_\theta^2\,d\varphi^2)\,, \tag{86}$$

where $s(\theta)$ is the general solution to Eq. (85), which can be given in the form

$$\pm\theta + \text{const} = \int ds(E - \lambda s^2 + 2D\ln s), \quad E = \text{const}. \tag{87}$$

A direct calculation of ρ and p through Eqs. (62) and (63) leads to

$$\rho = D(rs)^{-2} - 3b, \quad p = D(rs)^{-2} + 3b\,. \tag{88}$$

It is clear that the metric (86) is static ($\xi_1 = \partial/\partial t$) and translational invariant ($\xi_2 = \partial/\partial\varphi$). Moreover, if we take $s = \cos\theta$, $s = \theta$, or $s = \cosh\theta$ [which satisfy Eq. (85) with $D\equiv 0$] is very easy to see that the metric possesses SIS, PIS, or HIS, respectively, and the energy momentum possesses the same intrinsic symmetry because $p = -\rho = +3b$.

The standard energy conditions are satisfied for different values of the constant D and b. For instance, the electron $D \geqslant 0$ and $b \leqslant 0$ (with D, $b \not\equiv 0$ in order to have $\rho > 0$) gives physical solutions. In particular, the weak and dominant energy conditions are satisfied. For $D>0$ and $b\equiv 0$, one obtains a stiff equation of state.

(C) $\dot\gamma \neq 0$: Equation (41) implies

$$\beta = -\gamma\,, \tag{89}$$

after redefining r. The set of equations (40) can be rewritten as

$$\rho = -\{e^{-2\alpha}\dot\gamma^2 + e^{2\gamma}(2\gamma'' + 5\gamma'^2) + e^{-2\gamma}M^{-1}M_{\theta\theta}\}\,, \tag{90}$$

$$p - \rho = 2\{e^{-2\alpha}(\ddot\gamma + \dot\gamma^2 - \dot\alpha\dot\gamma)$$

$$+e^{2\gamma}[\alpha'' + \alpha'^2 + \alpha'\gamma' + 2(\gamma'' + 2\gamma'^2)]\}\,, \qquad (91)$$

$$2e^{-2\alpha}(\ddot{\gamma} + \dot{\gamma}^2 - \dot{\alpha}\dot{\gamma}) + e^{2\gamma}(\alpha'' + \alpha'^2 + \gamma'' + \gamma'^2)$$

$$-e^{-2\gamma}(\alpha_{\theta\theta} + \alpha_\theta^2 + M^{-1}M_{\theta\theta}) = 0\,, \qquad (92)$$

$$\dot{\gamma}' + 2\dot{\gamma}\gamma' - \dot{\gamma}\alpha' = 0\,. \qquad (93)$$

The integration of Eq. (93) leads to

$$e^\alpha = \dot{\gamma}e^{2\gamma}f(t,\theta,\varphi)\,, \qquad (94)$$

whereas Eqs. (4c) and (4d) are equivalent in this case to $\dot{\gamma}' + \dot{\gamma}\gamma' = 0$, whose general solution is

$$e^\gamma = a(t) + b(r)\,. \qquad (95)$$

Now, we can make a redefinition of t in such a way that $a \equiv t$. Thus, Eqs. (94) and (95) are rewritten as

$$e^\gamma = t + b(r), \quad e^\alpha = e^\gamma f(t,\theta,\varphi)\,. \qquad (96)$$

On the other hand, if we substitute Eqs. (96) into Eqs. (4e) and (4f) we obtain for f the structure (66) with $s(t,\varphi)$ and $b(t,\theta)$ satisfying Eq. (67). Then, two subcases can be studied: (Ci) $\alpha_\varphi \neq 0$ and (Cii) $\alpha_\varphi = 0$.

(Ci) $\alpha_\varphi \neq 0$: The analysis made in subcase (Bi) can be repeated—see Eqs. (70) and (71)—and we conclude that

$$f = M_k \sin\varphi\, a_1(t) + M_k \cos\varphi a_2(t) + a_3(t) \int d\theta\, M_k\,, \qquad (97)$$

where $M_k(\theta)$ is given by Eq. (7).

By substituting e^α and e^γ, given by expressions (96) and (97), into Eq. (92) we obtain the equivalent set

$$k = \pm 1: \quad e^\gamma(e^{3\gamma}b'' + k)f^3 - \ddot{f} - \dot{\gamma}\dot{f} = 0\,, \qquad (98)$$

$$k = 0: \quad e^{4\gamma}b''f^3 - \tfrac{1}{2}e^\gamma a_3(t)f^2 - \ddot{f} - \dot{\gamma}\dot{f} = 0\,. \qquad (99)$$

From Eq. (98) it is very easy to conclude that $e^{3\gamma}b'' + k = 0$ and this equation implies $\dot{\gamma} = 0$, which is impossible in case (C). Therefore, the possibilities $k = \pm 1$ do not exist. Equation (99) constitutes a polynomial in θ of degree 6; the vanishing of its coefficients leads to

$$a_3 \equiv 0, \quad b = \text{const}; \quad a_1 = d_1 t^{-1}, \quad a_2 = d_2 t^{-1},$$

$$d_1, d_2 = \text{const}\,. \qquad (100)$$

Summing up, we have found the following metric:

$$ds^2 = -(d\theta \sin\varphi)^2\, dt^2 + t^{-2}\, dr^2$$

$$+ t^2(d\theta^2 + \theta^2\, d\varphi^2)\,, \qquad (101)$$

$$d = \text{const}\,,$$

after a rescaling of t and φ. This metric possesses translational intrinsic symmetry (in particular PIS). On the other hand, Eqs. (90) and (91) allow us to calculate ρ and p

$$\rho = -(t\theta \sin\varphi)^{-2}, \quad p = \rho\,, \qquad (102)$$

i.e., the solution is unphysical because $\rho < 0$ for $t < \infty$.

(Cii) $\alpha_\varphi = 0$: The general solution of Eq. (67), taking into account the structure (96), is given by

$$e^\alpha = e^\gamma\left[f(t) + g(t)\int d\theta\, M\right], \quad e^\gamma = t + b(r)\,, \qquad (103)$$

where $f(t)$, $g(t)$, and $b(r)$ are arbitrary functions. If we substitute Eq. (103) into Eq. (92), the following is obtained:

$$-2\left\{e^\gamma\left[f + g\int d\theta\, M\right]\right\}^{\cdot\cdot} + 2b''e^{5\gamma}\left[f + g\int d\theta\, M\right]^3$$

$$-e^{-2\gamma}g\left[f + g\int d\theta\, M\right]^2$$

$$\times M^{-1}\left[M_\theta\left(\int d\theta\, M + \frac{f}{g}\right)\right]_\theta\,. \qquad (104)$$

Multiplying Eq. (14) by $e^{-2\gamma}$, and taking the derivative with respect to r and θ we obtain

$$-(e^{-2\gamma})' - (e^{-\gamma})'g^{-1}\dot{g}$$

$$+ 3(b''e^{3\gamma})'\left[f + g\int d\theta\, M\right]^2 = 0\,, \qquad (105)$$

and if we take the derivative with respect to θ in this last equation we conclude that $b'' = 0$, which implies—through Eqs. (103) and (105)—that $b' = 0$. So, we can take $b \equiv 0$ after rescaling the coordinate t. The substitution of this value ($e^\gamma = t$) into Eq. (104) leads to

$$-2(gt^2)^{-1}\left[f + g\int d\theta\, M\right] - 2(gt)^{-1}\left[\dot{f} + \dot{g}\int d\theta\, M\right]$$

$$= \left[f + g\int d\theta\, M\right]^2 M^{-1}\left[M_\theta\left(\int d\theta\, M + \frac{f}{g}\right)\right]_\theta\,. \qquad (106)$$

Two subcases can be considered: (Ciia) $(fg^{-1})^{\cdot} \neq 0$ and (Ciib) $(fg^{-1})^{\cdot} = 0$.

(Ciia) $(fg^{-1})^{\cdot} \neq 0$: Let us take the derivative with respect to θ in Eq. (106). Then one obtains ($y \equiv g^{-1}f + \int d\theta\, M$)

$$-2(g^3t^2)^{-1}(g + t\dot{g})$$

$$= 2Y\,[M_\theta + (M_{\theta\theta}/M)\,Y]$$

$$+ Y^2M^{-1}[M_\theta + (M_{\theta\theta}/M)\,Y]_\theta\,, \qquad (107)$$

whose derivative with respect to t gives

$$-\{(g^{-1}f)^{\cdot}\}^{-1}\{(gt)^{-2}(1 + tg^{-1}\dot{g})\}$$

$$= M_\theta + 2(M_{\theta\theta}/M)\,Y + Y/M$$

$$\times\left[M_\theta + \frac{M_{\theta\theta}}{M}\,Y\right]_\theta + \frac{Y^2}{2M}\left[\frac{M_{\theta\theta}}{M}\right]_\theta\,, \qquad (108)$$

and, by taking two derivatives with respect to t,

$$M_{\theta\theta}\, M^{-1} = A + B\int d\theta\, M, \quad A,B = \text{const}\,. \qquad (109)$$

Equation (109) can be substituted into Eq. (108)

$$-\{(g^{-1}f)^{\cdot}\}^{-1}\{(gt)^{-2}(1 + tg^{-1}\dot{g})\}$$

$$= M_\theta + 4y\left[A + B\int d\theta\, M\right] + \frac{3}{2}By^2\,, \qquad (110)$$

which obviously implies $B \equiv 0$. When this value and Eq. (109) are substituted into Eq. (107), it is very easy to conclude that $A \equiv 0$, and Eq. (109) implies

$$M = a\theta + d, \quad a,d = \text{const}\,. \qquad (111)$$

Thus, Eq. (107) is reduced to $g + t\dot{g} = 2aY = 0$, i.e.,

$$g = ct^{-1}, \quad c = \text{const}, \quad a = 0\,. \qquad (112)$$

Equations (106) and (112) imply $(fg^{-1})^{\cdot} = 0$, which is in contradiction to the initial assumption. Therefore this subcase is not possible.

(Ciib) $(fg^{-1})^{\cdot}=0$: Let us write $f=Ag$, $A=\text{const}$. Then Eq. (107) is reduced to

$$-2(g^3t^2)(g+tg)=2YM_\theta+4Y^2(M_{\theta\theta}/M)+Y^3M^{-1}(M_{\theta\theta}/M)_\theta=B=\text{const}\,, \tag{113}$$

which leads to $[s=s(\theta)]$; and, using Eq. (106),

$$t_g=[B\ln t+C)]^{-1/2},\quad C=\text{const}\,; \tag{114}$$

$$M=s_\theta,\quad s_{\theta\theta}=Ds^{-1}+Bs^{-1}\ln s,\quad D=\text{const}\,. \tag{115}$$

Summing up, we have found the following solution [see Eq. (103) with $b\equiv 0$, and Eqs. (114) and (115)]:

$$ds^2=-(B\ln t+C)^{-1}s^2\,dt^2+t^{-2}\,dr^2+t^2(d\theta^2+s_\theta^2\,d\varphi^2)\,, \tag{116}$$

where A, B, and C are arbitrary constants. This metric possesses the isometry (operating on the whole space-time) $\xi=\partial/\partial\varphi$, and regarding intrinsic symmetries, if $D=B\equiv 0$ (i.e., $s\equiv\theta$), there exits PIS.

On the other hand, Eqs. (90) and (91) give the following values for ρ and p:

$$\rho=(ts)^{-2}[D+B\ln s-C-B(1+\ln t)]\,, \tag{117a}$$

$$p=(ts)^{-2}[D+B\ln s-C-B\ln t\,]\,. \tag{117b}$$

In general, the standard energy conditions can be satisfied for certain intervals of the variables r and θ. For instance, $p<\rho$ will be satisfied for $B<0$. In particular, if $B=0$ and $D>C$ we obtain that $\rho>0$ and $p=\rho$, i.e., a stiff equation of state.

VI. CONCLUSIONS

The main purpose of this paper was to develop the idea of intrinsic symmetry to search for exact inhomogeneous solutions of Einstein's field equations. We have assumed from the beginning a certain structure for the metric, Eq. (2), which contains the physically interesting cases of plane, spherical, and hyperbolic intrinsic symmetry (PIS, SIS, and HIS, respectively), and resolved the field equations in the vacuum case and for a perfect fluid matter content with the flow orthogonal to the three-spaces $t=\text{const}$.

In the vacuum case we have obtained a generalization of Birkhoff's theorem: there is no possibility of having SIS, PIS, or HIS in vacuum if the t lines are orthogonal to the three-spaces $t=\text{const}$, because the metric must be the generalized Schwarzschild solution, Eq. (8). On the other hand, there exists a set of vacuum solutions, Eq. (25), which do not possess any such intrinsic symmetries, but with a translational isometry operating on the whole space-time.

As regards the nonvacuum exact solutions we have found, we mention that the only physical solution $(\rho>0)$ which does not possess isometries operating on the whole space-time is a solution found previously by Stephani[10] and Krasinski.[3] The rest of the nonvacuum solutions possess a translational isometry in the usual sense, and some of them also have PIS [see Eqs. (75), (86), and (116)], SIS, or HIS [see Eqs. (75) and (86)]. All of these latter solutions are physical, in the sense that the energy density is positive and at least one of the standard (weak, dominant, and strong) energy conditions are satisfied. In particular, a stiff equation of state is obtained in some cases [see Eqs. (81), (88), and (117)]. We also remark that in cases (A) and (B) the possibility of having a "radial" flow $(u^r\neq 0)$ has been examined and nothing different from the case $u^r=0$ has been obtained.

Finally, we want to emphasize the significance of the new technique of intrinsic symmetries in order to obtain new exact solutions. It appears that the idea is not trivial and can be successful in the search for models representing astrophysical or cosmological situations.

[1] G. F. R. Ellis, Gen. Relativ. Gravit. **11**, 925 (1979).
[2] G. F. R. Ellis, Ann. N. Y. Acad. Sci. **336**, 130 (1980).
[3] A. Krasinski, Gen. Relativ. Gravit. **13**, 1021 (1981).
[4] A. Krasinski, in *The Birth of the Universe*, edited by J. Audouze and T. Tranhvan, Proc. 17 Rencontre de Moriond (Frontières, Coif. sur Yvette, 1982), Vol. 34, p. 15.
[5] C. B. Collins, Gen. Relativ. Gravit. **10**, 925 (1979).
[6] C. B. Collins and D. A. Szafron, J. Math. Phys. **20**, 2347 (1979).
[7] C. B. Collins and D. A. Szafron, J. Math. Phys. **20**, 2354 (1979).
[8] C. B. Collins and D. A. Szafron, J. Math. Phys. **20**, 2362 (1979).
[9] P. Szekeres, Commun. Math. Phys. **41**, 55 (1975).
[10] H. Stephani, Commun. Math. Phys. **4**, 137 (1967).
[11] D. Kramer, M. Stephani, M. A. H. MacCallum, and H. Herlt, *Exact Solutions of Einstein's Field Equations*, edited by E. Schmutzer (VEB Deutscher Verlag der Wissenschaften, Berlin, 1980).
[12] G. D. Birkhoff, *Relativity and Modern Physics* (Harvard U. P., Cambridge, MA, 1923), p. 255.

Equilibrium configurations of degenerate fluid spheres

Patrick G. Whitman
Department of Physics, University of Southwestern Louisiana, Lafayette, Louisiana 70504

(Received 8 June 1984; accepted for publication 7 December 1984)

Equilibrium configurations of degenerate fluid spheres which assume a polytropic form in the ultrahigh-density regime are considered. We show that analytic solutions more general than those of Misner and Zapolsky exist which possess the asymptotic equation of state. Simple expressions are derived which indicate this nature of the fluids in the extreme relativistic limit, and the stability of these interiors is considered in the asymptotic region.

I. INTRODUCTION

Interest in the final state of matter for cold condensed stellar objects has led to a renewed effort on the part of investigators to ascertain the nature of physical systems at densities greater than that of nuclear matter.[1,2] A study of such superdense systems naturally leads to considerations of relativistic many-body calculations based on the nature of elementary particles and their interactions. Our knowledge of these is as yet incomplete, and the presently available experimental data is unable to unambiguously aid in the formulation of an appropriate equation of state in the ultrahigh-density regime. This situation has led to a number of plausible models to depict the final state, and only new experimental and observational data will be able to distinguish between the competing theories.

In this paper we consider the reverse problem. Since we are as yet unable to discriminate between models based on assumed particle populations and interactions, what of a general nature can be said of the superdense regime? We show that if one assumes only that the equation of state in the superdense region is asymptotic to the polytropic form for which the pressure becomes proportional to the density and that the pressure does not have an essential singularity, the field equations can be solved in closed form. The solutions thus obtained are those first discovered by Wyman[3] and later considered in more detail by Whitman,[4] Whitman and Redding[5] for the case of static equilibrium, and Whitman and Pizzo[6] when slow rotation is incorporated.

We examine the Sturm–Liouville equation for zeroth-order normal mode oscillations in the neighborhood of the origin for the complete family of solutions and show that all but one member are unstable. The instability, however, is reducible by replacing a small central region with a uniform density core in the same fashion as was done with the Misner–Zapolsky[7] solution. It is not necessary to introduce an outer envelope as was done by these authors. The solutions considered here have a pressure boundary, which is not the case for the Misner–Zapolsky solution. The remaining nonsingular solution is observed to satisfy the necessary part of the stability criteria over all density ranges, and to be completely stable in the low-density regime.

II. RELATIVISTIC INTERIORS IN THE HIGH-DENSITY LIMIT

For the case of a spherically symmetric fluid with density ρ and pressure P, the field equations of general relativity reduce to a set of three independent coupled ordinary differential equations in four unknowns. The underdetermined nature of this system allows for the introduction of an equation of state relating in some fashion to the thermodynamic variables throughout the interior.

The metric representing the static interior is taken to have the form

$$ds^2 = \gamma^2\, dt^2 - \tau^{-1}\, dr^2 - r^2\, d\Omega^2, \tag{2.1}$$

where $d\Omega^2$ represents the metric on the unit sphere. In this coordinate system, the condition that the pressure be isotropic forces a differential relation between the two metric functions γ and τ. The resulting equation is quasilinear and of second order in γ and first order in τ. The equation of isotropy is

$$2\tau r^2\gamma'' - r(2\tau - r\tau')\gamma' + (2 - 2\tau + r\tau')\gamma = 0, \tag{2.2}$$

where the prime refers to differentiation with respect to the radial coordinate.

The remaining equations relate the thermodynamic variables to the metric functions and their derivatives. The fluid pressure can be most simply expressed as

$$8\pi r^2 P = (1 + 2r\gamma'/\gamma)\tau - 1. \tag{2.3}$$

The fluid density is related only to the metric function τ. A single quadrature then leads to an expression for τ in terms of the fluid mass

$$\tau = 1 - 2m(r)/r, \quad m = 4\pi \int_0^r \rho r^2\, dr. \tag{2.4}$$

The remaining equation is the resultant Bianchi identity. This is given by

$$\gamma'/\gamma = -P'/(P + \rho). \tag{2.5}$$

Equation (2.5) is a consequence of the field equations and is not an independent expression. The usual approach is to select from Eqs. (2.2) through (2.5) any three and add to this the appropriate equation of state depicting the interior.

A. Limiting form for the interior

Most often considered in the ultrahigh-density regime are polytropic equations of state in which the pressure is proportional to the density. The constant of proportionality is the square of the sound speed within the fluid. It is not difficult to show that if such an equation of state is assumed to hold in the limit of infinitely large pressure and density, and if the pressure has a Laurent expansion with only a finite number of terms in negative powers of the radial coordinate,

 0022-2488/85/040792-06$02.50

then the leading term will be proportional to r^{-2}. This can be seen in the following fashion. Assume $P \approx P_0 r^n$ and introduce this, along with the condition that the density be proportional to the pressure, into Eq. (2.5) to determine the ratio γ'/γ. Placing the result into Eq. (2.3) shows that $n = -2$ if τ is both positive and nonsingular at the origin. This condition is guaranteed if the mass vanishes there at least as fast as r, as can be seen from Eq. (2.4). Introducing this form for τ into Eq. (2.2) leads to an expression for γ in this region

$$\gamma = ay^{i} + by^{j}, \tag{2.6a}$$

where $y = r/R$, the normalized radial coordinate, and the quantities i and j are, respectively, $1 + \alpha$ and $1 - \alpha$. The parameter α is related to the asymptotic form of τ through the relation $\tau = (2 - \alpha^2)^{-1}$. If the limiting form for τ is to be both positive and nonsingular, $\alpha < \sqrt{2}$. Examination of Eq. (2.6a) above indicates that γ will be nonsingular provided $|\alpha| \leqslant 1$. For α outside this range, the pressure will become negative somewhere within the interior. Note that by virtue of the symmetry inherent in Eq. (2.6a), there is no loss in generality if α is taken to be positive.

If $\alpha = 0$, Eq. (2.6a) degenerates. The second independent solution is logarithmic in y

$$\gamma = y(a + b \ln y), \tag{2.6b}$$

which is well behaved as y tends to zero.

B. Extrapolated general solution

By virtue of the above arguments, Eq. (2.6) depicts the behavior of all polytropic distributions in the vicinity of the singular region. Wyman,[3] Whitman,[4] and Whitman and Redding[5] observed the field equations could be solved exactly for this functional dependence of γ on the radial coordinate. None of these authors, however, noted the second independent solution associated with $\alpha = 0$.

Their solution for $\tau(y)$ is

$$\tau = s^{-1} - Cy^{2s/l}\,[aky^{2\alpha} + bl\,]^{-2s/lk}, \quad 0 < \alpha \leqslant 1, \tag{2.7a}$$

where $s = 2 - \alpha^2$, $l = 2 - \alpha$, and $k = 2 + \alpha$. In this expression, a, b, and C are constants of integration. Requiring continuity of these solutions leads to expressions for these constants in terms of the Schwarzschild parameter $\gamma_S^2 = 1 - 2M/R$, where M is the total integrated mass and R is the radius of the sphere. Determined in this fashion, the integration constants are

$$a = (1 - q\gamma_S^2)/4\alpha\gamma_S, \quad b = (n\gamma_S^2 - 1)/4\alpha\gamma_S,$$
$$C = (s^{-1} - \gamma_S^2)[(1 + \gamma_S^2)/2\gamma_S]^{2s/lk}, \tag{2.7b}$$

where $q = 3 - 2\alpha$ and $n = 3 + 2\alpha$. The solutions for $\alpha = 0$ can be obtained in the same manner. The function τ is given by

$$\tau = \tfrac{1}{2} - Cy^2[2a + b + 2b \ln y]^{-1}, \tag{2.7c}$$

with constants of integration

$$a = \gamma_S, \quad b = (1 - 3\gamma_S)/2\gamma_S,$$
$$C = (\tfrac{1}{2} - \gamma_S^2)(1 + \gamma_S)/2\gamma_S. \tag{2.7d}$$

The pressures and densities associated with these solutions are

$$8\pi R^2 y^2 P = \tau(any^{2\alpha} + bq)(ay^{2\alpha} + b)^{-1} - 1, \tag{2.8a}$$

$$8\pi R^2 y^2 \rho = 1 - \tau - 2(s\tau - 1)(ay^{2\alpha} + b)\times(aky^{2\alpha} + bl)^{-1}, \tag{2.8b}$$

and for the case $\alpha = 0$,

$$8\pi R^2 y^2 P = \tau(3a + 2b + 3b \ln y)(a + b \ln y)^{-1} - 1, \tag{2.8c}$$

$$8\pi R^2 y^2 \rho = 1 - \tau - 2(2\tau - 1)(a + b \ln y)\times(2a + b + 2b \ln y)^{-1}. \tag{2.8d}$$

These solutions are interesting in that they are extrapolations of the asymptotic form expected of degenerate equilibrium configurations in the region where general relativity plays its most important role in stellar structure. An examination of the solutions indicates they have many of the features expected of relativistic interiors even outside the asymptotic region. As pointed out by Whitman and Redding, the ratio of specific heats and the sound speed are well behaved throughout the interior provided the mass distribution is not near its Schwarzschild limit. Under this condition, the sound speed is causal throughout the interior, decreasing outward from the center for $\alpha \neq 1$. When α is unity, the opposite occurs and the sound speed increases outward. The pressure and density are well behaved, with the pressure vanishing at a finite boundary. If the inequality

$$\alpha^2 \geqslant (9\gamma_S^4 - 4\gamma_S^2 - 1)/4\gamma_S^4 \tag{2.9}$$

is satisfied, the surface value of the density will be positive, vanishing if the equality holds.

III. ASYMPTOTIC EQUATIONS OF STATE

The arguments of the previous section indicate the applicability of these solutions in the asymptotic region. In the ultrahigh-density limit, the form is identical to the Misner–Zapolsky solution and thus agrees with all equations of state that assume the form of a simple polytrope in this regime. In the high-density limit, the possible choices of physical equations of state become more varied, but there are a number which agree with this solution within the appropriate bounds. Two examples are the Fermi gas equation of state and the equation of state given by Leung and Wang.[8]

In this section, we develop the equations of state and in particular compare them to the degenerate Fermi gas.

The equations expressing the pressure and density, though in general somewhat complicated, can be expressed in the form $P(\rho)$ within the asymptotic region. If we ignore powers of y greater than 2α, the resulting equation of state is

$$P = \rho\left[\frac{1-\alpha}{1+\alpha} - \frac{4\alpha}{1-\alpha^2}\left(\frac{q\gamma_S^2 - 1}{n\gamma_S^2 - 1}\right)\left(\frac{\rho_0}{\rho}\right)^{\alpha}\right], \quad 0 < \alpha < 1. \tag{3.1a}$$

When α takes on the value zero, the appropriate expression is

$$P = \rho[1 - 4(3\gamma_S^2 - 1)(4 + (3\gamma_S^2 - 1)\ln(\rho/\rho_0))^{-1}]. \tag{3.1b}$$

In these expressions, ρ_0 is the limiting form of the product

ρy^2 and is given by $\rho_0 = (1-\alpha^2)/4\pi R^2(2-\alpha^2)$. For the case $\alpha = 1$, the equation of state takes on the form

$$P_0 - P = v_0^2(\rho_0 - \rho), \tag{3.1c}$$

where P_0, ρ_0, and v_0^2 are, respectively, the central values of the pressure, density, and sound speed squared.

Expressions for the asymptotic sound speed can be obtained from Eq. (3.1). Examination of these indicate the sound speed will be a decreasing function for $\gamma_S^2 \geqslant (3-2\alpha)^{-1}$. If equality holds or $\alpha = 1$, the sound speed will be uniform.

We observe that for $\alpha = \frac{1}{2}$, the equation of state (3.1a) becomes

$$\rho - 3P = \left(\frac{24}{7\pi}\right)^{1/2}\left(\frac{2\gamma_S^2 - 1}{4\gamma_S^2 - 1}\right)\left(\frac{1}{R}\right)\rho^{1/2}. \tag{3.2}$$

In the limit of high central densities this is precisely the form of the Fermi gas equation of state[9]

$$\rho - 3P = (16/\pi)^{1/2}\rho^{1/2}. \tag{3.3}$$

If we require Eq. (3.2) to *exactly* represent a Fermi gas in this limit, then

$$R = (\tfrac{3}{14})^{1/2}((2\gamma_S^2 - 1)/(4\gamma_S^2 - 1)). \tag{3.4}$$

Whitman and Redding showed that $\gamma_S^2 \leqslant 0.683$ if the surface density was non-negative for the $\alpha = \frac{1}{2}$ solution. A lower limit can as well be placed on γ_S^2 by demanding the exclusion of nonintegrable singularities. They indicated that this requires $\gamma_S^2 > 0.5$. In this range we find R lies within the interval

$$0 < R \leqslant 0.0978. \tag{3.5}$$

The upper limit on R corresponds to about 1.33 km (the unit of length is 1.36×10^6 cm). Numerical integration of the Fermi gas equation of state indicates however that $R \simeq 3.1$ km. Clearly then we cannot force agreement with the Fermi gas equation of state to this order, and at the same time retain an exact solution valid throughout the interior.

If what is desired is an analytic expression in the high-density limit, the constants in Eq. (3.1a) can be appropriately evaluated. One finds this identification requires (in terms of the integration constants a and b as well as the scale factor R)

$$a/b = -0.4969R. \tag{3.6}$$

This interior would then have to be joined numerically to the Fermi gas equation of state in the medium density regime. Nearer the origin, both equations of state take on the polytropic form $p = \frac{1}{3}\rho$, and correspondence is established in the ultrahigh-density regime.

We observe in passing that the upper limit on γ_S^2 corresponds to a vanishing surface density, which is the same boundary condition used for integration of the Fermi gas equation of state. Whitman and Redding showed the mass-to-radius ratio for this analytic solution is within 0.5% of the value obtained from numerical integration.

Though the M/R value is in good agreement with that obtained from the Fermi gas equation of state, the radius is considerably smaller. This indicates that the sphere is more collapsed, as one would expect. Recall that the asymptotic form of γ was extended to hold *throughout* the interior, not just in the high-density limit.

IV. THE MASS DEFECT

In order to ascertain further properties of the solutions discussed in this paper, an expression must be derived which depicts the distribution of particles throughout the sphere. A knowledge of this, the number density, allows one to determine the coefficient of gravitational packing, which determines the ratio between the gravitational and total energy, and the partial mass defect, which expresses how much of the rest mass energy is emitted during the formation of the star from initial dispersed material.

The proper number density is related to the pressure and density through the second law of thermodynamics with the assumption of constant entropy. A formal integration of this expression results in

$$n = K \exp\left[\int^r (P+\rho)^{-1}\, d\rho\right]. \tag{4.1}$$

The quantity K in Eq. (4.1) is a constant of integration, determined by the condition that the internal energy vanish at the surface.

In general, this integral is intractable for the full solutions, but simple expressions result for the special case $C = 0$ in Eq. (2.7). For the constant C to vanish, the boundary conditions require $\gamma_S^2 = 1/s$. This leads to the specific solution set

$$\tau = 1/s, \tag{4.2a}$$

$$8\pi R^2\rho y^2 = ij/s, \tag{4.2b}$$

$$8\pi R^2 P y^2 = \left(\frac{i^2 j^2}{s}\right)\left(\frac{1-y^{2\alpha}}{i^2 - j^2 y^{2\alpha}}\right). \tag{4.2c}$$

For the case $\alpha = 0$, Eq. (4.2c) becomes

$$16\pi R^2 P y^2 = \ln y/(\ln y - 2). \tag{4.2d}$$

The introduction of these expressions into Eq. (4.1) leads to

$$m_N n = (ij/16\pi R^2 \alpha s y^i)(i - j y^{2\alpha}), \quad \alpha \neq 0, \tag{4.3a}$$

$$m_N n = (1 - \ln y)/16\pi R^2 y, \quad \alpha = 0. \tag{4.3b}$$

The total number of particles contained within the sphere can be determined by integrating Eq. (4.3) over the proper volume

$$N = 4\pi R^3 \int_0^1 n(y) y^2\, dy/\sqrt{\tau}. \tag{4.4}$$

Introducing $n(y)$ from Eq. (4.3), we find

$$m_N N \equiv M_0 = 3(2-\alpha^2)^{1/2}(4-\alpha^2)^{-1}M, \quad 0 \leqslant \alpha \leqslant 1. \tag{4.5}$$

Having at our disposal an expression for the total number of particles within the volume, we may determine other bulk properties of the distribution. For a sphere with gravitational mass M given by

$$M = 4\pi R^3 \int_0^1 \rho y^2\, dy \tag{4.6}$$

and proper mass

$$M_1 = 4\pi R^3 \int_0^1 \frac{\rho y^2\, dy}{\sqrt{\tau}}, \tag{4.7}$$

we can obtain expressions which in the Newtonian limit correspond to the energy of motion W and the energy of self-gravitation U (see Ref. 10). These can be obtained from the total gravitational mass defect (gravitational binding energy)

$$\Delta_1 M \equiv M_1 - M, \quad (4.8)$$

and the partial mass defect given by the difference

$$\Delta_2 M \equiv M_0 - M. \quad (4.9)$$

In the classical limit, $\Delta_1 M = -U$ and $\Delta_2 M = -(W+U)$.

Two new quantities may be defined from Eqs. (4.8) and (4.9). Let

$$\alpha_1 \equiv \Delta_1 M/M, \quad (4.10a)$$

$$\alpha_2 \equiv \Delta_2 M/M. \quad (4.10b)$$

Equation (4.10a) is the coefficient of gravitational packing, and Eq. (4.10b) represents the fraction of rest mass energy emitted during the formation for the equilibrium distribution.

The quantities α_1 and α_2 can be calculated for these solutions under the assumption of an isentropic equation of state. The coefficient of gravitational packing is

$$\alpha_1 = \sqrt{s} - 1, \quad (4.11a)$$

and the fraction of rest mass-energy emitted during formation is

$$\alpha_2 = 3\sqrt{s}/lk - 1. \quad (4.11b)$$

Both the above coefficients are valid for the full range of the parameter α. Clearly both α_1 and α_2 are positive. The minimum value of these quantities is zero for the $\alpha = 1$ solution, which is as it should be. As the parameter $\alpha \to 1$, the solution tends to the vacuum metric for the special case $C = 0$.

Consider for a moment the expected signs of two quantities defined by Eq. (4.10). For any interior which satisfies the condition $\tau < 1$, Eq. (4.10a) must be positive. The expected sign of Eq. (4.10b), however, is not so obvious. It involves the difference $mn/\sqrt{\tau} - \rho$, and though $mn < \rho$ due to the energy of motion of nucleons and their interaction, $\tau < 1$. This being the case, the sign of α_2 is entirely model dependent. Clearly, if the interior is formed from the simple condensation of material from initially rarefied matter, $\alpha_2 > 0$. The physical interpretation of a negative α_2 is that simple condensation did not occur, one possibility being that nuclear reactions took place during the evolution to the final state.

There are models in the literature for which $\alpha_2 < 0$. One such model is a stellar model based on a real gas presented by Saakyan and Vartanyan.[11] Numerical calculations show that α_2 becomes negative when the central density exceeds some finite value, suggesting that singular interiors may share this property. The solutions of this paper offer a counterexample to such a conjecture.

V. STABILITY

Even though a solution corresponds to an equilibrium configuration, it need not be stable to small perturbations. The solution depicting a stellar interior is merely a statement specifying the magnitude of the pressure needed at a point to maintain the system in equilibrium for a given density. It does not follow that if small perturbations were to occur (either an increase or decrease in pressure) the system would return to its original configuration. Indeed, it could well occur that a small perturbation would grow, causing the distribution to either collapse or expand, depending upon the nature of the perturbation and characteristics of the fluid. Solutions which are not stable to these small perturbations could not remain static for a long time.

For a general relativistic fluid sphere, the approach to the question of stability is as follows.[12-14] To a solution which satisfies the static field equations, introduce a small radial perturbation which does not destroy the spherical symmetry. Neglect all quantities which are of the second order or higher in the motions of the fluid and retain only the linear terms. With the further assumption that only adiabatic processes are occurring, the time-dependent equations are investigated to this level of approximation around the static equilibrium configuration.

For the line element given by Eq. (2.1), the nth-order normal mode is expressible in terms of an amplitude $u_n(y)$ given by

$$\delta y(y,t) = y^{-2}\gamma(y)u_n(y)\exp(i\omega_n t). \quad (5.1)$$

To this order of approximation, the time-dependent equations reduce to the Sturm–Liouville system for u_n

$$\frac{d}{dy}\left[\underline{P}\frac{du_n}{dy}\right] + [\underline{Q} + \omega_n^2 \underline{W}]u_n = 0. \quad (5.2)$$

The functions $\underline{P}$, $\underline{Q}$, and $\underline{W}$ are expressed in terms of the equilibrium configurations for the star

$$\begin{aligned} \underline{P} &= \gamma^3\tau^{-1/2}y^{-2}\hat{\gamma}P \\ \underline{Q} &= -4\gamma^3\tau^{-1/2}y^{-3}\frac{dP}{dy} - 8\pi\gamma^3 P(P+\rho)\tau^{-3/2}y^{-2} \\ &\quad + \gamma^3\tau^{-1/2}y^{-2}(P+\rho)^{-1}\left(\frac{dP}{dy}\right)^2, \\ \underline{W} &= \gamma\tau^{-3/2}y^{-2}(P+\rho). \end{aligned} \quad (5.3)$$

The function $\hat{\gamma}$ is the adiabatic index expressed as

$$\begin{aligned} \hat{\gamma} &= P^{-1}(P+\rho)\left(\frac{\partial P}{\partial \rho}\right)_{\text{constant entropy}} \\ &= P^{-1}(P+\rho)v_S^2(y), \end{aligned} \quad (5.4)$$

where $v_S(y)$ is the sound speed throughout the fluid.

Solutions to Eq. (5.2) are sought which satisfy the boundary conditions

$$\begin{aligned} &u_n(y)|_{y=0} \sim y^3, \\ &\left(-\gamma y^{-2}\hat{\gamma}P\frac{du_n}{dy}\right)\bigg|_{y=1} = 0. \end{aligned} \quad (5.5)$$

The first condition guarantees that the perturbation will not displace the fluid at the center of the sphere, and the second that the Lagrangian change in the pressure vanishes at the surface.

Interest here is in whether or not the solution given by Eq. (2.8) is stable with respect to small perturbations. The approach is as follows. Integrate Eq. (5.2) with the trial value of ω_n set equal to zero. If, after the integration, the eigenfunction u_n has gone through zero N times, there are precisely N unstable modes of oscillations (0 through $N-1$). If after going through N nodes, the boundary conditions given by Eq. (5.5) at both endpoints are satisfied, then the modes 0 through $N-1$ are unstable and the Nth node is in neutral equilibrium. Hence, if the solution has zero nodes, the $N=0$ mode (the fundamental) is stable.

Again, due to the nonlinearity of the equation, Eq. (5.2) is not tractable in terms of known functions, and an approximate approach is necessary. A close examination of the pressure and density given by Eq. (2.8) indicates that to a good approximation, both behave as y^{-2} over a large portion of the sphere when $\alpha \neq 1$. Further, from the analysis of the last section, the $C=0$ solution has a density going precisely with this radial dependence. This being the case, the approach to be used will be to approximate both the pressure and density in this fashion and solve Eq. (5.2) in the vicinity of the origin. This will yield u_n in the region where u_n is expected to have the greatest variation.

Expanding Eq. (5.2) results in the expression

$$y^2u'' + y[3(y\gamma'/\gamma) - \tfrac{1}{2}(y\tau'/\tau) + y(\ln \hat{\gamma}P)' - 2]u' + v_S^{-2}[4(y\gamma'/\gamma) + (y\gamma'/\gamma)^2 - 8\pi y^2P/\tau]u = 0. \quad (5.6)$$

The approximate expressions for the interior, valid for small y and $\alpha \neq 1$ are

$$\begin{aligned} &\tau = 1/s, \quad 8\pi R^2y^2P = j^2/s, \\ &8\pi R^2y^2\rho = ij/s, \quad y\gamma'/\gamma = j, \quad v_S^2 = j/i. \end{aligned} \quad (5.7)$$

For $\alpha = 1$, both the pressure and density approach a constant.

Use of Eq. (5.7) in Eq. (5.6) leads to the expression for the fundamental mode of oscillation valid to this level of approximation

$$y^2u'' - (1+3\alpha)yu' + 4(1+\alpha)u = 0. \quad (5.8)$$

Equation (5.8) can be readily integrated. The resulting expression is

$$u(y) = y^h\,[C_1\cos(f\ln y) + C_2\sin(f\ln y)], \quad (5.9)$$

where $h = (2+3\alpha)/2$ and $f = (12 + 4\alpha - 9\alpha^2)^{1/2}/2$.

Note that $u(y)$ goes through infinitely many oscillations as the origin is approached. A direct calculation shows that the above expression can be made to satisfy the boundary conditions expressed by Eq. (5.5). A more accurate calculation leads to a $u(y)$ which is the same as Eq. (5.9) only multiplied by a benign polynomial in y, in no way altering the conclusion that the $\alpha \neq 1$ solutions are inherently unstable.

To this order of approximation, the solutions considered above for $\alpha \neq 1$ are identical to the Misner–Zapolsky solution. It is well known that their interior can be made stable. One simply deletes the central region of the sphere and replaces it with a small uniform high-density core. The size of the core is determined by carefully matching the solution to the uniform density configuration at the core boundary, then requiring that $u(y)$ expressed by Eq. (5.9) have zero nodes within the confines of the sphere.

The stabilizing method for the solutions considered in this paper follows the same general procedures outlined for the Misner–Zapolsky solution. However, as these more general solutions have a pressure boundary, it is *not* necessary to introduce an outer envelope in order to complete the solution, as is the case with the Misner–Zapolsky interior. The physical interpretation of the solid core is that at ultrahigh densities, neutron matter could form a solid.[15] Some of the pulsar data (in particular, the spin-down phenomenon of the vela pulsar) could best be understood in terms of such a solid core.

The $\alpha = 1$ solution[16] is the only nonsingular solution of the family, and as such requires a more detailed consideration. The density and pressure approach a constant near the origin. Upon forming their ratio, one sees the sound speed does also. Introducing this result into Eq. (5.6) with $\alpha = 1$, one finds only that $u \sim \text{const}\, xy^3$ near $y = 0$. Hence we find only that the appropriate boundary condition is satisfied there.

We can show that the solution satisfies the necessary part of the stability criteria

$$\frac{\partial M}{\partial \rho_0} \geqslant 0. \quad (5.10)$$

Upon taking the limit as α tends to one in the expression for the density, Eq. (2.8), with integration constants given by Eq. (2.7b), one finds

$$4\pi R^2\rho_0 = \frac{M}{R}\left(1 - \frac{M}{R}\right)^{2/3}\left(1 - \frac{5}{2}\frac{M}{R}\right)^{-5/3}, \quad (5.11)$$

and hence

$$\frac{\partial M}{\partial \rho_0} = 4\pi R^3\left(1 - \frac{M}{R}\right)^{1/3}\left(1 - \frac{5}{2}\frac{M}{R}\right)^{8/3} \times \left(3 - 4\frac{M}{R} + 10\left(\frac{M}{R}\right)^2\right)^{-1}. \quad (5.12)$$

We observe that Eq. (5.12) is positive definite for all values of M/R. The density is, however, singular for $M/R = \frac{2}{5}$, and one would expect the solution not to be stable near this mass-to-radius ratio.

A numerical integration of the stability equation shows that this is the case. One finds a maximum γ_S equal to 0.589. This corresponds to a maximum mass of 2.2 $M_\odot$ for a 10 km star.

VI. CONCLUSION

As stated earlier, the equations of state at ultrahigh density are not well known. Even in the limit of polytropic equations of state, the central sound speed is not agreed on. The value of this quantity depends critically on the nature of the interactions assumed to calculate it, and can take on any value between zero[17] and unity.[18] Observational evidence at this time is still not sufficiently accurate to aid in discriminating between the various models, but it is expected that pulsar data in the near future will shed light on this problem.

The solutions discussed in this paper are not based on particular assumptions with regard to the various interparticle interactions, but on the asymptotic behavior of such equations of state. In view of this, we make no claim that any of these solutions appropriately describe real interactions within a star. The asymptotic character of these interiors do, however, indicate that they may be of some use in the investigation of real fluids.

The Sturm–Liouville equation for the zeroth-order normal mode oscillations in the neighborhood of the origin for this family of solutions was solved. This demonstrated conclusively that all but one member was unstable in the central region. As pointed out by Whitman and Redding, the instability can be reduced by properly matching the solutions to a stable high-density core. Indeed, as pointed out earlier, some pulsar data presently available could best be understood with such a high-density uniform core. Recent calculations indicate that such a core could develop under extreme conditions of pressure and density. With regard to the solutions discussed here, it is not necessary to introduce an outer envelope as was done by Misner and Zapolsky, as the present interiors have a pressure boundary at finite radius. The $\alpha = 1$ solution was found stable, suggesting a maximum mass of 2.2 $M_{\odot}$ for a 10 km star.

[1]V. Canuto, Annu. Rev. Astron. Astrophys. **13**, 335 (1975).
[2]V. Canuto and J. Lodenquai, Phys. Rev. D **11**, 233 (1975).
[3]M. Wyman, Phys. Rev. **75**, 1930 (1949).
[4]P. G. Whitman, J. Math. Phys. **18**, 868 (1977).
[5]P. G. Whitman and R. W. Redding, Astrophys. J. **224**, 893 (1978).
[6]P. G. Whitman and J. F. Pizzo, Astrophys. J. **230**, 983 (1979).
[7]C. W. Misner and H. S. Zapolsky, Phys. Rev. Lett. **12**, 635 (1964).
[8]Y. C. Leung and C. G. Wang, Astrophys. J. **181**, 895 (1973).
[9]J. R. Oppenheimer and G. M. Volkoff, Phys. Rev. **55**, 374 (1939).
[10]Ya. B. Zel'dovich and I. D. Novikov, *Relativistic Astrophysics, Vol. 1: Stars and Relativity* (University of Chicago, Chicago, 1971), p. 286.
[11]G. S. Saakyan and Yu. L. Vartanyan, Astron. Zh. **41**, 193 (1964).
[12]S. Chandrasekhar, Phys. Rev. Lett. **12**, 114 (1964); Erratum, *ibid.* **12**, 437 (1964).
[13]J. M. Bardeen, K. S. Thorne, and D. W. Meltzer, Astrophys. J. **145**, 505 (1966).
[14]D. W. Meltzer and K. S. Thorne, Astrophys. J. **145**, 514 (1966).
[15]F. G. Smith, *Pulsars* (Cambridge U. P., New York, 1977), p. 40.
[16]R. Adler, J. Math Phys. **15**, 727 (1976); Erratum, *ibid.* **17**, 158 (1976).
[17]R. Hagedorn, Astron. Astrophys. **5**, 184 (1970).
[18]J. D. Walecka, Ann. Phys. **83**, 491 (1974).

Transient thermodynamics of electromagnetic media in general relativity

Claude Barrabès
Départment de Physique, Faculté des Sciences, 37200 Tours, France

(Received 24 May 1984; accepted for publication 23 November 1984)

This paper is devoted to a study of the constitutive equations of relativistic electromagnetic media, when submitted to transitory processes, following the causal thermodynamics of Müller–Israel. The formalism of differential geometry is used to describe the kinematical and dynamical evolution of the medium and the thermodynamical description is given in terms of local functional relations. According to the hypothesis of Müller–Israel, the second-order terms in the deviation from thermodynamical equilibrium are taken into account in the expression of the entropy current. The relaxation terms which then appear in the transport equations restore the causality. Several cases corresponding to various kinds of media are investigated.

I. INTRODUCTION

In the standard treatment of the transport phenomena in continuous media, the heat conduction and the viscosity are described by the laws of Fourier and Navier–Stokes or by their relativistic version proposed by Eckart.[1] It is well known that these laws suffer from the drawback of giving a noncausal description of the propagation phenomena. Such a result, which appears to be an offense to intuition in classical mechanics, becomes a real paradox in relativistic theory. Another unattractive feature arises when these laws are used to investigate the stability properties. As shown by Lindblom and Hiscock,[2] they allow the existence of generic short-wavelength secular instabilities when applied to rotating stellar models in general relativity.

The usual way to evade such difficulties is to introduce relaxation terms in the transport equations, however, until the 1960's the presence of these terms remained purely *ad hoc* and was only justified by means of heuristic arguments. In the last 20 years, attempts to give a proper explanation to the existence of the relaxation terms have been proposed both in classical and relativistic theory of continuous media. Among the approaches to solve this problem in a relativistic theory, three could be mentioned: the one of Maugin,[3] which follows the axiomatic formulation of hereditary processes in continuous media; the one of Bampi and Morro,[4] which uses the method of hidden variables; and the one of Carter,[5] which gives prominence to the canonical four-momentum associated with the entropy flux.

Another explanation was proposed by Müller[6] in classical mechanics and later rediscovered by Israel[7] in relativity. These authors showed that the existence of the relaxation terms is the consequence of the following hypothesis: besides the usual equilibrium parameters, the entropy also depends on the dissipative effects (heat flux and viscosity in a neutral fluid). They also pointed out that this hypothesis is particularly relevant in fast varying phenomena such as the transient phenomena accompanying shock waves. Their analysis makes clear that the usual theory of the irreversible quasistationary processes is not the appropriate framework for studying either the propagation phenomena or the stability problems as recently claimed by Hiscock and Lindblom.[8] Israel and Stewart[9] showed that this hypothesis is in agreement with the study of the fluid properties by means of the relativistic kinetic theory. Stewart[10] found a maximum value for the wave front of thermal disturbances which is equal to $(3/5)^{1/2}c$ and is then in accordance with the causality principle. Lebon, Jou, and Casas-Vazquez[11] proposed a generalization to the thermodynamics of Müller–Israel by assuming that the dissipative quantities are also included in the set of variables defining the internal energy. This assumption implies an extension of the Gibbs equation which finally leads to transport equations similar to the ones of Müller–Israel.

The purpose of this paper is to apply the thermodynamics of Müller–Israel to the relativistic electromagnetic continuous media. We also give a description of this theory within the framework of the axiomatic approach of the constitutive laws of continuous media as initiated by Truesdel and Noll.[12] We shall often refer to a series of papers by Maugin,[13] which presents a full description of the relativistic electromagnetic continuous media though it is limited to the irreversible quasistationary processes.

The electromagnetic media are of great importance in astrophysics and it is then particularly worthwhile to have at our disposal a coherent description in general relativity of their thermodynamical properties. Applications may, for instance, be found in pulsars, magnetospheres, black-hole accretion rings, and in the early stages of the cosmological evolution. Furthermore, the fact that fast varying phenomena may effectively occur in the above examples renders this study more urgent and necessary.

This paper is organized as follows. We first give a general description of the kinematical (Sec. II) and dynamical (See. III) evolutions of the medium by using the formalism of differential geometry. A nondetailed overview of the well-known properties of the reversible and irreversible quasistationary processes in electromagnetic media is given in Sec. IV in order to make easier the description of the transient thermodynamics which is performed in Sec. V. Our formalism is general enough to include electromagnetic solids as well as electromagnetic fluids, while a recent work of Israel and Stewart[14] only dealt with electromagnetic fluids. Finally, several cases corresponding to various kinds of media are investigated.

We have used, as much as possible, intrinsic tensorial

 0022-2488/85/040798-06$02.50

notations. In order to avoid any ambiguity we make precise the following conventions.

The (anti) symmetrized tensorial product is symbolized by $\otimes$ ($\hat{\otimes}$) and the contracted tensorial product by the dot $\cdot$. In the same way, $\tilde{t}$ ($\hat{t}$) represents the (anti) symmetrical part of the two-tensor t.

Here, ∇_v stands for the covariant derivative in the direction of the vector V. The covariant derivative with respect to the flow vector U will be noted $\dot{} = \nabla_U$.

The divergence operator of any tensor always applies to its last index, for instance, $(\operatorname{div} A)^{\alpha\beta} = A^{\alpha\beta\gamma}{}_{;\gamma}$, where ; is the usual covariant derivative symbol.

II. KINEMATICS OF CONTINUOUS MEDIA

Let (M,g) be the space-time manifold of general relativity equipped with the hyperbolic Riemannian metric tensor g. The local coordinates of the generic point p are x^α ($\alpha = 0, 1, 2, 3$) and we shall use for g the signature $(+---)$. Any description of matter in terms of a continuous medium is based on the use of a three-dimensional differential manifold $\mathscr{M}$ whose generic point P symbolizes a material point of the medium. The local coordinates on $\mathscr{M}$ will be called $X^K (K = 1,2,3)$.

The motion of the medium corresponds with the existence of a one-parameter, $s \in I \subset \mathbb{R}$, family of regular embeddings ψ_s of $\mathscr{M}$ in the space-time manifold such that (1) the image $\psi_s(\mathscr{M})$ is a three-dimensional spacelike hypersurface, $\forall s$, and (2) the point $p = \psi_s(P)$ describes as s varies, X being fixed, the timelike curve $\mathscr{C}_X$ in M which is the world line of the material point P. It is of practical use to identify s with the proper time so that the tangent vector to the curve $\mathscr{C}_X$, $U = \partial z/\partial s$, is the normalized flow vector which is associated with the motion. If $\perp$ is the projection of the tangent space to M at P onto the subspace of tangent vectors orthogonal to U, the composite mapping $\Phi_s = \perp \circ \psi_s'$ is, for any s, an isomorphism of the tangent vectors at $P \in \mathscr{M}$ onto the tangent vectors orthogonal to the flow at $p = \psi_s(P)$.

For any one-form θ at $p = \psi_s(P)$, its reciprocal image at P under the regular embedding ψ_s is $\mathbf{\theta} = \psi_s^* \theta$. For instance, the reciprocal image of the metric tensor g of M defines a metric tensor $\mathbf{G}_s = \psi_s^* g$ on $\mathscr{M}$ whose variation with respect to the evolution parameter s gives the strain tensor $\mathbf{e} = (\mathbf{G}_s - \mathbf{G}_{s'})/2$ at $P \in \mathscr{M}$.

An isomorphism between the one-forms at P and the one-forms orthogonal to U at $p = \psi_s(P)$ can also be obtained by using the canonical differentiable projection $\bar{\psi}$ such that any world line $\mathscr{C}_X$ is mapped onto the material point P ($\bar{\psi} \circ \psi_s = \mathrm{Id}_{\mathscr{M}}$, $\forall s$). It can be easily verified that the pullback by $\bar{\psi}$ of any one-form at P is a one-form orthogonal to U at $p \in \mathscr{C}_X$. When applied to the strain tensor e on $\mathscr{M}$ it gives the symmetrical orthogonal to the flow tensor $e = \bar{\psi}^* \mathbf{e}$, which is the strain tensor at $p \in M$.

It can be noticed that the mappings Φ_s and $\bar{\psi}^*$ are not the dual of each other, even though they, respectively, map vectors and one-forms at P. Their action on any mixed tensor $\mathbf{t}$ at $P \in \mathscr{M}$ gives the tensor t at $p \in M$ which is orthogonal to the flow. We shall use the following notation:

$$\Phi_s : \mathbf{t}(X) \mapsto t(x) = \Phi_s \mathbf{t}(X), \tag{1}$$

where the boldface notation will hereafter design any tensor field on $\mathscr{M}$ which is associated with an orthogonal to the flow tensor field on M by means of the above mappings.

The variation of a tensor field under the group of transformations generated by the flow vector U is given by its Lie derivative $\mathscr{L}_U$ with respect to U. It is, however, well known that this derivative does not preserve the orthogonality to U of the tensor to which it applies. It is then convenient to use the convective derivative, noted $[\,]^{\cdot}$, which is defined from the Lie derivative[15] $\mathscr{L}_U$ in the same way as the Fermi derivative with respect to U is built from the covariant derivative ∇_U. The components of the convective derivative of an arbitrary mixed two-tensor t are

$$[t]^{\cdot\alpha}{}_\beta = (\mathscr{L}_U t)^\alpha{}_\beta + t^\lambda{}_\beta \dot{U}_\lambda U^\alpha{}_\lambda U^\lambda \dot{U}_\beta. \tag{2}$$

If t is orthogonal to U the last term of the rhs vanishes and it can be verified that $[t]^{\cdot}$ is also orthogonal to U.

For any tensor t satisfying (1) it can be shown that

$$[t]^{\cdot} = \Phi_s\left(\frac{\partial \mathbf{t}}{\partial s}\right).$$

The convective derivative of t is thus the image under Φ_s of the partial derivative with respect to the evolution parameter of the material tensor $\mathbf{t}$. This corresponds to the objectivity property and it is the reason why the convective derivative has to be used whenever the principle of objectivity or rheological invariance is required.[3,15] The strain rate tensor d, which is an objective quantity, is such that

$$2d = [g]^{\cdot}.$$

Its trace, $\operatorname{tr} d$, describes the expansion rate and its trace-free part

$$\sigma = d - (\gamma/3)(\operatorname{tr} d), \tag{3}$$

gives the shear rate tensor, where γ is the projection tensor on the subspace orthogonal to U

$$\gamma_{\alpha\beta} = g_{\alpha\beta} - U_\alpha U_\beta. \tag{4}$$

III. DYNAMICS OF CONTINUOUS MEDIA

The balance law of any quantity A, represented by an r tensor on M, has the general form

$$\operatorname{div} A = \Lambda,$$

where Λ is the $(r-1)$ tensor describing the production of A. For instance, if N and T are, respectively, the pure mass current and the energy-momentum tensor, if there is no mass creation or annihilation, and if one calls f the external force density, then

$$\operatorname{div} N = 0, \quad \operatorname{div} T = f. \tag{5}$$

In the same way, the angular momentum tensor $\mathfrak{M}$, which can be decomposed as

$$\mathfrak{M} = x \hat{\otimes} T + \varphi,$$

φ being the spin tensor, admits the following balance law

$$\operatorname{div} \mathfrak{M} = L + x \hat{\otimes} f,$$

where L represents the external torque's density and $x \hat{\otimes} f$ the moment of the external forces. Thanks to the balance equation of T one gets

$$\operatorname{div} \varphi - \hat{T} = L, \tag{6}$$

showing that in spinless media ($\varphi = 0$) the antisymmetrical part of the energy-momentum tensor is given by the external torque's density L.

The dynamical quantities are more easily interpreted when using their decomposition in spatial and temporal components with respect to the flow vector U. In the Eckart decomposition, N and T take the following form:

$$N = \rho U, \quad T = \rho(1 + \epsilon_0) U \otimes U + U \otimes q + \bar{q} \otimes U + t, \tag{7}$$

where ρ is the pure mass density and ϵ_0 the internal energy density per unit mass. The spatial vectors q, $\bar{q}$ and the spatial tensor t will be related to the heat flux, the stresses, and eventually to electromagnetic effects.

As it is well known, there exist many possible expressions for the energy-momentum tensor of an electromagnetic medium. This corresponds to the various ways of splitting the total energy-momentum tensor, which appears in the Einstein equations, into matter and field parts. Instead of choosing one particular decomposition, we shall use here the expressions of the electromagnetic forces f and torques L densities which de Groot and Suttorp[116] directly derived from a microscopical approach

$$\begin{aligned} f^\alpha &= F^{\alpha\beta} J_\beta + \tfrac{1}{2}\pi_{\mu\nu} \nabla^\alpha F^{\mu\nu} + \rho \nabla_U \\ &\quad \times \left(\frac{U^\gamma (F^{\alpha\beta}\pi_{\beta\gamma} - \pi^{\alpha\beta}F_{\beta\gamma}) - U^\alpha U_\lambda F^{\lambda\mu}\pi_{\mu\nu} U^\nu}{\rho} \right), \\ L^{\alpha\beta} &= F^{\mu[\alpha}\pi^{\beta]}{}_\mu + U^{[\alpha}(F^{\beta]\lambda}\pi_{\lambda\gamma} - \pi^{\beta]\lambda}F_{\lambda\gamma}) U^\gamma. \end{aligned} \tag{8}$$

Here J is the electric current density, F the electromagnetic field tensor, and π the electromagnetic polarization tensor. These tensors admit the following decomposition with respect to U:

$$F^{\alpha\beta} = 2U^{[\alpha}E^{\beta]} + \eta^{\alpha\beta\mu\nu} B_\mu U_\nu,$$

$$\pi_{\alpha\beta} = 2U^{[\alpha}P^{\beta]} - \eta^{\alpha\beta\mu\nu} M_\mu U_\nu,$$

$$J = \rho_e U + j,$$

where ρ_e is the free charge density and where the (co) vectors E, B, P, M, and j represent, respectively, the electric field and magnetic induction, the electric and magnetic polarizations and the electric conduction current density ($j = 0$ in a perfect insulator). The four-tensor η is the usual antisymmetrical volume element tensor. The balance laws (5) and (6) are supplemented with Maxwell's equation which relate the electromagnetic tensors F and π to the current J.

When replaced in (5) and (6) the above expressions of f and L give the following equation of balance of energy for a spinless medium:

$$\begin{aligned} \rho\dot{\epsilon} &= (t - \mathscr{T})\cdot d - \operatorname{div} q + q\cdot\dot{U} + E\cdot j \\ &\quad + P\cdot[E]^{\cdot} + M\cdot[B]^{\cdot}, \end{aligned} \tag{9}$$

where the two-tensor $\mathscr{T}$ is given by

$$\mathscr{T} = E \otimes P + B \otimes M, \tag{10}$$

and where the internal energy density per unit mass

$$\epsilon = \epsilon_0 + (E\cdot P)/\rho,$$

contains the electric polarization energy per unit mass.

In some media it is more convenient to use the internal energy density

$$\tilde{\epsilon} = \epsilon_0 - (B\cdot M)/\rho = \epsilon - (\pi\cdot F)/2\rho,$$

which contains the magnetic polarization energy per unit mass. The balance law of ϵ is then

$$\begin{aligned} \rho\dot{\tilde{\epsilon}} &= (t - \mathscr{T})\cdot d - \operatorname{div} q + q\cdot\dot{U} + E \\ &\quad \cdot j + \rho E\cdot[\tilde{P}]^{\cdot} + \rho B\cdot[\tilde{M}]^{\cdot}, \end{aligned} \tag{11}$$

where $\tilde{P} = P/\rho$ ($\tilde{M} = M/\rho$) is the electric (magnetic) polarization per unit mass.

It must be noticed that the quantities E and B ($P, M, \tilde{P}$, and $\tilde{M}$) which appear in Eqs. (9) and (11) are vectors (covectors). The opposite convention would have given a different result because the action of raising or lowering the indices does not commute with the convective derivative. This convention will be used in the remaining part of this work.

IV. REVERSIBLE AND QUASISTATIONARY IRREVERSIBLE PROCESSES

The study of the motion of a charged continuous medium cannot be reduced to a problem of mechanics and electromagnetism. Every strain and every polarization are accompanied by thermal effects whose description requires new variables and new equations. Furthermore, among all the mechanically admissible processes only a limited number of them are allowed. This selection rule is a consequence of the entropy principle which states that the internal entropy production has to be null (positive) for any (ir)reversible process

$$\operatorname{dis} S \geqslant 0,$$

where S is the entropy current. The entropy inequality may also be written

$$\rho\dot{\eta} + \operatorname{div} s \geqslant 0, \tag{12}$$

if one uses the Eckart decomposition of S

$$S = \rho\eta U + s, \tag{13}$$

η being the entropy density per unit mass and s the entropy flux vector.

Actually, only equilibrium states and quasistatic or reversible processes are well defined. They correspond to the existence of a set of thermodynamical variables which fully determine the internal energy. For a charged medium, these variables are the entropy density η, a set of mechanical variables $Y_{\rm m}$, and a set of electromagnetic variables $Y_{\rm em}$ which characterize the mechanical and electromagnetic properties of the medium. The coordinates X^K of the material point also have to be taken into account if the medium is not homogeneous. The internal energy per unit mass ϵ is thus given by the functional relation

$$\epsilon = \mathscr{E}(\eta, Y_{\rm m}, Y_{\rm em}, X). \tag{14}$$

In some media it proves more convenient to start with the free energy density $\psi = \epsilon - \theta\eta$, where $\theta = \partial\epsilon/\partial\eta$ is the temperature. The functional relation which gives ψ is then

$$\psi = \mathscr{F}(\theta, Y_{\rm m}, Y_{\rm em}, X). \tag{15}$$

The choice of the variables $Y_{\rm m}$ and $Y_{\rm em}$ and the form of the functionals $\mathscr{E}$ and $\mathscr{F}$ determine the nature of the medium. This choice is, however, restricted by the principle of rheological invariance or objectivity which requires that the constitutive equations have to be independent of the observer.

The problem of the existence of Eq. (14) or Eq. (15) for any irreversible process, satisfying the entropy principle, has been widely debated and has not yet received a definite answer. For an irreversible quasistationary process, use is made of the axiom of the local equilibrium state which postulates that the functional relations (14) or (15) are valid at each point of the medium and at each time of its evolution.

Another important question arises when one wants to link the thermodynamical and mechanical properties of the medium and notably when one is looking for the relations between the entropy and the terms which appear in the balance equation of energy (9) or (11). For any reversible or irreversible quasistationary process it is assumed that the entropy density η is independent of the dissipative effects and that the entropy flux s is simply given by

$$s = q/\theta, \tag{16}$$

where q represents the heat flux in a spinless medium (see Maugin[3] for media with spin).

The entropy inequality, satisfied by any reversible or irreversible quasistationary process, is then

$$\rho\dot{\eta} + \operatorname{div}(q/\theta) \geqslant 0.$$

The term $\rho\dot{\eta}$ can be deduced from the balance equation of energy (9) or (11) and from the functional relations (14) or (15). The new inequality which one obtains is then used to determine the constitutive equations of the medium. We shall now briefly describe the well-known cases corresponding to the electromagnetic fluids or solids in order to make precise the notations which will be used in the next section. This will also make easier the comparison between the irreversible quasistationary processes and the transient ones.

The electromagnetic variables which we previously symbolized by $Y_{\rm em}$ will be chosen among the electromagnetic field vectors E, B and the electromagnetic polarization per unit mass covectors P, M. The first ones (E, B) will here be used in the case of fluids and the second ones (P, M) for solids, however, other choices may equally be performed (see Maugin).[17] Furthermore, as the axiom of objectivity requires the Lorentz invariance of the functional relations (14) and (15), only some special arrangements of these variables will appear. This will be made precise when dealing with fluids and solids.

The set of thermodynamical variables defining the internal energy density of a charged fluid is composed of the entropy density η, the volume per unit mass ρ^{-1}, and the Lorentz invariants $E^2, B^2, (E\cdot B)^2$. Let us now set

$$t_0 = \frac{\partial\epsilon}{\partial\rho^{-1}}\gamma = -p\gamma,\ \mathscr{T}_0 = E\otimes\rho\frac{\partial\epsilon}{\partial E} + B\otimes\rho\frac{\partial\epsilon}{\partial B},$$
$$t' = \hat{t} - t_0 - (\hat{\mathscr{T}} - \mathscr{T}_0),\ P' = P - \rho\frac{\partial\epsilon}{\partial E},$$
$$M' = M - \rho\frac{\partial\epsilon}{\partial B}, \tag{17}$$

where p is the pressure, γ the projection tensor (4), $\mathscr{T}$ is given by (10), and where $\hat{}$ is the symmetrization symbol. The two-tensor t' is thus symmetrical. The entropy inequality for an electromagnetic spinless fluid then gives

$$\tfrac{1}{2}t'\cdot[g]^{\cdot} + P'\cdot[E]^{\cdot} + M'\cdot[B]^{\cdot} - q\cdot(\theta^*/\theta) + E\cdot j \geqslant 0, \tag{18}$$

where θ^* is the relativistic temperature gradient

$$\theta^* = \gamma\cdot(\nabla\theta - \theta\dot{U}). \tag{19}$$

In the case of an electromagnetic elastic medium it is preferable to use the free energy density ψ because, one has to refer to a natural unstrained and unpolarized state for which ψ is minimum. Furthermore, the axiom of rheological invariance is satisfied by expressing all the quantities on the three-manifold $\mathscr{M}$, defining the medium, by means of the mapping Φ_s, see (1) of Sec. II. The thermodynamical variables which define ψ are the temperature θ, the metric tensor on $\mathscr{M}$, $\mathbf{G}_s = \Phi_s^* g$, and the electromagnetic polarization covectors on $\mathscr{M}$, $\tilde{\mathbf{P}} = \Phi_s^*\tilde{P}$ and $\tilde{\mathbf{M}} = \Phi_s^*\tilde{M}$. With similar notations to the ones used for fluids and by using the mapping Φ_s, we define the following tensors on $\mathscr{M}$:

$$\mathbf{t}_0 = 2\rho\frac{\partial\psi}{\partial\mathbf{G}_s},\ \tilde{\mathscr{T}}_0 = -\frac{\partial\psi}{\partial\tilde{\mathbf{P}}}\otimes\mathbf{P} - \frac{\partial\psi}{\partial\tilde{\mathbf{M}}}\otimes\mathbf{M},$$
$$\tilde{\mathbf{t}}' = \hat{\mathbf{t}} - \tilde{\mathbf{t}}_0 - (\hat{\mathscr{T}} - \tilde{\mathscr{T}}_0),\ \mathbf{E}' = \mathbf{E} + \frac{\partial\psi}{\partial\tilde{\mathbf{P}}},$$
$$\mathbf{B}' = \mathbf{B} + \frac{\partial\psi}{\partial\tilde{\mathbf{M}}}. \tag{20}$$

The entropy inequality for an electromagnetic spinless solid is then

$$\tfrac{1}{2}\tilde{\mathbf{t}}'\cdot\frac{\partial\mathbf{G}_s}{\partial s} - \rho\mathbf{E}'\cdot\frac{\partial\tilde{\mathbf{P}}}{\partial s} - \rho\mathbf{B}'\cdot\frac{\partial\tilde{\mathbf{M}}}{\partial s} + \frac{\mathbf{q}\cdot\boldsymbol{\theta}^*}{\theta} + \mathbf{E}\cdot\mathbf{j} \geqslant 0. \tag{21}$$

The inequalities (18) and (21) have to be satisfied by any reversible (= 0) or irreversible quasistationary (> 0) process. Let us briefly resume some of their main consequences for which a detailed description may be found in the review article by Maugin.[13]

First, they imply the existence of linear relations between the electromagentic polarization and electromagnetic field, because they require that $P' = M' = 0$ or $\mathbf{E}' = \mathbf{B}' = 0$. At equilibrium, fluids behave like perfect fluids $t = -p\gamma$, elastic media like hyperelastic ones, and there is neither heat nor electric conduction, $q = j = 0$.

Finally, if one considers an irreversible quasistationary process occuring in a homogeneous and isotropic fluid, the constitutive relations for q, j, and t have the following form:

$$\begin{aligned} &q = K\cdot\theta^* + \xi_q\cdot E,\ j = \sigma\cdot E + \xi_j\cdot\theta^*, \\ &t = -p\gamma + \zeta(\operatorname{tr} d)\gamma + 2\mu\sigma. \end{aligned} \tag{22}$$

For an homogeneous and isotropic thermoelastic charged medium, submitted to infinitesimal strains and thermal perturbations from equilibrium, the first two equations giving q and j keep the same from while the stress-strain relation is

$$t = \zeta_e(\operatorname{tr} e)\gamma + 2\mu_e e + K(\theta - \theta_0), \tag{23}$$

where θ_0 is the equilibrium temperature.

The coefficients K, σ, ξ_q, ξ_j, p, ζ_e, μ, μ_e, and K are scalar value functions of the thermodynamic variables. They describe respectively, Fourier's law, Ohm's law, the Peltier and Thomson effects, the pressure, the bulk viscosity, the shear viscosity, the elasticity, and thermoelasticity coefficients.

V. TRANSIENT PHENOMENA

Unlike the irreversible quasistationary processes which we studied in the preceding section, the transient processes are fast varying phenomena. The physical quantities which characterize them have very short length and time scales of variation and their space-time derivatives will therefore take large values. We shall here only consider the transient processes which correspond to small departures from an equilibrium state. Then, even though the associated perturbative terms are infinitesimal, their space-time derivatives are not and they cannot be neglected.

These considerations, which were pointed out by Müller[6] and Israel,[7] have to be taken into account in the expression of the entropy current S (see Ref. 18). If η' and s', respectively, indicate the perturbations in the entropy per unit mass and in the entropy flux, the entropy current is

$$S=\rho(\eta+\eta')U+q/\theta+s',$$

and the entropy principle gives the following inequality:

$$\rho\dot{\eta}+\operatorname{div}(q/\theta)+\rho\dot{\eta}'+\operatorname{div}s'>0, \tag{24}$$

where the derivatives $\dot{\eta}'$ and div s' now cannot be neglected.

The first two terms of this inequality correspond to the unperturbed state. They are given by the lhs of the inequalities (18) or (21), according to the nature of the medium, fluid or solid. It has been shown that these terms are of first order in the following dissipative quantities, which all vanish at equilibrium: q, j, t' or $\tilde{t}'$, P' or E', and M' or B'. Then, the last two terms of (24), which contain space-time derivatives of the perturbative terms η' and s', will also be of the first order provided that η' and s' are of the second order in the above dissipative quantities.

For a fluid, one then sets the expressions

$$\begin{aligned}\eta'&=\tfrac{1}{2}a_1q^2+\tfrac{1}{2}a_2j^2+\tfrac{1}{2}a_3t'^2+\tfrac{1}{2}a_4P'^2+\tfrac{1}{2}a_5M'^2,\\ s'&=b_1q\cdot t'+b_2j\cdot t'+b_3P'\cdot t'+b_4M'\cdot t',\end{aligned} \tag{25}$$

and for an elastic solid one uses tensors on the three-manifold $\mathscr{M}$ to write similar expressions

$$\begin{aligned}\eta'&=\tfrac{1}{2}A_1\mathbf{q}^2+\tfrac{1}{2}A_2\mathbf{j}^2+\tfrac{1}{2}A_3\tilde{\mathbf{t}}'^2+\tfrac{1}{2}A_4\mathbf{E}'^2+\tfrac{1}{2}A_5\mathbf{B}'^2,\\ s'&=B_1\mathbf{q}\cdot\mathbf{t}'+B_2\mathbf{j}\cdot\tilde{\mathbf{t}}'+B_3\mathbf{E}'\cdot\tilde{\mathbf{t}}'+B_4\mathbf{B}'\cdot\tilde{\mathbf{t}}'.\end{aligned} \tag{26}$$

The particular form of η' in (25) and (26) comes from the fact that the entropy density has to be maximum at equilibrium. As all the dissipative quantities are spacelike this condition is realized provided that the a_i's and A_i's are positive.[19] All the coefficients a_i, b_i, A_i, B_i are scalar value functions of the set of thermodynamical variables which define ϵ or ψ, and their space-time derivatives are negligible before the ones of the dissipative quantities.

Let us first consider the case of an electromagnetic fluid. By using (18) and (25) in (24), the entropy inequality may be written

$$t'\cdot X_{t'}+P'\cdot X_{p'}+M'\cdot X_{M'}+q\cdot X_q+j\cdot X_j>0,$$

where the two-tensor has been defined as

$$X_{t'}=d+\rho a_3t'+b_1\nabla q+b_2\nabla j+b_3\nabla P'+b_4\nabla M', \tag{27}$$

and the vectors as

$$\begin{aligned}X_{p'}&=[E]^{\cdot}+\rho a_4\dot{P}'+b_3\operatorname{div}t',\\ X_{M'}&=[B]^{\cdot}+\rho a_5\dot{M}'+b_4\operatorname{div}t',\\ X_q&=-\theta^*/\theta+\rho a_1\dot{q}+b_1\operatorname{div}t',\\ X_j&=E+\rho a_2\dot{j}+b_2\operatorname{div}t'.\end{aligned} \tag{28}$$

It is convenient to define the 12-components vectors $\mathfrak{X}=(X_{p'}, X_{M'}, X_q, X_j)$ and $\mathfrak{J}=(P', M', q, j)$ as also their scalar product

$$\mathfrak{J}\cdot\mathfrak{X}=p'\cdot X_{p'}+M'\cdot X_{M'}+q\cdot X_q+j\cdot X_j.$$

The vectors $\mathfrak{X}$ and $\mathfrak{J}$ generalize the thermodynamical forces and fluxes of the Onsager relations. The entropy inequality then takes the simple form

$$t'\cdot X_{t'}+\mathfrak{J}\cdot\mathfrak{X}>0,$$

and it will be satisfied provided that there exist two positive quadratic forms $\mathscr{A}$ and $\mathscr{B}$ acting, respectively, on the spacetime two-tensors and on the 12-components vectors, and such that

$$t'=\mathscr{A}\cdot X_{t'},\quad \mathfrak{J}=\mathscr{B}\cdot\mathfrak{X}. \tag{29}$$

Here we have used the same notation to represent the operators associated with the quadratic forms $\mathscr{A}$ and $\mathscr{B}$.

Equations (29) represent the transport or constitutive equations for an electromagnetic fluid when submitted to a transient phenomena. The first of these equations describe the stresses and the second ones the vectorlike dissipative effects (heat flux, electric conduction, polarization). From the definitions of $X_{t'}$ and $\mathscr{X}$ it appears that the transport equations will contain the time derivatives of the dissipative terms. This will precisely correspond to the relaxation terms thanks to which the propagation equations will become hyperbolic. It can be seen that the presence of these terms follows directly from the hypotheses concerning the introduction of the perturbative terms η' and s' in the entropy current.

Equations (27)–(29) also make clear the existence of couplings between the dissipative effects. The transport equations (29) constitute a set of coupled partial differential equations which presents a great complexity. Happily, in most applications, all the dissipative effects do not have the same importance and some of them may be neglected. Let us now consider some particular cases. In a dielectric fluid, there is no electric conduction and the magnetic effects may be neglected. The transport equations are then

$$\begin{aligned}&t'+\mathscr{A}_1\cdot\dot{t}'=\mathscr{A}_2\cdot d+\mathscr{A}_3\cdot\dot{q}+\mathscr{A}_4\cdot\nabla P',\\ &p+\chi_1\cdot\dot{P}=\chi_2\cdot E+\chi_3\cdot\theta^*+\chi_4\cdot\dot{q}+\chi_3\cdot\dot{E},\\ &q+\mathscr{K}_1\cdot\dot{q}=\mathscr{K}_2\cdot\theta^*+\mathscr{K}_3\cdot E+\mathscr{K}_4\cdot\dot{P}+\mathscr{K}_5\cdot\dot{E},\end{aligned}$$

where the $\mathscr{A}_i$'s are four-tensors and where the χ_i's and $\mathscr{K}_i$'s are two-tensors.

In the usual scheme of perfect magnetohydrodynamics the electric conduction is infinite, $E=0$, and the electric current j does not play any role. Because of the coupling between the different transport phenomena the influence of j now cannot be neglected even though the electric field and polarization are still negligible. We then have

$$\begin{aligned}&t'+\mathscr{A}_1\cdot\dot{t}'=\mathscr{A}_2\cdot d+\mathscr{A}_3\cdot\nabla q+\mathscr{A}_4\cdot\nabla j+\mathscr{A}_5\cdot\nabla M',\\ &M+\bar{\chi}_1\cdot\dot{M}=\bar{\chi}_2\cdot B+\bar{\chi}_3\cdot\theta^*+\bar{\chi}_4\cdot\dot{q}+\bar{\chi}_5\cdot\dot{j}+\bar{\chi}_6\cdot\dot{B},\\ &q+\mathscr{K}_1\cdot\dot{q}=\mathscr{K}_2\cdot\theta^*+\mathscr{K}_3\cdot\dot{j}+\mathscr{K}_4\cdot B\\ &\qquad+\mathscr{K}_5\cdot\dot{M}+\mathscr{K}_6\cdot\dot{B},\\ &j+\lambda_1\cdot\dot{j}=\lambda_2\cdot\theta^*+\lambda_3\cdot B+\lambda_4\cdot\dot{M}+\lambda_5\cdot\dot{B}.\end{aligned}$$

One may also easily verify that one recovers the equations of Israel and Stewart,[9] for a noncharged fluid,

$$t' + \mathscr{A}_1 \cdot \dot{t}' = \mathscr{A}_2 \cdot d + \mathscr{A}_3 \cdot \nabla q, \quad q + \mathscr{K}_1 \cdot \dot{q} = \mathscr{K}_2 \cdot \theta^*,$$

with $t' = t + p\gamma$, p being the pressure.

The determination of the transport equations for an elastic solid is achieved in a similar way, except that (1) we must use tensors on $\mathscr{M}$ in order to obey the axiom of rheological invariance, and (2) the electromagnetic variables are $\tilde{P}$ and $\tilde{M}$ instead of E and B. The transport equations are obtained from (21) and (26). They are analogous to the Eqs. (27)–(29) with the following substitutions: $a_i \to A_i$, $b_i \to B_i$, $t' \to \tilde{\mathbf{t}}'$, $P' \to \mathbf{E}'$, $M' \to \mathbf{B}'$, and $\dot{} \to \partial / \partial s$. These equations represent the transport equations on the three-manifold $\mathscr{M}$. Their image under the mapping Φ_s gives their space-time counterparts. By the way, the partial derivative $\partial / \partial s$ becomes the convective derivative $[\,]^{\cdot}$, as noticed in Sec. II. This is in agreement with the axiom of rheological invariance. Let us now consider the following examples which, as the perturbation is infinitesimal, may be treated in the linear approximation.

In the case of a neutral rigid body there remains the single equation of heat conduction

$$q + \mathscr{K}_1 \cdot [q]^{\cdot} = \mathscr{K}_2 \cdot \theta^*,$$

which has been already proposed by Maugin[3] and the author.[20] If the medium also possesses elastic properties then one has

$$t + \mathscr{A}_1 \cdot [t]^{\cdot} = \mathscr{C} \cdot e + \mathscr{A} \dot{\theta}_2 \cdot d + \mathscr{A}_3 \cdot \nabla q + \alpha_1 (\theta - \theta_0) + \alpha_2 \theta,$$

$$q + \mathscr{K}_1 \cdot [q]^{\cdot} = \mathscr{K}_2 \cdot \theta^* + \mathscr{K}_3 \operatorname{div} t,$$

where e is the strain tensor.

In magnetoelasticity, the electric polarization may be neglected and, if there is neither heat nor electric conduction, the transport equations are

$$t + \mathscr{A}_1 \cdot [t]^{\cdot} = \mathscr{C} \cdot e + \mathscr{A}_2 \cdot d + \mathscr{A}_4 \cdot \nabla M',$$

$$\tilde{M} + \bar{\chi}_1 \cdot [\tilde{M}]^{\cdot} = \bar{\chi}_2 \cdot B + \bar{\chi}_3 \cdot [B]^{\cdot} + \bar{\chi}_4 \cdot \operatorname{div} (t - \mathscr{C} \cdot e).$$

Similar transport equations may be obtained for a piezoelectric body (centrosymmetric anisotropic solids), by making the substitutions $M \to P$ and $B \to E$. Finally if one considers a nonconducting ($j = q = 0$) rigid body the only transport equations will concern the electromagnetic actions

$$\tilde{P} + \chi_1 \cdot [\tilde{P}]^{\cdot} + \chi_2 \cdot [\tilde{M}]^{\cdot} = \chi_3 \cdot E + \chi_4 \cdot B + \chi_5 \cdot [E]^{\cdot} + \chi_6 \cdot [B]^{\cdot},$$

$$\tilde{M} + \bar{\chi}_1 \cdot [\tilde{P}]^{\cdot} + \bar{\chi}_2 \cdot [\tilde{M}]^{\cdot} = \bar{\chi}_3 \cdot E + \bar{\chi}_4 \cdot B + \chi_5 \cdot [E]^{\cdot} + \chi_6 \cdot [B]^{\cdot}.$$

All the tensor coefficients $A_i, \chi_i, \ldots$, which appear in the foregoing transport equations, depend on the set of thermodynamical variables which characterize the reversible process tangent to the transient phenomenon under consideration. The values of their components cannot be deduced from the phenomenological approach that is followed in this paper, and they will have to be given by experiments. Finally, the propagation velocities will be determined by studying the characteristic surfaces of the set of the 18 first-order partial differential equations (29), supplemented by the five balance equations

$$\operatorname{div} N = 0, \quad \operatorname{div} T = f,$$

and also by the Maxwell equations.

[1] C. Eckart, Phys. **58**, 919 (1940).
[2] L. Lindblom and W. A. Hiscock, Astrophys. J. **267**, 383 (1983).
[3] G. A. Maugin, J. Phys. A **7**, 465 (1974).
[4] F. Bampi and A. Morro, J. Math. Phys. **21**, 1201 (1980); Phys. Lett. A **79**, 156 (1980).
[5] B. Carter, presented at Journées Relativistes, Bruxelles, 1976 (unpublished).
[6] I. Müller, Z. Phys. **198**, 329 (1967).
[7] W. Israel, Ann. Phys. (N.Y.) **100**, 310 (1976).
[8] W. A. Hiscock and L. Lindblom, Ann. Phys. (N.Y.) **151**, 466 (1983).
[9] W. Israel and J. M. Stewart, Ann. Phys. (N.Y.) **118**, 341 (1979).
[10] *Non-equilibrium Relativistic Kinetic Theory*, edited by J. M. Stewart (Springer, Berlin, 1971); Proc. R. Soc. London Ser. A **357**, 59 (1977).
[11] G. Lebon, D. Jou, and J. Casas-Vazquez, J. Phys. A **13**, 275 (1980).
[12] C. Truesdell and W. Noll, *Handbuch der Physik* (Flügge, Berlin, 1965), Vol. 3, 3rd ed.
[13] G. A. Maugin, J. Math. Phys. **19**, 1198, 1206, 1212, 1220 (1978).
[14] W. Israel and J. M. Stewart, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), p. 491.
[15] J. C. Olroyd, Proc. R. Soc. London Ser. A **316**, 1 (1970); B. Carter and H. Quintana, ibid. **331**, 57 (1972).
[16] *Foundations in Electrodynamics*, edited by S. R. de Groot and L. G. Suttorp (North-Holland, Amsterdam, 1972).
[17] See Ref. 13, last paper of the series.
[18] Following the results obtained in a microscopical approach of fluid mechanics we only consider here a modification of the entropy. Some authors[11] have also included the dissipative effects in the internal energy and an extension of these modifications to all the thermodynamical potentials, such as the free energy, could *a priori* be investigated. For instance, in polymers, quadratic terms in the fluxes appear in the expansion of the free energy from its equilibrium value.
[19] Equations (25) and (26) require some isotropy which is not necessarily the case for solids. Furthermore the behavior of all the terms written under time reversal has to be considered in order that the entropy production always has the correct sign. We are indebted to the referee for these comments.
[20] C. Barrabes, Nuovo Cimento B **28**, 377 (1975).

Simplicial minisuperspace I. General discussion

James B. Hartle
Department of Physics, University of California, Santa Barbara, California 93106

(Received 13 September 1984; accepted for publication 2 November 1984)

The use of the simplicial methods of the Regge calculus to construct a minisuperspace for quantum gravity and approximately evaluate the wave function of the state of minimum excitation is discussed.

I. INTRODUCTION

In the search for a conceptually clear and computationally manageable quantum theory of gravity, the sum-over-histories formulation of quantum mechanics has proved to be a powerful tool. For the investigation of conceptual issues, this formulation provides a direct route from classical action to quantum transition amplitude, and in a way which is easily accessible to formal manipulation.[1] Furthermore, the basic elements of the theory are often most clearly formulated in terms of the functional integrals which implement the sum-over-histories formulation. For example, a natural prescription for the wave function describing a closed cosmology in its state of minimum excitation is to take[2–4]

$$\Psi_0[\text{three-geometry}] = \sum_{\text{four-geometries},g} \exp(-I[g]/\hbar). \tag{1.1}$$

Here I is the Euclidean gravitational action including cosmological constant and the sum is over all compact Euclidean four-geometries which have the three-geometry as a boundary. It has been conjectured that this *is* the wave function of our universe.[3,5,6]

To investigate the consequences of a proposal like (1.1) or other consequences of a theory formulated in terms of integrals over four-geometries, one needs to evaluate these sums approximately. One can construct approximations to the sum by singling out a family of geometries described by only a few parameters or functions and carrying out the sum only over the geometries in this family. As the family of geometries is made larger and larger, one can hope to get a better and better approximation to the functional integral.

A restriction on the four-geometries which are included in the sum (1.1) will imply a restriction on the three-geometries which can occur as arguments of the resulting wave function. This is because the three-geometries must be embeddable in the four-geometries. The restriction thus reduces the configuration space on which the wave function is defined from the superspace of all three-geometries to a smaller space of three-geometries—a minisuperspace.[7] For this reason we call such approximations to the functional integral "minisuperspace approximations."

One way of constructing a minisuperspace approximation is to restrict the family of four-geometries to be only those with some symmetry. For example, one might restrict the four-geometries in (1.1) to have four-sphere topology and three-sphere symmetry and the three-geometries to be three-spheres of radius a_0. A single function of a single variable—radius as a function of polar angle—then describes these four-geometries. The sum over geometries reduces to a functional integral over this function. The wave function, which generally is a *functional* of the six components of the three-metric, reduces to a *function* of a single variable a_0. The minisuperspace is just half the real line.

Minisuperspace models based on symmetry have been used to explore quantum cosmology in the canonical theory of quantum gravity.[7,8] Minisuperspace approximations based on symmetry have been applied to the computation of the ground state wave function (1.1) in the functional integral formulation.[2,3,5,6] Minisuperspace approximations based on symmetries are simple to implement and generally easy to interpret. They do not, however, offer the possibility of systematic improvement because a general four-geometry is not well approximated by a symmetric one. It is therefore of interest to consider minisuperspace approximations which are not based on symmetries. The Regge calculus[9] provides an avenue to such approximations.

A general two-surface may be approximated by a surface made up of a net of flat triangles. The net of triangles is itself a two-geometry whose curvature is concentrated at the vertices where the triangles meet. The geometry of the surface is specified by the edge lengths of the triangles. Analogously, a general four-geometry may be approximated by a net of flat four-simplices. This net is a four-geometry specified by the edge lengths of the simplices. Its curvature is concentrated on the triangles in which these four-simplices intersect. Any geometrical quantity such as the curvature or the action can be expressed in terms of the edge lengths of the net. The conditions for the action of general relativity to be an extremum, with respect to variations of the edge lengths, give the simplicial analogs of Einstein's equations.

The simplicial methods adumbrated above were introduced into general relativity in a seminal paper by Regge.[9] These methods are usually called the "Regge calculus." They have been used in a number of interesting investigations in the classical theory of gravity. (See, for example, Refs. 10–19.) As Regge calculus is the natural lattice version of general relativity, it has also been extensively applied to the investigation of quantum theories of gravity (see, in particular, Refs. 20–26).

Simplicial approximation is a natural starting point for constructing a minisuperspace approximation to the functional integrals of quantum gravity. In this approximation one obtains the family of geometries integrated over in two steps. First fix a simplicial net, that is, specify the vertices of the net and the combinations of them that make up the one-simplices (edges), two-simplices (triangles), three-simplices (tetrahedra), and four-simplices. Second, assign lengths to

 0022-2488/85/040804-11$02.50

the edges and allow these lengths to range over values consistent with their making up the flat simplices of the net. There results a family of four-geometries parametrized by the n_1 edge lengths of the net. Suppose that m_1 of these edges lie in the boundary of the net. These edges define a simplicial three-geometry. The minisuperspace is that portion of $\mathbb{R}^{m_1}$ swept out as the m_1 squared edge lengths of the boundary range over values for which the simplicial inequalities for triangles and tetrahedra are satisfied. The functional integral (1.1) is approximated by an $(n_1 - m_1)$-dimensional multiple integral over the interior edge lengths. Schematically it has the form

$$\Psi_0(s_i, i \in \partial\Sigma_1) = \int_C d\Sigma_1 \exp\left[\frac{-I(s_i)}{\hbar}\right]. \tag{1.2}$$

Here, $\partial\Sigma_1$ denotes the edge lengths of the boundary, I is the Regge action, and $d\Sigma_1$ denotes an integration over the interior edge lengths on a contour C.

Simplicial minisuperspace approximations have a number of significant advantages over those constructed from symmetry.

(1) To represent the sum over the four-metrics on a given manifold there is a different simplicial minisuperspace approximation for each triangulation of the manifold. This is a much larger class than can be generated by symmetry. In particular, as the number of vertices n_0 is increased one expects an arbitrary four-geometry to become closely approximated by some simplicial geometry. Thus the simplicial minisuperspace approximations in principle permit an investigation of the continuum limit.

(2) The simplicial minisuperspace approximation leads directly to a numerical evaluation of the functional integral as a multiple integral. Approximations based on symmetry, by contrast, generally require a further discretization for explicit evaluation.

(3) The simplicial minisuperspace approximation allows a simple and direct discussion of the role topology may play in quantum gravity. Topological information is contained in an elementary way in the simplicial net. The simplicial approximation allows one to investigate different topologies efficiently with simple geometries by investigating different simplicial nets with small numbers of edge lengths. In general, the simplicial minisuperspace approximation permits the investigation of global questions with crude geometries. Moreover, it does this in a way which is accessible to systematic improvement of the approximation. At a time in the development of quantum gravity when qualitative results are often more instructive than precise quantitative calculations, this is an important advantage.

In this paper we shall begin an investigation of the use of simplicial minisuperspace to approximately evaluate the state of minimum excitation in quantum gravity constructed according to the prescription (1.1). The methods we shall discuss are certainly applicable to computations of other quantities in the theory,[27] but we shall focus on this one to obtain concreteness and because of its important role in the theory.

In Sec. II we shall discuss the minisuperspace approximation in greater detail and in particular the form of the action, the issues involved in the choice of the measure, and those issues connected with the choice of integration contour C. The implementation of the Regge calculus to evaluate the action in (1.2) requires a certain amount of algebraic technology. We shall collect and describe the necessary technology in Sec. III. In Sec. IV we shall discuss the semiclassical approximation to the integral in (1.2) and in Sec. V we shall describe how the diffeomorphism group of general relativity should be recovered in the limit of larger and larger simplicial nets. Section VI describes how sums over different topologies might be implemented in the simplicial minisuperspace approximation.

II. THE FUNCTIONAL INTEGRAL FOR THE WAVE FUNCTION

The Euclidean functional integral prescription for the wave function of a closed cosmology in its state of minimum excitation assigns an amplitude Ψ_0 to each possible compact three-geometry. The general compact three-geometry consists of disconnected compact connected three-manifolds $\partial M_1,\ldots, \partial M_n$ each without boundary, each perhaps with nontrivial topology, and three-metrics $h_1, h_2,\ldots,h_n$ on these pieces. The wave function Ψ_0 is a functional of these metrics, given by[2]

$$\Psi_0[h_1,\partial M_1;h_2,\partial M_2;\ldots;h_n,\partial M_n] = \sum_M \nu(M)\int_C \delta g \exp(-I[g,M]). \tag{2.1}$$

The sum is over a class of compact four-manifolds, each with boundary ∂M consisting of the pieces $\partial M_1,\ldots,\partial M_n$ and each contributing with weight $\nu(M)$. The functional integral is over physically distinct metrics on M which induce the metrics $h_1,\ldots,h_n$ on $\partial M_1,\ldots, \partial M_n$. I is the action for general relativity

$$l^2 I[g,M] = -2\int_{\partial M} d^3x\, h^{1/2}K - \int_M d^4x\, g^{1/2}(R-2\Lambda). \tag{2.2}$$

Here and for the rest of the paper we use units where $\hbar = c = 1$. Thus, $l = (16\pi G)^{1/2}$ is the Planck length. To complete the prescription, four further specifications are needed: the class of manifolds summed over, the weight $\nu(M)$ to be given each one, the measure on the space of metrics, and the contour of integration in the space of metrics.

If the transition amplitudes of quantum gravity can be constructed as integrals over Euclidean four-geometries on different manifolds, then the weight given each manifold and the measure on the space of metrics must be consistent with the composition law for quantum amplitudes: $\langle a|b\rangle = \Sigma_c \langle a|c\rangle\langle c|b\rangle$. For example, if one admits disconnected three-geometries, then one expects multiply connected four-geometries will be required for consistency. The composition of an amplitude with two disconnected three-geometries in its final state together with an amplitude with two disconnected three-geometries in its initial state would be represented by a sum over a multiply connected four-geometry. The measure must also be consistent with the composition laws. It may be that the class of manifolds, ν, and the measure are determined by these restrictions and in the case of the measure there are calculations to this effect.[28] The contour C must be chosen so that the integral correctly represents a complete sum over compact geometries and so that the integral is convergent.

To begin a discussion of these specifications in a simplicial approximation to (2.1), let us restrict attention until Sec. VI to the sum over metrics on a fixed manifold M with a single boundary. Suppressing the labels M and ∂M we write

$$\Psi_0[h] = \int \delta g \exp(-I[g]). \tag{2.3}$$

The simplicial approximation to (2.3) as described in Sec. I is

$$\Psi_0(s_i, i\in\partial\Sigma_1) = \int_C d\Sigma_1 \exp[-I(s_i)]. \tag{2.4}$$

Here, we are considering a specific triangulation of M with vertices Σ_0, edges Σ_1, triangles Σ_2, tetrahedra Σ_3, and four-simplices Σ_4. The simplices of the boundary and the interior we denote by $\partial\Sigma_\alpha$ and int Σ_α, respectively.

The action I is the Regge action[9] modified by the boundary term required by the composition law for quantum amplitudes and the classical limit[29]

$$I = -2\sum_{\sigma\in\partial\Sigma_2} A(\sigma)\psi(\sigma) - 2\sum_{\sigma\in\mathrm{int}\,\Sigma_2} A(\sigma)\theta(\sigma) + 2\Lambda\sum_{\tau\in\Sigma_4} V_4(\tau), \tag{2.5}$$

where $A(\sigma) = V_2(\sigma)$ is the area of triangle σ and $V_4(\tau)$ is the volume of the four-simplex τ. The angle $\theta(\sigma)$ is the deficit angle for triangle σ. It is given by

$$\theta(\sigma) = 2\pi - \sum_\tau \theta(\sigma,\tau), \tag{2.6}$$

where the sum is over all the four-simplices which contain σ. $\theta(\sigma,\tau)$ is the dihedral angle between the two tetrahedra of τ which have σ as a common face. [For a two-dimensional picture see Fig. 1; for more details on how to compute $A(\sigma)$, $V_4(\tau)$, and $\theta(\sigma,\tau)$ see Sec. III.] The angle $\psi(\sigma)$ necessary for calculating the boundary term is

$$\psi(\sigma) = \pi - \sum_\tau \theta(\sigma,\tau), \tag{2.7}$$

where the sum is over all four-simplices which intersect the boundary triangle σ. The surface term in (2.5) is just that

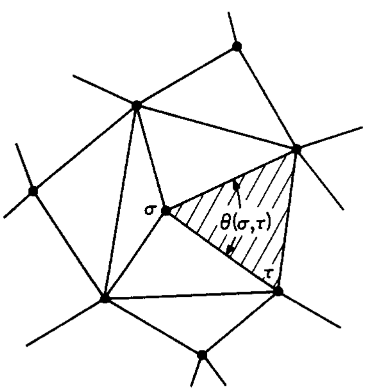


FIG. 1. A two-dimensional simplicial geometry is a net of flat triangles together with an assignment of lengths to their edges. The geometry includes both the interior points of the triangle as well as their edges. The distance between any two points can be determined in terms of the edge lengths. The curvature is concentrated at the vertices and is measured by the deficit angle. The deficit angle at a vertex σ is the 2π minus the sum of the dihedral angles $\theta(\sigma,\tau)$ over all triangles τ which have σ as a vertex.

necessary to ensure that the conditions for the extremum of the action correspond to the Regge equations

$$\sum_{\alpha\in\mathrm{int}\,\Sigma_2} \theta(\sigma)\frac{\partial A(\sigma)}{\partial s_i} = \Lambda\sum_{\tau\in\Sigma_4}\frac{\partial V_4(\tau)}{\partial s_i}, \tag{2.8}$$

which are the simplicial analogs of Einstein's equation.[9]

There are as yet no completely satisfactory arguments for the measure and the contour needed to complete the specification of the integral (2.4). For the measure we shall assume that the integration over edge lengths is restricted to values such that they define possible flat simplices. Thus the appropriate triangle, tetrahedral, and four-simplex inequalities are satisfied. Necessary and sufficient conditions for this are that the squared volumes of all simplices be positive or zero when expressed in terms of the squared edge lengths. To go further we can only proceed by analogy with the continuum case. There a number of possible measures have been put forward. DeWitt,[30] for example, has argued that an appropriate measure is

$$\prod_x \prod_{\mu<\nu} dg_{\mu\nu}(x). \tag{2.9}$$

Since the squares of edge lengths are linearly related to $g_{\mu\nu}$, the corresponding choice in the simplicial approximation is to take

$$d\Sigma_1 = \mu(s_i)\prod_{j\in\mathrm{int}\,\Sigma_1} ds_j, \tag{2.10}$$

where

$$\mu = \begin{cases} 1, & \text{with simplicial inequalities satisfied,} \\ 0, & \text{otherwise.} \end{cases} \tag{2.11}$$

In particular, this means that the measure does not vanish on zero-volume simplices.

The remaining specification in (2.4) is the contour C. At a minimum this must be chosen so that the integral (2.4) is convergent. For real edge lengths the action is no more positive in the simplicial approximation than it is in the continuum theory. For example, the action for the closed four-geometry of volume V_4, which is the surface of a five-simplex with all edges equal, is (see Ref. 26 or Paper II)

$$I(V_4) = -107.9(V_4/l^4)^{1/2} + 2\Lambda V_4/l^2. \tag{2.12}$$

For small V_4 and at the extremizing volume, this is negative. In the continuum theory the action can always be made negative by an appropriate conformal deformation of the metric.[31] The same is true in the simplicial approximation.[32] It seems likely that the action can be made arbitrarily negative by considering nets with near-zero four-volume but containing very-large-area triangles whose deficit angles are positive. If this is the case there is the presumption that the integrals in (2.4) will diverge in some large edge-length directions.

In the continuum theory of asymptotically flat spacetimes the functional integral can be made convergent by decomposing the integration over four-geometries into an integration over a conformal factor and one over conformal equivalence classes, and then rotating the conformal factor integration contour to complex values.[28] The same procedure applied in the linear theory of gravity does yield the correct ground state wave function[33] and arises naturally from the parametrization of the Hamiltonian path integral for that theory expressed in terms of the physical degrees of freedom.[34] One can exhibit analogous contours in the simplicial approximation along which the integral is convergent. In the absence of a compelling argument for one or the other, we shall not pursue their discussion further.

The information contained in the wave function Ψ_0 might be determined by carrying out the multiple integral (2.4) for a sampling of the space of edge lengths. In practice, this will be too much data to deal with since the dimension of the space of edge lengths can be very large. Equivalently and more usefully the information in Ψ_0 can be summarized by computing interesting expectation values[35] of Ψ_0:

$$\langle A \rangle = \frac{\int d(\partial\Sigma_1)\Psi_0(s_i)A(s_i)\Psi_0(s_i)}{\int d(\partial\Sigma_1)\Psi_0(s_i)\Psi_0(s_i)}, \tag{2.13}$$

where $d(\partial\Sigma_1)$ is the volume element on the boundary edge lengths analogous to (2.10). Certainly the information contained in Ψ_0 can be extracted from (2.13) by letting A range over appropriate filters sensitive to particular regions of edge lengths. Indeed, one imagines that the interesting physical questions can always be phrased in terms of the expectation value of an appropriate A. Monte Carlo numerical calculations will generally be much more feasible for expectation values of slowly varying A's than for the wave function itself. For these reasons we shall for the most part be concerned with calculating expectation values in what follows.

Assuming that the choice of measures in $d\Sigma_1$ and $d(\partial\Sigma_1)$ is compatible, Eq. (2.4) may be inserted into (2.13) to give the following expression for $\langle A \rangle$:

$$\langle A \rangle = \frac{\int_C d\Sigma_1 A(s_i)\exp[-I(s_i)]}{\int_C d\Sigma_1 \exp[-I(s_i)]}, \tag{2.14}$$

where the integral is now over the squared edge lengths of the compact, boundaryless, manifold formed by identifying M and a copy of itself at its boundary. At the risk of some confusion we shall also call this M in the following. The boundary term in the action may now be dropped so that

$$l^2 I = -(\mathscr{R} - 2\Lambda\mathscr{V}), \tag{2.15}$$

where $\mathscr{R}$ is the "curvature part" of the action

$$\mathscr{R} = 2\sum_{\sigma\in\Sigma_2} A(\sigma)\theta(\sigma), \tag{2.16}$$

and $\mathscr{V}$ is the total four-volume

$$\mathscr{V} = \sum_{\sigma\in\Sigma_4} V_4(\sigma). \tag{2.17}$$

Here, Σ_1, Σ_2, etc., now refer to the whole compact manifold. If convenient, the contour C may be further distorted so that the s_i on the joining boundary assume complex values.

We have introduced the question of the convergence of the integral (2.4) at large edge lengths. A natural complementary question is its behavior at small edge lengths. In the form (2.13), when the contour runs along real edge lengths, a simple answer to this question can be given. The volume part of the action (2.5) is evidently well behaved at small edge lengths. The curvature part may be bounded as follows. In a flat four-simplex the dihedral angle $\theta(\sigma,\tau)$ always lies between 0 and π. From (2.6) it follows that the deficit angle $\theta(\sigma)$ is bounded by

$$-\pi[k_4(\sigma) - 2] < \theta(\sigma) < 2\pi, \tag{2.18}$$

where $k_4(\sigma)$ is the number of four-simplices containing the triangle σ. Thus for the curvature action we find

$$-4\pi\mathscr{A} \leqslant -\mathscr{R} \leqslant 2\pi K\mathscr{A}, \tag{2.19}$$

where

$$\mathscr{A} = \sum_{\sigma\in\Sigma_2} A(\sigma), \tag{2.20}$$

and

$$K = \max_{\sigma\in\Sigma_2}[k_4(\sigma) - 2]. \tag{2.21}$$

For any particular complex, K is a finite number (for example, for the four-complex which is composed of the faces of a five-simplex, $K = 1$). The curvature action is thus bounded above and below by a multiple of $\mathscr{A}$, and is well behaved as any of the edge lengths go to zero, or at zero-volume simplices where the simplicial inequalities are saturated. This suggests, in particular, that, though one may recover short-distance (ultraviolet) divergences in the continuum limit, they are not present in any finite simplicial approximation.

III. PRACTICAL REGGE CALCULUS

To implement a computation of a functional integral in the simplicial minisuperspace approximation as described in the preceding section, the action must be expressed in terms of the squared edge lengths of the simplicial net. In turn this means that one must be able to express areas, deficit angles, and four-volumes in terms of these squared edge lengths. Further, to evaluate the simplicial analog of the field equations one needs the derivative of areas and four-volumes, with respect to squared edge length. In this section we shall set out the formulas the author has found most useful for these calculations and briefly describe a method for deriving them. By and large both these formulas and the method of derivation have appeared elsewhere in the literature on the Regge calculus[36] and one imagines that they could be tracked down in the older mathematical literature. We collect them here with their derivation in order to have a complete description of the tools with which to attack our problem.

An n-simplex is specified by giving its $n+1$ vertices $(0,1,\ldots,n)$ in flat space. Define the n vectors e_i which start with the vertex 0 and proceed to the vertex i. The vectors $e_1,\ldots,e_n$ span the n-simplex. The volume n-form associated with the n-simplex may be defined as

$$\omega_n = e_1 \wedge e_2 \wedge \cdots \wedge e_n. \tag{3.1}$$

The formulas for volumes, areas, angles, etc., become simple to express and simple to derive when these n-forms are manipulated like vectors. To this end, we introduce a scalar product between two n-forms by the definition

$$f\cdot g = (1/n!) f_{\alpha_1\cdots\alpha_n} g^{\alpha_1\cdots\alpha_n}. \tag{3.2}$$

Consider, for example, the squared volume of an n-simplex, V_n^2. The product $\omega_n\cdot\omega_n$ must be proportional to V_n^2 because, up to sign, ω_n is easily shown to be independent of the choice of the perferred vertex 0 and there is no other symmetrical invariant with the correct dimension. By evaluating the constant of proportionality in any special case, one has

$$V_n^2 = \omega_n \cdot \omega_n. \tag{3.3}$$

A more direct derivation may be provided as follows: Consider the n-simplex as made up of the $(n$-$1)$-simplex $(0,1,...,n-1)$ (the "base") and the vertex n. Evidently

$$\omega_n = \omega_{n-1} \wedge e_n. \tag{3.4}$$

Divide e_n into a part $e_n^\perp$ perpendicular to the base simplex and a part which lies in it. Insert this decomposition into (3.4) and this in turn in (3.2) to compute $\omega_n \cdot \omega_n$. One finds

$$\omega_n \cdot \omega_n = (1/n^2)(e_n^\perp \cdot e_n^\perp)(\omega_{n-1} \cdot \omega_{n-1}), \quad n > 1. \tag{3.5}$$

This formula can be used to prove inductively that $V_n^2 = \omega_n \cdot \omega_n$, because for each n this says $V_n = n^{-1}$ (height) $\times$ (volume of base).

To be computationally effective a formula like (3.3) must be expressed in terms of the squared edge lengths. This can be done by reexpressing the scalar product of two-volume n-forms $\omega_n = e_1 \wedge \cdots \wedge e_n$ and $\omega_n' = e_1' \wedge \cdots \wedge e_n'$ in terms of the scalar products of the constituent vectors as follows:

$$\omega_n \cdot \omega_n' = [1/(n!)^2] \det(e_i \cdot e_j'). \tag{3.6}$$

In particular, for the squared volume one has

$$V_n^2 = [1/(n!)^2] \det(e_i \cdot e_j), \tag{3.7}$$

where the $n \times n$ matrix of scalar products may be expressed in terms of the edge lengths s_{ij} between vertices i and j by

$$e_i \cdot e_j = \tfrac{1}{2}(s_{0i} + s_{0j} - s_{ij}). \tag{3.8}$$

Equations (3.7) and (3.8) express the volumes of the simplices of a net in terms of the squares of the edge lengths. Equations (3.7) and (3.8) do not show all the symmetries of the formula for V_n^2 fully expanded in terms of the edge lengths. This is because their construction involved a preferred vertex 0. Manifestly symmetric formulas can be constructed in terms of $(n+2) \times (n+2)$ bordered determinants (see e.g., Ref. 10) and these can be derived from Eq. (3.7) by a few simple determinantal manipulations. The evaluation of these symmetric formulas, however, involves more operations than does Eq. (3.7), which is therefore preferred for explicit computations.

The deficit angles are the remaining quantities needed to express the action in terms of the squared edge lengths. From Eq. (2.6) the computation of the deficit angle at a given triangle reduces to the computation of the dihedral angles at that triangle of the four-simplices which intersect it. We therefore consider this in detail and in slightly greater generality. Suppose we have an $(n+1)$-simplex which contains two n-simplices intersecting in a common $(n-1)$-simplex (the "hinge"). The dihedral angle θ of the $(n+1)$-simplex at the hinge is the angle between the normal to the hinge which lies in the first n-simplex and the normal to the hinge which lies in the second. It is given by

$$\cos\theta = \omega_n \cdot \omega_n' / (V_n V_n'), \tag{3.9}$$

where ω_n, ω_n' are the volume forms associated with the intersecting n-simplices and V_n, V_n' are the volumes of these simplices (the "lengths" of ω_n and ω_n', respectively). The correct sign for $\cos\theta$ will be obtained if ω_n and ω_n' are constructed as follows: Let ω_{n-1} be the volume form of the hinge. For the appropriate vectors e and e', write

$$\omega_n = \omega_{n-1} \wedge e, \quad \omega_n' = \omega_{n-1} \wedge e'. \tag{3.10}$$

Alternatively, ω_n and ω_n' must be oriented *oppositely* in a consistent orientation for the $(n+1)$-simplex.

Eq. (3.9) may be derived as follows: Decompose e and e' into parts orthogonal to and parallel to the hinge. Insert this decomposition into (3.10) and then into (3.9). One finds the right-hand side of (3.9) is $(e^\perp \cdot e'^\perp)/|e^\perp||e'^\perp|$. This, by definition, is the cosine of the dihedral angle.

Equation (3.9) may be expressed in terms of the edge lengths of the $(n+1)$-simplex through Eq. (3.6). Since each of the vectors e_i and e_i' lies along some edge of the $(n+1)$-simplex, their scalar product may be expressed in terms of the edge lengths through formulas analogous to Eq. (3.8).

A simple formula for $\sin\theta$ may be derived from the identity

$$(n+2)^2\omega^2(\omega \wedge a \wedge b)^2 = (n+1)^2\{(\omega \wedge a)^2(\omega \wedge b)^2 - [(\omega \wedge a)\cdot(\omega \wedge b)]^2\}, \tag{3.11}$$

valid for any n-form ω and one-forms a and b. Here, ω^2 means $\omega \cdot \omega$. Written out with ω_{n-1}, e, and e' as in Eq. (3.10) for ω, a, and b, respectively, decomposing e and e' into components parallel and orthogonal to ω_{n-1}, and using Eq. (3.3), one finds for $\sin\theta$

$$\sin\theta = [(n+1)/n][V_{n-1}V_{n+1}/(V_n V_n')]. \tag{3.12}$$

Here, V_n and V_n' are the volumes of the n-simplices intersecting in the $(n-1)$-hinge, V_{n-1} is the volume of the hinge, and V_{n+1} is the volume of the $(n+1)$-simplex spanned by the two n-simplices. This result for $\sin\theta$ is easily expressed in terms of the edge lengths of the $(n+1)$-simplex by expressing the volumes in terms of the edge lengths. From a computational point of view, however, it is not as useful as (3.9). The dihedral angle θ ranges from 0 to π. From a formula for $\cos\theta$ one can recover the angle itself whereas from a formula for $\sin\theta$ one cannot, and it is the angle which enters in the action.

Equations (3.3) and (3.9) and their expressions in terms of edge lengths are all that are needed to evaluate the action. To evaluate the simplicial field equations [Eq. (2.8)] one also needs as expression for the derivative of the volume of an n-simplex with respect to one of its squared edge lengths, keeping the other edge lengths fixed. One is straightforwardly worked out from Eqs. (3.3) and (3.2) by choosing the preferred vertex 0 so that it is not a vertex of the edge of interest and then considering the variation in the vectors e_i produced by a variation in this edge length. One finds

$$\partial V_n^2/\partial s_{ij} = (1/n^2)\omega_{n-1}\cdot\omega_{n-1}', \tag{3.13}$$

where ω_{n-1} and ω_{n-1}' are the volume forms for the $(n-1)$-simplices formed by the vertices $(0,...,i-1, i+1,...,n)$ and $(0,...,j-1, j+1,...,n)$, respectively. Equations (3.6) and (3.8) express this in terms of the squared edge lengths.

To find the action or field equations for a simplicial net one would proceed as follows: First, the net must be specified. This means specifying the vertices, edges, triangles, tetrahedra, and four-simplices of the net. One way of doing this is to give a list of all of the simplices in terms of their vertices. A second, and equivalent, way is to give the incidence matrices which specify which $(n-1)$-simplices make up an n-simplex. All the topological information is contained in this specification of the net.

With a net in hand one can now proceed to specify the squared edge lengths and calculate the action. Not every assignment of edge lengths is consistent with the simplices having flat interiors. The triangle inequalities and their analogs for tetrahedra and four-simplices must be satisfied. Necessary and sufficient conditions for this are that the squared volumes of all the triangles, tetrahedra, and four-simplices in the net must have a positive squared volume, i.e.,

$$(V_2)^2 \geqslant 0, \quad (V_3)^2 \geqslant 0, \quad (V_4)^2 \geqslant 0, \tag{3.14}$$

for the whole net. To see this in the case of triangles, for example, fix two of the edges and consider $(V_2)^2$ as a function of the remaining squared edge length, s. As s is varied from a value where the triangle inequality is satisfied to one where it is saturated, $(V_2)^2$ ranges from a positive value to zero. Since $(V_2)^2$ is quadratic in s no further regions of positive $(V_2)^2$ exist outside the range where the triangle inequality is satisfied. The generalization to higher simplices is straightforward. The conditions (3.14) are independent. One can find, for example, squared edge lengths for which $(V_3)^2$ is positive but the triangle inequalities are violated.

For an assignment of squared edge length which does satisfy (3.14) one can compute the action and field equations by evaluating the necessary volumes and their derivitives using Eqs. (3.3), (3.9), and (3.13) and expressing these relations in terms of the edge lengths via Eqs. (3.7) and (3.8). From the specification of the net one can compute which four-simplices are incident on a given triangle. Their dihedral angles at the triangle may be found from (3.9) and the deficit angle from the sum in (2.6). By doing this for all triangles and performing the sums in (2.5) and (2.8), one arrives at the action and field equations for the net.

IV. THE SEMICLASSICAL APPROXIMATION

Considerable insight into the qualitative behavior of the ground state wave function and its expectation values may be obtained by evaluating these quantities in the semiclassical approximation. In the simplicial approximation this means carrying out the defining multiple integral (2.4) by the method of steepest descents.

To apply the method of steepest descents, one first locates the stationary points of the action in the space of complex edge lengths by solving the simplicial field equations (2.8). One then attempts to distort the contour of integration in (2.4) so that it runs through one or more of these stationary configurations and elsewhere follows a contour along which $|\exp(-I)|$ decreases as rapidly as possible away from these stationary configurations. The asymptotic behavior of Ψ_0 as $\hbar \to 0$ is then given by the integral (2.4) in the neighborhood of one or more of the stationary configurations or in the neighborhood of the boundaries of the contour. It may or may not be possible to distort the contour C to pass through any particular stationary point. The stationary point of smallest $\operatorname{Re} I$ therefore does not always give the semiclassical behavior of the wave function.[37] Even if the dominant stationary point can be identified, one should still check whether it or the behavior near a boundary of the contour dominates the integral. In the present case the contour has boundaries because of the simplicial inequalities (see Sec. III). Because of the necessity of a classical limit, however, it is a reasonable expectation that the semiclassical approximation will be given by one or more stationary configurations. If the stationary configurations have complex edge lengths they will contribute to the semiclassical approximation in complex conjugate pairs since the original integral was real.

To see the form of the semiclassical approximation let us consider for simplicity the case when a single real stationary configuration provides the dominant contribution. Let s_i^0 be the squared edge lengths of the stationary configuration. Evaluate the measure on this configuration, expand the exponent in (2.4) to quadratic order in small deviations of the edge lengths from this configuration, and evaluate the resulting Gaussian integral to find

$$\Psi_0 \approx N\mu(s_i^0)\left[\det\left(\frac{\partial^2 I}{\partial s_i\, \partial s_j}\right)_{s_i = s_i^0}\right]^{-1/2} \exp[-I(s_i^0)], \tag{4.1}$$

for some constant N. The expectation value of a quantity A will be $A(s_i^0)$ in this approximation.

To make the further discussion of possible stationary configurations more concrete let us focus on the integrals over the boundaryless simplicial geometries which define the expectation values in the ground state [Eq. (2.14)]. The continuum theory gives some guide as to when one can expect stationary configurations with real edge lengths. There, the analogous problem is to solve the Euclidean Einstein equation

$$R_{\alpha\beta} = \Lambda g_{\alpha\beta}, \tag{4.2}$$

for real metrics $g_{\alpha\beta}$ on a compact manifold.[38] For positive Λ there is the four-sphere metric on S^4, the product of equal radii two-sphere metrics on $S^2 \times S^2$, and the Fubini-Study metric on CP^2. For the case of $S^1 \times S^3$, T^4, and $K3$ real solutions with either sign of Λ are ruled out by the inequalities[39]

$$\chi > 0, \quad \chi > \tfrac{3}{2}|\tau|, \tag{4.3}$$

which are necessary conditions for solutions to (4.1) with $\Lambda \neq 0$. $S^1 \times S^3$ and T^4 have $\chi = 0$ while $K3$ has $\chi = 24$ and $\tau = -16$.

One would expect the situation regarding the existence of solutions to the Regge equations to be similar to that for their continuum limit. The Regge equations offer the opportunity for approximately addressing questions still open in the continuum case (e.g., the existence of complex solutions) through an analysis of a finite number of algebraic equations.

V. RECOVERY OF THE DIFFEOMORPHISM GROUP

Diffeomorphisms are an invariance of general relativity. On a given manifold M, if two metrics g and g' are diffeomorphic they have the same physical consequences. The Einstein action which summarizes the theory is preserved by diffeomorphisms. This is analogous to the preservation of the action by the gauge group in a gauge theory[40] and has important and well-known consequences for a formulation of the quantum theory in terms of functional integrals. Consider, for example, the expression for the ground-state expectation value of a physical quantity $A[g]$ which is the natural generalization of those in field theories without gauge symmetries

$$\langle A \rangle = \frac{\int_C \delta g A[g] \exp(-I[g])}{\int_C \delta g \exp(-I[g])}. \tag{5.1}$$

By $\int \delta g$ is meant the sum over all metrics on the manifold M in some class C. Two diffeomorphic metrics contribute identically in both numerator and denominator of (5.1) since both A and I are invariant under diffeomorphisms. Each integral is therefore the volume of the diffeomorphism group times a sum over physically distinct metrics. Since the volume of the diffeomorphism group is infinite each integral diverges. The divergent factor of the diffeomorphism group formally cancels between the numerator and denominator of (5.1) to give a finite answer for $\langle A \rangle$, but to implement this cancellation in a practical way all the familiar techniques of gauge fixing and ghosts are required.

Simplicial geometries are simplicial manifolds with a metric. By simplicial manifold we shall mean a piecewise linear manifold[41] made up of simplices. We stress that by the manifold we mean the points interior to the four-simplices and on their boundaries, and not only the vertices of the net. A metric is determined by the edge lengths of the net and a flat metric in the interior of each simplex. With this information the distance between two points on any curve threading the simplicial geometry could be computed.

There are piecewise diffeomorphisms of simplicial manifolds exactly as there are diffeomorphisms in the continuum case. They are the one-to-one invertible maps from a simplicial manifold to itself[42] which are smooth on each simplex. Relabeling the vertices and smooth diffeomorphisms of the interior of simplices are two trivial examples. The action of a piecewise diffeomorphism on a metric gives a new metric which is physically equivalent to the old one. For a general curved simplicial geometry one expects a diffeomorphism to leave the edge lengths unchanged or to change them only according to a trivial relabeling of the vertices. This is because a nontrivial reassignment of edge lengths will in general correspond to different curvatures and a different geometry. The example of flat space, however, shows that there can be cases where diffeomorphisms lead to nontrivial reassignment of edge lengths.[43] Imagine a simplicial net obtained by distributing vertices about flat space, connecting them to form a simplicial net, and assigning edge lengths which are the flat distances between them. By moving the location of the vertices in flat space, one can find a different assignment of edge lengths on the same simplicial net which represents the same flat geometry. There is thus a $4n_0$-parameter family of transformations of edge lengths in flat space which leads to different metrics on the simplicial manifold which are piecewise diffeomorphic.

It is not easy to give an algorithm for deciding when two simplicial metrics are piecewise diffeomorphic any more than it is in the continuum case. Necessary conditions are certainly that any curvature invariant be the same, and in a certain sense these conditions are sufficient as well.[44] Intuitively, it seems reasonable to suppose that different assignments of edge lengths correspond to different geometries except in the case of flat space. In this sense there are no gauge transformations in the Regge calculus.[45]

If nontrivially different assignments of edge lengths correspond to different simplicial geometries we would expect the multiple integrals defining the expectation values in Eq. (2.14) not to diverge as do those in (5.1). Thus in the simplicial approximation no additional gauge-fixing machinery should be needed to effect a sum over geometries. This is a considerable convenience. One would expect, however, to recover the divergence of these integrals associated with the diffeomorphism group in the continuum limit, that is, in the limit of large simplicial nets. In the following we shall describe how this comes about.

A given continuum geometry may be approximated by a simplicial geometry on an appropriate net. Consider a family of nets with an increasingly large number of vertices obtained by repeatedly subdividing the original net. As the number of vertices n_0 becomes large there will be more and more simplicial geometries (i.e., more and more assignments of edge lengths) which approximate the given continuum geometry to a fixed level of accuracy. For large n_0 all these simplicial approximations contribute approximately equally to the multiple integrals in (2.14). As n_0 becomes large, both numerator and denominator will, therefore, be approximately a large factor times a sum over physically distinct geometries. The factor should cancel between the numerator and the denominator. Thus, while the numerator and denominator will diverge as n_0 becomes large, $\langle A \rangle$ should tend to a definite value. This behavior can be illustrated more precisely in the semiclassical approximation.

By way of illustration let us suppose that one stationary configuration with squared edge lengths s_i^0 gives the dominant semiclassical contribution to both numerator and denominator. The classical approximation to the denominator of (2.14) would read

$$N\mu(s_i^0)\exp[-I(s_i^0)] \times \int \left(\prod_i d\xi_i\right) \exp\left[-\left(\frac{\partial^2 I}{\partial s_i\,\partial s_j}\right)_{s_i = s_i^0} \xi_i \xi_j\right], \tag{5.2}$$

where the ξ_i are the deviations in the squared edge lengths from their stationary values. Suppose that as n_0 becomes large the stationary simplicial geometries approach a continuum geometry. Then, since we expect many different simplicial geometries which *approximate* a given continuum geometry we should expect to find directions λ_i in the space of edge lengths along which the action is *approximately* stationary,

$$\lambda_i \left[\frac{\partial}{\partial s_i}\left(\frac{\partial I}{\partial s_j}\right)\right]_{s_k = s_k^0} \approx 0. \tag{5.3}$$

In fact, we can identify what these directions are.

The curvature of the stationary configuration must be characterized by the only scale in the Regge equations (2.8), that set by the cosmological constant Λ. As n_0 becomes large the characteristic squared edge length s_i^0 in the stationary configuration will become small compared to Λ^{-1} as $\Lambda^{-1} f^2(n_0)$, where f is some rapidly decreasing function of n_0 dependent on the subdivision process. On scales small com-

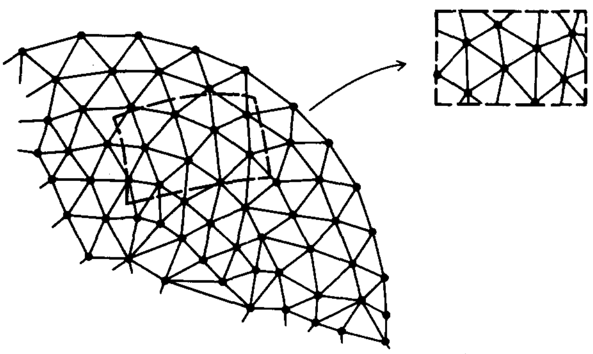


FIG. 2. The origin of the approximate diffeomorphism group. The figure shows a two-dimensional simplicial geometry whose net is sufficiently refined that the characteristic edge lengths are much smaller than the scale of the curvature. Local regions of this geometry will be approximately flat. Variations in the edge lengths which correspond to those induced by motions of the vertices in two-dimensional flat space and which are small on the curvature scale will leave the geometry approximately unchanged. For this net there are thus many different assignments of edge lengths which approximately correspond to the same geometry.

pared to $\Lambda^{-1/2}$ but large eventually compared to $\Lambda^{-1/2}f(n_0)$ the net will approximate flat space. In flat space there is a $4n_0$-parameter family of variations of the edge lengths which leave the geometry flat. In *approximately* flat space there will be a $4n_0$-parameter family of variations of the edge lengths which leave the geometry *approximately* flat (cf. Fig. 2). We thus expcet $4n_0$ directions in which (5.3) is approximately satisfied. Put differently, if a geometry is approximated by increasingly subdivided simplicial nets we expect $4n_0$ of the n_1 eigenvalues of $\partial^2 I/\partial s_i\, \partial s_j$ to be a number near zero times Λ/l^2, while the rest are a number of order unity times Λ/l^2.

The multiple integral in (5.2) is easily carried out along the $4n_0$ directions in which the action is approximately stationary. While the action is approximately stationary in these directions in the vicinity of the stationary configuration, we expect it to remain stationary only for deviations ξ^i which are of order of the curvature scale Λ^{-1}. Beyond that the deviations represent physically distinct geometries. Thus, for increasingly subdivided simplicial nets we expect the semiclassical approximation (5.2) to behave as

$$N(n_0, s_i^0)\exp[-I(s_i^0)]$$
$$\times(l^{-2}\Lambda^{-1})^{4n_0}\left\{\det'\left[\left(\frac{\partial^2 I}{\partial s_i\,\partial s_j}\right)_{s_k=s_k^0}\right]\right\}^{-1/2}, \qquad (5.4)$$

where $N(n_0, s_i^0)$ is a slowly varying function of n_0. Here det$'$ denotes the determinant over the $n_1 - 4n_0$ directions in which the action is not approximately zero, i.e., the product of the nonsmall eigenvalues of $\partial^2 I/\partial s_i\, \partial s_j$.

While the above discussion has been illustrated using the integral in the denominator of (2.14) the situation is similar with the numerator. If A is a quantity which is not sensitive to scales much smaller than the curvature scale, the integral for the numerator may be divided into a sum of pieces, each one of which locally behaves like (5.2).

Equation (5.4) implies that for large n_0 the integrals in both numerator and denominator of (2.14) will diverge as $(l^{-2}\Lambda^{-1})^{4n_0}$. ($l^{-2}\Lambda^{-1}$ is greater than unity when the semiclassical approximation is valid.) This is the degree of divergence associated with the diffeomorphism group—four divergent factors for each point. These divergent factors cancel between the numerator and denominator of (2.14) to give a nondivergent expression for $\langle A\rangle$ for those A's which are not sensitive to arbitrarily small scales. Thus in the limit of large n_0 we recover the behavior of the functional integrals arising from the diffeomorphism group of general relativity.

VI. SUMMING OVER TOPOLOGIES

Our discussion up to this point has proceeded as though the Euclidean prescription for the ground state wave function were to take a fixed and particular compact manifold and sum exp($-$ action) over the possible geometries on this manifold. This has been convenient for the exposition, but there is no compelling physical reason to make such a restriction and none is proposed. Indeed, to have the laws of physics fix the topology of the manifold but allow all possible geometries on it would seem to be assigning a very different status in physics to two closely related elements of geometry. Further, there are attractive physical reasons for considering four-geometries with different topology. In the set of ideas evoked by the words "space-time foam"[46] one would ask for quantum transition amplitudes between states specified by disconnected as well as connected three-geometries and multiply connected as well as simply connected ones. Unitarity would then suggest that the Euclidean functional integral prescription for these amplitudes contain a sum over the topologically nontrivial four-geometries into which these topologically nontrivial three-geometries can be embedded.

In this section we shall discuss how sums over different topologies might be implemented in the simplicial minisuperspace approximation. What we want to give practical meaning to might be written schematically as

$$\langle A\rangle = \frac{\Sigma_M \nu(M)\Sigma_{g \text{ on } M} A\,[g,M\,]\exp(-I\,[g,M\,])}{\Sigma_M \nu(M)\Sigma_{g \text{ on } M}\exp(-I\,[g,M\,])}. \qquad (6.1)$$

The first sum is over some class of compact four-manifolds. A weighting $\nu(M)$ which depends on the topological invariants of the manifold M, e.g., its Euler number, signature, fundamental group, etc., would be part of the prescription. The second sum is over physically distinct (i.e., nondiffeomorphic) metrics on M with the action I for general relativity.

The first step in turning the schema (6.1) into a computable procedure is to restrict the sum over four-manifolds to a sum of simplicial manifolds (i.e., simplicial complexes which are piecewise linear manifolds). The second step is to implement the sum over metrics by an integral over edge lengths. The latter sum has been the subject of the preceding parts of this paper. Including only simplicial manifolds is a restriction because not every four-manifold is triangulable. However, the author is not aware of anything physically interesting lost by this. We now discuss the problems involved in implementing a sum over compact simplicial four-manifolds. We shall not be able to answer every question, but we will be able to provide a framework for discussion and some practical proposals to try out.

To specify a sum over simplicial manifolds we must effectively have a procedure for listing those to be included in the sum. One way would simply be to list famous compact four-manifolds, e.g., $S^4, S^2\times S^2, S^1\times S^3, T^4, CP^2, K3$, etc., to find triangulations of these, and to add together sums over geometries on these of the kind we have been discussing. This can hardly be the basis for a general principle. A better approach is to sum over all simplicial four-manifolds or some generally specified subclass.

It is not possible to classify *all* four-manifolds. That is to say roughly, an algorithm for deciding when two four-manifolds are the same does not exist.[47] Subclasses of four-manifolds may be classifiable. For example, simply connected four-manifolds with spin structure are classified by their Euler number and signature up to a finite number of connected sums.[48] That all four-manifolds are not classifiable does not mean that they cannot be enumerated. One can imagine a procedure (and we shall describe one below) which would generate an exhaustive list of four-manifolds. A given four-manifold would occur on the list more than once but there would be no universal algorithm to decide when two entries on the list are the same manifold. To implement the sum over simplicial manifolds one therefore has the following choice: One can sum over a class of four-manifolds which are classifiable by virtue of having more structure and assign a weight to each. Alternatively, one can devise a procedure for listing all four-manifolds and accept the weighting implied by the procedure.

To describe concretely how a list of manifolds can be prepared let us first consider how a single manifold is specified.[41] A simplicial complex is a collection of simplices such that whenever a simplex lies in the collection then so does each of its faces, and whenever two simplices of the collection intersect they do so in a common face. The dimension of a complex is the largest dimension of a simplex in it. A four-dimensional simplicial complex may be specified by giving the vertices of each four-simplex in the complex. From the list of four-simplices the vertices of the edges, triangles, and tetraheda may be computed. These must be such that the conditions of the definition are satisfied. Figure 3 shows a two-dimensional example.

Not every simplicial complex is a piecewise linear manifold (i.e., such that every point has a neighborhood which is piecewise linearly homeomorphic to an open subset of R^4 or a half-space of R^4). (From now on we shall omit the qualification "piecewise linear" when referring to manifolds, homeomorphisms, etc.) A two-dimensional example is shown in Fig. 4. There is a necessary and sufficient condition for a complex to be a manifold. To state it we shall need the notions of the *star* and the *link* of a simplex in the complex. They are illustrated in two dimensions in Fig. 5. The star of a simplex σ is the collection of simplices which contain σ together with all their faces. The link of a simplex σ is the collection of simplices in the star of σ which do not meet σ. The necessary and sufficient condition that an n-dimensional complex be an n-manifold is that the link of every simplex of dimension k be a triangulation of (i.e., homeomorphic to) an $(n-k-1)$-sphere.[41] Given a complex, one might imagine checking it to see if it is a manifold. Were space-time

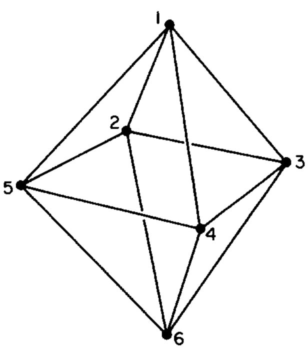


FIG. 3. A two-dimensional simplicial manifold. The six vertices are connected as for an octohedron. The simplicial complex can be specified by giving a list of vertices of the triangles. In the present case the list would be

123 134 236 346
125 145 256 456 .

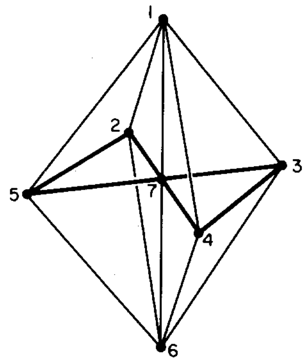


FIG. 4. A two-dimensional simplicial complex which is not a manifold. This complex consists of the triangles

125 147 346 134 256 467
127 157 367 137 267 567 .

This complex is not a manifold because local neighborhoods of vertices 1, 6, and 7 as well as the edges (1,7) and (6,7) are not homeomorphic to a region of R^2. Neighborhoods of vertex 7, for example, are homeomorphic to regions in the intersection of two planes. The link of vertices 1 and 6 is the heavy curve and is not topologically a circle. If edge lengths were assigned to the complex by embedding it in a flat three-dimensional space as suggested by the figure, then the deficit angle of vertex 7 would be -2π. The two-dimensional gravitational action defined by $-2\pi l^{-2}\Sigma$(deficit angles) is $-4\pi/l^2$ for this complex. This is larger than the value $-8\pi/l^2$ for manifolds which are topologically two-spheres such as the example in Fig. 3.

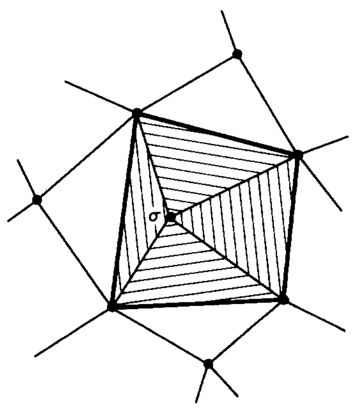


FIG. 5. Link and star of a vertex. The figure shows a portion of a two-dimensional simplicial manifold. The star of a given vertex consists of the interior, edges, and vertices of those triangles which intersect the given vertex. The star of vertex σ consists of the four shaded triangles. The link of a vertex consists of those simplices of its star which do not themselves intersect the vertex. The link of σ is the heavily drawn edges and their vertices in the figure. The link of σ is topologically S^1.

three-dimensional, this would in principle be possible to do. The one- and two-dimensional manifolds are classifiable by their homology groups and these are in principle computable. Were space-time five-dimensional, it would be in principle impossible to decide whether a complex was a manifold. In particular, one would need to decide whether the link of a vertex was a four-sphere and this problem is known to be unsolvable (i.e., roughly, it can be shown that there is no algorithm to do it).[47] In four dimensions the difficult problem would be to decide whether the link of a vertex is a three-sphere. Even assuming the Poincaré conjecture one would still need a procedure for deciding whether the link was simply connected. It seems that at the time of writing both of these decision problems are still open.[49] Thus one cannot list a family of manifolds to sum over by listing all the simplicial complexes with, say, a fixed number of vertices, and then discarding those which are not manifolds. There is no way to check. We must create the list by a different route.

A natural way to generate a list of manifolds is to create simplicial complexes from smaller units, "building blocks," which are locally and explicitly known to be manifolds. We begin with a fixed number of vertices n_0 with the idea of eventually allowing n_0 to become large. In view of the approximate recovery of the diffeomorphism group discussed in Sec. V there is no purpose to be served in considering lists with different numbers of vertices. One expects the manifolds with the largest n_0 to dominate both the numerator and denominator of (6.1) for large n_0. To create the list we start by enumerating all the four-dimensional simplicial complexes with n_0 vertices. This is a matter of enumerating all possible lists of four-simplices and checking that when two intersect they do so in a common face. We now discard from the list all complexes for which the link of every vertex is not one of a finite list of known triangulations of a three-sphere. For example, one might require that the link of every vertex be either the boundary of a four-simplex or of a 600-cell. The complexes remaining on the list are therefore known to be manifolds. If a sufficient number of basic building blocks is taken, lists containing all manifolds can be generated in this way. For example, in two dimensions it suffices to require the links to be five-, six- or, seven-gons.

A procedure such as described above would generate a list of manifolds with which to define a sum over topologies. For each member the sum over geometries would be carried out as described in the previous sections of this paper. One could assign relative weights to the different members of the list on the basis of some computable topological invariant, but the simplest assignment would be the weighting generated by the procedure itself. It then becomes an interesting mathematical question to ask with what multiplicity a given manifold occurs on the list. In particular, for the procedure to make sense, the large multiplicities should be independent of the particular basic building blocks chosen out of some general class.

We have described a procedure for summing over four-dimensional manifolds which at least can be tried out. One could ask: "Why restrict attention to four dimensions?" or "Why consider only manifolds?" The familiar answers that space-time seems to be four-dimensional and that a manifold is the mathematical implementation of the principle of equivalence are possibly too unadventurous. For example, the Regge action extends naturally to simplicial complexes which are not manifolds. (See the example in Fig. 4.) It also generalizes naturally to higher dimensions. One is thus invited to sum over complexes which are not manifolds and over other dimensions. The simplicial minisuperspace methods described here provide a framework for investigating such questions. Before embarking on such a journey, however, it would be useful to know if there is a way back to the familiar four-dimensional space-time on large scales.

ACKNOWLEDGMENTS

The author has benefited from useful conversations and correspondence with many colleagues, among them B. Allen, G. Gibbons, S. W. Hawking, and T. Piran. Any mathematics of interest in the paper the author has almost surely learned from G. Horowitz and M. Scharlemann. Special thanks are due to Professor Scharlemann for his generous and patient response to the author's many mathematical questions. Section V arose from questions and suggestions put forward by R. Geroch and B. DeWitt. The author would like to thank R. Williams for a fruitful correspondence on the Regge calculus, for many of the references, and for checking some of the formulas.

This paper was supported in part by the National Science Foundation under Grant No. PHY 81-07384.

[1]See, for example, the review articles of B. S. DeWitt and S. W. Hawking, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge, 1979); and C. Teitelboim, Phys. Rev. D **25**, 3159 (1982); D **28**, 297 (1983); D **28**, 310 (1983).

[2]J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).

[3]S. W. Hawking, in *Relativity, Groups and Topology* II, edited by B. S. DeWitt and R. Stora (North-Holland, Amsterdam, 1984).

[4]Related ideas have been put forward by A. Vilenkin, Phys. Lett. B **117**, 25 (1983); Phys. Rev. D **27**, 2848 (1983); preprint (1984); by D. Atkatz and H. Pagels, Phys. Rev. D **25**, 2065 (1982); and by T. Padmanabhan, Int. J. Theor. Phys. **22**, 1023 (1983); J. V. Narlikar and T. Padmanabhan, Phys. Rep. **100**, 151 (1983).

[5]S. W. Hawking, in *Astrophysical Cosmology*, edited by H. A. Brück, G. V. Coyne, and M. S. Longair (Pontificiae Academie Scientiarum Scripta Varia, Vatican, 1982).

[6]S. W. Hawking, Nucl. Phys. B **239**, 257 (1984); S. W. Hawking and J. C. Luttrell, Phys. Lett. B **143**, 83 (1984); S. W. Hawking and J. C. Luttrell, Nucl. Phys. B **247**, 250 (1984).

[7]C. W. Misner, in *Magic Without Magic: John Archibald Wheeler*, edited by J. R. Klauder (Freeman, San Francisco, 1972).

[8]See, e.g., B. S. DeWitt, Phys. Rev. **160**, 1113 (1967); C. Isham and J. E. Nelson, Phys. Rev. D **10**, 3226 (1974); W. Blyth and C. Isham. Phys Rev. D **11**, 768 (1975); M. Ryan, *Hamiltonian Cosmology* (Springer, New York, 1972); T. Christodoulakis and J. Zanelli, Phys. Lett. A **102**, 227 (1984).

[9]T. Regge, Nuovo Cimento **19**, 558 (1961).

[10]J. A. Wheeler, in *Relativity, Groups and Cosmology*, edited by C. DeWitt and B. DeWitt, (Gordon and Breach, New York, 1964).

[11]C-Y. Wong, J. Math. Phys. **12**, 70 (1971).

[12]P. A. Collins and R. Williams, Phys. Rev. D **5**, 1908 (1972); Phys. Rev. D **7**, 965 (1973); Phys. Rev. D **10**, 3537 (1974).

[13]R. Sorkin, Phys. Rev. D **12**, 385 (1975).

[14]R. Sorkin, J. Math. Phys. **16**, 2432 (1975).

[15]G. Ellis and R. Williams, Gen. Relativ. Gravit. **13**, 361 (1981).

[16]S. Lewis, Phys. Rev. D **25**, 306 (1982).

[17]W. Miller and J. A. Wheeler, in *Proceedings of the 1983 Rome Workshop on Galactic and Extragalactic Dark Matter* (to be published).

[18]R. Friedberg and T. D. Lee, Nucl. Phys. B **242**, 145 (1984).
[19]W. Miller (to be published).
[20]G. Ponzano and T. Regge, in *Spectroscopic and Group Theoretical Methods in Physics*, edited by F. Bloch, S. Cohen, A. De Shalit, S. Sambursky, and I. Talmi (North-Holland, Amsterdam, 1968).
[21]B. Hasslacher and M. Perry, Phys. Lett. B **103**, 21 (1981).
[22]M. Rocek and R. M. Williams, Phys. Lett. B **104**, 31 (1981); Z. Phys. C **21**, 371 (1984); and in *Quantum Structure of Space and Time*, edited by M. J. Duff and C. J. Isham (Cambridge U. P., Cambridge, 1982).
[23]N. P. Warner, Proc. R. Soc. London A **383**, 359 (1982).
[24]M. Martellini, in *Geometrical Methods in Theoretical Physics*, edited by G. Denardo and H. Dobner (World Scientific, Singapore, 1983).
[25]G. Feinberg, R. Friedberg, T. D. Lee, and H. C. Ren, "Lattice Gravity Near the Continuum Limit," Columbia University preprint, 1984.
[26]H. Hamber and R. M. Williams, Nucl. Phys. B **248**, 145 (1984).
[27]For example partition functions, see also Refs. 25 and 26.
[28]H. Leutwyler, Phys. Rev. B **134**, 1155 (1964); E. S. Fradkin and G. A. Vilkovisky, Phys. Rev. D **8**, 4241 (1973); E. S. Fradkin and G. A. Vilkovisky, CERN Report 2332 (1977) (unpublished).
[29]J. B. Hartle and R. Sorkin, Gen. Relativ. Gravit. **13**, 541 (1981).
[30]B. DeWitt, in *Magic Without Magic: John Archibald Wheeler*, edited by J. R. Klauder (Freeman, San Francisco, 1972), and in *Recent Developments in Gravitation*, Cargese, 1978, edited by M. Levy and S. Deser (Plenum, New York, 1979).
[31]G. W. Gibbons, S. W. Hawking, and M. Perry, Nucl. Phys. B **138**, 141 (1978).
[32]For the definition of conformal transformations in simplicial nets, see Refs. 22. For an explicit numerical example see Paper II in this series.
[33]J. B. Hartle, Phys. Rev. D **29**, 2730 (1984).
[34]J. B. Hartle (unpublished).
[35]In order to avoid introducing yet another obscure term we have called the result of (2.13) an "expectation value" even though it is not an expectation value in the sense of the traditional canonical theory. These would be calculated on a hypersurface of superspace corresponding to a fixed "time." They are, however, the expectation values in the approach of C. Teitelboim in Ref. 1.
[36]See, for example, Refs. 10, 11, 14, 17, and especially J. Cheeger, W. Müller, and R. Schrader, Commun. Math. Phys. **92**, 405 (1984), where a method essentially the same as used here is described. The author would like to thank Dr. R. Williams for providing these references.
[37]See Ref. 2 for an example.
[38]For a summary of the present status of the existence of Euclidean Einstein metrics on compact manifolds see S–T. Yau, in *Seminar on Differential Geometry*, edited by S–T. Yau (Princeton U. P., Princeton, 1982).
[39]See, e.g., S. W. Hawking, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge, 1979).
[40]The analogy, however, especially viewed in the Hamiltonian framework is not complete. See J. B. Hartle and K. Kuchar, J. Math. Phys. **25**, 57 (1984); in *Quantum Theory of Gravity*, edited by S. Christensen (Hilger, Bristol, 1984); C. Teitelboim, Phys. Rev. D **25**, 3159 (1982); Ann. Phys. (N.Y.) **143**, 127 (1982).
[41]See, e.g., C. P. Rourke and B. J. Sanderson, *Introduction to Piecewise Linear Topology* (Springer, Berlin, 1972).
[42]One could consider piecewise diffeomorphisms between different simplicial manifolds. Those between a simplicial manifold and its subdivisions would provide examples. The present restriction will, however, simplify our discussion.
[43]This has been stressed by Rocek and Williams, Ref. 22.
[44]See, e.g., the discussion for the continuum case in A. Karlhede, Gen. Relativ. Gravit. **12**, 693 (1980).
[45]Rocek and Williams (Ref. 22) use the term "gauge transformation"for four-parameter transformations of the edge lengths which leave the action invariant. Infinitesimally in the continuum theory this would correspond to variations in the metric for which

$$\delta I = \int d^4x(g)^{1/2} G_{\mu\nu}(x)\delta g^{\mu\nu}(x) = 0, \tag{*}$$

while in the simplicial theory this would correspond to variations in the squared edge lengths for which

$$\delta I = \sum_i \frac{\delta I}{\delta s_i}\delta s^i = 0. \tag{**}$$

In this paper we use the term "gauge transformation" to mean transformations which not only preserve the action but all other physical quantities as well. For example, electromagnetic gauge transformations preserve the field as well as the action. For simplicial nets by gauge transformations we would mean transformations of the edge lengths which preserve the geometry as well as the action. Distinct geometries may have the same action. Thus in general there will be many more transformations which preserve the action than which preserve the geometry. This can be seen directly from (*) and (**) where it is not difficult to find parametric families of $\delta g^{\mu\nu}$ and δs^i which satsify these relations.
[46]See, e.g., J. A. Wheeler in *Battelle Rencontres*, edited by C. M. DeWitt and J. A. Wheeler (Benjamin, New York, 1968).
[47]See, e.g., W. Haken in *Word Problems*, edited by W. W. Boone, F. B. Cannonito, and R. C. Lyndon (North-Holland, Amsterdam, 1973).
[48]See, e.g., C. T. C. Wall, J. London Math. Soc. **39**, 131 (1964).
[49]See, however, W. Haken (to be published).

The statistics of interacting dumbbells on a $2\times N$ lattice space

D. Walikainen and R. B. McQuistan
Department of Physics and Laboratory for Surface Studies, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin 53201

(Received 18 April 1984; accepted for publication 14 September 1984)

A set theoretic argument is utilizied to determine a 31-term recursion that describes exactly the composite nearest-neighbor degeneracy for indistinguishable dumbbell particles distributed on a rectangular $2\times N$ lattice space. The associated generating functions, the expectation of the lattice coverage, and the density of occupied nearest-neighbor pairs are also determined.

I. INTRODUCTION

The statistical mechanical treatment of such physical and chemical phenomena as adsorption, magnetism, superconductivity, and crystallization often involves a consideration of the occupational degeneracy of particles distributed on a lattice space. Of particular interest and difficulty are those situations that arise when correlation exists, either within the particle itself (as with dumbbell particles) or when the particles interact (as with nearest-neighbor interaction).

In the present paper we consider a problem in which both kinds of correlation exist, i.e., the nearest-neighbor interaction of dumbbell particles on a $2\times N$ rectangular lattice space.

Recently,[1] a recursion was developed that yields exactly the composite nearest-neighbor degeneracy for simple, indistinguishable particles distributed on a $2\times N$ lattice. The present paper is an attempt to extend such results to more complex particles.

Specifically, we will treat the following problem: Given a $2\times N$ rectangular lattice space on which are distributed q indistinguishable dumbbells with nonidentifiable ends, what is the multiplicity of those arrangements, characterized by the stipulation of N,q as well as n_{11}, n_{01}, and n_{00} (the number of occupied, mixed, and vacant nearest-neighbor pairs, respectively)? (See Fig. 1.) An occupied nearest-neighbor pair occurs when adjacent sites are occupied by ends of different dumbbells. We do not distinguish between a 0-1 pair and a 1-0 pair.

It should be pointed out that n_{11}, n_{01}, and n_{00} are not independent. Their sum must equal the total number of nearest-neighbor pairs on the space, i.e.,

$$3N-q-2=n_{11}+n_{01}+n_{00}. \tag{1}$$

In light of this constraint we will, in the remainder of the present paper, ignore n_{01}, the number of mixed nearest-neighbor pairs.

If, from each end of each of the dumbbells, we draw lines to the nearest-neighbor sites, except for the sites occupied by the other end of the dumbbells (see Fig. 2), there will be $4q$ lines. A line will go either to a vacant site (indicating a mixed nearest-neighbor pair) or to an occupied site (indicating an occupied nearest-neighbor pair), i.e.,

$$4q=2n_{11}+n_{01}+\delta_1, \tag{2}$$

where $\delta_1 \leqslant 4$ denotes the number of lines that go off the ends of the space.

A similar process can be carried out for the vacant sites, yielding

$$3[2N-2q]=2n_{00}+n_{01}+\delta_0, \tag{3}$$

where $\delta_0 \leqslant 4$ is the number of lines that go off the ends of the space.

Adding Eqs. (2) and (3) yields the expected result [when Eq. (1) is considered]

$$\delta_0+\delta_1=4. \tag{4}$$

Note that if N, q, n_{11}, and n_{00} are specified, n_{01} and the δ's are determined uniquely. We conclude that N, q, n_{11}, and n_{00} completely determine the (composite) nearest-neighbor degeneracy.

II. RECURSION RELATION

In the present section we derive a recursion for $A\,[N,q,n_{11},n_{00}]$, the total number of independent arrangements that can be created when q indistinguishable dumbbell particles are distributed on a $2\times N$ lattice to form n_{11} occupied *and* n_{00} vacant nearest-neighbor pairs.

To establish a recursion for $A\,[N,q,n_{11},n_{00}]$, we first differentiate between an $\alpha(N)$-space (which consists of two aligned rows of N equivalent rectangular sites) and a $\beta(N)$-space [an $\alpha(N)$-space from which one lattice site in the lower left-hand side has been deleted] (see Fig. 3).

The set of all arrangements of q dumbbell particles on an $\alpha(N)$-space, that contain n_{11} occupied an n_{00} vacant nearest-neighbor pairs, is designated $a[N,q,n_{11},n_{00}]$. Thus $\#a[N,q,n_{11},n_{00}]$, the number of elements of the set $a[N,q,n_{11},n_{00}]$, is the desired $A\,[N,q,n_{11},n_{00}]$.

Let $a_j\,[N,q,n_{11},n_{00}]$ $(j=0,...,3)$ be subsets of $a[N,q,n_{11},n_{00}]$. Each of the a_j's is characterized by the state of

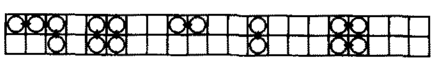


FIG. 1. An arrangement of eight dumbbells on a 2×21 lattice, giving rise to five occupied nearest neighbors, 21 mixed neighbors, and 27 vacant nearest neighbors.

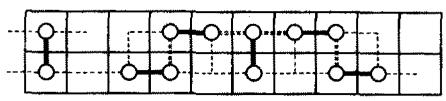


FIG. 2. On this 2×10 space there are six particles, so there are 24 dashed lines. There are four occupied nearest neighbor pairs, each of which is associated with two dashed lines, and 14 mixed nearest neighbor pairs associated with a single dashed line; $\delta_1=2$.

 0022-2488/85/040815-07$02.50

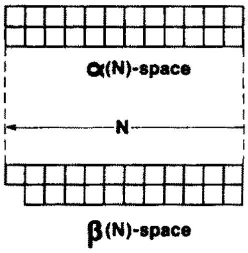


FIG. 3. (a) Definition of an $\alpha(N)$-space. (b) Definition of a $\beta(N)$-space.

occupation of the two sites in the Nth column (see Fig. 4). Every arrangement in a_j differs from every arrangement in a_k $(j \neq k)$, i.e.,

$$a_j[N,q,n_{11},n_{00}] \cap a_k[N,q,n_{11},n_{00}] = \phi, \tag{5}$$

a null set. In addition every element $a[N,q,n_{11},n_{00}]$ will be found in one of the a_j's, i.e.,

$$a[N,q,n_{11},n_{00}] = \bigcup_{j=0}^{3} a_j[N,q,n_{11},n_{00}]. \tag{6}$$

We conclude that $\# a[N,q,n_{11},n_{00}]$ is given by

$$\# a[N,q,n_{11},n_{00}] = \sum_{j=0}^{3} \# a_j[N,q,n_{11},n_{00}]$$

or

$$A[N,q,n_{11},n_{00}] = \sum_{j=0}^{3} A_j[N,q,n_{11},n_{00}], \tag{7}$$

where, for example, $\# a_2[N,q,n_{11},n_{00}] \equiv A_2[N,q,n_{11},n_{00}]$ is the number of arrangements possible when the Nth column is occupied by a single dumbbell particle.

The top-to-bottom reflection of the arrangements, contained in set a_1, results in a degeneracy factor of 2. The degeneracy factor for the rest of the a_j sets is unity.

Similarly, we let $b[N,q,n_{11},n_{00}]$ be the set of all arrangements of q dumbbell particles on a $\beta(N)$-space that exhibit prescribed numbers of occupied *and* vacant nearest-neighbor pairs.

On the basis of the state of occupation of the sites furthest to the left of both rows of a $\beta(N)$-space, we define five subsets $b_j[N,q,n_{11},n_{00}]$ $(j = 0,...,4)$ (see Fig. 5). Thus, for example, $b_2[N,q,n_{11},n_{00}]$ is the set of all arrangements of q dumbbell particles on a $\beta(N)$-space exhibiting n_{11} occupied *and* n_{00} vacant nearest neighbor pairs, when a dumbbell occupies the two lower left-hand sites *and* the upper left-hand site is vacant.

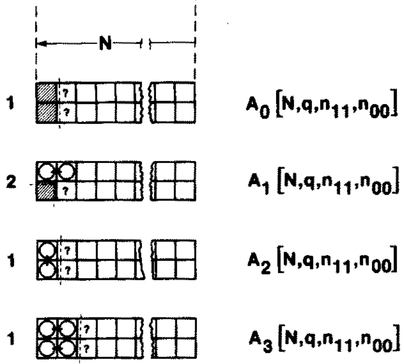


FIG. 4. Four α-spaces are characterized by an occupation of the Nth column. The top-to-bottom degeneracy factors are indicated at the left of the space.

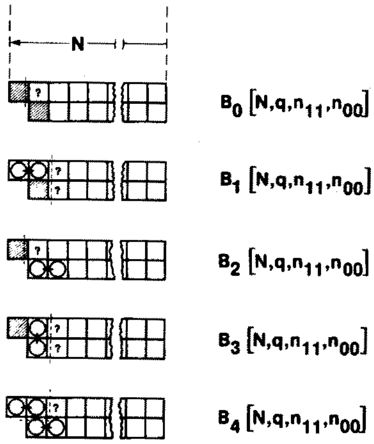


FIG. 5. Five β-spaces are characterized by the state of occupation of the sites furthest to the left in each row.

Since

$$b[N,q,n_{11},n_{00}] = \bigcup_{j=1}^{4} b_j[N,q,n_{11},n_{00}] \tag{8}$$

and

$$b_j[N,q,n_{11},n_{00}] \cap b_k[N,q,n_{11},n_{00}] = \phi \quad (j \neq k), \tag{9}$$

we conclude that

$$\begin{aligned} B[N,q,n_{11},n_{00}] &\equiv \#[N,q,n_{11},n_{00}] \\ &= \sum_{j=0}^{4} \# b_j[N,q,n_{11},n_{00}] \\ &= \sum_{j=0}^{4} B_j[N,q,n_{11},n_{00}], \end{aligned} \tag{10}$$

where Eq. (10) serves to define the B_j's.

Next we decompose the four a_j sets and the five b_j sets. The basis for this decomposition is the state of occupation of those sites in which a question mark appears. (See Figs. 4 and 5.)

For example, consider the $a_0[N,q,n_{11},n_{00}]$ set in Fig. 4 (see Fig. 6); if both sites in which a question mark appears are vacant, an $a_0[N-1,q,n_{11},n_{00}-3]$ set is formed because q dumbbell particles must be arranged on an $\alpha(N-1)$-space to form n_{11} occupied and $n_{00}-3$ vacant nearest-neighbor pairs. If one of the two sites is occupied *and* the other is vacant, an $a_1[N-1,q,n_{11},n_{00}-2]$ set is formed. Note that

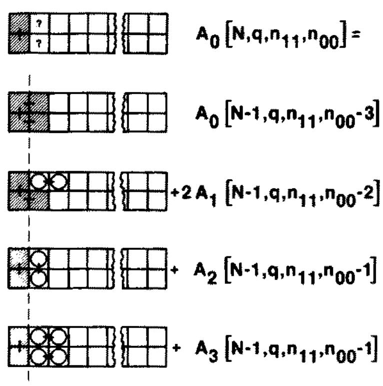


FIG. 6. This figure shows the decomposition of the set of $a_0[N,q,n_{11},n_{00}]$ into four mutually exclusive subsets on the basis of the state of occupation of the $N-1$ column.

because of the top-to-bottom symmetry, this can occur in two ways.

If both sites are filled by a single dumbbell particle, an $a_2[N-1,q,n_{11},n_{00}-1]$ set is formed.

An $a_3[N-1,q,n_{11},n_{00}-1]$ set is formed if both sites are occupied by parts of two different dumbbell particles.

Thus we may write

$$\begin{aligned} a_0[N,q,n_{11},n_{00}] &= a_0[N-1,q,n_{11},n_{00}-3] \\ &\quad \cup a_1[N-1,q,n_{11},n_{00}-2] \\ &\quad \cup a_1[N-1,q,n_{11},n_{00}-2] \\ &\quad \cup a_2[N-1,q,n_{11},n_{00}-1] \\ &\quad \cup a_3[N-1,q,n_{11},n_{00}-1]\,. \end{aligned} \tag{11a}$$

Similarly, we may write

$$\begin{aligned} a_1[N,q,n_{11},n_{00}] &= \mathrm{b}_1[N,q,n_{11},n_{00}-1] \\ &\quad \cup b_4[N,q,n_{11},n_{00}]\,, \end{aligned} \tag{11b}$$

$$\begin{aligned} a_2[N,q,n_{11},n_{00}] &= a_0[N-1,q-1,n_{11},n_{00}] \\ &\quad \cup a_1[N-1,q-1,n_{11}-1,n_{00}] \\ &\quad \cup a_1[N-1,q-1,n_{11}-1,n_{00}] \\ &\quad \cup a_2[N-1,q-1,n_{11}-2,n_{00}] \\ &\quad \cup a_3[N-1,q-1,n_{11}-2,n_{00}], \end{aligned} \tag{11c}$$

$$\begin{aligned} a_3[N,q,n_{11},n_{00}] &= a_0[N-2,q-2,n_{11}-2,n_{00}] \\ &\quad \cup a_1[N-2,q-2,n_{11}-3,n_{00}] \\ &\quad \cup a_1[N-2,q-2,n_{11}-3,n_{00}] \\ &\quad \cup a_2[N-2,q-2,n_{11}-4,n_{00}] \\ &\quad \cup a_3[N-2,q-2,n_{11}-4,n_{00}]\,. \end{aligned} \tag{11d}$$

The b_j sets may also be decomposed as follows:

$$\begin{aligned} b_0[N,q,n_{11},n_{00}] &= a_0[N-1,q,n_{11},n_{00}-1] \\ &\quad \cup a_1[N-1,q,n_{11},n_{00}]\,, \end{aligned} \tag{12a}$$

$$\begin{aligned} b_1[N,q,n_{11},n_{00}] &= a_0[N-2,q-1,n_{11},n_{00}-1] \\ &\quad \cup a_1[N-2,q-1,n_{11}-1,n_{00}-1] \\ &\quad \cup a_1[N-2,q-1,n_{11},n_{00}] \\ &\quad \cup a_2[N-2,q-1,n_{11}-1,n_{00}] \\ &\quad \cup a_3[N-2,q-1,n_{11}-1,n_{00}]\,, \end{aligned} \tag{12b}$$

$$\begin{aligned} b_2[N,q,n_{11},n_{00}] &= a_1[N-1,q,n_{11},n_{00}-1] \\ &\quad \cup a_3[N-1,q,n_{11},n_{00}]\,, \end{aligned} \tag{12c}$$

$$\begin{aligned} b_3[N,q,n_{11},n_{00}] &= a_0[N-2,q-1,n_{11},n_{00}] \\ &\quad \cup a_1[N-2,q-1,n_{11}-1,n_{00}] \\ &\quad \cup a_1[N-2,q-1,n_{11}-1,n_{00}] \\ &\quad \cup a_2[N-2,q-1,n_{11}-2,n_{00}] \\ &\quad \cup a_3[N-2,q-1,n_{11}-2,n_{00}]\,, \end{aligned} \tag{12d}$$

$$\begin{aligned} b_4[N,q,n_{11},n_{00}] &= b_1[N-1,q-1,n_{11}-1,n_{00}] \\ &\quad \cup b_4[N-1,q-1,n_{11}-2,n_{00}]. \end{aligned} \tag{12e}$$

An examination of Eqs. (11) and (12) indicates that we need only six equations [Eqs. (11a)–(11d), (12b), and (12e)] to represent the entire decomposition regime.

We note, for example, that in Eq. (11b) every element in $b_1[N,q,n_{11},n_{00}-1]$ differs from every element in $b_4[N,q,n_{11},n_{00}]$ by the state of occupation of the lower site in column $(N-1)$, i.e.,

$$b_1[N,q,n_{11},n_{00}-1] \cap b_4[N,q,n_{11},n_{00}] = \phi. \tag{13}$$

In addition, every element in $a_1[N,q,n_{11},n_{00}]$ will be found either in $\mathrm{b}_1[N,q,n_{11},n_{00}-1]$ or $b_4[N,q,n_{11},n_{00}]$, i.e.,

$$a_1[N,q,n_{11},n_{00}] = b_1[N,q,n_{11},n_{00}-1] \cup b_4[N,q,n_{11},n_{00}]. \tag{14}$$

We conclude that

$$\begin{aligned} \#a_1[N,q,n_{11},n_{00}] &= \#b_1[N,q,n_{11},n_{00}-1] \\ &\quad + \#b_4[N,q,n_{11},n_{00}], \end{aligned} \tag{15}$$

or

$$\begin{aligned} A_1[N,q,n_{11},n_{00}] &= B_1[N,q,n_{11},n_{00}-1] \\ &\quad + B_4[N,q,n_{11},n_{00}]. \end{aligned} \tag{16}$$

Thus, Eqs. (11a)–(11d), (12b), and (12e) can be written in terms of the number of elements of the sets, i.e., in terms of the degeneracies A_j and B_j.

To determine the recursion for $A\,[N,q,n_{11},n_{00}]$ we utilize shift operators R, S, T, and U, with the understanding that

$$\begin{aligned} &A_j\,[N-r,q-s,n_{11}-t,n_{00}-u] \\ &\quad = R^r S^s T^t U^u A\,[N,q,n_{11},n_{00}]\,. \end{aligned} \tag{17}$$

Then Eqs. (11a)–(11d), (12b), and (12e), when written in terms of the degeneracies A_j and B_j, may be written in matrix form:

$$\begin{pmatrix} RU^3-1 & 2RU^2 & RU & RU & 0 & 0 \\ 0 & -1 & 0 & 0 & U & 1 \\ RS & 2RST & RST^2-1 & RST^2 & 0 & 0 \\ R^2S^2T^2 & 2R^2S^2T^3 & R^2S^2T^4 & R^2S^2T^4-1 & 0 & 0 \\ R^2SU & R^2STU+R^2S & R^2ST & R^2ST & -1 & 0 \\ 0 & 0 & 0 & 0 & RST & RST^2-1 \end{pmatrix} \begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \\ B_1 \\ B_4 \end{pmatrix} = 0. \tag{18}$$

For this system of equations to have a nontrivial solution, we require that the determinant of the shift operator matrix should annihilate the solution space (A_0, A_1, A_2, A_3, B_1, or B_4). Thus, the determinant, operating on any of the A_j's (or B_j's) yields the same recursion relation. If, however, the same recursion describes the composite nearest-neighbor degeneracy for A_0, A_1, A_2, and A_3, it describes the degeneracy for q dumbbell particles on an $\alpha(N)$-space, regardless of the state of occupation of the sites at the left-hand end. The only factors that differentiate one A_j from another (or from the B_j's) are the different initial conditions for each of the degeneracies.

The determinant of the matrix in Eq. (18) yields the following recursion:

$$\begin{aligned}A[N,q,n_{11},n_{00}] = {}& A[N-1,q,n_{11},n_{00}-3] + 2A[N-1,q-1,n_{11}-2,n_{00}] + 2A[N-2,q-1,n_{11},n_{00}-1] \\ & + A[N-2,q-1,n_{11}-1,n_{00}-2] - 2A[N-2,q-1,n_{11}-2,n_{00}-3] + A[N-3,q-1,n_{11},n_{00}-4] \\ & - A[N-3,q-1,n_{11}-1,n_{00}-5] + A[N-3,q-2,n_{11}-1,n_{00}] + A[N-3,q-2,n_{11}-2,n_{00}-1] \\ & - 2A[N-3,q-2,n_{11}-3,n_{00}-2] - A[N-3,q-3,n_{11}-6,n_{00}] - A[N-4,q-2,n_{11},n_{00}-2] \\ & + 4A[N-4,q-2,n_{11}-1,n_{00}-3] - 5A[N-4,q-2,n_{11}-2,n_{00}-4] + 2A[N-4,q-2,n_{11}-3,n_{00}-5] \\ & + A[N-4,q-3,n_{11}-3,n_{00}] - 2A[N-4,q-3,n_{11}-4,n_{00}-1] + A[N-4,q-3,n_{11}-6,n_{00}-3] \\ & - A[N-5,q-3,n_{11}-1,n_{00}-1] + 3A[N-5,q-3,n_{11}-2,n_{00}-2] - 3A[N-5,q-3,n_{11}-3,n_{00}-3] \\ & + A[N-5,q-3,n_{11}-4,n_{00}-4] + A[N-5,q-4,n_{11}-5,n_{00}] - 2A[N-5,q-4,n_{11}-6,n_{00}-1] \\ & + A[N-5,q-4,n_{11}-7,n_{00}-2] - A[N-6,q-4,n_{11}-3,n_{00}-1] + 4A[N-6,q-4,n_{11}-4,n_{11}-2] \\ & - 6A[N-6,q-4,n_{11}-5,n_{11}-3] + 4A[N-6,q-4,n_{11}-6,n_{11}-4] - A[N-6,q-4,n_{11}-7,n_{00}-5].\end{aligned} \tag{19}$$

It is interesting to note that when Eq. (2) is subtracted from Eq. (3),

$$3N - 5q = n_{00} - n_{11} + (\delta_0 - \delta_1)/2, \tag{20}$$

any simultaneous change in N, q, n_{00}, and n_{11} must conform to

$$3(\Delta N) + (\Delta n_{11}) = 5(\Delta q) + (\Delta n_{00}). \tag{21}$$

An examination of the arguments of A in Eq. (19) reveals that Eq. (21) is obeyed.

III. GENERATING FUNCTIONS

We form the polynomials

$$f_{N,q}(x,y) \equiv \sum_{n_{11}} \sum_{n_{00}} A[N,q,n_{11},n_{00}]\, x^{n_{11}} y^{n_{00}}. \tag{22}$$

If the value of n_{11} is specified, then the sum over n_{00} contains five terms, as can be seen from the elimination of n_{01} in Eqs. (2) and (3)

$$3N - 5q + n_{11} = n_{00} + (\delta_0 - \delta_1)/2, \tag{23}$$

where $(\delta_0 - \delta_1)/2$ can take on the values -2, -1, 0, 1, and 2.

When Eq. (19) is substituted into Eq. (22) we obtain the following relationship for $f_{N,q}(x,y)$:

$$\begin{aligned}f_{N,q}(x,y) = {}& [y^3] f_{N-1,q}(x,y) + [2x^2] f_{N-1,q-1}(x,y) + y[2 + xy - 2x^2y^2] f_{N-2,q-1}(x,y) \\ & + y^4[1-xy] f_{N-3,q-1}(x,y) + x[1-xy][1+2xy] f_{N-3,q-2}(x,y) - x^6 f_{N-3,q-3}(x,y) \\ & - y^2[(1-xy)^3 - xy(1-xy)^2] f_{N-4,q-2}(x,y) + x^3(1 - 2y + x^3y^3) f_{N-4,q-3}(x,y) \\ & - xy[1-xy]^3 f_{N-5,q-3}(x,y) + x^5[1-xy]^2 f_{N-5,q-4}(x,y) - x^3y[1-xy]^4 f_{N-6,q-4}(x,y).\end{aligned} \tag{24}$$

Equation (24), combined with the following initial conditions:

$$\begin{aligned}& f_{0,0} = 1,\ \ f_{1,0} = y,\ \ f_{1,1} = 1,\ \ f_{2,0} = y^4 f_{2,1} = 4y^4,\ \ f_{2,2} = 2x^2,\ \ f_{3,0} = y^7,\ \ f_{3,1} = y^2 + 4y^3 + 2y^4, \\ & f_{3,2} = y + 2x + 4xy + 4x^2y,\ \ f_{3,3} = 3x^4,\ \ f_{4,0} = y^{10},\ \ f_{4,1} = 4y^5 + 4y^6 + 2y^7, \\ & f_{4,2} = 4y^2 + 4y^3 + y^4 + 8xy^2 + 6xy^3 + 2x^2y^2 + 4x^2y^4,\ \ f_{4,3} = 4x^2 + 6x^2y + 2x^3 + 8x^3y + 6x^4y, \\ & f_{4,4} = 5x^6,\ \ f_{5,0} = y^{13},\ \ f_{5,1} = 7y^8 + 4y^9 + 2y^{10}, \\ & f_{5,2} = y^3 + 8y^4 + 6xy^4 + 12y^5 + 12xy^5 + 4x^2y^5 + 4y^6 + 4xy^6 + y^7 + 4x^2y^7, \\ & f_{5,3} = 4x^2y^4 + 6x^4y^4 + 4xy^3 + 12x^2y^3 + y^2 + 4xy + 8x^3y^3 + 4x^2y + 12xy^2 + 24x^2y^2 + 12x^3y^2 + 3x^4y^2, \\ & f_{5,4} = 12x^4 + 2x^5 + 18x^4y + 12x^5y + 10x^6y,\ \ f_{5,5} = 8x^8,\end{aligned} \tag{25}$$

will generate, as coefficients of the various powers of x and y, the required degeneracies.

The grand canonical [bivariate-generating] function is

$$g_N(x,y,z) = \sum_{q=0}^{N} f_{N,q}(x,y) z^q. \tag{26}$$

Then $g_N(x,y,z)$ may be found by substituting Eq. (24) into (26):

$$g_N(x,y,z)=g_{N-1}(x,y,z)[2x^2z+y^3]+g_{N-2}(x,y,z)[z(2y+xy^2-2x^2y^3)]+g_{N-3}(x,y,z)$$
$$\times[-x^6z^3+x^2z^2y+y^4z-xzy^5+z^2(x-2x^3y^2)]+g_{N-4}(x,y,z)[x^6y^3z^3+z^3(x^3-2yx^4)+z^2(-5x^2y^4$$
$$+4xy^3+2x^3y^5-y^2)]+g_{N-5}(x,y,z)[z^3(x^4y^4-3x^3y^3+3x^2y^2-xy)+z^4(x^5+x^7y^2-2x^6y)]$$
$$+g_{N-6}(x,y,z)z^4[4x^6y^4-6x^5y^3-x^7y^5+4x^4y^2-x^3y]\,, \tag{27}$$

where the initial conditions are

$$g_1=y+z,\quad g_2=y^4+4y^4z+2x^2z^2,\quad g_3=y^7+(y^2+4y^3+2y^4)z+(y+2x+4xy+4x^2y)z^2+3x^4z^3,$$
$$g_4=y^{10}+(4y^5+4y^6+2y^7)z+(4y^2+4y^3+y^4+8xy^2+5xy^3+2x^2y^2+4x^2y^4)z^2$$
$$+(4x^2+6x^2y+2x^3+8x^3y+6x^4y)z^3+5x^6z^4, \tag{28}$$
$$g_5=y^{13}+(7y^8+4y^9+2y^{10})z+(y^3+8y^4+6xy^4+12y^5+12xy^5+4x^2y^5+4y^6+4xy^6+y^7+4x^2y^7)z^2$$
$$+(4x^2y^4+6x^4y^4+4xy^3+12x^2y^3+y^2+4xy+8x^3y^3+4x^2y+12xy^2+24x^2y^2+12x^3y^2+3x^4y^2)z^3$$
$$+(12x^4+2x^5+18x^4y+12x^5y+10x^6y)z^4+8x^8z^5.$$

To obtain an explicit relation for $g_N(x,y,z)$ we first form the polynomials

$$h(x,y,z,\eta)=\sum_{N=1}^{\infty}g_N(x,y,z)\eta^N=\eta\frac{\gamma(\eta)}{d(\eta)}, \tag{29}$$

where

$$\gamma(\eta)=-g_1+\eta\{-g_2+g_1(2x^2z+y^3)\}+\eta^2\{-g_3-g_2(2x^2z+y^3)+g_1[2yx+xy^2-2x^2y^3z]\}$$
$$+\eta^3\{-g_4+g_3(2x^2z+y^3)+g_2[2y+xy^2-2x^2y^3]z+g_1[-x^6z^3+x^2yz^2+zy^4+z^2(x-2x^3y^2)-zxy^5]\}$$
$$+\eta^4\{-g_5+g_4(2x^2z+y^3)+g_3[2yz+xy^2z-2x^2y^3z]+g_2[-x^6z^3+x^2yz^2+zy^4+z^2(x-3x^3y^2)-zxy^5]$$
$$+g_1[x^6y^3z^3+(x^3-2yx^4)z^3+z^2(-5x^2y^4+4xy^3+2x^3y^5-y^2)]\}+\eta_5\{-g_6+g_5[2x^2z+y^3]+g_4[2yz+xzy^2-2x^2y^3z]$$
$$+g_3[-x^6z^3+x^2yz^2+zy^4+z^2(x-2x^3y^2)-zxy^5]+g_2[x^6y^3z^3+z^3(x^3-2yx^4)+(-5x^2y^4+4xy^3+2x^3y^5-y^2)z^2]$$
$$+g_1[(x^4y^4-3x^3y^3+3x^2y^2-xy)z^3+(x^5+x^7y^2-2x^6y)z^4]\}+\eta^6\{-g_6+g_5[2x^2z+y^3]+g_4[2yz+xy^2z-2x^2y^3z]$$
$$+g_3[-x^6z^3+x^2yz^2+zy^4+z^2(x-2x^3y^2)-zxy^5]+g_2[x^6y^3z^2+z^3(x^3-2yx^4)+z^2(-5x^2y^4+4xy^3+2x^3y^5-y^2)]$$
$$+g_1[z^3(x^4y^4-3x^3y^3+3x^2y^2-xy)+z^4(x^5+x^7y^2-2x^6y)]\}, \tag{30}$$

and

$$d(\eta)=-1+\eta\{2x^2z+y^3\}+\eta^2\{z(2y+xy^2-2x^2y^3)\}+\eta^3\{-x^6z^3+x^2z^2y+zy^4+(x-2x^3y^2)z^2-zxy^5\}$$
$$+\eta^4\{x^6y^3z^3+z^3(x^3-2yx^4)+z^2(-5x^2y^4+4xy^3+2x^3y^5-y^2)\}$$
$$+\eta^5\{z^3(x^4y^4-3x^3y^3+3x^2y^2-xy)+z^4(x^5+x^7y^2-2x^6y)\}$$
$$+\eta^6\{z^4(4x^6y^4-6x^5y^3-x^7y^5+4x^4-x^3y)\}. \tag{31}$$

Using a partial fraction expansion of $h(x,y,z,\eta)$ we obtain

$$g_N(x,y,z)=\sum_{j=1}^{6}K_jR_j^{\,N}, \tag{32}$$

where the K_j's are constants and the R_j's are the reciprocals of the roots of $d(\eta)$. If η_1 is the smallest root, then, as $N\to\infty$, Eq. (32) becomes

$$g_N(x,y,z)\simeq K_1R_1^{\,N}, \tag{33}$$

where $R_1^{-1}=\eta_1$.

IV. COVERAGE

We may determine the expectation of θ, the coverage,

$$\langle\theta\rangle_N=\langle q\rangle_N/N, \tag{34}$$

where

$$\langle q\rangle_N=\frac{\Sigma_{q=0}^{N}qf_{N,q}(x,y)z^q}{\Sigma_{q=0}^{N}g_N(x,y,z)}, \tag{35}$$

or

$$\langle q_N\rangle=z\frac{\partial\ln g_N}{\partial z}. \tag{36}$$

Now, from Eq. (33) and the fact that

$$\langle\theta\rangle=\lim_{N\to\infty}\langle\theta\rangle_N=\lim_{N\to\infty}\langle\theta\rangle_{N-1}=\lim_{N\to\infty}\langle\theta\rangle_{N-2}+\cdots, \tag{37}$$

(where the limit is taken as both N and q increase in such a way that the ratio $\theta\equiv q/N$ remains constant), we have

$$\langle \theta \rangle = -\frac{z}{\eta_1}\frac{\partial \eta_1}{\partial z}. \tag{38}$$

Using $d(\eta)$ to find $\partial\eta/\partial z$, we have

$$\langle \theta \rangle = \frac{\sum_{j=1}^{6} a_j(x,y,z)\eta^j}{\sum_{j=1}^{6} jb_j(x,y,z)\eta^j}, \tag{39}$$

where

$$\begin{aligned}
a_1(x,y,z) &\equiv 2x^2z,\\
a_2(x,y,z) &\equiv z(2y + xy^2 - 2x^2y^3),\\
a_3(x,y,z) &\equiv z[-3x^6z^2 + 2x^2yz + y^4\\
&\quad + 2z(x - 2x^3y^2) - xy^5],\\
a_4(x,y,z) &\equiv z[3z^2x^6y^3 + 3z^2(x^3 - 2yx^4)\\
&\quad + 2z(-5x^2y^4 + 4xy^3 + 2x^3y^5 - y^2)],\\
a_5(x,y,z) &\equiv [3z^3(x^4y^4 - 3x^3y^3 + 3x^2y^2 - xy)\\
&\quad + 4z^4(x^5 + x^7y^2 - 2x^6y)],\\
a_6(x,y,z) &\equiv 4z^4(4x^6y^4 - 6x^5y^3 - x^7y^5 + 4x^4y^2 - x^3y),
\end{aligned} \tag{40}$$

$$\begin{aligned}
b_1(x,y,z) &\equiv 2x^2z + y^3,\\
b_2(x,y,z) &\equiv (2y + xy^2 - 2x^2y^3)z,\\
b_3(x,y,z) &\equiv (-x^6z^3 + x^2yz^2 + zy^4\\
&\quad + z^2(x - 2x^3y^2) - zxy^5),\\
b_4(x,y,z) &\equiv [x^6y^3z^3 + (x^3 - 2x^4y)z^3\\
&\quad + z^2(-5x^2y^4 + 4xy^3 + 2x^3y^5 - y^2)],\\
b_5(x,y,z) &\equiv [z^3(x^4y^4 - 3x^3y^3 + 3x^2y^2 - xy)\\
&\quad + z^4(x^5 + x^7y^2 - 2x^6)],\\
b_6(x,y,z) &\equiv z^4(4x^6y^4 - 6x^5y^3 - x^7y^5 + 4x^4y^2 - x^3y).
\end{aligned} \tag{41}$$

If we let $y = 1$, $x = 1$, and $z = 1$, with no interaction energy and no chemical potential, then

$$\langle \theta \rangle = 0.606\,492\,711,$$

which agrees with a previous result.[2] Figure 7 is a plot of $\langle \theta \rangle$ versus the ln z for several values of x.

V. OCCUPIED NEAREST-NEIGHBOR DENSITY

Now we shall calculate the expectation of θ_{11}, where

$$\theta_{11} = n_{11}/(3N - q - 2). \tag{42}$$

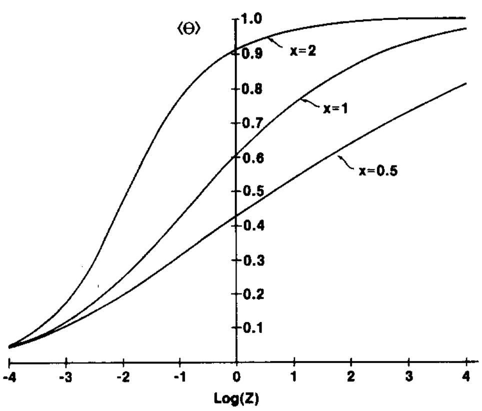


FIG. 7. The coverage as a function of log z for various values of x.

Thus,

$$\langle \theta_{11} \rangle = [1/(3N - q - 2)]\langle n_{11} \rangle, \tag{43}$$

with

$$\langle n_{11} \rangle = \frac{\sum_{q=0}^{N} \sum_{n_{11}} n_{11} A\,[N,q,n_{11},n_{00}]\, x^{n_{11}} y^{n_{00}} z^q}{\sum_{q=0}^{N} \sum_{n_{11}} A\,[N,q,n_{11},n_{00}] x^{n_{11}} y^{n_{00}} z^q}, \tag{44}$$

$$\langle n_{11} \rangle = x\frac{\partial(\ln g_N)}{\partial x},$$

and

$$g_N = K_1 R_1^{\,N}.$$

We have

$$\lim_{N\to\infty} \langle \theta_{11} \rangle = -\left(\frac{xN}{\eta_1}\right)\left(\frac{\partial \eta_1}{\partial x}\right)\left(\frac{1}{3N - q}\right). \tag{45}$$

Finally, using $d(\eta_1) = 0$, we may find $\langle \theta_{11} \rangle$

$$\langle \theta_{11} \rangle = \frac{x}{3-\theta}\,\frac{\sum_{j=1}^{6} c_j(x,y,z)\eta^j}{\sum_{j=1}^{6} jb_j(x,y,z)\eta^j}, \tag{46}$$

where $\theta \equiv q/N$, and

$$\begin{aligned}
c_1(x,y,z) &\equiv 4xz,\\
c_2(x,y,z) &\equiv z(y^2 - 4xy^3),\\
c_3(x,y,z) &\equiv -6x^5z^3 + 2xyz^2 + z^2(1 - 6x^2y^2) - zy^5,\\
c_4(x,y,z) &\equiv z^3(6x^5y^3 + 3x^2 - 8yx^3)\\
&\quad + z^2(-10xy^4 + 4y^3 + 6x^2y^5),\\
c_5(x,y,z) &\equiv z^3(4x^3y^4 - 9x^2y^3 + 6xy^2 - y)\\
&\quad + z^4(5x^4 + 7x^6y^2 - 12x^5y),\\
c_6(x,y,z) &\equiv z^4(24x^5y^4 - 30x^4y^3 - 7x^6y^5\\
&\quad + 16x^3y^2 - 3x^2y).
\end{aligned} \tag{47}$$

Note Fig. 8 for a plot of $\langle \theta_{11} \rangle$ vs $\langle \theta \rangle$, for several values of x.

VI. CONCLUSION

We have determined exactly the recursion relation that generates the number of ways of having q dumbbells on a $2\times N$ array, such tht there are n_{00} vacant pairs and n_{11} occupied nearest-neighbor pairs. We have calculated the grand canonical partition function, which enabled us to find the coverage and occupied nearest-neighbor density. We also provide plots of coverage as a function of the various activities and the nearest-occupied-neighbor density versus coverage.

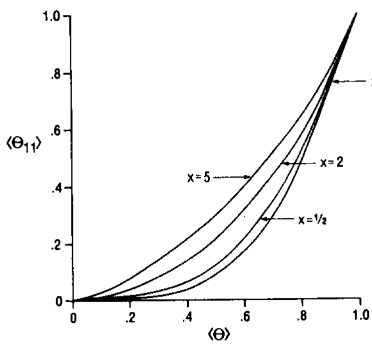


FIG. 8. The occupied nearest-neighbor density as a function of coverage for various values of x.

It should be noted from Eqs. (38) and (45), that all the statistically relevant quantities can be calculated from the determinant of the shift operator matrix [see Eq. (18)]. We make the association of x, y, z, and η with R, S, T, and U, respectively, and thereby obtain $d(\eta)$, the denominator of $h(x, y, z, \eta)$ [see Eq. (29)]. From $d(\eta)$ we may determine $\langle\theta\rangle$ and $\langle\theta_{11}\rangle$ by implicit differentiation. Equation (38) becomes

$$\langle\theta\rangle = +\left\{\frac{z(\partial d/\partial z)}{\eta(\partial d/\partial\eta)}\right\}_{\eta=\eta_1},$$

and Eq. (45) may be written

$$\langle\theta_{11}\rangle = +\left\{\frac{x(\partial d/\partial x)}{\eta(\partial d/\partial\eta)}\right\}_{\eta=\eta_1}\left(\frac{1}{3-\theta}\right).$$

ACKNOWLEDGMENT

The research reported in this paper has been supported in part by the Air Force Office of Scientific Research under Grant No. 81-0192.

[1]R. B. McQuistan and J. L. Hock, J. Math. Phys. **25**, 261 (1984).
[2]R. B. McQuistan and S. J. Lichtman, J. Math. Phys. **11**, 3095 (1970).

Lie–Bäcklund transformations for the massive Thirring model

Paul H. M. Kersten and Ruud Martini
Twente University of Technology, Department of Applied Mathematics, P.O. Box 217, 7500 AE Enschede, The Netherlands

(Received 26 June 1984; accepted for publication 24 August 1984)

Second- and third-order Lie–Bäcklund transformations of the massive Thirring model are computed by symbolic integration.

I. INTRODUCTION AND GENERAL

In two recent papers, Gragert *et al.*[1] and Kersten and Gragert[2] demonstrated how the infinitesimal symmetries of certain partial differential equations can be computed in a semiautomatic way using symbolic integration.

These symmetries have to satisfy the condition

$$\mathscr{L}_V \subset I, \tag{1.1}$$

where V is a vector field, $\mathscr{L}_V$ denotes the Lie derivative with respect to the vector field V, and I denotes a closed ideal of differential forms describing the partial differential equation.[3]

Using the local jet bundle formalism,[4] (1.1) can be generalized to Lie–Bäcklund transformations which have to satisfy

$$\mathscr{L}_V(D^\infty I) \subset D^\infty I, \tag{1.2}$$

where $D^\infty I$ is the infinitely prolonged ideal in the jet bundle J^∞ obtained by total partial differentiation.

Now if

$$\mathscr{L}_V(I) \subset D^\infty I, \tag{1.3}$$

then, since the total derivative fields D_x, D_t commute, V satisfies (1.2); i.e., condition (1.3) is equivalent to condition (1.2).

Moreover, due to the commuting of these vector fields, and since D_x, D_t satisfy (1.2) in an obvious way, we can restrict our search for Lie–Bäcklund transformations to vertical vector fields.[4]

It can be shown that, assuming the first component of such a vector field depends on $x,t,\ldots,u_i$ $[i=(i_1,i_2),$ $|i| = i_1 + i_2 = k]$, i.e., is dependent on the variables up to the kth order derivatives in J^∞, (1.3) reduces to the condition

$$\mathscr{L}_V I \subset D^k I, \tag{1.4}$$

leading to conditions equivalent to those of Anderson and Ibragimov[5] (pp. 62 and 63).

Condition (1.4) results in an overdetermined system of partial differential equations for the defining coefficients of the Lie–Bäcklund transformation V.

In Kersten[6] the method of determining Lie–Bäcklund transformations is demonstrated for Burgers' equation, leading to results which are in agreement with those of Vinogradov.[7]

In Sec. II Lie–Bäcklund transformations for the massive Thirring model[8] are computed. In the derivation of the results, which seem to be new, we introduced a grading of the equations. For more details of the computation we refer to Kersten.[6]

II. LIE–BÄCKLUND TRANSFORMATIONS FOR THE MASSIVE THIRRING MODEL

We shall establish the existence of Lie–Bäcklund transformations for the massive Thirring model, which can be defined as the following system of partial differential equations for the unknown functions $u_1(x,t), v_1(x,t), u_2(x,t), v_2(x,t)$:

$$\begin{aligned}
-\frac{\partial u_1}{\partial x} + \frac{\partial u_1}{\partial t} &= m v_2 - (u_2^2 + v_2^2)v_1,\\
\frac{\partial u_2}{\partial x} + \frac{\partial u_2}{\partial t} &= m v_1 - (u_1^2 + v_1^2)v_2,\\
\frac{\partial v_1}{\partial x} - \frac{\partial v_1}{\partial t} &= m u_2 - (u_2^2 + v_2^2)u_1,\\
-\frac{\partial v_2}{\partial x} - \frac{\partial v_2}{\partial t} &= m u_1 - (u_1^2 + v_1^2)u_2.
\end{aligned} \tag{2.1}$$

The ideal I in

$$\mathbb{R}^{10} = \{(x,t,u_1,u_2,v_1,v_2,u_{1x},u_{2x},v_{1x},v_{2x})\}$$

describing (2.1) is generated by the four one-forms

$$\begin{aligned}
\alpha_1 &= du_1 - u_{1x}dx - G(11)dt,\\
\alpha_2 &= du_2 - u_{2x}dx - G(21)dt,\\
\alpha_3 &= dv_1 - v_{1x}dx - G(31)dt,\\
\alpha_4 &= dv_2 - v_{2x}dx - G(41)dt,
\end{aligned} \tag{2.2}$$

where $G(11),\ldots,G(41)$ are defined by

$$\begin{aligned}
G(11) &= u_{1x} + m v_2 - (u_2^2 + v_2^2)v_1,\\
G(21) &= -u_{2x} + m v_1 - (u_1^2 + v_1^2)v_2,\\
G(31) &= v_{1x} - m u_2 + (u_2^2 + v_2^2)u_1,\\
G(41) &= -v_{2x} - m u_1 + (u_1^2 + v_1^2)u_2.
\end{aligned} \tag{2.3}$$

The 2 times prolonged ideal $D^2 I$ in $\mathbb{R}^{18}$ is given by

$$\alpha_1,\alpha_2,\alpha_3,\alpha_4 \tag{2.2}$$

$$\begin{aligned}
\alpha_5 &= du_{1x} - u_{1xx}dx - G(12)dt,\\
\alpha_6 &= du_{2x} - u_{2xx}dx - G(22)dt,\\
\alpha_7 &= dv_{1x} - v_{1xx}dx - G(32)dt,\\
\alpha_8 &= dv_{2x} - v_{2xx}dx - G(42)dt,\\
\alpha_9 &= du_{1xx} - u_{1xxx}dx - G(13)dt,\\
\alpha_{10} &= du_{2xx} - u_{2xxx}dx - G(23)dt,\\
\alpha_{11} &= dv_{1xx} - v_{1xxx}dx - G(33)dt,\\
\alpha_{12} &= dv_{2xx} - v_{2xxx}dx - G(43)dt,
\end{aligned} \tag{2.4}$$

where the coefficients $G(**)$ are derived by total partial differentiation of $G(11), G(21), G(31), G(41)$ with respect to x.

Now, the vector field

$$V = F(3)\partial_{u_1} + F(4)\partial_{u_2} + F(5)\partial_{v_1} + F(6)\partial_{v_2} + \text{pr}, \tag{2.5}$$

 0022-2488/85/040822-04$02.50

where "pr" represents the prolongation of V^4, is a Lie–Bäcklund transformation for (2.1) if

$$\mathscr{L}_V I \subset D^2 I, \tag{2.6}$$

which leads to an overdetermined system of $2\times 4 = 8$ partial differential equations for $F(3)$, $F(4)$, $F(5)$, and $F(6)$, the defining coefficients of the vector field V (2.5).

If we want to search for the vertical vector fields equivalent to the classical symmetries of (2.1), we require $F(3),...,F(6)$ to be dependent on $x,t,u_1,...,v_{2x}$.

A straightforward computation then leads to the following four symmetries:

$$X_i = X_i^3\,\partial_{u_1} + X_i^4\,\partial_{v_1} + X_i^5\,\partial_{u_2} + X_i^6\,\partial_{v_2} \quad (i = 1,...,4),$$

where

$$\begin{aligned}
X_1^3 &= \tfrac{1}{2}(-mv_2 + v_1(u_2^2 + v_2^2)),\\
X_1^4 &= \tfrac{1}{2}(2u_{2x} - mv_1 + v_2(u_1^2 + v_1^2)),\\
X_1^5 &= \tfrac{1}{2}(mu_2 - u_1(u_2^2 + v_2^2)),\\
X_1^6 &= \tfrac{1}{2}(2v_{2x} + mu_1 - u_2(u_1^2 + v_1^2));\\
X_2^3 &= \tfrac{1}{2}(2u_{1x} + mv_2 - v_1(u_2^2 + v_2^2)),\\
X_2^4 &= \tfrac{1}{2}(mv_1 - v_2(u_1^2 + v_1^2)),\\
X_2^5 &= \tfrac{1}{2}(2v_{1x} - mu_2 + u_1(u_2^2 + v_2^2)),\\
X_2^6 &= \tfrac{1}{2}(-mu_1 + u_2(u_1^2 + v_1^2));\\
X_3^3 &= u_{1x}(x+t) + mv_2x + \tfrac{1}{2}u_1 - v_1(u_2^2 + v_2^2)x,\\
X_3^4 &= u_{2x}(-x+t) + mv_1x - \tfrac{1}{2}u_2 - v_2(u_1^2 + v_1^2)x,\\
X_3^5 &= v_{1x}(x+t) - mu_2x + \tfrac{1}{2}v_1 + u_1(u_2^2 + v_2^2)x,\\
X_3^6 &= v_{2x}(-x+t) - mu_1x - \tfrac{1}{2}v_2 + u_2(u_1^2 + v_1^2)x;\\
X_4^3 &= v_1;\quad X_4^4 = v_2;\quad X_4^5 = -u_1;\quad X_4^6 = -u_2.
\end{aligned} \tag{2.7}$$

In order to find Lie–Bäcklund transformations for (2.1) we introduced a grading for (2.1) in the following way:

$$\begin{aligned}
&\deg(x) = -2,\\
&\deg(t) = -2,\\
&\deg(u_1) = \deg(u_2) = \deg(v_1) = \deg(v_2) = 1,\\
&\deg(m) = 2.
\end{aligned} \tag{2.8}$$

By (2.8) each term in (2.1) is of degree 3.

Now, since the total partial differentiation operators are graded $\deg(D_x) = \deg(D_t) = 2$, solutions of the Lie–Bäcklund symmetry condition are graded correspondingly.

We introduce the following notation:

$$\begin{aligned}
&[u] \text{ refers to } u_1,u_2,v_1, \text{ or } v_2,\\
&[u]_x \text{ refers to } u_{1x},u_{2x},v_{2x}, \text{ or } v_{2x}.
\end{aligned} \tag{2.9}$$

In the search for Lie–Bäcklund transformations we did not construct the general solution of the overdetermined system of partial differential equations but restricted our search to those, induced by (2.8) and (2.9), which motivated us to seek a solution in the following way:

$$\begin{aligned}
F(J) := [u]_{xx} &+ ([u]^2 + [m])[u]_x\\
&+ ([u]^5 + [m][u]^3 + [m]^2[u]) \quad (J = 3,...,6),
\end{aligned} \tag{2.10}$$

i.e., a vector field whose defining coefficients are of degree 5.

In fact, we only introduced the maximal power of the coefficient of $[u]_x$, i.e., $[u]^2$, and the maximal power of $[u]$, i.e, $[u]^5$, into the overdetermined system of partial differential equations.[6]

Using an integration package,[2] we found two Lie–Bäcklund transformations X_5,X_6 given by

$$\begin{aligned}
X_5^3 &= \tfrac{1}{4}\{2u_{2x}(-m + 2v_1v_2) - 4v_{2x}u_2v_1 - mv_2(R_1 + R_2)\\
&\quad - 2mv_1R + v_1(R_2^2 + 2R_1R_2)\},\\
X_5^4 &= \tfrac{1}{4}\{-4v_{2xx} + 2u_{1x}(-m + 2u_1u_2)\\
&\quad + 4u_{2x}(R_1 + R_2) + 4v_{1x}u_2v_1 - mv_1(R_1 + R_2)\\
&\quad - 2mv_2R + v_2(R_1^2 + 2R_1R_2)\},
\end{aligned} \tag{2.11a}$$

$$\begin{aligned}
X_5^5 &= \tfrac{1}{4}\{2v_{2x}(-m + 2u_1u_2) - 4u_{2x}u_1v_2 + mu_2(R_1 + R_2)\\
&\quad + 2mu_1R - u_1(R_2^2 + 2R_1R_2)\},\\
X_5^6 &= \tfrac{1}{4}\{4u_{2xx} + 2v_{1x}(-m + 2v_1v_2)\\
&\quad + 4v_{2x}(R_1 + R_2) + 4u_{1x}u_1v_2\\
&\quad + mu_1(R_1 + R_2) + 2mu_2R - u_2(R_1^2 + 2R_1R_2)\},
\end{aligned}$$

and

$$\begin{aligned}
X_6^3 &= \tfrac{1}{4}\{4v_{1xx} + 2u_{2x}(-m + 2u_1u_2)\\
&\quad + 4u_{1x}(R_1 + R_2) + 4v_{2x}u_1v_2\\
&\quad + mv_2(R_1 + R_2) + 2mv_1R - v_1(R_2^2 + 2R_1R_2)\},\\
X_6^4 &= \tfrac{1}{4}\{2u_{1x}(-m + 2v_1v_2) - 4v_{1x}u_1v_2 + mv_1(R_1 + R_2)\\
&\quad + 2mv_2R - v_2(R_1^2 + 2R_1R_2)\},
\end{aligned} \tag{2.11b}$$

$$\begin{aligned}
X_6^5 &= \tfrac{1}{4}\{-4u_{1xx} + 2v_{2x}(-m + 2v_1v_2) + 4v_{1x}(R_1 + R_2)\\
&\quad + 4u_{2x}u_2v_1 - mu_2(R_1 + R_2) - 2mu_1R\\
&\quad + u_1(R_2^2 + 2R_1R_2)\},\\
X_6^6 &= \tfrac{1}{4}\{2v_{1x}(-m + 2u_1u_2) - 4u_{1x}u_2v_1 - mu_1(R_1 + R_2)\\
&\quad - 2mu_2R + u_2(R_1^2 + 2R_1R_2)\},
\end{aligned}$$

whereas in (2.11a) and (2.11b)

$$R = u_1u_2 + v_1v_2,\quad R_1 = u_1^2 + v_1^2,\quad R_2 = u_2^2 + v_2^2. \tag{2.12}$$

In order to find the third-order Lie–Bäcklund transformation we have to prolong the ideal D^2I once more.

In the search for third-order results we restricted ourselves to vector fields whose defining coefficients are schematically given by

$$\begin{aligned}
F(J) = [u]_{xxx} &+ ([u]^2 + [m])[u]_{xx} + [u][u]_x^2\\
&+ ([u]^4 + [u]^2[m] + [m]^2)[u]_x\\
&+ ([u]^7 + [u]^5[m] + [u]^3[m]^2 + [u][m]^3)\\
&(J := 3:6).
\end{aligned} \tag{2.13}$$

After a massive amount of computations we obtained two additional Lie–Bäcklund transformations X_7, X_8; i.e.,

$$\begin{aligned}
X_7^3 = \tfrac{1}{8}\{ & 8u_{2xx}u_2v_1 + 4v_{2xx}(2v_1v_2 - m) - 4u_{2x}^2 v_1 + 4u_{2x}(m(R_1+R_2+v_1^2+v_2^2) - 3v_1v_2(R_1+R_2)) - 4v_{2x}^2 v_1 \\
& + 4v_{2x}(-(u_1v_1+u_2v_2)m + 3u_2v_1(R_1+R_2)) + 4u_{1x}mR - 2m^2v_1(R_1+R_2) - 4v_2m^2R + 4v_1mR\,(R_1+2R_2) \\
& + v_2m(R_1^2+4R_1R_2+R_2^2) - v_1(R_2^3 + 6R_2^2R_1 + 3R_2R_1^2)\}, \\
X_7^4 = \tfrac{1}{8}\{ & 8u_{2xxx} + 12v_{2xx}(R_1+R_2) + 8u_{1xx}u_1v_2 + 4v_{1xx}(2v_1v_2 - m) - 12u_{2x}^2v_2 + 24u_{2x}v_{2x}u_2 \\
& + 2u_{2x}(10mR - 3R_1^2 - 12R_1R_2 - 3R_2^2) + 12v_{2x}^2v_2 + 24v_{2x}u_{1x}u_1 + 24v_{2x}v_{1x}v_1 + 8u_{1x}^2v_2 \\
& + 4u_{1x}(m(R_1+R_2+u_1^2+u_2^2) - 3u_1u_2(R_1+R_2)) + 8v_{1x}^2v_2 + 4v_{1x}(m(u_1v_1+u_2v_2) - 3u_2v_1(R_1+R_2)) \\
& - 4m^2v_1R - 2m^2v_2(R_1+R_2) + mv_1(R_2^2+4R_1R_2-R_1^2) + 4mv_2R\,(R_2+2R_1) - v_2(R_1^3+6R_1^2R_2+3R_1R_2^2)\}, \\
X_7^5 = \tfrac{1}{8}\{ & -8v_{2xx}v_2u_1 - 4v_{2xx}(u_1u_2 - m) + 4v_{2x}^2u_1 + 4v_{2x}(m(R_1+R_2+u_1^2+u_2^2) - 3u_1u_2(R_1+R_2)) + 4u_{2x}^2u_1 \\
& + 4u_{2x}(-(u_1v_1+u_2v_2)m + 3v_2u_1(R_1+R_2)) + 4v_{1x}mR + 2m^2u_1(R_1+R_2) + 4u_2m^2R \\
& - 4u_1mR\,(R_1+2R_2) - u_2m(R_1^2+4R_1R_2+R_2^2) + u_1(R_2^3+6R_2^2R_1+3R_2R_1^2)\}, \\
X_7^6 = \tfrac{1}{8}\{ & 8v_{2xxx} - 12u_{2xx}(R_1+R_2) + 8v_{1xx}u_2v_1 - 4u_{1xx}(2u_1u_2 - m) + 12v_{2x}^2u_2 - 24u_{2x}v_{2x}v_2 \\
& + 2v_{2x}(10mR - 3R_1^2 - 12R_1R_2 - 3R_2^2) - 12u_{2x}^2u_2 - 24u_{2x}v_{1x}v_1 + 24u_{2x}u_{1x}u_1 - 8v_{1x}^2u_2 \\
& + 4v_{1x}(m(R_1+R_2+v_1^2+v_2^2) - 3v_1v_2(R_1+R_2)) - 8u_{1x}^2u_2 + 4u_{1x}(m(u_1v_1+u_2v_2) - 3u_1v_2(R_1+R_2)) \\
& + 4m^2u_1R + 2m^2u_2(R_1+R_2) - mu_1(R_2^2+4R_1R_2+R_1^2) \\
& - 4mu_2R\,(R_2+2R_1) + u_2(R_1^3+6R_1^2R_2+3R_1R_2^2)\},
\end{aligned} \tag{2.14}$$

while X_8^i $(i = 3,...,6)$ can be derived from X_7^j $(j = 3,...,6)$ by the transformation[6]

$$\begin{aligned} &T{:}u_1\to u_2,\ u_2\to u_1,\ v_1\to v_2,\ v_2\to v_1,\ \partial_x\to -\partial_x \\ &(R_1\to R_2,\ R_2\to R_1,\ R\to R) \end{aligned} \tag{2.15}$$

in the following way:

$$\begin{aligned} &X_8^3 = -T(X_7^4);\quad X_8^4 = -T(X_7^3); \\ &X_8^5 = -T(X_7^6);\quad X_8^6 = -T(X_7^5). \end{aligned} \tag{2.16}$$

III. THE LIE ALGEBRA STRUCTURE OF THE LIE–BÄCKLUND TRANSFORMATIONS $X_1,...,X_8$

The Lie bracket for the vertical vector fields X_i, defined by

$$X_i = X_i^3\,\partial_{u_1} + X_i^4\,\partial_{u_2} + X_i^5\,\partial_{v_1} + X_i^6\,\partial_{v_2} + \mathrm{pr}, \tag{3.1}$$

is given by

$$[X_j,X_k]^1 = X_j(X_k^1) - X_k(X_j^1) \quad (1 = 3,...,6). \tag{3.2}$$

In (3.2) only the $\partial_{u_1},...,\partial_{v_2}$ components of the commutator of two vector fields are defined, while the other components are derived by total differentiation[4].

Computation of (3.2) for the vector fields $X_1,...,X_8$, given in (3.7), (2.11), (2.14), and (2.16) results in the following nonzero commutators:

$$\begin{aligned} &[X_1,X_3] = X_1,\quad [X_2,X_3] = -X_2, \\ &[X_3,X_5] = -2X_5 - (m^2/2)X_4 \\ &[X_3,X_6] = 2X_6 - (m^2/2)X_4, \\ &[X_3,X_7] = -3X_7 + m^2(X_1+X_2), \\ &[X_3,X_8] = 3X_8 - m^2(X_1+X_2). \end{aligned} \tag{3.3}$$

Transformation of the basis-vector fields of the Lie algebra by

$$\begin{aligned} &Y_1 = X_1;\quad Y_2 = X_2;\quad Y_3 = X_3;\quad Y_4 = X_4; \\ &Y_5 = X_5 + (m^2/4)X_4;\quad Y_6 = X_6 - (m^2/4)X_4; \\ &Y_7 = X_7 - (m^2/2)X_1 - (m^2/4)X_2; \\ &Y_8 = X_8 - (m^2/4)X_1 - (m^2/2)X_2, \end{aligned} \tag{3.4}$$

then leads to Table I. From (3.3) and Table I we see $[Y_i,Y_j] = 0$ $(i,j = 1,2,5,6,7,8)$; i.e., the Lie–Bäcklund transformations commute.

IV. CONCLUSION

By symbolic integration and grading of (2.1) we obtained second- and third-order Lie–Bäcklund transformations for the massive Thirring model.

TABLE I. Commutators of Y_1, ..., Y_8.

		$j\to$							
	$[Y_i,Y_j]$	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8
	Y_1	*	0	Y_1	0	0	0	0	0
	Y_2		*	$-Y_2$	0	0	0	0	0
	Y_3			*	0	$-2Y_5$	$2Y_6$	$-3Y_7$	$3Y_8$
i	Y_4				*	0	0	0	0
$\downarrow$	Y_5					*	0	0	0
	Y_6						*	0	0
	Y_7							*	0
	Y_8								*

[1]P. K. H. Gragert, P. H. M. Kersten, and R. Martini, Acta Appl. Math. **1**, 43 (1983).
[2]P. H. M. Kersten and P. K. H. Gragert, "Symbolic integration of overdetermined systems of partial differential equations," Memorandum 430, Twente University of Technology, 1983.
[3]B. K. Harrison and F. B. Estabrook, J. Math. Phys. **12**, 653 (1971).
[4]F. A. E. Pirani, D. C. Robinson, and W. F. Shadwick, "Local Jet Bundle Formulation of Bäcklund Transformations," *Mathematical Physics Studies 1, A Supplementary Series to Letters in Mathematical Physics* (Reidel, Dordrecht, 1979), Vol. 1.
[5]R. L. Anderson and N. H. Ibragimov, *Lie–Bäcklund Transformations in Applications* (SIAM, Philadelphia, 1979).
[6]P. H. M. Kersten, "Lie–Bäcklund symmetries for the Massive Thirring Model and Burgers' equation," Memorandum 469, Twente University of Technology, 1984.
[7]A. M. Vinogradov, Acta Appl. Math. **2**, 21 (1984).
[8]D. J. Kaup and A. C. Newell, Nuovo Cimento Lett. **20**, 325 (1977).

Radially separated classical lumps in non-Abelian gauge models

Jürgen Burzlaff[a)]
School of Theoretical Physics, Dublin Institute for Advanced Studies, 10 Burlington Road, Dublin 4, Ireland

(Received 7 March 1984; accepted for publication 16 November 1984)

We search for smooth and time-independent finite-energy solutions to Yang–Mills–Higgs theory with an arbitrary compact gauge group. Excluding the monopole solutions which have been studied before, we concentrate on configurations with no long-range fields, which include the saddle points corresponding to noncontractible (hyper-) loops. It is shown that if the radial dependence of the fields is factorized, only one solution satisfies all these conditions. This solution is the one which has been studied before by Dashen, Hasslacher, and Neveu and by Boguta, and whose existence has recently been proved rigorously. Formulas for the asymptotic behavior of this solution are given.

I. INTRODUCTION

Classical lumps, i.e., smooth and nondissipative finite-energy solutions, are very interesting objects from a mathematical point of view. As long as the corresponding quantum field theory is missing, they also provide one of the few clues to an understanding of the nonperturbative sector of gauge field theories. Recently, Manton[1] has argued that the existence of noncontractible loops of finite-energy configuration in the classical, bosonic Weinberg–Salam theory probably implies the existence of an unstable classical lump and the breakdown of perturbation theory in the TeV region. Since many gauge models have noncontractible (hyper-) loops[2] and therefore, in this sense, nontrivial topology, we search for the corresponding saddle points, restricting our attention to time-independent fields whose radial dependence is factorized. The time independence guarantees that the solution is nondissipative. Separability of the *Ansatz* seems to be necessary to separate the equations, solve the angular equations, and prove existence of a solution of the radial equations.

Configurations which satisfy all these conditions and furthermore have long-range fields, have been studied by Shankar,[3] by Michel *et al.*,[4] and by O'Raifeartaigh and Rawnsley.[5] These authors find different solutions for a $(2n+1)$-tuple $(n=1, 2, \ldots)$ of real Higgs fields, thus generalizing the 't Hooft–Polyakov monopole solution[6] which is the first in this series. For configurations which do not have long-range fields and therefore do not describe magnetic monopoles, at least one more finite-energy solution has to be added. This solution, with a complex Higgs doublet, has been studied by Dashen *et al.*[7] and by Boguta,[8] and its existence has recently been proved rigorously.[9] Boguta argues that this solution is responsible for the anomalons produced in heavy-ion collisions. The purpose of this paper is to extend the analysis of Michel *et al.*[4] to configurations with no long-range fields and find out whether the solution just mentioned is also the first in a whole series of saddle points.

To this end, we separate the equations of motion into equations for angular and radial functions. Because this is simple and illustrative and makes the paper self-contained we go through the derivation instead of specializing the formulas of Michel *et al.*,[4] which up to this point are still general and thus contain the case of no long-range fields. We then solve the angular equations in this case and find that they have only one solution. This proves the uniqueness of the classical lump with a complex Higgs doublet under the conditions of time independence, finite energy, separability, and vanishing long-range fields. We end our discussion by extracting the asymptotic behavior of this classical lump from the radial equations.

II. SEPARATION OF THE YANG–MILLS–HIGGS EQUATIONS

We consider gauge theories with arbitrary compact Lie group G and Lagrangian density

$$\mathscr{L} = -\tfrac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a - \tfrac{1}{2}(D_\mu\phi)^T(D^\mu\phi) - V(\phi) \tag{2.1}$$

[$\mu, \nu = 0, 1, 2, 3$, metric diag $(+1, -1, -1, -1)$]. The gauge fields are defined as

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + e[A_\mu, A_\nu], \tag{2.2}$$

in terms of the potentials $A_\mu = A^a_\mu \tau_a$, which belong to the adjoint representation given by the antisymmetric matrices τ_a. The matrix elements of τ_a are the structure constants of G:

$$(\tau_b)_{ac} = C_{abc}. \tag{2.3}$$

Without loss of generality we assume that the Higgs fields ϕ belong to an arbitrary real (not necessarily irreducible) representation of G. For an l-tuple of complex Higgs fields this amounts to formulating the theory in terms of the corresponding $2l$-tuple of real Higgs fields. We denote the antisymmetric matrices of the representation to which ϕ belongs by t_a and define the covariant derivative by

$$D_\mu\phi = \partial_\mu\phi + eA^a_\mu t_a\phi, \tag{2.4}$$

with coupling constant e. Furthermore we assume that the potential $V(\phi)$ in (2.1) is a fourth-degree polynomial in ϕ which is bounded below and invariant under G, and that all the group invariants in $V(\phi)$ tend to finite constants for large r.

From the Lagrangian density (2.1) we derive the Yang–Mills–Higgs equations

$$D_\mu D^\mu\phi = -\frac{\partial V}{\partial\phi}, \tag{2.5}$$

[a)] Feodor Lynen Fellow of the Alexander von Humboldt Foundation.

 0022-2488/85/040826-05$02.50

$$D_\mu F^{\nu\mu} := \partial_\mu F^{\nu\mu} + e[A_\mu, F^{\nu\mu}] = -e\phi^T t_a D^\nu \phi \tau_a. \tag{2.6}$$

We seek regular static finite-energy solutions to these equations with $A_0 = 0$ for which the radial dependence of the fields is factorized in the Coulomb gauge $\partial_i A_i = 0$. In particular we assume that

$$\lim_{r\to\infty} \phi(r,\omega) = V_0 \varphi(\omega) \tag{2.7}$$

and

$$\lim_{r\to\infty} erA_i(r,\omega) = a_i(\omega) \tag{2.8}$$

exist which is compatible with the finite-energy condition. Then $|\varphi|^2 = 1$ holds without loss of generality because the group invariant $|\phi|^2$ was assumed to tend to a finite constant at infinity. Together with the separability condition this leads to the *Ansatz*

$$\phi(r,\omega) = [h(r)/er]\varphi(\omega), \quad |\varphi|^2 = 1, \tag{2.9}$$

$$A_i(r,\omega) = \{[2 - K(r)]/2er\} a_i(\omega), \tag{2.10}$$

with the boundary conditions

$$\lim_{r\to\infty} [h(r)/er] = V_0, \quad \lim_{r\to\infty} K(r) = 0. \tag{2.11}$$

For nonvanishing fields

$$f_{ij} = r\,\partial_i a_j - r\,\partial_j a_i + [a_i a_j] + \hat{x}_j a_i - \hat{x}_i a_j \tag{2.12}$$

on the sphere at infinity, all regular finite-energy solutions of the form (2.9) and (2.10) have been found.[4] We can therefore restrict our attention to the case $f_{ij} = 0$. In this case, a_i can be written

$$a_i = -r(\partial_i \Omega)\Omega^{-1}, \quad \Omega(\omega) \in G. \tag{2.13}$$

Using the finite-energy condition in the form

$$r\,\partial_i \varphi + a_i^b t_b \varphi = 0, \tag{2.14}$$

we conclude that

$$\varphi(\omega) = W(\omega)\varphi_0, \quad -r\,\partial_i W = a_i^b t_b W \tag{2.15}$$

holds. Our choice of gauge furthermore yields

$$\partial_i\,\partial_i \Omega - (\partial_i \Omega)\Omega^{-1}(\partial_i \Omega) = 0. \tag{2.16}$$

The *Ansatz* (2.9), (2.10), (2.13), (2.15), and (2.16) reduces the equation of motion (2.5) to

$$-\tfrac{1}{4} h K^2 (L^2 W)\varphi_0 = [r^2 h'' - \lambda r^2 h U(h/er)] W\varphi_0, \tag{2.17}$$

where L_i is the orbital angular momentum operator

$$L_i = -\epsilon_{ijk} x_j \partial_k, \quad L^2 = L_i L_i, \tag{2.18}$$

and U is of the form[4]

$$U(h/er) = (h/er - V_0)(\sigma^2 h/er + V_0), \tag{2.19}$$

with a dimensionless parameter $\sigma \geqslant 1$. Equation (2.17) yields for nonvanishing K

$$L^2 W\varphi_0 = -l(l+1)W\varphi_0, \quad l = 0, 1, 2, \ldots, \tag{2.20}$$

$$r^2 h'' = \tfrac{1}{4} l(l+1) h K^2 + \lambda r^2 h U(h/er). \tag{2.21}$$

For $G = SU(2)$ and a four-tuple of real Higgs fields, Eq. (2.20) is enough to show that $l = 0$, 1 and that there is only one nontrivial solution.[10] In our general case we also have to consider Eq. (2.6).

To reduce the equation of motion (2.6) for our *Ansatz* we use the condition (2.16) and obtain

$$\begin{aligned} r^2 K''\,\partial_i \Omega\Omega^{-1} &= \tfrac{1}{2} K(K-2) r^2 [\partial_j(\partial_i \Omega\Omega^{-1}), \partial_j \Omega\Omega^{-1}] \\ &\quad + \tfrac{1}{4} K(K-2)^2 r^2 [\partial_j \Omega\Omega^{-1}, [\partial_i \Omega\Omega^{-1}, \partial_j \Omega\Omega^{-1}]] \\ &\quad - h^2 K(\varphi_0^T W^T t_a\, \partial_i W\varphi_0)\tau_a. \end{aligned} \tag{2.22}$$

For the $\alpha_i^b(\omega)$, which are orthogonal to the $a_i^b(\omega)$, this equation implies

$$\begin{aligned} &\tfrac{1}{2} K(K-2) r^2 \operatorname{tr}\{\alpha_i^b \tau_b [\partial_j(\partial_i \Omega\Omega^{-1}), \partial_j \Omega\Omega^{-1}]\} \\ &\quad + \tfrac{1}{4} K(K-2)^2 r^2 \operatorname{tr}\{\alpha_i^b \tau_b [\partial_j \Omega\Omega^{-1}, \\ &\quad \times [\partial_i \Omega\Omega^{-1}, \partial_j \Omega\Omega^{-1}]]\} \\ &\quad - h^2 K(\varphi_0^T W^T t_a\, \partial_i W\varphi_0) \operatorname{tr}(\alpha_i^b \tau_b \tau_a) = 0. \end{aligned} \tag{2.23}$$

Since h goes like r asymptotically and the different powers of K are linearly independent functions for nonvanishing K, we conclude from Eq. (2.23) that their r independent coefficient functions vanish. This is true for all orthogonal $\alpha_i^b(\omega)$ and therefore

$$\begin{aligned} -r^2\,\partial_j\,\partial_j(r\,\partial_i \Omega\Omega^{-1}) &= r^3[\partial_j(\partial_i \Omega\Omega^{-1}), \partial_j \Omega\Omega^{-1}] \\ &= L(L+1) r\,\partial_i \Omega\Omega^{-1}, \end{aligned} \tag{2.24}$$

$$\begin{aligned} (a_j a_j)^b_c a_i^c \tau_b &= r^3[\partial_j \Omega\Omega^{-1}, [\partial_i \Omega\Omega^{-1}, \partial_j \Omega\Omega^{-1}]] \\ &= (\nu^2/N) r\,\partial_i \Omega\Omega^{-1}, \end{aligned} \tag{2.25}$$

$$(\varphi_0^T W^T t_a r\,\partial_i W\varphi_0)\tau_a = -[l(l+1)/\nu^2] r\,\partial_i \Omega\Omega^{-1} \tag{2.26}$$

follows. Here we have written the coefficient functions also in a different form to show that $L = 0, 1, 2, \ldots$, and $N > 0$ hold. We have that ν^2 is positive because

$$\nu^2 = -(1/\kappa^2) r^2 \operatorname{tr}(\partial_i \Omega\Omega^{-1} \partial_i \Omega\Omega^{-1}), \quad \operatorname{tr}(\tau_a \tau_b) = -\kappa^2 \delta_{ab} \tag{2.27}$$

follows from

$$r^2(\varphi_0^T W^T t_a\, \partial_i W\varphi_0) \operatorname{tr}(\partial_i \Omega\Omega^{-1}\tau_a) = -\kappa^2 \varphi_0^T W^T L^2 W\varphi_0 \tag{2.28}$$

and (2.20).

So far, we have three different sets (L, N, ν^2) of functions of ω for $i = 1, 2, 3$ in Eqs. (2.24)–(2.26), whose index i we have suppressed. Using these equations we therefore can derive from (2.22) three equations

$$\begin{aligned} r^2 K'' &= \tfrac{1}{2} L(L+1) K(K-2)[1 + (\nu^2/2NL(L+1))(K-2)] \\ &\quad + h^2 K(l(l+1)/\nu^2) \end{aligned} \tag{2.29}$$

for the radial functions K and h. Taking the difference of each pair of equations (2.29) and using the linear independence of the different powers of K again, we derive first that the three different sets (L, N, ν^2) are the same. By subtracting Eq. (2.29) at $\omega = \omega_1$ from the same equation at $\omega = \omega_2$ we conclude second that (L, N, ν^2) are independent of ω. This completes the separation of variables. What remains to be done is to solve Eqs. (2.20) and (2.24)–(2.26) for the angular functions and the Eqs. (2.21) and (2.29) for the radial functions.

III. SOLUTION OF THE ANGULAR EQUATIONS

To solve the angular equations we first derive the conditions that

$$u = -\tfrac{1}{2}L_i a_i = -(r/2)\epsilon_{ijk}x_i\,\partial_j\Omega\Omega^{-1}\,\partial_k\Omega\Omega^{-1} \qquad (3.1)$$

has to satisfy. Using this definition, the identity

$$(L_iL_j + r\,\partial_i r\,\partial_j)f(\omega) = (\delta_{ij} - \hat{x}_i\hat{x}_j)L^2 f(\omega), \qquad (3.2)$$

and Eq. (2.24), we find

$$a_i = [2/L(L+1)]L_i u. \qquad (3.3)$$

Excluding only the trivial case, we assume $L \neq 0$ and $u \neq 0$. Therefore, u satisfies

$$L^2 u = -L(L+1)u, \quad L = 1, 2, ..., \qquad (3.4)$$

and can be expanded in terms of spherical harmonics $Y^L_m(\omega)$:

$$u^a\tau_a = \sum_{m=-L}^{L} t^a_m Y^L_m(\omega)\tau_a, \quad \bar{t}_m = (-1)^m t_{-m}. \qquad (3.5)$$

Equations (3.3) and (3.4) yield, on the one hand,

$$\mathrm{tr}(a_ia_i) = [4/L(L+1)]\,\mathrm{tr}\,u^2. \qquad (3.6)$$

On the other hand, Eqs. (2.25) and (2.27) yield

$$\mathrm{tr}\,u^2 = (v^2/8N)\mathrm{tr}(a_ia_i) = -\kappa^2 v/8N. \qquad (3.7)$$

Hence the number of free parameters is further reduced by the relation

$$2L(L+1) = v^2/N. \qquad (3.8)$$

Using this relation, Eqs. (2.24) and (2.25), and the fact that the totally antisymmetric tensor $\epsilon_{[ijk\,]}v_l$ vanishes for every vector v, we derive

$$r\,\partial_i u + [a_i,u] = -\tfrac{1}{2}L(L+1)\epsilon_{ijk}\hat{x}_j a_k. \qquad (3.9)$$

If we substitute for a_i its expression (3.3) in terms of u, we obtain

$$r\,\partial_i u = [1/L(L+1)][u, L_i u], \qquad (3.10)$$

and are left with the task of solving Eqs. (3.7) and (3.10) for a u of the form (3.5).

Up to an irrelevant factor 2, Eqs. (3.4), (3.7), and (3.10) for u are identical to the equations Michel *et al.*[4] derive for f, which in our case is zero and trivially satisfies these equations. We can therefore use their analysis to determine u. The essential steps in this derivation are the following: We first solve Eqs. (3.7) and (3.10) for

$$t^a_m t^a_{m'} = (-1)^m[2\pi NL^2(L+1)^2/(2L+1)]\delta_{m,-m'}, \qquad (3.11)$$

and

$$\begin{aligned}&[(m'-m)/L(L+1)]C_{abc}t^b_m t^c_{m'}\\ &\quad= (-1)^L\sqrt{\frac{4\pi}{2L+1}}\sum_J(\langle JM|LLmm'\rangle\\ &\qquad+\langle JM|LLm'm\rangle)\int d\omega\, Y^J_M t^a_\mu\,\partial_3 Y^L_\mu,\\ &M = -\mu = m + m'. \end{aligned}\qquad (3.12)$$

(Cf. Ref. 11 and see, e.g., Ref. 12 for the definition of the Clebsch–Gordan coefficients.) We then define the projection operator

$$P_{ab} = \frac{2L+1}{2\pi NL^2(L+1)^2}\sum_m(-1)^m t^a_m t^b_{-m}, \qquad (3.13)$$

and show that the solutions u^a span the $(2L+1)$-dimensional subalgebra projected out by P_{ab}. Since it can also be shown that SO(3) is a group of automorphisms which acts irreducibly on this subalgebra, the subalgebra is SO(3) itself. Hence $L = 0, 1$ holds and the only nontrivial solution to Eqs. (2.24) and (2.25) is

$$a_i = 2\epsilon_{ijk}\hat{x}_j\tau_k, \quad (\tau_k)_{\alpha\beta} = \epsilon_{\alpha k\beta}. \qquad (3.14)$$

We now have to solve Eqs. (2.20) and (2.26) for a_i of the form (3.14). From (3.14), (2.16), and (2.20) follows

$$(\hat{x}_it_i)^2 W\varphi_0 = [\tfrac{1}{4}l(l+1) - t(t+1)]W\varphi_0 =: -j^2 W\varphi_0, \qquad (3.15)$$

where t is defined by

$$t_it_i = -t(t+1)I, \quad t = n/2, \quad n = 0, 1, 2 \dots . \qquad (3.16)$$

Equation (3.15) shows that there are not solutions for all t. For $t = 1$, e.g., the eigenvalues of $(\hat{x}_it_i)^2$ are -1 and 0, neither of which can be expressed in the form $l(l+1)/4 - 2$ with integer l.

To find further conditions, we take the derivative $t_i\,\partial_i$ of Eq. (3.15). Using Eq. (2.15) we obtain

$$[1 - 2t(t+1) + 2j^2]\hat{x}_it_i W\varphi_0 = 0. \qquad (3.17)$$

Thus, $W\varphi_0$ is either an eigenvector of $\hat{x}_it_i$ with eigenvalue zero, in which case

$$rt_i\,\partial_i(\hat{x}_jt_jW\varphi_0) = t(t+1)W\varphi_0 = 0, \qquad (3.18)$$

and therefore $t = 0$ holds, or for $j^2 \neq 0$,

$$l = 1, \quad j^2 = t(t+1) - \tfrac{1}{2} \qquad (3.19)$$

follows. For integer t, this condition cannot be satisfied.

For half-integer $t = (2n-1)/2$, the condition that $-j^2$ is an eigenvalue of $(\hat{x}_it_i)^2$ of the form (3.19) reads

$$m = \tfrac{1}{2}(2n - 1 \pm \sqrt{4n^2 - 3}) = 0, 1, 2, ..., n-1. \qquad (3.20)$$

This condition can only be satisfied for $m = n - 1 = 0$. In this case, we can use the identity

$$t_it_j = -\tfrac{1}{4}\delta_{ij} + \tfrac{1}{2}\epsilon_{ijk}t_k \qquad (3.21)$$

to prove

$$\varphi = 2\hat{x}_it_i\varphi_0. \qquad (3.22)$$

We have thus shown that the *Ansatz* of Dashen *et al.*[7] and Boguta[8] is the only one which is compatible with the conditions of time independence for $A_0 = 0$, finite energy, separability, and vanishing f_{ij}.

The angular equation which is left for us to discuss is (2.26). Equation (2.26) yields

$$\epsilon_{ijk}\hat{x}_j(\varphi_0^T W^T t_a t_k W\varphi_0) = 0, \quad a = 4, 5, ..., \qquad (3.23)$$

i.e., all currents corresponding to generators other than the three generators t_i of the SO(3) subalgebra vanish. If we can allow for different coupling constants in our theory these currents can of course also vanish because the corresponding coupling constants are zero or because the complex Higgs doublets come in pairs with opposite coupling constants. None of these conditions hold for the standard electroweak model[13] without fermions. In fact, for the generator of weak hypercharge t_0, Eq. (3.23) implies

$$(\varphi_0^T t_0 t_i \varphi_0) = 0, \qquad (3.24)$$

and finally $\varphi_0 = 0$. This argument can be generalized to grand unified theories.

IV. ASYMPTOTIC BEHAVIOR OF THE SOLUTION TO THE RADIAL EQUATIONS

We are now only left with a discussion of the two radial equations

$$r^2K'' = K(K-1)(K-2) + \tfrac{1}{4}h^2K, \tag{4.1}$$

$$r^2h'' = \tfrac{1}{2}K^2h + (\lambda/e^2)(h - eV_0r)(\sigma^2h + eV_0r)h. \tag{4.2}$$

These equations have a solution which minimizes the energy

$$E = \frac{4\pi V_0}{e}\int_0^\infty \frac{d\xi}{\xi^2}\Big\{\xi^2\Big(\frac{dK}{d\xi}\Big)^2 + \frac{1}{2}\Big(\xi\frac{dh}{d\xi} - h\Big)^2 + \frac{1}{2}K^2(K-2)^2 + \frac{1}{4}h^2K^2 + \frac{\lambda}{e^2}V(\sigma^2, \xi, \mathrm{h})\Big\}, \quad \xi = eV_0r, \tag{4.3}$$

in the topologically nontrivial sector of the submodel given by (4.3) (see Ref. 9). A numerical analysis suggests that this solution is the only one with finite energy.[10]

Since we do not know this solution in closed form we want to study here its asymptotic behavior. At the origin, all derivatives of K and h exist[9] which permits us to write K and h as Taylor series:

$$K = 2 + \sum_{m=2}^{\infty} K_m r^m, \quad h = \sum_{m=2}^{\infty} h_m r^m. \tag{4.4}$$

The coefficients h_2 and K_2 are determined by the asymptotic behavior at infinity. The other coefficients ($m > 2$) are given by the recursion relations

$$K_m = \frac{1}{m(m-1)-2}\Big[\sum_{m_1, m_2 \geqslant 2}(3K_{m_1}K_{m_2} + \frac{1}{2}h_{m_1}h_{m_2})\delta_{m, m_1+m_2} + \sum_{m_1, m_2, m_3 \geqslant 2}(K_{m_1}K_{m_2} + \frac{1}{4}h_{m_1}h_{m_2})K_{m_3}\delta_{m, m_1+m_2+m_3}\Big], \tag{4.5}$$

$$h_m = \frac{1}{m(m-1)-2}\Big[2\sum_{m_1, m_2 \geqslant 2} K_{m_1}h_{m_2}\delta_{m, m_1+m_2} + \frac{1}{2}\sum_{m_1, m_2, m_3 \geqslant 2} K_{m_1}K_{m_2}h_{m_3}\delta_{m, m_1+m_2+m_3} + \frac{\lambda}{e^2}\Big\{\sigma^2\sum_{m_1, m_2, m_3 \geqslant 2} h_{m_1}h_{m_2}h_{m_3}\delta_{m, m_1+m_2+m_3} + (1-\sigma^2)eV_0\sum_{m_1, m_2 \geqslant 2} h_{m_1}h_{m_2}\delta_{m, m_1+m_2+1} - e^2V_0^2h_{m-2}\Big\}\Big]. \tag{4.6}$$

Notice that all odd terms vanish for $\sigma^2 = 1$.

As in Ref. 12, we can prove here by induction that

$$|K_m| < M^m/(m+1)^2, \quad |h_m| < M^m(m+1)^2 \tag{4.7}$$

hold for sufficiently large $M > 1$. To estimate the sums in (4.5) and (4.6) we use inequalities of the following type:

$$\sum_{m_1=2}^{m-2}\frac{1}{(m_1+1)^2(m-m_1+1)^2} \leqslant \int_{3/2}^{m-3/2} dx \frac{1}{(x+1)^2(m-x+1)^2}. \tag{4.8}$$

The inequalities (4.7) guarantee convergence of the Taylor series (4.4) for $r \leqslant 1/M$. The Taylor expansions can therefore be used as a starting point for a numerical analysis.

To derive asymptotic formulas for the behavior at infinity we use the arguments and the theorems of Plohr.[14] First, we establish that for $\lambda \neq 0$, K and $H = eV_0r - h$ vanish at infinity. The vanishing of K follows from the inequality

$$\frac{1}{2}[K^2(r) - K^2(r_0)] \leqslant 2\Big[\int_{r_0}^\infty dr\, K'^2 \int_{r_0}^\infty \frac{K^2H^2}{r^2}\Big]^{1/2} < \infty, \tag{4.9}$$

which holds for r_0 large enough to guarantee $H/r \geqslant \frac{1}{2}$ for $r \geqslant r_0$. Since $C_0^2H^2$ is the leading term in the expansion of the potential V in the asymptotic region we can find a constant $C > 0$ such that

$$V(r) \geqslant CH^2, \quad r \geqslant r_0, \tag{4.10}$$

holds for large enough r_0. Then

$$\int_{r_0}^\infty dr\, V \geqslant \mathrm{C}\int_{r_0}^\infty dr\, H^2 \tag{4.11}$$

shows that H is square integrable at infinity. Then so is $dH/d\xi$ because $dH/d\xi - H/\xi$ is square integrable for finite energy. This implies the existence of $\lim_{r\to\infty} H(r)$ and the vanishing of H at infinity.

To prove exponential decay for K and H we use Eqs. (4.1) and (4.2) which we write as

$$K'' = [\tfrac{1}{4}e^2V_0^2 + 2/r^2 + \kappa(r)]K, \tag{4.12}$$

$$H'' = [\lambda(1+\sigma^2)V_0^2 + \eta(r)]H - g, \quad g = (eV_0/2r)K^2. \tag{4.13}$$

Since κ and η vanish at infinity, K and H are of the form

$$K \underset{r\to\infty}{\sim} \alpha\exp(-\tfrac{1}{2}eV_0r),$$

$$H \underset{r\to\infty}{\sim} \beta\exp(-\min\{\sqrt{\lambda(1+\sigma^2)}, e\}V_0r). \tag{4.14}$$

This even yields exponential decay for κ and η which enables us to express the asymptotic behavior of K and H in terms of modified Bessel functions K_ν of order ν.

The asymptotic behavior of K is

$$K \underset{r\to\infty}{\sim} \alpha\sqrt{r}K_{3/2}(\tfrac{1}{2}eV_0r), \quad K' \underset{r\to\infty}{\sim} -\tfrac{1}{2}eV_0K, \tag{4.15}$$

as Eq. (4.12) shows. To find the asymptotic behavior of H we have to solve the homogeneous equation corresponding to (4.13), construct from these solutions $u_\pm$ a particular solution to (4.13),

$$u_p = \frac{1}{2\sqrt{\lambda(1+\sigma^2)}V_0}\int_{r_0}^r dr'\, u_-(r)u_+(r')g(r') + \frac{1}{2\sqrt{\lambda(1+\sigma^2)}V_0}\int_r^\infty dr'\, u_+(r)u_-(r')g(r'), \tag{4.16}$$

and study its asymptotic behavior. This leads to (cf. Ref. 14)

(a) $$h \underset{r\to\infty}{\sim} eV_0r - \beta\sqrt{r}K_{1/2}(\sqrt{\lambda(1+\sigma^2)}V_0),$$

$$h' \underset{r\to\infty}{\sim} -\sqrt{\lambda(1+\sigma^2)}V_0(h - eV_0r), \quad e^2 > \lambda(1+\sigma^2); \tag{4.17}$$

(b) $h \underset{r\to\infty}{\sim} eV_0 r - \frac{1}{2}\alpha^2 K_{3/2}^2(\frac{1}{2}eV_0 r)$,

$$h' \underset{r\to\infty}{\sim} -eV_0(h - eV_0 r), \quad e^2 = \lambda(1+\sigma^2); \tag{4.18}$$

(c) $h \underset{r\to\infty}{\sim} eV_0 r - \{e\alpha^2/2[\lambda(1+\sigma^2) - e^2]V_0\}K_{3/2}^2(\frac{1}{2}eV_0 r)$,

$$h' \underset{r\to\infty}{\sim} -eV_0(h - eV_0 r), \quad e^2 < \lambda(1+\sigma^2). \tag{4.19}$$

Here we had to distinguish the three cases $e^2 > \lambda(1+\sigma^2)$, $e^2 = \lambda(1+\sigma^2)$, and $e^2 < \lambda(1+\sigma^2)$.

For $\lambda = 0$, we cannot conclude from (4.11) that H is square integrable. We know however that H/r and therefore κ and η vanish at infinity. Thus,

$$K \underset{r\to\infty}{\sim} \alpha e^{-(1/2)eV_0 r}, \quad h \underset{r\to\infty}{\sim} eV_0 r + \beta \tag{4.20}$$

hold, and we can expand K and h in terms of exponential functions with polynomially bounded coefficient functions:

$$K(r) = \sum_{m=0}^{\infty} K_m(r) e^{-(m+1)\gamma r}, \quad \gamma = \frac{1}{2}eV_0, \tag{4.21}$$

$$h(r) = \sum_{m=0}^{\infty} h_m(r) e^{-m\gamma r}, \quad h_0 = eV_0 r + \beta. \tag{4.22}$$

Here, $\widetilde{K}_0 = K_0 \exp(-\gamma r)$ is the exponentially decreasing solution to

$$\widetilde{K}_0'' = [\gamma^2 + \beta\gamma/r + (\beta^2 + 8)/4r^2]\widetilde{K}_0, \tag{4.23}$$

which is Whittaker's equation of the second kind. The other coefficient functions can be found by solving recursively the differential equations

$$\begin{aligned}
K_m'' &- 2(m+1)\gamma K_m' + \left[m(m+2)\gamma^2 - \frac{\beta\gamma}{r} - \frac{\beta^2+8}{4r^2}\right]K_m \\
&= \left(\frac{\gamma}{r} + \frac{\beta}{2r^2}\right)\sum_{\substack{m_1 > 1\\ m_2 > 0}} h_{m_1} K_{m_2} \delta_{m,\, m_1 + m_2} \\
&- \frac{3}{r^2} \sum_{m_1,\, m_2 > 0} K_{m_1} K_{m_2} \delta_{m,\, m_1 + m_2 + 1} \\
&+ \frac{1}{r^2} \sum_{\substack{m_1,\, m_2 > 1\\ m_3 > 0}} h_{m_1} h_{m_2} K_{m_3} \delta_{m,\, m_1 + m_2 + m_3} \\
&+ \frac{1}{r^2} \sum_{m_1,\, m_2,\, m_3 > 0} K_{m_1} K_{m_2} K_{m_3} \delta_{m,\, m_1 + m_2 + m_3 + 2},
\end{aligned} \tag{4.24}$$

$$\begin{aligned}
h_m'' &- 2m\gamma h_m' + m^2\gamma^2 h_m \\
&= \frac{1}{2r^2} \sum_{m_1,\, m_2,\, m_3 > 0} K_{m_1} K_{m_2} h_{m_3} \delta_{m,\, m_1 + m_2 + m_3 + 2}.
\end{aligned} \tag{4.25}$$

Using these solutions the convergence of the series (4.21) and (4.22) can be proved by induction. Since all of this has been done in a similar case in Ref. 15, we omit the technical details here.

V. CONCLUDING REMARKS

We have proved that the Dashen–Hasslacher–Neveu–Boguta solution is unique under certain conditions and we have studied its asymptotic behavior. This type of solution only occurs in a very special model which may be relevant to nuclear physics.[8] Even if the model is relevant the instability of the solution[9] casts some doubt on the relevance of the solution itself. As far as current models in particle physics are concerned, our analysis shows that no solution of this kind occurs in these models.

We might therefore try to relax the conditions we imposed. Besides smoothness, time independence, and finiteness of the energy, these conditions were the vanishing of A_0 and the separability of the *Ansatz*. If we start out with a nonvanishing A_0, the finite-energy condition in the form

$$(D_0\phi)_\infty = A_0\varphi = 0 \tag{5.1}$$

implies that φ is an eigenvector to A_0 with eigenvalue zero. This already excludes all $A_0 \neq 0$ with half-integer eigenvalues, in particular the case studied in Sec. IV. It is therefore not very likely that we can find finite-energy solutions with nonvanishing electric field besides the dyon excitation of the monopoles mentioned in the Introduction.

The problem of proving this conjecture rigorously as well as the discussion of nonseparable *Ansätze* deserves further attention. Dashen *et al.*[7] have given the *Ansatz*

$$\phi = [f_1(r)\hat{x}_j\sigma_j + if_2(r)]\begin{pmatrix}1\\0\end{pmatrix}, \tag{5.2}$$

$$A_i = a_1(r)\epsilon_{ijk}\hat{x}_j\sigma_k + a_2(r)\hat{x}_i\hat{x}_j\sigma_j + a_3(r)\sigma_i. \tag{5.3}$$

Although this *Ansatz* is not separable in the sense of Eqs. (2.9) and (2.10) it is compatible with the equations of motion. It is not known whether it leads to a new finite-energy solution.

ACKNOWLEDGMENTS

I would like to thank Professor O'Raifeartaigh and Dr. Tchrakian for helpful discussions, and Dr. Maison and Professor Rühl for introducing me some time ago to some of the techniques used in Sec. IV.

[1] N. S. Manton, Phys. Rev. D **28**, 2019 (1983).
[2] P. Forgács and Z. Horváth, CERN Preprint Ref. TH. 3743.
[3] R. Shankar, Phys. Rev. D **14**, 1107 (1976).
[4] L. Michel, L. O'Raifeartaigh, and K. C. Wali, Phys. Lett. B **67**, 198 (1977); Phys. Rev. D **15**, 3641 (1977).
[5] L. O'Raifeartaigh and J. Rawnsley, Phys. Lett. B **72**, 465 (1978).
[6] G. 't Hooft, Nucl. Phys. B **79**, 276 (1974); A. M. Polyakov, JETP Lett. **20**, 194 (1974).
[7] R. F. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D **10**, 4138 (1974).
[8] J. Boguta, Phys. Rev. Lett. **50**, 148 (1983).
[9] J. Burzlaff, Nucl. Phys. B **233**, 262 (1984).
[10] J. Burzlaff, Phys. Lett. B **133**, 95 (1983).
[11] E. Cremmer, F. Schaposnik, and J. Scherk, Phys. Lett. B **65**, 78 (1976).
[12] D. M. Brink and G. R. Satchler, *Angular Momentum* (Oxford U. P., Oxford, 1968).
[13] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967); A. Salam, *Proceedings of the 8th Nobel Symposium*: *Elementary Particle Theory*, edited by N. Svartholm (Wiley, New York, 1968), p. 367.
[14] B. Plohr, Doctoral Dissertation, Princeton University, October 1980; J. Math. Phys. **22**, 2184 (1981).
[15] J. Burzlaff, Phys. Rev. D **23**, 1329 (1981).

Generalization of Weyl's gauge group

Arthur Komar
Physics Department, Yeshiva University, New York, New York 10033 and Physics Department, New York University, New York, New York 10003

(Received 18 September 1984; accepted for publication 16 November 1984)

Weyl's gauge transformations in a general n-dimensional Riemannian manifold are extended from the conformal group to GLn(R). The gauge-covariant field generalizing that of the Maxwell tensor is determined. The relationship between Weyl gauging and Yang–Mills gauging is developed. It is shown that the two processes are not equivalent, but can be made compatible.

I. INTRODUCTION

In order to unify the Einstein theory of gravitation with electromagnetism, Weyl[1] proposed an extension of the coordinate symmetry group of general relativity to include conformal mappings of the space-time. Although the proposed unification did not prove to be successful or convincing, the consequent introduction of the concept of gauge symmetry for the Maxwell theory turned out to be very fruitful and important. In our recent work[2] we suggested a new application of this conformal symmetry to facilitate the regularization of the operators which occur when one seeks to quantize gravitation theory. However, the conformal freedom is too modest an extension to be adequate for that purpose. Therefore, we have developed recently an enlarged gauge symmetry for gravitation theory which is a natural and essentially unique extension of the Weyl gauge group.

Once Weyl had introduced his conformal symmetry for the metric, it was natural for him to consider other geometric structures which transformed covariantly or remained invariant under such transformations. It is in this fashion that the vector potential and its related gauge transformation was introduced. As this ultimately proved to be the most lasting consequence of the Weyl theory, it became of interest to us to see whether similar geometric structures are available for our enlarged gauge group. In the present paper we shall show that this is indeed the case, and that the resulting fields are very intimately connected with those of electromagnetic theory. It is not clear at this point how the additional degrees of freedom which now occur are to be identified in the physical world.

II. REVIEW OF THE WEYL FORMALISM

From the geometric point of view, there is nothing special in our generalization which singles out four-dimensional manifolds. We will present therefore the new geometric formalism in n dimensions. Of course, for application to physics, one will eventually restrict one's considerations to four dimensions.

The Weyl conformally covariant geometry can be summarized as follows. We are given as our fundamental geometric quantities under the group of general curvilinear coordinate transformations a nonsingular symmetric covariant tensor field $g_{\mu\nu}$ and a nonvanishing scalar field ϕ. We employ the tensor field in the usual fashion as a Riemannian metric to raise and lower tensor indices, to form the Christoffel symbols $\Gamma^{\alpha}_{\beta\gamma}$ (which we use for covariant differentiation), the curvature tensor $R^{\alpha}_{\beta\gamma\delta}$, etc. In addition, we have a gauge group described by an arbitrary nonvanishing scalar field $\Omega(x)$, such that under this group the basic geometric quantities are required to transform thus:

$$\bar{g}_{\alpha\beta} = \Omega^2 g_{\alpha\beta} \tag{2.1}$$

and

$$\bar{\phi}(x) = \Omega^{-1}\phi(x)\,. \tag{2.2}$$

With regard to the metric tensor this is evidently a conformal transformation. The effect of the conformal freedom is to reduce the degrees of freedom of the theory back to that of Riemannian geometry. The consequent transformation law for the Christoffel symbols is somewhat more intricate, but if we form the affine connection

$$\Lambda^{\alpha}_{\beta\gamma} = \Gamma^{\alpha}_{\beta\gamma} + \phi^{-1}(\phi_{,\beta}\delta^{\alpha}_{\gamma} + \phi_{,\gamma}\delta^{\alpha}_{\beta} - \phi_{,\mu}g^{\mu\alpha}g_{\beta\gamma})\,, \tag{2.3}$$

then, by employing the first logarithmic derivative of Eq. (2.2),

$$\bar{\phi}^{-1}\bar{\phi}_{,\mu} = \phi^{-1}\phi_{,\mu} - \Omega^{-1}\Omega_{,\mu}\,, \tag{2.4}$$

it is easy to confirm that $\Lambda^{\alpha}_{\beta\gamma}$ is conformally invariant; that is,

$$\bar{\Lambda}^{\alpha}_{\beta\gamma} = \Lambda^{\alpha}_{\beta\gamma}\,. \tag{2.5}$$

Although now formally conformally covariant, the structure thus far obtained is evidently reducible to Riemannian geometry. However, an irreducible geometric structure can be obtained by Weyl's procedure. Rather than employing the scalar field ϕ, we introduce as an additional geometric quantity a covariant vector field A_{μ}, and we require that it transform under the gauge group as does $\phi^{-1}\phi_{,\mu}$, namely,

$$\bar{A}_{\mu} = A_{\mu} - \Omega^{-1}\Omega_{,\mu}\,. \tag{2.6}$$

We then replace $\phi^{-1}\phi_{,\mu}$ by A_{μ} in the definition of $\Lambda^{\alpha}_{\beta\gamma}$, Eq. (2.3). This preserves the gauge invariant relation, Eq. (2.5). Thus

$$\Lambda^{\alpha}_{\beta\gamma} \equiv \Gamma^{\alpha}_{\beta\gamma} + A_{\beta}\delta^{\alpha}_{\gamma} + A_{\gamma}\delta^{\alpha}_{\beta} - A_{\mu}g^{\mu\alpha}g_{\beta\gamma}\,. \tag{2.7}$$

Defining the field strength tensor

$$F_{\mu\nu} \equiv A_{\mu\nu} - A_{\nu,\mu}\,, \tag{2.8}$$

we evidently have under the Ω gauge group

$$\bar{F}_{\mu\nu} = F_{\mu\nu}\,. \tag{2.9}$$

As is well known, the vanishing of of $F_{\mu\nu}$ is the necessary and sufficient condition that A_{μ} can be written in the trivial form

$\psi^{-1}\psi_{,\mu}$. Selecting as a gauge transformation

$$\Omega = \psi\,, \tag{2.10}$$

A_μ can be made to vanish. Thus, nontrivial vector potentials A_μ are introduced into the conformal formalism by the requirement that the (conformal) gauge-invariant tensor field $F_{\mu\nu}$ not vanish.

We end out brief synopsis of Weyl's theory at this point since in this paper we are only interested in generalizing the kinematics of Weyl's construction. The applications of the kinematics to dynamical considerations will be presented elsewhere.

III. THE GENERAL LINEAR FORMALISM

We have extended[3] the Weyl formalism by defining as our fundamental geometric quantities the metric tensor $g_{\mu\nu}(x)$ (employed as in the previous section) and a mixed second rank tensor (under the curvilinear coordinate group) $b^{\alpha}_{\ \beta}(x)$ of nonvanishing determinant. We extend the Weyl conformal gauge group to the local GLn(R) group. An arbitrary element of that group is described by a nonsingular matrix field, $\Omega^{\alpha}_{\ \beta}(x)$, the group operation being matrix multiplication on each fiber. We require that under the coordinate group, the $\Omega^{\alpha}_{\ \beta}$ transform as a mixed second-rank tensor. Thus the total effective group of this geometry is a semidirect product of the coordinate group and the gauge group. The fundamental geometric quantities are postulated to transform under the gauge group via the relations

$$\bar{g}_{\alpha\beta}(x) = g_{\mu\nu}(x)\Omega^{\mu}_{\ \alpha}(x)\Omega^{\nu}_{\ \beta}(x) \tag{3.1}$$

and

$$\bar{b}^{\alpha}_{\ \beta}(x) = b^{\mu}_{\ \beta}(x)\Theta^{\alpha}_{\ \mu}(x)\,, \tag{3.2}$$

where $\Theta^{\alpha}_{\ \beta}(x)$ is the inverse matrix of $\Omega^{\alpha}_{\ \beta}(x)$; that is,

$$\Omega^{\alpha}_{\ \mu}\Theta^{\mu}_{\ \beta} = \delta^{\alpha}_{\ \beta}\,. \tag{3.3}$$

In addition, it proves convenient to denote the inverse matrix of $b^{\alpha}_{\ \beta}(x)$ by $a^{\alpha}_{\beta}(x)$; that is,

$$b^{\alpha}_{\ \mu}a^{\mu}_{\beta} = \delta^{\alpha}_{\beta}\,. \tag{3.4}$$

It follows, of course, that under the gauge group, a^{α}_{β} transforms thus:

$$\bar{a}^{\alpha}_{\beta} = a^{\alpha}_{\mu}\Omega^{\mu}_{\ \beta}\,. \tag{3.5}$$

It is evident that, as in the previous section, the apparently additional degrees of freedom introduced into the geometry by the field $b^{\alpha}_{\ \beta}$ have been removed by the introduction of the gauge group. Defining the field $\Lambda^{\ \alpha}_{\beta\gamma}(x)$ as

$$\Lambda^{\ \alpha}_{\beta\gamma} \equiv \Gamma^{\ \alpha}_{\beta\gamma} + a^{\alpha}_{\rho}b^{\ \rho}_{(\beta;\gamma)} + a^{\alpha}_{\rho}a^{\lambda}_{\sigma}b^{\mu}_{\ \beta}b^{\ \nu}_{[\lambda;\gamma]}g^{\rho\sigma}g_{\mu\nu} + a^{\alpha}_{\rho}a^{\lambda}_{\sigma}b^{\mu}_{\ \gamma}b^{\ \nu}_{[\lambda;\beta]}g^{\rho\sigma}g_{\mu\nu}\,, \tag{3.6}$$

where

$$b^{\ \rho}_{(\beta;\gamma)} \equiv \tfrac{1}{2}(b^{\ \rho}_{\beta,\gamma} + b^{\ \rho}_{\gamma,\beta})$$

and

$$b^{\ \rho}_{[\beta;\gamma]} \equiv \tfrac{1}{2}(b^{\ \rho}_{\beta;\gamma} - b^{\ \rho}_{\gamma;\beta})\,,$$

it is evident that under the coordinate group $\Lambda^{\ \alpha}_{\beta\gamma}$ transforms as an affine connection. It can be shown by direct substitution of the previous relations that $\Lambda^{\ \alpha}_{\beta\gamma}$ is invariant under the gauge group, i.e.,

$$\bar{\Lambda}^{\ \alpha}_{\beta\gamma}(x) = \Lambda^{\ \alpha}_{\beta\gamma}(x)\,. \tag{3.7}$$

Indeed, if we choose a gauge for which $b^{\alpha}_{\ \beta}$ has the form

$$b^{\alpha}_{\ \beta}(x) = \phi(x)\delta^{\alpha}_{\beta} \tag{3.8}$$

and restrict all further gauge transformations to the subgroup which preserves this form; namely, $\Omega^{\alpha}_{\ \beta}$ of the form

$$\Omega^{\alpha}_{\ \beta}(x) = \Omega(x)\delta^{\alpha}_{\beta}\,, \tag{3.9}$$

then it is easily confirmed that Eqs. (3.1), (3.2), and (3.6) reduce to Eqs. (2.1), (2.2), and (2.3). Thus our geometric structure is a natural generalization of that of Weyl. It is also the smallest generalization in the following sense: If we ask for an extension of the conformal gauge group in such a fashion that the resulting combined symmetry group remains a semidirect product of the group of curvilinear coordinate transformations and the gauge group, the only possible choices for the gauge group smaller than local GLn(R) are the continuous normal subgroups of GLn(R). This is due to the fact that the coordinate group acts locally on each fiber as a matrix of GLn(R). There are only two such nontrivial normal subgroups available, local SLn(R) and the subgroup given by Eq. (3.9). The latter subgroup yields the Weyl conformal theory, while the former introduces the $n^2 - 1$ new field components of our generalization.

The structure we have thus far presented is again completely reducible to that of Riemannian geometry. Indeed that was our intention when introducing this formalism. (The easiest way to demonstrate the reducibility of the structure is to express all gauge-invariant geometric quantities in terms of the gauge-invariant metric $g_{\mu\nu}b^{\mu}_{\ \alpha}b^{\nu}_{\ \beta}$. Alternatively, we can always choose a gauge where $b^{\alpha}_{\ \beta} = \delta^{\alpha}_{\beta}$.) We now pose the question which is the heart of this paper: *Is it possible to generalize Weyl's procedure of the introduction of the vector potential thereby obtaining a nontrivial geometric structure irreducible under the GLn(R) gauge group?*

IV. WEYL GAUGING

Weyl gauging is quite distinct from the more familiar Yang–Mills gauging.[4] In the latter, an affine connection is introduced having a transformation law so contrived that it compensates for the lack of commutativity of ordinary differentiation with local gauge transformation. Weyl gauging, on the contrary (the only extant exemplar thus far being that of the conformal theory described in Sec. II above), requires the introduction of a tensorial object which transforms affinely under the gauge group in such a manner that the already occurring connection remains invariant under the gauge transformations. In order to accomplish this we first seek the appropriate generalization of Eq. (2.4). Let us therefore consider the expression

$$C^{\ \alpha}_{\beta\gamma} \equiv a^{\rho}_{\beta}b^{\ \alpha}_{\rho|\gamma}\,, \tag{4.1}$$

where the solidus denotes covariant differentiation employing the affine connection $\Lambda^{\ \alpha}_{\beta\gamma}$ of Eq. (3.6). We can, of course, reexpress $C^{\ \alpha}_{\beta\gamma}$ in terms of the usual semicolon covariant differentiation which employs the metric connection $\Gamma^{\ \alpha}_{\beta\gamma}$, but the resulting expression is more complicated and consequently less illuminating. It is easily confirmed that under the gauge transformation, Eqs. (3.1) and (3.2), $C^{\ \alpha}_{\beta\gamma}$ transforms thus:

$$\bar{C}^{\alpha}_{\beta\gamma} = \Theta^{\alpha}_{\mu}\Omega^{\nu}_{\beta}C^{\mu}_{\nu\gamma} - \Theta^{\alpha}_{\mu}\Omega^{\mu}_{\beta|\gamma} . \tag{4.2}$$

This generalization of Eq. (2.4), although strikingly reminiscent, does not retain the feature of being independent of the metric fields. But that cannot be helped as we insist that $C^{\alpha}_{\beta\gamma}$ be a tensor.

We now parallel the Weyl procedure [viz. Eq. (2.6)] and introduce a tensor potential $A^{\alpha}_{\beta\gamma}$ which, under the GLn(R) gauge group, transforms as does $C^{\alpha}_{\beta\gamma}$, namely

$$\bar{A}^{\alpha}_{\beta\gamma} = \Theta^{\alpha}_{\mu}\Omega^{\nu}_{\beta}A^{\mu}_{\nu\gamma} - \Theta^{\alpha}_{\mu}\Omega^{\mu}_{\beta|\gamma} . \tag{4.3}$$

We are thereby assured that the transformation law provides a realization of the gauge group. We note at this point that if we contract Eq. (4.3) on α and β and identify $A^{\mu}_{\mu\gamma}$ with A_{γ} and Ω with det Ω^{α}_{β}, we obtain precisely Eq. (2.6). Our construction is therefore a true generalization of Weyl's theory.

The final and most difficult step of the Weyl procedure is to replace the auxiliary field b^{α}_{β} everywhere in the definition of $\Lambda^{\alpha}_{\beta\gamma}$, Eq. (3.6), by the tensor potential $A^{\alpha}_{\beta\gamma}$ in such a fashion that the connection remains invariant under the Ω^{α}_{β} gauge group. After considerable algebra it can be verified that the required redefinition of $\Lambda^{\alpha}_{\beta\gamma}$ is remarkably simple:

$$\Lambda^{\alpha}_{\beta\gamma} \equiv \Gamma^{\alpha}_{\beta\gamma} + A^{\alpha}_{(\beta\gamma)} + A^{\nu}_{[\mu\gamma]}g^{\alpha\mu}g_{\beta\nu} + A^{\nu}_{[\mu\beta]}g^{\alpha\mu}g_{\gamma\nu} . \tag{4.4}$$

Employing this new expression for $\Lambda^{\alpha}_{\beta\gamma}$ in Eq. (4.3) we find that the transformation law for $A^{\alpha}_{\beta\gamma}$ becomes

$$\begin{aligned}\bar{A}^{\alpha}_{\beta\gamma} &= A^{\alpha}_{(\beta\gamma)} + A^{\nu}_{[\mu\beta]}g^{\alpha\mu}g_{\gamma\nu} + A^{\nu}_{[\mu\gamma]}g^{\alpha\mu}g_{\beta\gamma}\\ &\quad + \Theta^{\alpha}_{\mu}\Omega^{\rho}_{\beta}(A^{\mu}_{[\rho\gamma]} - A^{\nu}_{[\sigma\gamma]}g^{\mu\sigma}g_{\rho\nu} - A^{\nu}_{[\sigma\rho]}g^{\mu\sigma}g_{\gamma\nu})\\ &\quad - \Theta^{\alpha}_{\mu}\Omega^{\mu}_{\beta;\gamma} .\end{aligned} \tag{4.5}$$

In this form, the transformation of $A^{\alpha}_{\beta\gamma}$ seems considerably more complicated and one would not have suspected that it induces a realization of the gauge group. Such an intricate expression would not have been postulated *ab initio*. Surprisingly, we find that it is partially reducible, as $A^{\alpha}_{[\beta\gamma]}$ evidently transforms independent of $A^{\alpha}_{(\beta\gamma)}$. At this point, we can, if we wish, dispense with the heuristic development which we have thus far pursued and postulated the transformation laws, Eqs. (4.4) and (4.5), together with Eq. (3.1). It then follows by straightforward substitution that Eq. (3.7) is satisfied. That Eq. (4.5) is indeed a generalization of the gauge transformation of Maxwell's theory can be substantiated by restricting our gauge transformations to that of the conformal subgroup, Eq. (3.9). Equation (4.5) then yields an expression more familiar in appearance:

$$\bar{A}^{\alpha}_{\beta\gamma} = A^{\alpha}_{\beta\gamma} - \Omega^{-1}\Omega_{,\gamma}\delta^{\alpha}_{\beta} . \tag{4.6}$$

This expression suggests that, if there exists a gauge in which $A^{\alpha}_{\beta\gamma}$ takes the form

$$A^{\alpha}_{\beta\gamma} = \delta^{\alpha}_{\beta}A_{\gamma} , \tag{4.7}$$

we could identify A_{γ} with the usual vector potential. This form is then correctly preserved under the conformal subgroup. However, it is not preserved under the full gauge group. Clearly, as we have already indicated, the invariant and consequently preferred identification is $A^{\mu}_{\mu\gamma} = A_{\gamma}$.

V. CURVATURE

We have introduced the solidus covariant differentiation employing the gauge-invariant symmetric affine connection $\Lambda^{\alpha}_{\beta\gamma}$. We therefore have the associated gauge-invariant (in general, non-Riemannian) curvature tensor $\Lambda^{\alpha}_{\beta\gamma\delta}$ obtained by commuting the solidus derivative; thus

$$\xi^{\alpha}_{|\beta\gamma} - \xi^{\alpha}_{|\gamma\beta} \equiv \xi^{\mu}\Lambda^{\alpha}_{\mu\beta\gamma} , \tag{5.1}$$

whereby

$$\Lambda^{\alpha}_{\beta\gamma\delta} = \Lambda^{\alpha}_{\beta\gamma,\delta} - \Lambda^{\alpha}_{\beta\delta,\gamma} + \Lambda^{\mu}_{\beta\gamma}\Lambda^{\alpha}_{\delta\mu} - \Lambda^{\mu}_{\beta\delta}\Lambda^{\alpha}_{\gamma\mu} . \tag{5.2}$$

Clearly under arbitrary gauge transformations we have

$$\bar{\Lambda}^{\alpha}_{\beta\gamma\delta} = \Lambda^{\alpha}_{\beta\gamma\delta} . \tag{5.3}$$

Another curvature expression, analogous to the Maxwell field, can be obtained by determining the conditions for $A^{\alpha}_{\beta\gamma}$ to be "pure gauge"; that is, when a gauge frame can be found in which the tensor potential vanishes. That is evidently the case when $A^{\alpha}_{\beta\gamma}$ has the form of $C^{\alpha}_{\beta\gamma}$ of Eq. (4.1). It is easy to confirm that this form is preserved under arbitrary gauge transformations provided it is understood that b^{α}_{β} transforms according to Eq. (3.2) [and consequently, its reciprocal, a^{α}_{β}, according to Eq. (3.5)].

We shall therefore call $A^{\alpha}_{\beta\gamma}$ trivial if there exists a nonsingular tensor field b^{α}_{β}, with reciprocal a^{α}_{β}, such that

$$A^{\alpha}_{\beta\gamma} = a^{\mu}_{\beta}b^{\alpha}_{\mu|\gamma} . \tag{5.4}$$

[This relation is somewhat more complicated if written out in terms of the metric (i.e., semicolon) covariant derivative as, it should be recalled, the solidus derivative contains $A^{\alpha}_{\beta\gamma}$ terms in its definition.] An integrability condition for this equation is obtained in the usual fashion by performing a second solidus differentiation, commuting the two operations and employing Eq. (5.1). In this manner we obtain

$$F^{\alpha}_{\beta\gamma\delta} = \Lambda^{\alpha}_{\beta\gamma\delta} - a^{\mu}_{\beta}b^{\alpha}_{\rho}\Lambda^{\rho}_{\mu\gamma\delta} , \tag{5.5}$$

where we have defined

$$F^{\alpha}_{\beta\gamma\delta} \equiv A^{\alpha}_{\beta\gamma|\delta} - A^{\alpha}_{\beta\delta|\gamma} + A^{\alpha}_{\rho\gamma}A^{\rho}_{\beta\delta} - A^{\alpha}_{\rho\delta}A^{\rho}_{\beta\gamma} . \tag{5.6}$$

Contracting Eqs. (5.5) and (5.6) on α and β we obtain as a necessary condition for triviality the familiar gauge-invariant result

$$F^{\mu}_{\mu\gamma\delta} = A^{\mu}_{\mu\gamma,\delta} - A^{\mu}_{\mu\delta,\gamma} = 0 . \tag{5.7}$$

Furthermore, for tensor potentials of the form given in Eq. (4.7), $F^{\alpha}_{\beta\gamma\delta}$ becomes

$$F^{\alpha}_{\beta\gamma\delta} = \delta^{\alpha}_{\beta}(A_{\gamma,\delta} - A_{\delta,\gamma}) . \tag{5.8}$$

Thus, as our notation was chosen to indicate, $F^{\alpha}_{\beta\gamma\delta}$ is an appropriate candidate for the generalization of the Maxwell field tensor. Unfortunately it has a rather complicated transformation property under the gauge group. One may confirm that for transformations described by arbitrary Ω^{α}_{β} we have

$$\bar{F}^{\alpha}_{\beta\gamma\delta} = \Theta^{\alpha}_{\nu}\Omega^{\sigma}_{\beta}(F^{\nu}_{\sigma\gamma\delta} - \Lambda^{\nu}_{\sigma\gamma\delta}) + \Lambda^{\alpha}_{\beta\gamma\delta} , \tag{5.9}$$

so that only the trace $F^{\mu}_{\mu\gamma\delta}$ is gauge invariant. In view of Eq. (5.3), it is therefore desirable to define the tensor field

$$K^{\alpha}_{\beta\gamma\delta} \equiv F^{\alpha}_{\beta\gamma\delta} - \Lambda^{\alpha}_{\beta\gamma\delta} , \tag{5.10}$$

which transforms much more simply under gauge transformations:

$$\bar{K}^{\alpha}_{\beta\gamma\delta} = \Theta^{\alpha}_{\nu}\Omega^{\sigma}_{\beta}K^{\nu}_{\sigma\gamma\delta} . \tag{5.11}$$

In terms of this new tensor the integrability condition for triviality, Eq. (5.5), becomes

$$K^{\alpha}_{\beta\gamma\delta} = -a^{\mu}_{\beta}b^{\alpha}_{\rho}\Lambda^{\rho}_{\mu\gamma\delta} . \tag{5.12}$$

We are now in a position to obtain a sequence of gauge-invariant conditions upon $A^{\alpha}_{\beta\gamma}$, all of which are necessary for triviality. By successive multiplications and contractions of Eq. (5.12) we find

$$K^{\mu}_{\mu\gamma\delta} = -\Lambda^{\mu}_{\mu\gamma\delta} = 0 , \tag{5.13a}$$

$$K^{\mu}_{\nu\alpha\beta}K^{\nu}_{\mu\gamma\delta} = \Lambda^{\mu}_{\nu\alpha\beta}\Lambda^{\nu}_{\mu\alpha\delta} , \tag{5.13b}$$

$$\vdots$$

$$K^{\mu}_{\nu\alpha\beta}K^{\nu}_{\sigma\alpha\delta}\cdots K^{\rho}_{\mu\epsilon\zeta} = (-1)^n\Lambda^{\mu}_{\nu\alpha\beta}\Lambda^{\nu}_{\sigma\alpha\delta}\cdots\Lambda^{\rho}_{\mu\epsilon\zeta} . \tag{5.13n}$$

One can confirm that for the generic case in n dimensions, the satisfaction of the first n sets of equations is also sufficient for triviality. In singular situations it is necessary to obtain further invariant relations by means of the solidus differentiation of Eq. (5.12).

We shall call a space flat if there exists a combined gauge and coordinate frame such that $A^{\alpha}_{\beta\gamma} = 0$ and $g_{\mu\nu} = \eta_{\mu\nu}$ (i.e., the Minkowski metric). [Note that for real Ω^{α}_{β} the signature of the metric is invariant under the transformation law, Eq. (3.1).] Clearly, for flatness it is necessary that

$$\Lambda^{\alpha}_{\beta\gamma\delta} = 0 \tag{5.14}$$

and

$$K^{\alpha}_{\beta\gamma\delta} = 0 , \tag{5.15}$$

as these are tensorial gauge invariant relations which, if valid in one frame, must remain valid in all. These two relations are also sufficient as we shall now prove.

If Eqs. (5.14) and (5.15) are satisfied, then the integrability condition for triviality, Eq. (5.12), is identically valid. It follows that $A^{\alpha}_{\beta\gamma}$ has the form of Eq. (5.4) and can be made to vanish by means of a gauge transformation (i.e., choose $\Omega^{\alpha}_{\beta} = b^{\alpha}_{\beta}$). In this gauge frame Eq. (5.14) becomes

$$0 = \Lambda^{\alpha}_{\beta\gamma\delta} = R^{\alpha}_{\beta\gamma\delta} , \tag{5.16}$$

and the Riemannian metric $g_{\mu\nu}$ is flat in the conventional sense. As the potential $A^{\alpha}_{\beta\gamma}$ transforms tensorially under coordinate transformations, it continues to vanish as we perform the coordinate transformation which reduces the metric to the Minkowski form.

VI. YANG–MILLS GAUGING

We wish to determine the relationship between the Weyl method of gauging that we have developed and the more familiar Yang–Mills[4] methods of gauging of the same group, GLn(R), on a curvilinear Riemannian manifold. Let us therefore consider a space-time vector field ψ_μ which transforms under our gauge group in the same manner as the covariant indices of the metric tensor; thus,

$$\bar{\psi}_\mu = \Omega^{\rho}_{\mu}\psi_\rho . \tag{6.1}$$

(We choose not to suppress the vector index, as is customary, because we wish to emphasize that it is not merely an internal index but also a space-time index.) The Yang–Mills procedure requires the introduction of a covariant differentiation on ψ_μ which commutes with the gauge transformations, Eq. (6.1). However, we are assuming that we already have a Riemannian metric $g_{\mu\nu}$ with its associated affine connection $\Gamma^{\alpha}_{\beta\gamma}$. It is therefore convenient, without loss of generality, to introduce this connection explicitly into the gauge-covariant differential operator. We therefore consider the covariant expression

$$D^{\mu}_{\alpha\beta}\psi_\mu \equiv \psi_{\alpha;\beta} + Y^{\mu}_{\alpha\beta}\psi_\mu , \tag{6.2}$$

where $Y^{\mu}_{\alpha\beta}$, the (modified) Yang–Mills potential, is evidently a tensor under coordinate transformations. In order to obtain its transformation properties under the GLn(R) gauge group we impose the customary demand

$$\bar{D}^{\mu}_{\alpha\beta}\bar{\psi}_\mu = \Omega^{\nu}_{\alpha}D^{\mu}_{\nu\beta}\psi_\mu . \tag{6.3}$$

After some straightforward, but tedious, algebra we obtain

$$\bar{Y}^{\alpha}_{\beta\gamma} = \Theta^{\alpha}_{\nu}\Omega^{\mu}_{\beta}Y^{\nu}_{\mu\gamma} - \Theta^{\alpha}_{\mu}\Omega^{\mu}_{[\beta;\gamma]} + \Theta^{\alpha}_{\rho}\Theta^{\delta}_{\sigma}\Omega^{\mu}_{\beta}\Omega^{\nu}_{[\delta;\gamma]}g^{\rho\sigma}g_{\mu\nu}$$
$$+ \Theta^{\alpha}_{\rho}\Theta^{\delta}_{\sigma}\Omega^{\mu}_{\gamma}\Omega^{\nu}_{[\delta;\beta]}g^{\rho\sigma}g_{\mu\nu} . \tag{6.4}$$

Contracting Eq. (6.4) on α and β yields

$$\bar{Y}^{\mu}_{\mu\gamma} = Y^{\mu}_{\mu\gamma} . \tag{6.5}$$

We see that the Yang–Mills structure contains a gauge-invariant vector field. As no such quantity occurs in the Weyl gauge structure, the two procedures are not equivalent. However, if we set

$$Y^{\mu}_{\mu\gamma} = 0 , \tag{6.6}$$

we can inquire if it is possible to realize the transformation law, Eq. (6.4), with the structures available in the Weyl gauge theory. It may be confirmed by direct computation that this is uniquely the case for the following expression:

$$Y^{\alpha}_{\beta\gamma} \equiv A^{\alpha}_{[\beta\gamma]} - A^{\sigma}_{[\delta\gamma]}g^{\alpha\delta}g_{\beta\sigma} - A^{\sigma}_{[\delta\beta]}g^{\alpha\delta}g_{\gamma\sigma} . \tag{6.7}$$

Curiously, only the antisymmetric part of the Weyl potential contributes to the Yang–Mills potential. Thus, for the gauge group under consideration, the two methods of gauging are not equivalent, but they can be made compatible. As a consequence of Eq. (6.7) we can express the Yang–Mills covariant derivative, Eq. (6.2), in the particularly simple form

$$D^{\mu}_{\alpha\beta}\psi_\mu = \psi_{\alpha|\beta} + A^{\mu}_{\alpha\beta}\psi_\mu . \tag{6.8}$$

Had we started directly with an ansatz of this form for the gauge-covariant differential operator, however, it would have obscured the fact that Eq. (6.6) was imposed as an extra requirement on the Yang–Mills structure. Furthermore, when written in terms of the solidus differentiation, it is by no means evident that only $A^{\mu}_{[\alpha\beta]}$ occurs in Eq. (6.8), thereby vastly reducing the number of independent components of the effective Yang–Mills potential $Y^{\alpha}_{\beta\gamma}$.

Applying the $D^{\mu}_{\alpha\beta}$ operator to ψ_μ twice and anticommuting the two operators we obtain in the usual fashion the Yang–Mills curvature tensor

$$(D^{\nu}_{\gamma\beta}D^{\mu}_{\nu\alpha} - D^{\nu}_{\gamma\alpha}D^{\mu}_{\nu\beta})\psi_\mu = (F^{\mu}_{\gamma\alpha\beta} - \Lambda^{\mu}_{\gamma\alpha\beta})\psi_\mu = K^{\mu}_{\gamma\alpha\beta}\psi_\mu . \tag{6.9}$$

It may also be confirmed by direct computation that

$$D_\gamma g_{\alpha\beta} \equiv D^{\mu}_{\alpha\gamma}g_{\mu\beta} + D^{\mu}_{\beta\gamma}g_{\alpha\mu} = 0 , \tag{6.10}$$

a relation of fundamental importance which can serve as the starting point for a reformulation of the entire formalism.

Indeed, it is this relation which in effect restricts the Yang–Mills structure to the form of Eq. (6.7). Employing Eqs. (6.9) and (6.10) it is easy to confirm that $K^{\ \alpha}_{\beta\gamma\delta}$ satisfies the familiar identity

$$K_{\alpha\beta\gamma\delta} + K_{\beta\alpha\gamma\delta} = 0\,. \tag{6.11}$$

That the tensor $K^{\ \mu}_{\gamma\alpha\beta}$ should occur, rather than $F^{\ \mu}_{\gamma\alpha\beta}$, as the Yang–Mills curvature tensor is not surprising given its simple transformation properties under the gauge group. The consequent observation that $K^{\ \alpha}_{\beta\gamma\delta}$ is independent of $A^{\ \alpha}_{(\beta\gamma)}$ is, however, unexpected.

We observed earlier that the transformation law for $A^{\ \alpha}_{\beta\gamma}$, Eq. (4.5), was partially reducible, that $A^{\ \alpha}_{[\beta\gamma]}$ transformed independently under gauge transformations. It is therefore possible to have a self-contained Yang–Mills-type theory. However, the fact that the transformation law is only partially reducible requires us to retain $A^{\ \alpha}_{(\beta\gamma)}$ if we are to avail ourselves of the more general gauge-invariant structures developed in this paper, such as $\Lambda^{\ \alpha}_{\beta\gamma}$ and $\Lambda^{\ \alpha}_{\beta\gamma\delta}$. These quantities do not occur in a purely Yang–Mills formulation.

VII. CONCLUSION

We have demonstrated that Weyl's procedure of constructing a gauge theory based on conformal transformations can be extended to the local GLn(R) group. This procedure does not seem to work for most other groups. It is therefore far more restrictive than the Yang–Mills procedure. Even when it is possible to employ Weyl gauging, we have shown that it is, in general, not equivalent to that of Yang–Mills gauging. However, in the present case the two procedures can be made compatible, although not equivalent, by requiring what are in effect $\frac{1}{2}n^2(n+1)$ additional conditions for the Yang–Mills structure.

Our interest in developing this structure is to apply it to the physical problem of the quantization of gravitation theory. In the present paper, we presented the geometry, which, when restricted to four dimensions, is essentially the kinematics we intend to employ. For the dynamics, we require a gauge-invariant Lagrangian density constructed out of those elements. We regard the fact that Weyl gauging appears to be a much more restrictive requirement than Yang–Mills gauging as a virtue of this approach. The presentation of the dynamical theory, however, is beyond the intended scope of the present paper.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under Grant No. PHY 84-06114.

[1]H. Weyl, Ann Phys. **59**, 101 (1919).
[2]A. Komar, Phys. Rev. D **27**, 2277 (1983).
[3]A. Komar, Phys. Rev. D **30**, 305 (1984).
[4]C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

Gauge fields in algebraically special space-times

G. F. Torres del Castillo
Departamento de Física Matemática, Instituto de Ciencias de la Universidad Autónoma de Puebla, Puebla, Mexico and Departamento de Física, Centro de Investigación y de Estudios Avanzados del IPN, Apartado Postal 14-740, 07000 México, D.F., Mexico

(Received 25 July 1984; accepted for publication 12 October 1984)

It is shown that in an algebraically special space-time which admits a congruence of null strings, a source-free gauge field aligned with the congruence is determined by a matrix potential which has to satisfy a second-order differential equation with quadratic nonlinearities. The Einstein–Yang–Mills equations are then reduced to a scalar and two matrix equations. In the case of self-dual gauge fields in a self-dual space-time, the existence of an infinite set of conservation laws, of an associated linear system, and of infinitesimal Bäcklund transformations is demonstrated. All the results apply for an arbitrary gauge group.

I. INTRODUCTION

In recent years there has been a great interest in the classical Yang–Mills fields and much effort has been devoted to understand their properties. In the particular case of the self-dual fields in flat space-time, several remarkable results have been obtained such as the reduction of the self-duality conditions to a single wavelike matrix equation,[1–5] the proof of the existence of an infinite set of nonlocal conservation laws,[5,6] the existence of Bäcklund transformations,[4,5,7–9] and the equivalence of the self-duality conditions with a linear system[10,11] and with certain vector bundles.[12,7] Most of these works are based on considering complex space-time where there exist two-dimensional totally null planes on which, owing to the self-duality of the gauge field, the field strength vanishes.

In the present work we show that in a certain class of curved space-times which possess an appropriate generalization of the two-dimensional totally null planes existing in complex flat space-time, the source-free Yang–Mills equations can be reduced to a wavelike matrix equation, provided that the field strength vanishes on those surfaces. Such a class of curved space-times has been brought to light in the study of exact solutions of the Einstein equations,[13–17] where it has been found that the existence of a single family of two-dimensional, totally null, geodesic surfaces (called null strings), and the use of coordinates adapted to this foliation, allows us to integrate some of the equations involved in a strikingly simple way and, in general, to reduce the equations considerably. We also consider the case of the self-dual gauge fields, showing that some of the results mentioned above can be extended to the self-dual space-times.

This paper is organized as follows. In Sec. II we show that if the space-time admits a congruence of null strings defined by a repeated principal spinor of the conformal curvature, then the source-free gauge fields aligned along the congruence are determined by the solutions of a second-order partial differential equation with quadratic nonlinearities for a matrix potential. The case of the electromagnetic field is then considered with some more detail, establishing the connection with previous results. In Sec. III we employ the reduced form of the Einstein equations found in Ref. 16 together with the results of Sec. II, to give a reduced expression for the coupled Einstein–Yang–Mills equations with cosmological constant, in the case where the Yang–Mills field is aligned along a congruence of null strings. In Sec. IV, we show that for a self-dual gauge field in a self-dual space-time there exists an infinity of conservation laws, or continuity equations, from which an equivalent linear system can be constructed. Then, following Ref. 9, the solution of this linear system is used to define infinitesimal transformations which produce new self-dual gauge fields. We use throughout the spinorial notation with the conventions $\psi_A = \epsilon_{AB}\psi^B$, $\psi^B = \psi_A\epsilon^{AB}$ for *all* the spinorial indices and similarly for the dotted ones.

II. REDUCTION OF THE YANG–MILLS EQUATIONS

In the case where the space-time is real, we shall assume that its conformal curvature is algebraically special and that it admits a shear-free congruence of null geodesics along the repeated principal null direction. As is well known, if the Ricci tensor vanishes (i.e., the Einstein vacuum field equations are satisfied), then each of these conditions implies the other (Goldberg–Sachs theorem). Denoting by l the repeated principal null direction of the conformal curvature tensor and expressing its spinorial components as $l_A\ l_{\dot B}$, then, as a consequence of the fact that l is geodesic and shear-free, the vectors of the form $l_A\ m_{\dot B}$, for any spinor $m_{\dot B}$, are tangent to a congruence of geodesic, totally null, two-dimensional surfaces (null strings). Analogously, the vectors of the form $m_A\ l_{\dot B}$, for any spinor m_A, are tangent to another congruence of null strings. These two congruences intersect each other along the shear-free congruence of null geodesics which have l as tangents,[13] in fact, l defines the only real direction tangent to each of these null strings.

Actually, the derivation presented below only requires the local existence of a spinor field $l_{\dot A}$ such that the vectors of the form $m_A\ l_{\dot B}$, with m_A variable, are tangent to a congruence of null strings and which satisfies the condition $l^{\dot A} l^{\dot B} l^{\dot C} C_{\dot A\dot B\dot C\dot D} = 0$, where $C_{\dot A\dot B\dot C\dot D}$ denotes the spinorial components of the anti-self-dual part of the conformal curvature tensor; without having to assume analogous conditions for an undotted spinor l_A and without any explicit restriction on the Ricci tensor. Under these assumptions there exists locally a set of (complex) coordinates q^A, p^A such that the metric has the form[14,16,18]

$$g = 2\phi^{-2}\, dq^A \underset{s}{\otimes} (dp_A + Q_{AB}\, dq^B), \tag{2.1}$$

 0022-2488/85/040836-05$02.50

where ϕ and Q_{AB} are complex-valued functions with $Q_{AB} = Q_{BA}$. In terms of these coordinates the null strings whose tangent vectors are of the form $m_A\, l_{\dot B}$ are given by $dq^A = 0$.

A convenient basis (null tetrad) for the tangent space is

$$\partial_{A\dot 1} = \sqrt{2}\frac{\partial}{\partial p^A} \equiv \sqrt{2}\partial_A,$$
$$\partial_{A\dot 2} = \sqrt{2}\phi^2\left(\frac{\partial}{\partial q^A} + Q_{AB}\partial^B\right) \equiv \sqrt{2}\phi^2 D_A, \tag{2.2}$$

which satisfies $\partial_{A\dot B}\cdot\partial_{C\dot D} = -2\epsilon_{AC}\epsilon_{\dot B\dot D}$ (however, it does not satisfy the Hermiticity condition $\partial_{A\dot B} = \overline{\partial_{B\dot A}}$; therefore, with respect to this basis the dotted components are not the complex conjugates of the undotted ones). Clearly, the vector fields ∂_A are tangent to the congruence of null strings $dq^A = 0$. It is easy to verify that the following commutation relations hold:

$$\partial^A\partial_A = 0, \tag{2.3a}$$
$$\partial^A D_B = D_B\partial^A + (\partial^A Q_{BC})\partial^C, \tag{2.3b}$$
$$D^A D_A = (D^A Q_{AB})\partial^B. \tag{2.3c}$$

In terms of the spinorial notation, the source-free Yang–Mills equations are expressed by

$$\nabla^A{}_{\dot R}\, F_{AB} + [A^A{}_{\dot R}, F_{AB}] = 0,$$

and

$$\nabla_R{}^{\dot A} F_{\dot A\dot B} + [A_R{}^{\dot A}, F_{\dot A\dot B}] = 0, \tag{2.4}$$

with

$$F_{AB} = \nabla_{(A}{}^{\dot R} A_{B)\dot R} + A_{(A}{}^{\dot R} A_{B)\dot R},$$
$$F_{\dot A\dot B} = \nabla^R{}_{(\dot A} A_{|R|\dot B)} + A^R{}_{(\dot A} A_{|R|\dot B)} \tag{2.5}$$

(round brackets denote symmetrization on the indices enclosed). F_{AB} and $F_{\dot A\dot B}$ are matrices which represent the spinorial components of the self-dual and the anti-self-dual part of the field strength, respectively, while $A_{A\dot B} \equiv A\,(\partial_{A\dot B})$, where A denotes the matrix-valued potential one-form and $\partial_{A\dot B}$ is a basis of the tangent space labeled with spinorial indices; $\nabla_{A\dot B}$ is the covariant derivative with respect to the Levi-Cività connection along $\partial_{A\dot B}$. With respect to the tetrad (2.2), using the corresponding components of the connection,[16] Eqs. (2.4)–(2.5) are given explicitly by

$$\sqrt{2}\partial^A(\phi^{-2}F_{AB}) + [A^A{}_{\dot 1}, \phi^{-2}F_{AB}] = 0, \tag{2.6a}$$

$$\sqrt{2}\{D^A(\phi^{-2}F_{AB}) - (\partial_B Q^{AR})\phi^{-2}F_{AR}\} + [\phi^{-2}A^A{}_{\dot 2}, \phi^{-2}F_{AB}] = 0, \tag{2.6b}$$

$$\sqrt{2}\{\partial_R(\phi^{-2}F_{\dot 1\dot 2}) - (D_R + \partial^B Q_{BR})F_{\dot 1\dot 1}\} + [A_{R\dot 1}, \phi^{-2}F_{\dot 1\dot 2}] - [\phi^{-2}A_{R\dot 2}, F_{\dot 1\dot 1}] = 0, \tag{2.7a}$$

$$\sqrt{2}\{\partial_R(\phi^{-4}F_{\dot 2\dot 2}) - D_R(\phi^{-2}F_{\dot 1\dot 2}) - (D^B Q_{BR})F_{\dot 1\dot 1}\} + [A_{R\dot 1}, \phi^{-4}F_{\dot 2\dot 2}] - [\phi^{-2}A_{R\dot 2}, \phi^{-2}F_{\dot 1\dot 2}] = 0, \tag{2.7b}$$

$$\phi^{-2}F_{AB} = \sqrt{2}\{\partial_{(A}\,\phi^{-2}A_{B)\dot 2} - D_{(A}A_{B)\dot 1} + \partial_{(A}Q_{B)}{}^R A_{R\dot 1}\} + A_{(A|\dot 1}\phi^{-2}A_{B)\dot 2} - \phi^{-2}A_{(A|\dot 2|}A_{B)\dot 1}, \tag{2.8}$$

$$F_{\dot 1\dot 1} = \sqrt{2}\partial^R A_{R\dot 1} + A^R{}_{\dot 1}A_{R\dot 1}, \tag{2.9a}$$

$$2\phi^{-2}F_{\dot 1\dot 2} = \sqrt{2}\{\partial^R(\phi^{-2}A_{R\dot 2}) + (D^R - \partial_B Q^{BR})A_{R\dot 1}\} + A^R{}_{\dot 1}\,\phi^{-2}A_{R\dot 2} + \phi^{-2}A^R{}_{\dot 2}\,A_{R\dot 1}, \tag{2.9b}$$

$$\phi^{-4}F_{\dot 2\dot 2} = \sqrt{2}\{D^R(\phi^{-2}A_{R\dot 2}) - (D_B Q^{BR})A_{R\dot 1}\} + \phi^{-2}A^R{}_{\dot 2}\,\phi^{-2}A_{R\dot 2}. \tag{2.9c}$$

[The fact that the conformal factor ϕ appears in Eqs. (2.6)–(2.9) only through certain combinations with the gauge fields is related with the conformal invariance of the Yang–Mills equations.]

We shall now restrict the gauge field by aligning it along the congruence of null strings, that is, we shall impose the condition $F_{\dot 1\dot 1} = 0$. Written in a covariant form, this condition amounts to $l^{\dot A}l^{\dot B}F_{\dot A\dot B} = 0$. From Eq. (2.9a) it follows that the condition $F_{\dot 1\dot 1} = 0$ is locally equivalent to the existence of a nonsingular matrix M such that (cf., Ref. 1)

$$A_{R\dot 1} = \sqrt{2}M^{-1}\,\partial_R\, M. \tag{2.10}$$

Substituting this expression, together with $F_{\dot 1\dot 1} = 0$, into Eq. (2.7a) we obtain $\partial_R(M\phi^{-2}F_{\dot 1\dot 2}M^{-1}) = 0$, which means that

$$\phi^{-2}F_{\dot 1\dot 2} = 2M^{-1}\epsilon M, \tag{2.11}$$

where ϵ is a matrix which depends on q^A only. On the other hand, from Eq. (2.9b), using (2.10) and the commutation relation (2.3b), we have $2\phi^{-2}F_{\dot 1\dot 2} = \sqrt{2}M^{-1}\partial^R\{M(\phi^{-2}A_{R\dot 2} - \sqrt{2}M^{-1}D_R\, M)M^{-1}\}M$; therefore, comparing with (2.11), it follows that there exists locally a matrix H such that

$$\phi^{-2}A_{R\dot 2} = \sqrt{2}\{M^{-1}D_R\, M - M^{-1}(\epsilon p_R - \partial_R\, H)M\}. \tag{2.12}$$

Substituting (2.10)–(2.12) into Eq. (2.7b), we find $\partial_R\{M\phi^{-4}F_{\dot 2\dot 2}\,M^{-1} - 2\epsilon_{,A}\,p^A - 2[H,\epsilon]\} = 0$, where $\epsilon_{,A} \equiv \partial\epsilon/\partial q^A$, hence

$$\phi^{-4}F_{\dot 2\dot 2} = 2M^{-1}\{\epsilon_{,A}\,p^A + [H,\epsilon] + \delta\}M, \tag{2.13}$$

where δ is a matrix which depends on q^A only. Similarly, from (2.9c), using the commutation relation (2.3c), we obtain

$$\phi^{-4}F_{\dot 2\dot 2} = 2M^{-1}\{\epsilon_{,A}\,p^A + D^A\,\partial_A\, H + \partial^A H\,\partial_A H + [\,p^A\,\partial_A\, H,\epsilon]\}M. \tag{2.14}$$

Comparison of (2.13) with (2.14) shows that H must satisfy the condition

$$D^A\,\partial_A\, H + \partial^A H\,\partial_A\, H + [\,p^A\,\partial_A\, H - H,\epsilon] = \delta. \tag{2.15}$$

The components F_{AB} can be now obtained by substituting (2.10) and (2.12) into (2.8), which gives

$$\phi^{-2}F_{AB} = 2M^{-1}(\partial_A\,\partial_B\, H)M, \tag{2.16}$$

then, Eq. (2.6a) is automatically satisfied while Eq. (2.6b) holds if and only if condition (2.15) does.

Thus, if a source-free gauge field satisfies the condition $F_{\dot 1\dot 1} = 0$, then the field is given in terms of two arbitrary matrices, ϵ and δ, depending on q^A only and a matrix potential, H, which must fulfill a second-order differential equation with quadratic nonlinearities [Eq. (2.15)]. The alignment condition $l^{\dot A}l^{\dot B}F_{\dot A\dot B} = 0$ means that the covariant derivatives with respect to the gauge field along two commuting vector fields tangent to a null string defined by $l_{\dot A}$ commute, i.e., the connection defined by the gauge field restricted to each null string is flat.

The potential is given by

$$A = A_\mu \, dx^\mu = M^{-1} dM - M^{-1}(\epsilon p_R - \partial_R H) M \, dq^R. \tag{2.17}$$

Under a gauge transformation given by a matrix L, A is replaced by $L^{-1}AL + L^{-1}dL$, which corresponds to substitute ML in place of M in Eq. (2.17), leaving ϵ, δ, and H unchanged. The matrix M represents (the inverse of) a gauge transformation which makes $A_{R\dot{1}} = 0$ [see Eq. (2.10)]. However, for a real field, M does not belong to the gauge group; its presence in the matrix-valued one-form A is necessary in order to assure the correct behavior for A [i.e., $A(v)$ must belong to the Lie algebra of the gauge group for any real tangent vector v]. For a fixed gauge, M is not uniquely defined by Eq. (2.10) but it can be replaced by ζM, where ζ is an arbitrary nonsingular matrix depending on q^A only. Under this last transformation, ϵ, δ, and H are not invariant; they are replaced by $\zeta\epsilon\zeta^{-1}$, $\zeta\delta\zeta^{-1}$ and $\zeta H \zeta^{-1} - \zeta_{,A} p^A \zeta^{-1}$, respectively.

Solutions of (2.15) determine locally all the source-free Yang–Mills fields aligned with the congruence of null strings, considering these fields as perturbations, or test fields, on a given background. In Sec. III we show that the coupled Einstein–Yang–Mills equations can be simultaneously reduced, consistently with the condition of the existence of a congruence of null strings and the alignment of the gauge field.

In the case of the electromagnetic field, Eq. (2.15) reduces to $D^A \partial_A H = \delta$. Hence, if we define $\widetilde{H} \equiv H + \gamma_A p^A$, where $\gamma_A = \gamma_A(q^R)$ is such that $\gamma^A{}_{,A} = \delta$, then we have

$$D^A \partial_A \widetilde{H} = 0, \tag{2.18}$$

and $A = d \ln M + (\epsilon p^R + \gamma^R) dq_R + \partial_R \widetilde{H} \, dq^R$. The self-dual part of the field is then also given by

$$F_{AB} = 2\phi^2 \partial_A \partial_B \widetilde{H}. \tag{2.19}$$

These expressions were previously given in Ref. 15. A peculiar feature of the electromagnetic (i.e., Abelian) case is that Eqs. (2.18)–(2.19) actually represent, locally, the (self-dual part of the) most general solution of the source-free Maxwell equations. Eqs. (2.18)–(2.19) are equivalent to the expressions found by Cohen and Kegeles[19] and Wald.[20] The proof that this is indeed the most general solution was given in Refs. 15 and 16.

Expressed in a covariant way, Eqs. (2.18) and (2.19), respectively, take the form (rescaling if necessary the function $\widetilde{H}$)

$$\nabla_{B(\dot{A}} \phi^{-2} \nabla^{B\dot{D}} \widetilde{H} l_{\dot{C})} l_{\dot{D}} = 0 \tag{2.20}$$

and[17]

$$F_{AB} = \nabla_{(A}{}^{\dot{C}} \phi^{-2} \nabla_{B)}{}^{\dot{D}} \widetilde{H} l_{\dot{C}} l_{\dot{D}}. \tag{2.21}$$

Thus, an arbitrary *real* electromagnetic field without sources in a real algebraically special space-time which admits a geodesic and shear-free principal null congruence, referred to an arbitrary null tetrad such that $\partial_{A\dot{B}} = \overline{\partial_{B\dot{A}}}$, is given by

$$F_{AB} = \nabla_{(A}{}^{\dot{R}} A_{B)\dot{R}}, \quad F_{\dot{A}\dot{B}} = \nabla^R{}_{(\dot{A}} A_{|R|\dot{B})}, \tag{2.22}$$

with

$$A_{B\dot{R}} = \phi^{-2} \nabla_B{}^{\dot{S}} \widetilde{H} l_{\dot{R}} l_{\dot{S}} + \text{Hermitian conjugate},$$

where $\widetilde{H}$ satisfies Eq. (2.20). The potential given in (2.22) fulfills $l^{\dot{B}} l^{\dot{R}} A_{B\dot{R}} = 0$, i.e., $l^\mu A_\mu = 0$. It is perhaps worthwhile to point out that a duality rotation corresponds to replace $\widetilde{H}$ by $e^{i\psi}\widetilde{H}$.

To close this section, we express the self-dual part of the Yang–Mills field (2.16) in a covariant form[17]

$$F_{AB} = M^{-1} \{ \nabla_{(A}{}^{\dot{C}} \phi^{-2} \nabla_{B)}{}^{\dot{D}} H l_{\dot{C}} l_{\dot{D}} \} M. \tag{2.23}$$

III. REDUCTION OF THE EINSTEIN–YANG–MILLS EQUATIONS

In Ref. 16 it was shown that if the space-time admits a congruence of null strings defined by a spinor $l_{\dot{A}}$ (Ref. 18) and the energy-momentum tensor of the matter has a constant trace and satisfies the condition $l^{\dot{C}} l^{\dot{D}} T_{AB\dot{C}\dot{D}} = 0$, where $T_{AB\dot{C}\dot{D}}$ are the spinorial components of the traceless part of the energy-momentum tensor, then the Einstein field equations reduce to a single second-order partial differential equation with quadratic nonlinearities. Since the energy-momentum tensor of a Yang–Mills field is traceless and $T_{AB\dot{C}\dot{D}} = -(1/\pi)\text{Tr}\, F_{AB} F_{\dot{C}\dot{D}}$, where Tr represents the trace, the condition $l^{\dot{C}} l^{\dot{D}} T_{AB\dot{C}\dot{D}} = 0$ is satisfied when $l^{\dot{A}} l^{\dot{B}} F_{\dot{A}\dot{B}} = 0$ (i.e., when the Yang–Mills field is aligned along the congruence of null strings). Thus, under this restriction, using the results of Sec. II, the coupled Einstein–Yang–Mills equations can be simultaneously reduced. This is a generalization of the results obtained in Ref. 15 for the Einstein–Maxwell equations under analogous restrictions.

It may be pointed out that, if $l_{\dot{A}}$ defines a congruence of null strings then, assuming that Einstein's equations hold, condition $l^{\dot{C}} l^{\dot{D}} T_{AB\dot{C}\dot{D}} = 0$ implies $l^{\dot{A}} l^{\dot{B}} l^{\dot{C}} C_{\dot{A}\dot{B}\dot{C}\dot{D}} = 0$,[13] and therefore the covariant derivatives of a vector field *tangent* to a null string along two commuting vector fields tangent to that null string do commute, i.e., the Levi-Cività connection restricted to the tangent space of each null string is flat.

From Eqs. (2.11), (2.13), and (2.16) we find, assuming $F_{\dot{1}\dot{1}} = 0$, that the nonvanishing components of the energy-momentum tensor of the gauge field are given by

$$\begin{aligned} T_{AB\dot{1}\dot{2}} &= (1/8\pi)\phi^4 \partial_A \partial_B T_0, \\ T_{AB\dot{2}\dot{2}} &= -(1/8\pi)\phi^6 \partial_{(A} T_{B)}, \end{aligned} \tag{3.1}$$

where

$$\begin{aligned} T_0 &= -32\, \text{Tr}\, \epsilon H, \\ T_A &= 32\, \text{Tr}\{(\epsilon_{,B} p^B + [H,\epsilon] + \delta)\partial_A H - \epsilon_{,A} H\}, \end{aligned} \tag{3.2}$$

hence, by virtue of Eq. (2.15),

$$\tfrac{1}{2}\partial^A T_A + D^A \partial_A T_0 = -32\, \text{Tr}\, \epsilon\delta. \tag{3.3}$$

Turning now to the reduced Yang–Mills equations (2.15), using (2.3b) we have

$$\begin{aligned} 0 = \partial_A \{ & D^A H + H \partial_B Q^{BA} \\ & + \tfrac{1}{2}[\partial^A H, H] + [H,\epsilon] p^A - \tfrac{1}{2}\delta p^A \} \\ & + 3[\epsilon, H] - H \partial_A \partial_B Q^{AB}. \end{aligned} \tag{3.4}$$

Condition $l^{\dot{C}} l^{\dot{D}} T_{AB\dot{C}\dot{D}} = 0$ implies $\phi = J_A p^A + \kappa$, where J_A and κ depend on q^A only. Now, the reduced form of the Einstein equations given in Ref. 16 is valid only in a restricted class of coordinate systems in which, in the case where J_A

$= 0$ (called case I in Refs. 14–16), κ is equal to one; and in the case where $J_A \neq 0$ (case II), J_A and κ are constant. Then, in case I,[16]

$$Q_{AB} = -\partial_{(A} B_{B)} - \tfrac{2}{3} L_{(A}\, p_{B)} + (\lambda/3) p_A\, p_B, \qquad B_A \equiv G_A + \partial_A \theta, \tag{3.5}$$

where G_A is any solution of $\partial_A G^A = 4T_0$, θ is some function, $L_A = L_A(q^R)$ and λ is a cosmological constant. Thus, from Eq. (3.4) we find

$$0 = \partial_A \{ D^A H - H(2\partial^A T_0 + L^A - \lambda p^A) + \tfrac{1}{2}[\partial^A H, H] + [H,\epsilon] p^A - \tfrac{1}{2}\delta p^A + 3[\epsilon, F^A] - 2\lambda F^A \}, \tag{3.6}$$

where the matrices F^A satisfy $\partial_A F^A = H$, hence there exists locally a matrix χ such that the expression between braces in Eq. (3.6) is equal to $\partial^A \chi$.

Choosing $G_A = -2^7 \operatorname{Tr} \epsilon F_A$ and using Eqs. (3.2), (3.3), and (3.6), we find that the right-hand side of Eq. (3.33b) of Ref. 16 is equal to $\partial_A \{32 \operatorname{Tr}[(\epsilon_{,B}\, p^B + \delta) H + 2\epsilon\chi]\}$, therefore Einstein's equations reduce in this case to

$$\begin{aligned} &D^A B_A - \tfrac{1}{2}\partial^A B^B \partial_{(A} B_{B)} - (L^A - \lambda p^A) B_A - \lambda\theta \\ &\quad + \tfrac{1}{18}(L_A\, p^A)^2 - \tfrac{1}{6} L_{A,B}\, p^A p^B + 2(32 \operatorname{Tr} \epsilon H)^2 \\ &= 32 \operatorname{Tr}\{(\epsilon_{,A}\, p^A + \delta) H + 2\epsilon\chi\} + N_A\, p^A + \gamma, \end{aligned} \tag{3.7a}$$

where N_A and γ are arbitrary functions of q^A only and χ is defined through

$$\begin{aligned} \partial_A \chi = {} & D_A H + H\{2\partial_A(32 \operatorname{Tr} \epsilon H) - L_A + \lambda p_A\} \\ & + \tfrac{1}{2}[\partial_A H, H] \\ & + [H,\epsilon]\, p_A - \tfrac{1}{2}\delta p_A + 3[\epsilon, F_A] - 2\lambda F_A. \end{aligned} \tag{3.7b}$$

In case II, the presence of ϵ makes a difficult task to write Eq. (3.4) as a continuity equation. We shall restrict ourselves to the subcase where ϵ vanishes (i.e., $F_{\dot{1}\dot{2}} = 0$); then $T_0 = 0$ [see Eq. (3.2)], and from Eqs. (3.24) and (3.27a) of Ref. 16 we have

$$Q_{AB} = -\phi^3 \partial_{(A}\, \phi^{-2} \partial_{B)} W + (\mu\phi^3 + \lambda/6) K_A\, K_B, \tag{3.8}$$

where W is some function, $\mu = \mu(q^A)$, K_A is a constant spinor such that $K^A J_A = 1$ and λ is the cosmological constant. Thus, in case II with $\epsilon = 0$, Eq. (3.4) becomes

$$\begin{aligned} 0 = \partial_A \{ & D^A H + H(3\mu\phi^2 K^A - 2J_B \partial^A \partial^B W) \\ & + \tfrac{1}{2}[\partial^A H, H] - \tfrac{1}{2}\delta p^A - 6\mu F^A \}, \end{aligned} \tag{3.9}$$

where the matrices F^A are required to satisfy $\partial_A F^A = \phi H$. Therefore, there exists locally a matrix χ such that

$$\begin{aligned} \partial_A \chi = {} & D_A H + H(3\mu\phi^2 K_A + 2J^B \partial_A \partial_B W) \\ & + \tfrac{1}{2}[\partial_A H, H] - \tfrac{1}{2}\delta p_A - 6\mu F_A \end{aligned} \tag{3.10a}$$

and from Eqs. (3.2), (3.3) above, and Eq. (3.33a) of Ref. 16, we find that Einstein's equations give

$$\begin{aligned} &\tfrac{1}{2}\phi^4(\partial^A \phi^{-2} \partial^B W) \partial_A\, \phi^{-2}\, \partial_B W \\ &\quad + \phi^{-1} \partial_A W^{,A} - (\lambda/6)\phi^{-1} K^A \partial_A\, K^B \partial_B\, W \\ &\quad - \mu\phi^4 K^A \partial_A\, \phi^{-1} K^B \partial_B\, \phi^{-1} W \\ &\quad + \tfrac{1}{2} K^A p_A [K^B p_B\, J^C - (\phi + \kappa) K^C] \mu_{,C} \\ &= 32 \operatorname{Tr} \delta H + N_A\, p^A + \gamma, \end{aligned} \tag{3.10b}$$

where N_A and γ are arbitrary functions of q^A only.

IV. SELF-DUAL GAUGE FIELDS IN SELF-DUAL SPACE-TIMES

From Eqs. (2.11) and (2.13) we see that an arbitrary self-dual gauge field corresponds to $\epsilon = \delta = 0$. Then Eq. (2.15) reduces to

$$D^A \partial_A\, H + \partial^A H\, \partial_A\, H = 0. \tag{4.1}$$

On the other hand, the metric of a self-dual space-time is locally given by Eq. (2.1), with $\phi = 1$ and

$$Q_{AB} = -\partial_A\, \partial_B \theta, \tag{4.2a}$$

where θ is a solution of[21,22]

$$\tfrac{1}{2}(\partial_A \partial_B \theta)(\partial^A \partial^B \theta) + \partial_A \theta^{,A} = 0. \tag{4.2b}$$

It is easy to verify that Eqs. (4.2) are equivalent to the commutation relations[23] [cf., Eqs. (2.3)]

$$\partial^A D_A = D_A \partial^A, \tag{4.3a}$$

$$D^A D_A = 0. \tag{4.3b}$$

Therefore, for a self-dual gauge field in a self-dual space-time, from Eqs. (4.1) and (4.3), we have [cf., Eq. (3.4)]

$$\partial_A \{ D^A H + (\partial^A H) H \} = 0, \tag{4.4}$$

which implies the local existence of a matrix $\chi^{(2)}$ such that

$$D^A H + (\partial^A H) H = \partial^A \chi^{(2)}. \tag{4.5}$$

We now define inductively $\chi^{(n)}$ by

$$(D^A + \partial^A H)\chi^{(n-1)} = \partial^A \chi^{(n)}. \tag{4.6}$$

It can be easily verified that the integrability conditions of (4.6) are indeed satisfied since, using (4.3a) and (4.6),

$$\begin{aligned} \partial_A \{ (D^A + \partial^A H)\chi^{(n-1)} \} &= (D^A + \partial^A H)\partial_A\, \chi^{(n-1)} \\ &= (D^A + \partial^A H)(D_A + \partial_A H)\chi^{(n-2)}, \end{aligned}$$

which vanishes by virtue of (4.3) and (4.1). Equation (4.6) holds for $n = 1,2,\ldots$, if we define $\chi^{(1)} \equiv H, \chi^{(0)} \equiv 1$; then (4.5) becomes a special case of (4.6).

Thus, due to the simplicity of the commutation relations (4.3), we have been able to obtain an infinity of continuity equations from the self-duality condition (4.1); each of these continuity equations being locally equivalent to the existence of a corresponding potential.

Following Chau Wang *et al.*[11] we now consider

$$\chi \equiv \sum_{n=0}^{\infty} \lambda^n \chi^{(n)}, \tag{4.7}$$

where λ is a complex variable. Then, from Eq. (4.6) it follows that χ obeys the equation

$$\partial_A\, \chi = \lambda\, (D_A + \partial_A\, H)\chi. \tag{4.8}$$

Conversely, the integrability conditions of the linear system $\partial_A\, \chi = \lambda\, (D_A + A_A)\chi$, lead [assuming (4.3)] to the local existence of a matrix H such that $A_A = \partial_A\, H$, which must satisfy Eq. (4.1). In the case of flat space-time, Eq. (4.8) is equivalent to the linear system obtained by other authors.[10,11]

It must be pointed out that Eq. (4.2b), which locally determines all the self-dual space-times, also leads to an infinite number of conservation laws and of corresponding potentials.[22,24]

The matrix χ can be used to define infinitesimal transformations which map a given solution of (4.1) into a new one.[9] Such an infinitesimal transformation is given by

$$\delta H = \chi T \chi^{-1}, \tag{4.9}$$

where T is an infinitesimal constant matrix. From Eqs. (4.9) and (4.8) it follows that $\partial_A \delta H = \lambda \{D_A \delta H + [\partial_A H, \delta H]\}$, therefore, assuming that H is a solution of (4.1) and omitting second-order terms, we have

$$\begin{aligned} D^A\partial_A(H+\delta H) &+ \partial^A(H+\delta H)\partial_A(H+\delta H) \\ &= \partial_A\{D^A\delta H + [\partial^A H, \delta H]\} \\ &= \partial_A(\lambda^{-1}\partial^A \delta H) = 0. \end{aligned} \tag{4.10}$$

This means that, to first order in δH, $H + \delta H$ is also a solution of Eq. (4.1), thus defining another self-dual gauge field.

Equation (4.1) can be written in an equivalent form in terms of a matrix J defined by

$$J^{-1} D_A J = \partial_A H. \tag{4.11}$$

It is easy to verify, using (4.3b), that the integrability condition of (4.11) yields Eq. (4.1). In terms of J, the self-duality condition becomes

$$\partial^A(J^{-1}D_A J) = 0, \tag{4.12}$$

which follows from (4.11) by applying ∂^A. In the case of flat space-time (with $D_A = \partial/\partial q^A$), this form of the self-duality conditions is the one used by several authors.[4–6,8,9,11] Also, the Einstein equations for the stationary axisymmetric vacuum fields have been expressed by Ward[25] in a form analogous to Eq. (4.12).

V. FINAL REMARKS

The results given here show that the existence of a null foliation and the use of coordinates and bases adapted to the null strings lead to a rather straightforward integration of the Yang–Mills equations when the field is suitable aligned with the foliation. However, due to the fact that the null tetrad (2.2) does not satisfy the Hermiticity condition, from the components of the field referred to that tetrad it is not easy to see what conditions must be imposed on the matrices ϵ, δ, and H in order to get a real field. Nevertheless, this problem can be solved by working with covariant expressions, such as that given by Eq. (2.23). [Note that the expression of the potential one-form A given in (2.17) is basis-independent but it is written in terms of a particular set of coordinates.]

It seems reasonable to expect that the reduced form of the Yang–Mills equations (2.15), which determines non-self-dual fields, also leads to an infinite set of conservation laws, which is a common feature of various nonlinear systems.

ACKNOWLEDGMENTS

The author wishes to thank Professor J. F. Plebański for helpful conversations.

[1] C. N. Yang, Phys. Rev. Lett. **38**, 1377 (1977).
[2] E. T. Newman, Phys. Rev. D **18**, 2901 (1978).
[3] K. P. Tod, Twistor Newsletter No. 6 (Dec. 1977).
[4] Y. Brihaye, D. B. Fairlie, J. Nuyts, and R. G. Yates, J. Math. Phys. **19**, 2528 (1978).
[5] K. Pohlmeyer, Commun. Math. Phys. **72**, 37 (1980).
[6] M. K. Prasad, A. Sinha, and L.-L. Chau Wang, Phys. Lett. B **87**, 237 (1979).
[7] E. F. Corrigan, D. B. Fairlie, R. G. Yates, and P. Goddard, Commun. Math. Phys. **58**, 223 (1978).
[8] M. K. Prasad, A. Sinha, and L.-L. Chau Wang, Phys. Rev. Lett. **43**, 750 (1979).
[9] L.-L. Chau Wang, M.-L. Ge, and Y.-S. Wu, Phys. Rev. D **25**, 1086 (1982).
[10] A. A. Belavin and V. E. Zakharov, Phys. Lett. B **73**, 53 (1978).
[11] L.-L. Chau Wang, M. K. Prasad, and A. Sinha, Phys. Rev. D **24**, 1574 (1981); P. Forgács, Z. Horváth, and L. Palla, Phys. Rev. D **23**, 1876 (1981).
[12] R. S. Ward, Phys. Lett. A **61**, 81 (1977); M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977).
[13] J. F. Plebański and S. Hacyan, J. Math. Phys. **16**, 2403 (1975). For a more detailed discussion on null strings see J. F. Plebański and K. Rózga, J. Math. Phys. **25**, 1930 (1984).
[14] J. F. Plebański and I. Robinson, Phys. Rev. Lett. **37**, 493 (1976). For a spinorial version see J. D. Finley, III and J. F. Plebański, J. Math. Phys. **17**, 2207 (1976).
[15] A. García, J. F. Plebański, and I. Robinson, Gen. Relativ. Gravit. **8**, 841 (1977). For a spinorial derivation see J. D. Finley, III and J. F. Plebański, J. Math. Phys. **18**, 1662 (1977).
[16] G. F. Torres del Castillo, J. Math. Phys. **24**, 590 (1983).
[17] G. F. Torres del Castillo, J. Math. Phys. **25**, 342 (1984).
[18] We employ the same notation as in Refs. 14–17, only changing undotted indices to dotted ones and vice versa.
[19] J. M. Cohen and L. S. Kegeles, Phys. Rev. D **10**, 1070 (1974).
[20] R. M. Wald, Phys. Rev. Lett. **41**, 203 (1978).
[21] J. F. Plebański, J. Math. Phys. **16**, 2395 (1975).
[22] C. P. Boyer and J. F. Plebański, J. Math. Phys. **18**, 1022 (1977).
[23] Actually, from (4.3) it follows that the left-hand side of (4.2b) is a function of q^A only which is absorbed by redefining θ, without altering Q_{AB}. See Ref. 21.
[24] C. P. Boyer and J. F. Plebański, J. Math. Phys. **26**, 229 (1985).
[25] R. S. Ward, Gen. Relativ. Gravit. **15**, 105 (1983).

Nonabelian progressive waves

C. H. Oh
Department of Physics, National University of Singapore, Kent Ridge, Singapore 0511, Republic of Singapore

Rosy Teh
MARA Institute of Technology, Arau, Perlis, Malaysia

(Received 20 June 1984; accepted for publication 17 August 1984)

We present two families of nonabelian wave solutions of the Yang–Mills field equations, some of the previous known solutions occur as special cases of our solutions.

I. INTRODUCTION

Several years ago, Coleman[1] constructed plane wave solutions of the Yang–Mills (YM) field equations in Minkowski space-time, since then there has been some interest in nonabelian waves.[2–9] Unlike wave solutions in electrodynamics which are extremely useful, the physical relevance of nonabelian waves is certainly not immediate and clear. However, as gauge field theories are playing an increasingly vital role in elucidating fundamental interactions of elementary particles, it may not be without interest to investigate these solutions further (and besides future work may shed light on them). The authors in Ref. 4 suggested that wave solutions may be relevant in our understanding of the structure of quantum vacuum and asymptotic states of the YM theory.

In this paper, we present two families of nonabelian wave solutions which are of different nature. It turns out that the previous known solutions[1,2] are special cases of our general solutions. For the first family of solutions, the gauge field potential A_μ and the field strengths $F_{\mu\nu}$ satisfy

$$[A_\mu, A_\nu] = [A_\mu, F^{\mu\nu}] = 0,$$

but they are essentially nonabelian since

$$[A_\mu, F^{\alpha\beta}] \neq 0.$$

The energy and momentum densities are equal in magnitude, and the Lagrangian density and $F^a_{\mu\nu}F^{b\mu\nu}$ vanish. The nonabelian electric and magnetic fields and the direction of wave propagation are mutually perpendicular to each other but the magnitude of the field strengths vary over each wave front. This type of wave is referred to as a guided wave,[2] and solutions of Ref. 2 are included here. Because of the vanishing of nonlinear terms $[A_\mu, A_\nu]$ and $[A_\mu, F^{\mu\nu}]$, the full strength of the nonabelian effect may not be realized in the first family of solutions just described. For this reason we introduce a new ansatz to obtain a second family of solutions such that the nonlinear terms are nonzero. In contrast with the first family of solutions, here the energy and momentum densities are not equal in magnitude, meaning that the wave field when quantized describes a particle of nonzero mass. The waves propagate along a fixed direction and the wave fronts are plane. However, it turns out that the phase velocity v_p can be greater than, less than, or equal to c, the speed of light. When $v_p < c$, the wave solutions have singularities whereas when $v_p > c$, they are regular throughout the whole space-time and possess finite energy density everywhere. Our solutions become the Coleman plane waves[1] when $v_p = c$. Note that both families of our progressive wave solutions are non-self-dual or non-self-anti-dual. Self-dual or self-anti-dual solutions[6] cannot describe progressive waves since the Poynting vector vanishes. In the next section we introduce our notations. In Sec. III the first family of solutions is given, the second family of solutions is exhibited in Sec. IV, and we end in Sec. V with some remarks.

II. NOTATIONS

The SU(2) YM field equations are

$$D_\mu F^{\mu\nu} = 0, \tag{1a}$$

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu + [A^\mu, A^\nu], \tag{1b}$$

$$A_\mu = g A^a_\mu \sigma^a/(2i), \tag{1c}$$

where g is the gauge field coupling constant, σ_a the Pauli matrices, and our metric is $\mathrm{diag}(g_{\mu\nu}) = (-\ +\ +\ +)$. The nonabelian electric and magnetic fields are, respectively, given by

$$E^i = F^{0i},\ B^i = \tfrac{1}{2}\epsilon^{ijk}F_{jk}. \tag{2}$$

The energy-momentum tensor is written as

$$T_{\mu\nu} = F^a_{\mu\alpha}F^{\alpha a}_{\ \nu} + g_{\mu\nu}\mathscr{L}, \tag{3}$$

where the Lagrangian density is

$$\mathscr{L} = -\tfrac{1}{4}F^a_{\mu\nu}F^{a\mu\nu}. \tag{4}$$

The energy density can be expressed as

$$T^{00} = \tfrac{1}{2}(E^a_i E^a_i + B^a_i B^a_i), \tag{5}$$

and the Poynting vector S^i is related to the momentum density T^{0i} by

$$S^i = \epsilon_0 c^2 T^{0i} = \epsilon_0 c^2(\mathbf{E}^a \wedge \mathbf{B}^a)^i, \tag{6}$$

where ϵ_0 is the premittivity and we shall set the speed of light c equal to 1.

III. NONABELIAN WAVES WITH NULL FIELD STRENGTHS

To construct the first family of nonabelian progressive waves for the YM equations, we prescribe the following ansatz for the YM potential in Minkowski space:

$$A^a_\mu = (\psi_\lambda f^a_\lambda(U) + h^a(U))\partial_\mu U. \tag{7}$$

Here U and ψ_λ are functions of x_μ, f^a_λ and h^a depend on U

 0022-2488/85/040841-04$02.50

only, and the repeated index $\bar{\lambda}$ means summation from 1 to 4; furthermore, the function $h^a(U)$ can be gauge-transformed away. The field strengths are given by

$$F^a_{\mu\nu} = (\partial_\nu U\, \partial_\mu \psi_{\bar{\lambda}} - \partial_\mu U\, \partial_\nu \psi_{\bar{\lambda}}) f^a_{\bar{\lambda}}(U). \tag{8}$$

The YM potential (7) will be a solution of Eq. (1) if the following equations are fulfilled throughout the whole space-time:

$$\partial^\mu U\, \partial_\mu U = 0, \tag{9a}$$

$$\partial^\mu U\, \partial_\mu \psi_{\bar{\lambda}} = 0, \tag{9b}$$

$$\partial_\nu\, \partial_\mu U\, \partial^\mu \psi_{\bar{\lambda}} + \partial_\nu U \Box \psi_{\bar{\lambda}} - \Box U\, \partial_\nu \psi_{\bar{\lambda}} - \partial^\mu U\, \partial_\mu U\, \partial_\nu U\, \partial_\nu \psi_{\bar{\lambda}} = 0, \tag{9c}$$

where $\Box = \partial_\mu\, \partial^\mu$. However if Eqs. (9) are satisfied everywhere except at some region of space-time, then in general the solution (7) requires the presence of an external source at that region.

Conditions (9) imply that the field strengths are null,

$$F^a_{\mu\nu} F^{b\,\mu\nu} = 0, \tag{10a}$$

$$\epsilon_{\mu\nu\alpha\beta} F^{a\,\mu\nu} F^{a\,\alpha\beta} = 0, \tag{10b}$$

thus the action for the solution (7) is zero. The energy and momentum densities can be readily evaluated by using Eq. (3); we get

$$T_{00} = \tau\, \partial_0 U\, \partial_0 U, \tag{11a}$$

$$T_{0i} = \tau\, \partial_0 U\, \partial_i U, \tag{11b}$$

$$\tau = \partial_\mu \psi_{\bar{\lambda}}\, \partial^\mu \psi_{\bar{\eta}}\, f^a_{\bar{\lambda}} f^a_{\bar{\eta}}. \tag{11c}$$

They are, in general, nonvanishing and hence solution (7) is neither self-dual nor self-anti-dual. Expression (11b) indicates that the waves are progressive along the direction defined by $\partial_i U$. From condition (9a) we find that the energy and momentum densities are equal in magnitude. The field strengths can be easily written down as

$$E^a_i = (\partial_i U\, \partial_0 \psi_{\bar{\lambda}} - \partial_0 U\, \partial_i \psi_{\bar{\lambda}}) f^a_{\bar{\lambda}}, \tag{12}$$

$$B^a_i = \epsilon_{ijk}\, \partial_k U\, \partial_j \psi_{\bar{\lambda}}\, f^a_{\bar{\lambda}}. \tag{13}$$

It is not difficult to show that the field strengths and the direction of wave propagation are mutually perpendicular to each other,

$$E^a_i B^a_i = E^a_i T_{0i} = B^a_i T_{0i} = 0. \tag{14}$$

The conditions (9) lead to $[A_\mu, A_\nu] = [A_\mu, F^{\mu\nu}] = 0$, but $[A_\mu, F^{\alpha\beta}] \neq 0$ which implies that the solution (7) is essentially nonabelian. The solution (7) will indeed become abelian if $f^a_{\bar{\lambda}} = F_{\bar{\lambda}} g^a$, where g^a is a constant vector in the internal group space and $F_{\bar{\lambda}}$ is a function.

From the ansatz (7), nonabelian wave solutions are obtained whenever we can find functions $U(x)$ and $\psi_{\bar{\lambda}}(x)$ that fulfill conditions (9). In the following we shall exhibit various solutions derived from the ansatz (7) by employing appropriate U and $\psi_{\bar{\lambda}}$.

(a) Nonabelian spherical-fronted waves. The first example that fulfills conditions (9) is given by

$$U(x) = r \pm t, \quad r^2 = x_1^2 + x_2^2 + x_3^2, \tag{15a}$$

$$\psi_1(x) = \phi, \quad \psi_2(x) = \ln \tan(\theta/2), \tag{15b}$$

$$\psi_3(x) = \psi_4(x) = 0, \tag{15c}$$

where θ and ϕ are the spherical coordinates. This gives rise to nonabelian waves which propagate radially and the magnitude of field strengths varies over each wave front and hence the waves are known as spherical-fronted waves.[10] This solution has also been discussed in Ref. 2 and it demands an external line source lying along the x_3-axis to sustain itself.[11]

(b) Nonabelian plane-fronted waves. The following choice will satisfy conditions (9):

$$U(x) = x_3 \pm t, \tag{16a}$$

$$\psi_1(x) = x_1, \; \psi_2(x) = x_2, \tag{16b}$$

$$\psi_3(x) = \ln \rho, \; \psi_4(x) = \phi, \tag{16c}$$

where $\rho^2 = x_1^2 + x_2^2$. The solution resulting from this choice describes waves traveling along the x_3-direction with plane fronts. The magnitude of the field strengths varies across each plane front and hence the solution depicts nonabelian plane-fronted waves.[10] As in the case (a), a line source lying along the x_3-axis is needed for this solution. The solution in case (a) and the present solution have also been discussed in Ref. 11 in the context of color radiation from a nonabelian source. Note that when $f^a_3 = f^a_4 = 0$ and f^a_1 and f^a_2 are bounded everywhere, the Coleman solution[1] is recovered.

(c) Nonabelian plane waves. Selecting

$$U(x) = |x_3| - t, \tag{17a}$$

$$\psi_1(x) = x_1, \;\; \psi_2(x) = x_2, \tag{17b}$$

$$\psi_3(x) = \psi_4(x) = 0, \tag{17c}$$

we arrive at another kind of plane wave solution different from the Coleman plane waves. The direction of propagation is along the x_3-direction when $x_3 > 0$ and along the negative x_3-direction when $x_3 < 0$. In other words the plane waves here travel along opposite directions with the plane $x_3 = 0$ as the dividing plane. An external source j^a_μ is present on the dividing plane, $j^a_0(x) = 0, j^a_i(x) = -2\delta(x_3)(\delta^1_i f^a_1 + \delta^2_i f^a_2)$.

(d) Many other different expressions can be constructed for U and $\psi_{\bar{\lambda}}$ such that conditions (9) are fulfilled but the resulting solutions have less obvious physical portrait. For example, the following will give rise to YM potentials which satisfy the YM equations (1) everywhere except at some regions: (d.1),

$$U(x) = \ln R - \sinh^{-1}(x^0/R), \;\; R^2 = x_i x_i - x_0^2, \tag{18a}$$

$$\psi_1(x) = \phi, \tag{18b}$$

$$\psi_2(x) = \ln \tan(\theta_2/2), \;\; \theta_2 = \cos^{-1}(x_3/r), \tag{18c}$$

$$\psi_3(x) = \psi_4(x) = 0; \tag{18d}$$

and (d.2),

$$U(x) = \ln \bar{\rho} + \tanh^{-1}(x^0/x_3), \;\; \bar{\rho}^2 = x_3^2 - x_0^2, \tag{19a}$$

$$\psi_1(x) = x_1, \;\; \psi_2(x) = x_2, \tag{19b}$$

$$\psi_3(x) = \ln \rho, \;\; \psi_4(x) = \phi. \tag{19c}$$

IV. NONABELIAN PERIODIC PLANE-FRONTED WAVES WITH BOUNDED ENERGY DENSITY[12]

As mentioned in Sec. I, the solutions presented in Sec. III may not contain the full strength of the nonabelian effect because of the vanishing of nonlinear terms such as $[A_\mu, A_\nu] = [A_\mu, F^{\mu\nu}] = 0$. In the present section, we devise a different ansatz for the YM potential so as to retain the nonlinear terms. We find that the Coleman plane wave solution[1] emerges under special conditions. The ansatz we introduce is[13]

$$A_\mu^a = \delta_1^a \Phi(u_1, u_2) p_\mu + \delta_3^a \Psi(u_1, u_2) q_\mu, \tag{20a}$$

where

$$U_A = s_A^\mu x_\mu + e_A, \; A = 1, 2, \tag{20b}$$

e_A are arbitrary constants, and the four vectors $p_\mu, q_\mu, s_{1\mu}, s_{2\mu}$ are orthogonal to each other. Substituting the above ansatz into the YM field equations (1), the following coupled nonlinear equations result:

$$\Box\Phi - g^2 q^2 \Phi \Psi^2 = 0, \tag{21a}$$

$$\Box\Psi - g^2 p^2 \Psi \Phi^2 = 0, \tag{21b}$$

where $p^2 = p_\mu p^\mu$ and $q^2 = q_\mu q^\mu$. Imposing

$$\Phi(u_1, u_2) = \Psi(u_1, u_2) \text{ and } p^2 = q^2, \tag{22}$$

Eqs. (21) become

$$\Box\Phi - g^2 p^2 \Phi^3 = 0. \tag{23a}$$

When the function Φ depends only on one variable $u_1 = u$, Eq. (23a) can be rewritten as

$$\Phi''(u) - (g^2 p^2 / s^2) \Phi^3(u) = 0, \tag{23b}$$

where prime means d/du, $u = sx + e$, and $s^2 = s_\mu s^\mu$. Solutions of Eq. (23b) can be expressed in terms of the Jacobi elliptic functions and we find[14]

$$\Phi_{11}(u) = \mathrm{cn}(u,k), \; s^2 = -g^2 p^2, \tag{24a}$$

$$\Phi_{12}(u) = \mathrm{sd}(u,k), \; s^2 = -2g^2 p^2, \tag{24b}$$

$$\Phi_{21}(u) = \mathrm{nc}(u,k), \; s^2 = g^2 p^2, \tag{25a}$$

$$\Phi_{22}(u) = \mathrm{ds}(u,k), \; s^2 = \tfrac{1}{2} g^2 p^2. \tag{25b}$$

Here the parameter k is fixed such that $k^2 = \frac{1}{2}$. Owing to the orthogonality, the vectors p_μ and q_μ must be spacelike. Consequently $g^2 p^2$ is always positive. For solutions (24), s_μ is necessarily timelike, whereas for solutions (25), it is spacelike. Solutions (24) and (25) are periodic in u and cannot be linearly superposed.

The energy-momentum tensor $T_{\mu\nu}(x)$ of the YM field specified by the ansatz (20) can be written as

$$\begin{aligned} T_{\mu\nu} = \{ & p^2 \partial_\mu \Phi \partial_\nu \Phi + q^2 \partial_\mu \Psi \partial_\nu \Psi + p_\mu p_\nu \partial_\alpha \Phi \partial^\alpha \Phi \\ & + q_\mu q_\nu \partial^\alpha \Psi \partial_\alpha \Psi + g^2 (p^2 q_\mu q_\nu + q^2 p_\mu p_\nu) \Phi^2 \Psi^2 \} \\ & - \tfrac{1}{2} g_{\mu\nu} (p^2 \partial_\alpha \Phi \partial^\alpha \Phi + q^2 \partial_\alpha \Psi \partial^\alpha \Psi \\ & + g^2 p^2 q^2 \Phi^2 \Psi^2). \end{aligned} \tag{26}$$

When conditions (22) and $\Phi = \Phi(u)$ hold, this is reduced to

$$\begin{aligned} T_{\mu\nu} = {} & p^2 s_\mu s_\nu (T_1 \Phi^4 - T_2) + (p_\mu p_\nu + q_\mu q_\nu) \\ & \times s^2 (\tfrac{1}{2} T_1 \Phi^4 - \tfrac{1}{2} T_2) - g_{\mu\nu} p^2 s^2 (T_1 \Phi^4 - \tfrac{1}{2} T_2), \end{aligned} \tag{27}$$

where $T_1 = g^2 p^2 s^{-2}$ and $T_2 = g^{-2} p^{-2} s^2 d$. Here d is a constant and it takes the value 1 for the periodic solutions $\{\Phi_{AB}(u){:}A,B = 1,2\}$. In contrast to the periodic solutions obtained in Ref. 5, the energy-momentum tensor here is not a constant but periodic. As a result, solutions (24) and (25) are not the gauge transform of the periodic solutions found in Ref. 5.

The solutions $\{\Phi_{1A}(u){:}A = 1,2\}$ lead to gauge fields which are real and regular everywhere. By letting the vectors $p_\mu = (0,0,1,0)$, $q_\mu = (0,1,0,0)$, and $s_\mu = (s_0,0,0,1)$ such that $|s_0| > 1$, the momentum density can be written as

$$T_{0i} = s_0 s_i \left(\frac{g^2}{s^2} \Phi_{1A}^4 - \frac{s^2}{g^2} \right), \; A = 1,2. \tag{28}$$

The energy density takes the form

$$T_{00} = \frac{g^2}{s^2} \Phi_{1A}^4 + \frac{s_0^4 - 1}{2g^2}, \; A = 1,2. \tag{29}$$

Since the direction of the Poynting vector, as given by Eq. (28), is constant and the energy density (29) is bounded throughout space-time, solutions (24) can be interpreted as nonabelian plane waves propagating along the fixed direction δ_i^3.

Evaluating the electric and magnetic fields for the solutions $\{\Phi_{1A}(u){:}A = 1,2\}$, we obtain

$$E_i^a = \{\delta_i^a \delta_i^2 + \delta_3^a \delta_i^1\} s_0 \Phi'_{1A}(u), \tag{30a}$$

$$B_i^a = \{-\delta_1^a \delta_i^1 + \delta_3^a \delta_i^2\} \Phi'_{1A}(u) + g \delta_2^a \delta_{3i} \Phi_{1A}^2(u), \tag{30b}$$

which are real and regular throughout space-time. In contrast to the field strengths of the plane waves obtained in Ref. 1, the magnetic fields (30b) are nonlinear in the YM potentials. The electric fields being linear in the YM potentials are transverse to the propagating direction. However, the nonlinear magnetic field is not perpendicular to the propagating direction since

$$B_i^a s_i = g \delta_2^a \Phi_{1A}^2(u). \tag{31}$$

This suggests that the nonlinear terms in the YM equations can be regarded as arising from a medium.[15]

The solutions (24) correspond to well-defined special waveforms consisting of regular series of similar periodic waves. As the vector s_μ for the solutions (24) is timelike, the phase velocity is greater than the speed of light. The phase velocity here is, however, not the same as the velocity of propagation. This is due to the proposition that the nonlinearities of the YM equations correspond to the presence of a medium. In a medium other than the vacuum, the phase velocity tells us only how the phase of the plane waves is delayed by interaction with the medium, but tells us nothing about the process of propagation.

When an infinitely long plane wave propagates through a homogeneous medium, the velocity of energy transport v_i can be defined by the relation[16]

$$T_{0i} = v_i T_{00}. \tag{32}$$

Thus the velocity of energy transport of solutions (24) is

$$\begin{aligned} V_i = {} & \delta_i^3 s_0 (A - A^{-1} \Phi_{1A}^4) \{ (s_0^2 + 1)(2A)^{-1} \\ & - A^{-1} \Phi_{1A}^4 \}^{-1}, \; A = 1,2. \end{aligned} \tag{33}$$

The velocity v_i is defined for all values of $|s_0| > 1$. In addition, we also note that $|v_i| < 1$ when $|s_0| > 1$. Hence the velocity of energy transport is always less than 1, the velocity of light, and solutions (24) can be interpreted as plane waves propagating through a medium with a phase velocity greater than 1. The corresponding quantized field will describe a massive particle.

The real gauge fields of solutions $\{\Phi_{2A}(u){:}A = 1, 2\}$ are periodic and regular everywhere except when

$u(x) = (2N+1)k$ and $u(x) = 2NK$, respectively. Here N is any integer and k is the complete elliptic integral of the first kind. Although the solutions (25) also propagate along the fixed directions s_i when the time components of p_μ and q_μ vanish, the energy density is not bounded throughout space-time. Hence the solutions (25) cannot be regarded as plane waves.

Another solution to the simplified YM equations when s_μ is spacelike is

$$\Phi_{23}(u) = u^{-1}, \quad s^2 = \tfrac{1}{2}g^2p^2. \tag{34}$$

With the CFtHW ansatz,[17] the scalar funciton $\Phi_{23}(u)$ leads to a self-dual gauge field. However, with ansatz (20), $\Phi_{23}(u)$ leads to a non-self-dual gauge field as the energy-momentum tensor

$$T_{\mu\nu} = g^2p^2\{p^2s^{-2}s_\mu s_\nu + \tfrac{3}{2}(p_\mu p_\nu + q_\mu q_\nu) - p^2 g_{\mu\nu}\}\Phi_{23}^4 \tag{35}$$

does not vanish. As in the periodic solution (25), the solution (34) travels in the fixed direction s_i when $p_0 = q_0 = 0$ with a speed less than 1. Since the energy density is not bounded, solution (34) does not describe a plane wave.

Referring again to ansatz (20), Φ and Ψ are functions of variables, the gradients of which are orthogonal to p_μ and q_μ. Therefore when both p_μ and q_μ are lightlike and orthogonal to each other, we can set $p_\mu = q_\mu$, and Φ and Ψ are also functions of $p_\mu x^\mu$ in addition to the variables u_1 and u_2. In this case, the YM equations (21) linearize to

$$\Box\Phi = \Box\Psi = 0. \tag{36}$$

Choosing $p_\mu = \delta_\mu^3 - \delta_\mu^0$, $s_{1\mu} = \delta_\mu^1$, and $s_{2\mu} = \delta_\mu^2$, a solution of Eqs. (36) is

$$\Phi = x_1 f(px) \text{ and } \Psi = x_2 g(px), \tag{37}$$

where f and g are arbitrary functions of px. Solution (37) gives rise to the YM potential

$$A_\mu^a = (\delta_\mu^3 - \delta_\mu^0)(x_1\delta_1^a f(px) + x_2\delta_3^a g(px)). \tag{38}$$

We recognize that this is the Coleman plane wave.[1]

So, starting from the ansatz (20) we are able to derive three different classes of solutions which are real in the Minkowski space. The first class are nonabelian plane waves propagating with a phase velocity greater than that of light. These solutions are regular throughout space-time and are hence true sourceless solutions of the YM theory. The second class of solutions travels at a phase speed less than that of light and can be static. This class of solutions is, however, singular and does not correspond to plane waves. The third class of solutions propagate at the speed of light. The Coleman plane waves[1] are included in this class.

V. REMARKS

We make a few remarks.

(a) The ansatz (20) can also be employed to construct real Euclidean space solutions when the simplified YM equation (23a) is solved by using a Euclidean metric. Obviously solutions $\{\Phi_{2i}(u), i = 1, 2, 3\}$ can be carried over to Euclidean space. Another possible set of Euclidean space solutions is the set of cylindrical solutions[18]

$$\Phi(x) = \left(\frac{-b}{ag^2}\right)^{1/2} \frac{1}{\rho} E\left(\frac{1}{\sqrt{a}}\tan^{-1}\frac{x_2}{x_1}, k\right), \tag{39}$$

where E is a Jacobi elliptic function, $0 \leqslant k \leqslant 1$ and a, b are constants of the function E. In this case

$$p_\mu = \delta_\mu^3, \; q_\mu = \delta_\mu^4, \; s_{1\mu} = \delta_\mu^4, \text{ and } s_{2\mu} = \delta_\mu^2.$$

(b) The Coleman plane wave solution[1] can be deduced either from the first family of solutions in Sec. III or from the second family of solution in Sec. IV.

(c) Nonabelian wave solutions are useful in the investigation of the color radiation problem in the classical YM theory.[19,11]

(d) Following the approach of Sec. IV,[12] nonabelian plane waves in the Higgs model have been constructed in Ref. 20.

(e) Solution (17) is generated by a nonabelian current source f_i^a on the x_1-x_2 plane. The plane source can be infinite or finite in its extent by choosing the functions f_1^a and f_2^a appropriately. Similar considerations apply to the line sources of solutions (15) and (16).

ACKNOWLEDGMENTS

We wish to thank W. K. Koo for helpful discussions.

[1] S. Coleman, Phys. Lett. B **70**, 59 (1977).

[2] W. B. Campbell and T. A. Morgan, Phys. Lett. B **84**, 87 (1979).

[3] A. Actor, Lett. Math. Phys. **2**, 275 (1978); R. Casalbuoni, G. Domokos, and S. Kövesi-Domokos, Phys. Lett. B **81**, 45 (1979).

[4] G. Z. Baseyan, S. G. Matinyan, and G. K. Sarvidi, Sov. Phys. JETP Lett. **29**, 587 (1979).

[5] C. H. Oh and R. Teh, Phys. Lett. B **87**, 83 (1979).

[6] E. Kovacs and S. Y. Lo, Phys. Rev. D **19**, 3649 (1979); S. Y. Lo, P. Desmond and E. Kovas, Phys. Lett. B **90**, 419 (1980); E. Kovas, J. Math. Phys. **23**, 834 (1982).

[7] R. Güven, Phys. Rev. D **19**, 471 (1979).

[8] R. P. Mondaini, Nuovo Cimento B **68**, 315 (1982).

[9] H. Arodz, Phys. Rev. D **27**, 1903 (1983).

[10] W. B. Bonner, Nature **225**, 932 (1970).

[11] C. H. Oh, C. H. La, and R. Teh, "Color radiation in the classical Yang-Mills theory," submitted for publication.

[12] Section IV is adopted from an unpublished CERN preprint: R. Teh and C. H. Oh, "Non-Abelian plane waves," TH 3276-CERN, 1982.

[13] After the completion of our work, we were informed that the ansatz (20) was also discussed by R. P. Treat, Nuovo Cimentro A **6**, 121(1971).

[14] Solution (24a) was also obtained by J. Ishida and A. Hosoya, Lett. Nuovo Cimento **13**, 237 (1975), but not the rest of solutions (24) and (25).

[15] R. Maciejoko, J. Math. Phys. **19**, 2491 (1978).

[16] L. Brillouin, *Wave Propagation and Group Velocity* (Academic, New York, 1960).

[17] E. Corrigan and D. B. Fairlie, Phys. Lett. B **67**, 69 (1977); G. 't Hooft (unpublished); F. Wilczek, *Quark Confinement and Field Theory*, edited by D. Stump and D. Weingarten (Wiley, New York, 1977).

[18] C. H. Oh and R. Teh, Phys. Rev. D **21**, 531 (1980).

[19] J. Tafel and A. Trautman, J. Math. Phys. **24**, 1087 (1983).

[20] Y. Brihaye, Lett. Nuovo Cimento **36**, 275 (1983).

A technique for evaluation of the strong potential Born approximation for electron capture

N. C. Sil[a] and J. H. McGuire
Department of Physics, Kansas State University, Manhattan, Kansas 66506

(Received 1 June 1984; accepted for publication 4 October 1984)

A technique is presented for evaluating differential cross sections in the strong potential Born (SPB) approximation. Our final expression is expressed as a finite sum of one-dimensional integrals, expressible as a finite sum of derivatives of hypergeometric functions.

I. INTRODUCTION

Significant progress in the theory of electron capture has been achieved[1–4] with the development of the strong potential Born (SPB) approximation. The SPB aproximation consists of expanding the scattering amplitude to first order in a weak potential and keeping a strong potential to all orders so that important second-Born-approximation terms[5,6] (with one weak and one strong interaction) are retained. Because some other approximations may be derived[7] from the SPB approximation, the SPB method has been useful in unifying understanding of the theory of electron capture for heavy projectiles.

In principle, the SPB approximation, without further peaking approximations, should be valid at high collision velocities v for systems both symmetric and asymmetric in the projectile and target charges Z_p and Z_T. Since the SPB approximation is valid only through first order in the weak potential V_p, the SPB approximation contains errors[8,9] of order $(Z_p/nv)^2$ where n is the principal quantum number of the projectile state. However, unless further approximations are used, it has been necessary to evaluate numerically a three-dimensional integral containing sharp singularities to calculate SPB amplitudes. Consequently, various peaking approximations[3,9] (which introduce further error) have been used to evaluate SPB amplitudes. In this paper we present details of an improved approximation for evaluating the SPB amplitude. Our method contains errors of order $(Z_p/nv)^2$, i.e., the same order as errors intrinsic to the SPB approximation itself. Furthermore, our method, unlike previous methods,[1,3] can be used to evaluate amplitudes for electron capture to and from excited initial and final atomic states.

II. SPB APPROXIMATION

A. Formalism

Exact solutions to our problem may be expressed as eigenfunctions of the full Hamiltonian

$$H = H_0 + V = H_0 + V_T + V_p + V_{pT}, \tag{2.1}$$

where H_0 gives the total kinetic energy, and the total interaction V is a sum of Coulomb potentials

$$\begin{aligned} V_p &= -Z_p/r_p, \\ V_T &= -Z_T/r_T, \\ V_{pT} &= Z_p Z_T/R, \end{aligned} \tag{2.2}$$

where $\mathbf{r}$ and $\mathbf{R}$ are electronic and internuclear coordinates. It is also convenient to consider

$$\overline{V}_T = V - V_T = V_p + V_{pT} \tag{2.3}$$

and

$$\overline{V}_p = V - V_p = V_T + V_{pT},$$

presenting the perturbations when the electron is, respectively, around the target (initially) and the projectile (finally).

Useful Green's functions are

$$\begin{aligned} G^{\pm}(E) &= 1/(E - H \pm i\eta) \\ &= 1/(E - H_0 - V \pm i\eta), \quad \eta \to 0^+, \\ G_T^{\pm}(E) &= 1/(E - H_0 - V_T \pm i\eta), \\ G_p^{\pm}(E) &= 1/(E - H_0 - V_p \pm i\eta). \end{aligned} \tag{2.4}$$

There are two forms of the transition operator $T^{\pm}$. The matrix elements are equal on the energy shell, and as such the physical scattering amplitudes are the same for either T^{fi+} or T^{fi-}, namely,

$$T^{fi+} = \langle \psi_f | T^+ | \psi_i \rangle = \langle \psi_f | \overline{V}_p | \Psi_i^+ \rangle \tag{2.5}$$

and

$$T^{fi-} = \langle \psi_f | T^- | \psi_i \rangle = \langle \Psi_f^- | \overline{V}_T | \psi_i \rangle, \tag{2.6}$$

where

$$\Psi_j^{\pm} = \psi_j + G^{\pm} \overline{V}_{\substack{T \\ p}} \psi_j = \psi_j + G_{\substack{T \\ p}}^{\pm} \overline{V}_{\substack{T \\ p}} \Psi_j^{\pm}, \quad j = i, f. \tag{2.7}$$

Equation (2.6) can be written

$$\begin{aligned} T^{fi-} &= \langle (1 + G^- \overline{V}_p) \psi_f | \overline{V}_T | \psi_i \rangle, \\ &= \langle \psi_f | (1 + \overline{V}_p G^+) \overline{V}_T | \psi_i \rangle, \end{aligned} \tag{2.8}$$

where ψ_i and ψ_f are the initial and final asymptotic states corresponding to

$$\begin{aligned} \psi_i &= e^{i\mathbf{K}_i \cdot \mathbf{R}_T} \phi_i(\mathbf{r}_T), \\ \psi_f &= e^{i\mathbf{K}_f \cdot \mathbf{R}_p} \phi_f(\mathbf{r}_p), \end{aligned} \tag{2.9}$$

where ϕ_i and ϕ_f are bound state wave functions (assumed known) with coordinates described in Fig. 1.

The SPB aproximation may now be simply derived from expressions for the exact T^+ and T^- transition operators, given by

$$T^+ = \overline{V}_p (1 + G^+ \overline{V}_T) = \overline{V}_p + \overline{V}_p G^+ \overline{V}_T \tag{2.10}$$

and

---

[a] Permanent address: Indian Association for the Cultivation of Science, Calcutta, India.

 0022-2488/85/040845-09$02.50

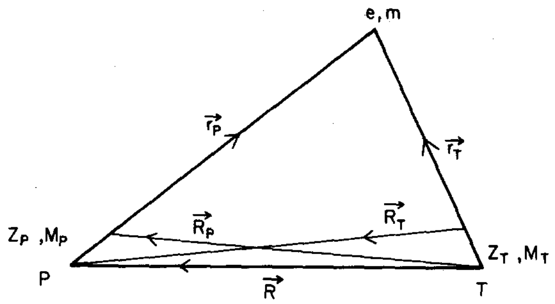


FIG. 1. Various coordinates appearing in electron capture calculations.

$$T^- = (1 + \overline{V}_p G^+)\overline{V}_T = \overline{V}_T + \overline{V}_p G^+ \overline{V}_T. \quad (2.11)$$

It is noted that knowing the exact Green's function G^+ suffices to determine exact solutions. That is, knowing the intermediate states through which the system passes gives a full solution.

The SPB approximation[3,10] results from dropping two terms in the exact transition matrix. First

$$V_{pT} \to 0, \quad (2.12)$$

so that $V \to V_p + V_T$, $\overline{V}_T \to V_p$, and $\overline{V}_p \to V_T$. For projectiles with a mass M_p large compared to the mass of electron m the contributions[11] from V_{pT} are of order $m/M_p \sim 10^{-3}$. Hence this requirement is easily met. Second, the weaker of V_p and V_T is kept only to first order (retaining the strong potential to all orders).

Specifically, if $Z_T > Z_p$, then V_p is neglected in the exact G^+, so that $G^+ \cong G_T^+$ and from (2.11),

$$\begin{aligned} T^-_{\rm SPB} &= V_p + V_T G_T^+ V_p \\ &= (1 + V_T G_T^+) V_p \quad (Z_p < Z_T). \end{aligned} \quad (2.13)$$

On the other hand, if $Z_p > Z_T$ then V_T is neglected in the exact $G^+ \cong G_p^+$ and from (2.10)

$$\begin{aligned} T^+_{\rm SPB} &= V_T + V_T G_p^+ V_p \\ &= V_T (1 + G_p^+ V_p) \quad (Z_T < Z_p). \end{aligned} \quad (2.14)$$

Thus in the SPB approximation the intermediate states propagate as Coulomb waves in the field of the Coulomb potential V_T or V_p, whichever is stronger. And, as in usual first-order perturbation theory, the SPB approximation contains errors of order $(\int V_< \, dt)^2$, i.e., order $(Z_< /nv)^2$, where $V_<$ (or $Z_<$) is the smaller of V_p and V_T (or Z_p and Z_T).

B. Basic SPB amplitude

In this section and the next we present a brief derivation[3,10] of the basic expressions for the T matrix element in the SPB approximation. We hereafter use atomic units ($e^2 = \hbar = m_e = 1$).

Consider here the case for $Z_p < Z_T$. The case for $Z_T < Z_p$ is similar, with (2.13) replaced by (2.14). We first note[12] that, as shown in Fig. 1,

$$\begin{aligned} \mathbf{R}_p &= (1 - \alpha\beta)\mathbf{r}_T + \beta \mathbf{R}_T, \\ \mathbf{r}_p &= \alpha \mathbf{r}_T - \mathbf{R}_T, \\ \mathbf{P}\cdot\mathbf{R}_p + \mathbf{p}\cdot\mathbf{r}_p &= \mathbf{P}_1\cdot\mathbf{R}_T + \mathbf{p}_2\cdot\mathbf{r}_T, \end{aligned} \quad (2.15)$$

with

$$\mathbf{P}_1 = \beta\mathbf{P} - \mathbf{p}, \ \mathbf{p}_2 = \alpha\mathbf{p} + (1 - \alpha\beta)\mathbf{P},$$
$$\alpha = M_T/(1 + M_T), \quad \beta = M_p/(M_p + 1).$$

Then defining

$$\tilde{\phi}_f(\mathbf{p}) = \frac{1}{(2\pi)^{3/2}} \int e^{i\mathbf{p}\cdot\mathbf{r}_p} \phi_f(\mathbf{r}_p) d^3 r_p \quad (2.16)$$

and

$$\int e^{i(\mathbf{P} - \mathbf{K}_f)\cdot\mathbf{R}_p} d^3 R_p = (2\pi)^3 \delta(\mathbf{P} - \mathbf{K}_f),$$

we have, using a complete set of intermediate plane waves in (2.13)

$$\begin{aligned} T^{fi-}_{\rm SPB} = &\frac{1}{(2\pi)^{3/2}} \int d^3p \, d^3P \, \delta(\mathbf{P} - \mathbf{K}_f) \tilde{\phi}_f(\mathbf{p}) \\ &\times \langle e^{i(\mathbf{P}_1\cdot\mathbf{R}_T + \mathbf{p}_2\cdot\mathbf{r}_T)} | (1 + V_T G_T^+) V_p | e^{i\mathbf{K}_i\cdot\mathbf{R}_T} \phi_i(\mathbf{r}_T)\rangle \end{aligned} \quad (2.17)$$

Now we note that

$$[1/(2\pi)^{3/2}] \langle e^{i\mathbf{p}_2\cdot\mathbf{r}_T} (1 + V_T G_T^+(\epsilon))| = \langle \psi^{cT-}_{\mathbf{p}_{2,\epsilon}}|, \quad (2.18)$$

where

$$\begin{aligned} \psi^{cT-}_{\mathbf{p}_{2,\epsilon}} &= \frac{1}{(2\pi)^{3/2}} (1 + G_T^- V_T) e^{i\mathbf{p}_2\cdot\mathbf{r}_T} \\ &= \frac{1}{(2\pi)^{3/2}} \left(1 + \frac{1}{\epsilon - H_0 - V_T - i\eta} V_T\right) e^{i\mathbf{p}_2\cdot\mathbf{r}_T}. \end{aligned}$$

Here $\psi^{cT-}_{\mathbf{p}_{2,\epsilon}}$ is an intermediate state Coulomb wave function that is off-the-energy shell since $p_2^2/2 - \epsilon \neq 0$ in general.

Then, for $Z_p < Z_T$, from (2.13) we have, after integration over $\mathbf{P}$,

$$T^-_{\rm SPB} = \int d^3p \, \tilde{\phi}(\mathbf{p}) \langle e^{i\mathbf{P}_1\cdot\mathbf{R}_T} \psi^{cT-}_{\mathbf{P}_{2,\epsilon}}(\mathbf{r}_T) | V_p | e^{i\mathbf{K}_i\cdot\mathbf{R}_T} \phi_i(\mathbf{r}_T)\rangle, \quad (2.19)$$

where $\epsilon = E - P_1^2/2\nu_i$, with $\nu_i = M_p(1 + M_T)/(1 + M_p + M_T)$. It may be noted that $\mathbf{P}_1$ and $\mathbf{p}_2$ are defined in (2.6) in terms of $\mathbf{P}$ and $\mathbf{p}$, and $\mathbf{P} = \mathbf{K}_f$ because of the δ function in (2.6). This basic SPB expression for the electron capture amplitude is equal to an integral of a first-order amplitude for direct Coulomb ionization $\langle \mathbf{P}_1, \Psi^{cT-}_{\mathbf{p}_{2,\epsilon}} | V_p | \phi_i(\mathbf{r}) \phi_i(\mathbf{R})\rangle$ weighted by the momentum distribution of the final state $\tilde{\phi}_f(\mathbf{p})$. This may be regarded as a two-step process: namely, ionization in the Coulomb field of V_T followed by overlap onto the final state of the projectile as illustrated in Fig. 2. It is noted that in Eq. (2.19), $p_2^2/2 \to \epsilon$ as $\mathbf{p} \to 0$, i.e., the integrand is peaked about the on-shell limit of the off-shell Coulomb amplitude since $\tilde{\phi}_f(\mathbf{p})$ is peaked about $\mathbf{p} = 0$. It is therefore significant that this on-shell limit is singular.

C. Macek's expression

Defining the momentum transfers

$$\begin{aligned} \mathbf{K} &= \beta\mathbf{K}_f - \mathbf{K}_i, \\ \mathbf{J} &= \alpha\mathbf{K}_i - \mathbf{K}_f, \end{aligned} \quad (2.20)$$

and integrating over $\mathbf{R}_T$, namely,

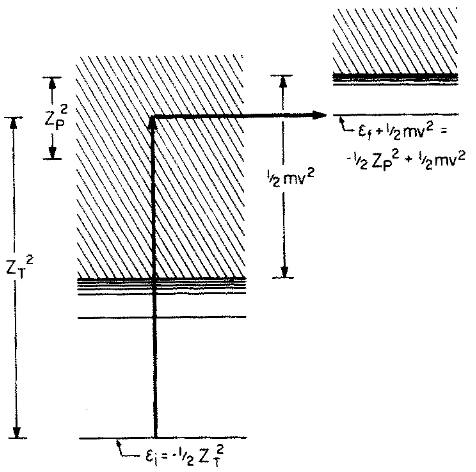


FIG. 2. Illustration of Eq. (2.18). In the strong potential Born approximation, the electron capture amplitude is expressed as the overlap of a first-order amplitude for ionization with the momentum distribution of the final bound state moving with speed v. This corresponds to a two-step process where intermediate states propagate in the strong Coulomb field of the target charge, Z_T.

$$\int e^{i(\mathbf{K}_i - \mathbf{p}_1)\cdot \mathbf{R}_T} V_p \, d^3R_T = -Z_p \int \frac{e^{i(\mathbf{p}-\mathbf{K})\cdot \mathbf{R}_T}}{|\alpha \mathbf{r}_T - \mathbf{R}_T|} d^3\mathbf{R}_T = -4\pi Z_p^2 \frac{e^{i\alpha(\mathbf{p}-\mathbf{K})\cdot \mathbf{r}_T}}{|\mathbf{K}-\mathbf{p}|^2}, \tag{2.21}$$

one obtains (with $\alpha \cong 1$),

$$T_{\rm SPB}^{if} = -4\pi Z_p \int d^3p\, \tilde{\phi}_f(\mathbf{p}) \frac{1}{|\mathbf{K}-\mathbf{p}|^2} \times \langle \Psi^{cT\,-}_{\mathbf{p}+\mathbf{v},\epsilon} | e^{i(\mathbf{p}-\mathbf{K})\cdot \mathbf{r}} | \phi_i(\mathbf{r}) \rangle, \tag{2.22}$$

where to order $(1/M)$ conservation of momentum is given[3,10] by

$$\mathbf{K} + \mathbf{J} + \mathbf{v} = 0, \tag{2.23}$$

and conservation of energy by[3,10]

$$K^2 + 2\epsilon_i = J^2 + 2\epsilon_f, \tag{2.24}$$

where ϵ_i and ϵ_f are the binding energies initially and finally of the electron. Using these expressions with $\epsilon = E - P_1^2/2\nu_i$ gives

$$\epsilon = \tfrac{1}{2}v^2 + \mathbf{v}\cdot\mathbf{p} + \epsilon_f. \tag{2.25}$$

Evaluation of the off-energy-shell ionization matrix element is discussed in detail by Macek and Alston,[3] based on the expression for $\Psi^{cT\,-}_{\mathbf{p}_2,\epsilon}$ given by Kelsey and Macek.[13] Taking $\mathbf{k}_1 = \mathbf{p} - \mathbf{K}$ and $\mathbf{k}_2 = \mathbf{p}_2 = \mathbf{p} + \mathbf{v}$, Macek and Alson obtain an expression valid near $\epsilon = \frac{1}{2}k_2^2$, i.e., near the on-shell limit, valid to order $(p/v)^2$, for hydrogenic $1s$ initial wave functions, namely,

$$\langle \Psi^{cT\,-}_{\mathbf{k}_2,\epsilon} | e^{i\mathbf{k}_1\cdot\mathbf{r}} | \phi_i \rangle \underset{\epsilon \approx \frac{1}{2}k_2^2}{\approx} e^{\pi\tilde{\nu}/2}\Gamma(1+i\tilde{\nu}) \left(\frac{1-2\epsilon/k_2^2}{4}\right)^{-i\tilde{\nu}} \times \left[\frac{-4\pi^2 N_i}{(2\pi)^{3/2}} e^{\pi\tilde{\nu}/2}\Gamma(1-i\tilde{\nu}) \times \frac{\partial}{\partial\mu}\left\{\frac{(K^2 - v^2 + \mu^2 + 2\mathbf{J}\cdot\mathbf{p} - 2ik_2\mu)^{-i\tilde{\nu}}}{(\mu^2+J^2)^{1-i\tilde{\nu}}}\right\}\right]. \tag{2.26}$$

Here $\tilde{\nu} = Z_T/\sqrt{v^2 + 2\mathbf{v}\cdot\mathbf{p} + p^2}$ becomes the usual Coulomb phase as $p/v \to 0$, $N_i = Z_T^{3/2}/\pi^{1/2}$ is the normalization of $\phi_i(r)$, and $\mu \to Z_T$ after performing the differentiation with respect to μ. The term in square brackets is the on-shell matrix element. Hence the off-shell contribution in (2.27) factors. Note that this off-shell factor is singular, i.e., asymptotically equal to $(0)^{i\nu}$, as $\epsilon \to \frac{1}{2}k_2^2$. This singularity disappears if the absolute square is taken. However, in our application to Eq. (9), there is a contribution from the off-shell singularity in the on-shell limit, because the amplitude is integrated with $\tilde{\phi}_f(\mathbf{p})$ in (2.22).

Taking $k_2 = |\mathbf{p} + \mathbf{v}|$ and ignoring order $(p/v)^2$ terms, one may obtain, setting $\tau = iZ_T/\sqrt{v^2 + 2\mathbf{v}\cdot\mathbf{p} + p^2} = i\tilde{\nu}$,

$$\langle \Psi^{cT\,-}_{\mathbf{p}+\mathbf{v},\epsilon} | e^{i(\mathbf{p}-\mathbf{K})\cdot\mathbf{r}} | \phi_i \rangle = \frac{4\pi^2}{(2\pi)^{3/2}} N_i \tau \frac{e^{i\pi\tau}}{\sin \pi\tau} \frac{\partial}{\partial\mu}\left\{\frac{e^{-2\pi i\tau}}{\mu^2+J^2}\left(\frac{[(\mu - iv)^2 + K^2 + 2\mathbf{p}\cdot(\mathbf{J} - i\mu\hat{v})](p^2 + p_0^2)}{4(v^2 + 2\mathbf{v}\cdot\mathbf{p} + p^2)(\mu^2 + J^2)}\right)^{-\tau}\right\}. \tag{2.27}$$

Here $p_0^2 = Z_T^2/n^2$.

Then using Eq. (2.27) in Eq. (2.18), one has

$$T_{\rm SPB}^{if} = -4\pi Z_p \int d^3p\, \tilde{\phi}_f^*(\mathbf{p}) |\mathbf{p} - \mathbf{K}|^{-2} \left(\frac{-4\pi^2}{(2\pi)^{3/2}} N_i \tau \frac{e^{-i\pi\tau}}{\sin\pi\tau} \frac{\partial}{\partial\mu}\left\{\frac{1}{\mu^2+J^2} \times \left(\frac{[(\mu - iv)^2 + K^2 + 2\mathbf{p}\cdot(\mathbf{J} - i\mu\hat{v})](p^2 + p_0^2)}{4(v^2 + 2\mathbf{v}\cdot\mathbf{p} + p^2)(\mu^2 + J^2)}\right)^{-\tau}\right\}\right). \tag{2.28}$$

This is Macek's expression, given[14] by Eq. (4.17) of Ref. 3. In the development from Eq. (7) errors of order $(p/v)^2$ have been introduced. However, due to the presence of $\tilde{\phi}_f$, $p \sim Z_p/n$. Hence $(p/v)^2$ is the same order as $(Z_p/nv)^2$ and the error is the same order as in Eq. (3). That is, Eq. (15) is accurate to the error of $(Z_p/nv)^2$ intrinsic to the SPB approximation itself.

III. REDUCTION OF SPB AMPLITUDE

In this section we reduce Macek's expression for the SPB amplitude, (2.28), to final form for capture from a hydrogenic $1s$ target state to a hydrogenic $1s$ final state. The method used here may also be applied to initial and final states with arbitrary quantum numbers nlm. Our technique is to expand $|\mathbf{p}-\mathbf{K}|^{-2}$, τ, and $(v^2+2\mathbf{v}\cdot\mathbf{p}+p^2)^\tau$ in a power series in p/v, keeping the first two terms so that the error is order $(p/v)^2$. Note that $p\sim Z_p/n$ due to $\tilde{\phi}_f$, so that our error is order $(Z_p/nv)^2$. We are careful to retain the term $[(\mu-iv)^2+K^2+2\mathbf{p}\cdot(\mathbf{J}-i\mu\hat{v})]$ exactly. It is this term, arising from the Green's function, that gives rise to the Thomas peak[15–17] when $\mathrm{Re}[(\mu-iv)^2+K^2]\to 0$ at high velocity. If one ignores the $2\mathbf{p}\cdot(\mathbf{J}-i\mu\hat{v})$ term then errors of order Z_p/Z_T are introduced, which we are careful to avoid.

A. Expansion in p/v

In Eq. (2.28) we redefine the term containing the Thomas peak as

$$T\equiv(\mu-iv)^2+K^2+2\mathbf{p}\cdot(\mathbf{J}-i\mu\hat{v})\equiv A+2\mathbf{L}\cdot\mathbf{p}. \tag{3.1}$$

This term is not approximated. Other terms in Eq. (2.28) are expanded through first order in $\mathbf{p}\cdot\mathbf{v}/v^2$ and $\mathbf{p}\cdot\mathbf{K}/K^2$ ($p\sim Z_p/n$ and K is order v or larger). For this expansion it is useful to define

$$x=\mathbf{p}\cdot\mathbf{v}/v^2,\quad y=\mathbf{p}\cdot\mathbf{K}/K^2,\quad \kappa^2=\mu^2+J^2,$$
$$s=(p^2+Z_p^2/n^2)/4v^2,\quad \nu=Z_T/v,$$
$$A=(\mu-iv)^2+K^2,\quad \mathbf{L}=\mathbf{J}-i\mu\hat{v},\quad L=(L^2)^{1/2},$$
$$c=(Z_T/2)^{3/2}Z_p16\pi,\quad p_0=Z_p/n. \tag{3.2}$$

Then Macek's expression (2.28) becomes

$$T_{\mathrm{SPB}}=c\int d^3p\tilde{\phi}\,\tilde{f}^*(\mathbf{p})|\mathbf{p}-\mathbf{K}|^{-2}\frac{\tau e^{-i\pi\tau}}{\sin\pi\tau}\times\frac{\partial}{\partial\mu}\left\{\kappa^{-2}\left[\frac{Ts}{\kappa^2(1+2x+p^2/v^2)}\right]^{-\tau}\right\}. \tag{3.3}$$

It is straightforward to expand $|\mathbf{p}-\mathbf{K}|^{-2}$, τ, and $\tau e^{-i\pi\tau}/\sin\pi\tau$ to first order in x and y, namely to order $(p/v)^2$

$$|\mathbf{p}-\mathbf{K}|^{-2}\cong(1+2y)/(K^2+p^2)\cong(1+2y)/K^2, \tag{3.4}$$

$$\tau\cong i\nu(1-x), \tag{3.5}$$

and

$$\tau e^{-i\pi\tau}/\sin\pi\tau\cong h_0+h_1x,$$

where

$$h_0=\frac{2\nu}{1-e^{-2\pi\nu}},\qquad h_1=\frac{-2\nu}{1-e^{-2\pi\nu}}\left\{1-\frac{2\pi\nu e^{-2\pi\nu}}{1-e^{-2\pi\nu}}\right\}. \tag{3.6}$$

The term $[Ts/\kappa^2(1+2x)]^{-\tau}$ may be expanded by writing

$$[f(x)]^{-\tau(x)}=\exp(-\tau(x)\ln f(x))\simeq\exp(-(i\nu-i\nu x)\ln f(x))=\exp(-(i\nu-i\nu x)[\ln Ts/\kappa^2-\ln(1+2x)])$$

so that with $\ln(1+2x)\cong 2x$ and $\exp(A+Bx)\cong\exp(A)(1+Bx)$,

$$\left[\frac{Ts}{\kappa^2(1+2x)}\right]^{-\tau}\cong\left[\frac{Ts}{\kappa^2}\right]^{-i\nu}\left(1+i\nu x(\ln\frac{Ts}{\kappa^2}+2)\right). \tag{3.7}$$

Taking the derivative with respect to μ in (3.3) using (3.7) we have

$$\begin{aligned}\frac{\partial}{\partial\mu}\left\{\kappa^{-2}\left[\frac{Ts}{\kappa^2(1+2x)}\right]^{-\tau}\right\}&\cong\frac{\partial}{\partial\mu}\left\{\kappa^{-2}\left[\frac{Ts}{\kappa^2}\right]^{i\nu}\left(1+i\nu x\left(\ln\frac{Ts}{\kappa^2}+2\right)\right)\right\}\\
&=s^{-i\nu}\left\{\left(-i\nu\frac{\partial T}{\partial\mu}T^{-i\nu-1}\kappa^{2(i\nu-1)}+(i\nu-1)\frac{\partial\kappa^2}{\partial\mu}T^{-i\nu}\kappa^{2(i\nu-2)}\right)\cdot\left(1+i\nu x\left(\ln\frac{Ts}{\kappa^2}+2\right)\right)\right.\\
&\quad\left.+T^{-i\nu}\kappa^{2(i\nu-1)}(i\nu x)\left(\frac{\partial T}{\partial\mu}\frac{1}{T}-\frac{\partial\kappa^2}{\partial\mu}\frac{1}{\kappa^2}\right)\right\}\\
&=s^{-i\nu}\left\{(2\mu(i\nu-1)T^{-i\nu}\kappa^{2i\nu-4}-2i\nu(\mu-iv)T^{-i\nu-1}\kappa^{2i\nu-2}\right.\\
&\quad+2i\nu(i\mathbf{p}\cdot\hat{v})T^{-i\nu-1}\kappa^{2i\nu-2})\left(1+i\nu x\left(\ln\frac{Ts}{\kappa^2}+2\right)\right)\\
&\quad\left.+i\nu x\left(-\frac{2\mu}{\kappa^2}T^{-i\nu}\kappa^{2i\nu-2}+2(\mu-iv)T^{-i\nu-1}\kappa^{2i\nu-2}-2i\mathbf{p}\cdot\hat{v}T^{-i\nu-1}\kappa^{2i\nu-2}\right)\right\},\end{aligned} \tag{3.8}$$

where (3.1) and (3.2) were used in the last step.

This completes the expansion in (p/v).

B. Reduction of general expression

In expression (3.8) there are a number of terms independent of the integration variable $\mathbf{p}$. It is useful to collect these terms into common coefficients so that

$$\begin{aligned}\frac{\partial}{\partial\mu}\left\{\kappa^{-2}\left[\frac{Ts}{\kappa^2(1+2x)}\right]^{-\tau}\right\}&\cong s^{-i\nu}\kappa^{2i\nu-2}\{T^{-i\nu}(d_{10}+xd_{1x}+x\ln Td_{1\,x\ln T})+T^{-i\nu-1}(d_{20}+xd_{2x}+x\ln Td_{2\,x\ln T})\\
&\quad+\mathbf{p}\cdot\hat{v}T^{-i\nu-1}(d_{30}+xd_{3x}+x\ln Td_{3\,x\ln T})\},\end{aligned} \tag{3.9}$$

where

$$d_{10}=2\mu(i\nu-1)/\kappa^2,\quad d_{20}=-2i\nu(\mu-i\nu),\quad d_{30}=-2\nu,\quad d_{1x}=d_{10}(i\nu\ln s/\kappa^2+2i\nu)-2i\nu\mu/\kappa^2,$$
$$d_{2x}=d_{20}(i\nu\ln s/\kappa^2+2i\nu)+2i\nu(\mu-i\nu),\quad d_{3x}=d_{30}(i\nu\ln s/\kappa^2+2i\nu)-2i\nu, \tag{3.10}$$
$$d_{1\,x\ln T}=i\nu d_{10},\quad d_{2\,x\ln T}=i\nu d_{20},\quad d_{3\,x\ln T}=i\nu d_{30}.$$

Using (3.1) together with (3.4)–(3.6) in (3.3) we have

$$T_{\rm SPB}=c\int d^3p\,\tilde\phi_f^*(\mathbf p)\Big(\frac{1+2y}{K^2+p^2}\Big)(h_0+h_1x)s^{-i\nu}\kappa^{2i\nu-2}\{T^{-i\nu}(d_{10}+xd_{1x}+x\ln T d_{1\,x\ln T})$$
$$+T^{-i\nu-1}(d_{20}+xd_{2x}+x\ln Td_{2\,x\ln T})+\mathbf p\cdot\hat v T^{-i\nu-1}(d_{30}+xd_{3x}+x\ln Td_{3\,x\ln T}). \tag{3.11}$$

Keeping terms through order (p/v) one obtains

$$T_{\rm SPB}=\int d^3pf(\mathbf p)\{(a_{10}T^{-i\nu}+a_{20}T^{-i\nu-1}+a_{30}\mathbf p\cdot\hat vT^{-i\nu-1})+y(a_{1y}T^{-i\nu}+a_{2y}T^{-i\nu-1}+a_{3y}\mathbf p\cdot\hat vT^{-i\nu-1})$$
$$+x(a_{1x}T^{-i\nu}+a_{2x}T^{-i\nu-1}+a_{3x}\mathbf p\cdot\hat vT^{-i\nu-1})$$
$$+x\ln T\,(a_{1\,x\ln T}T^{-i\nu-1}+a_{2\,x\ln T}T^{-i\nu-1}+a_{3\,x\ln T}\mathbf p\cdot\hat vT^{-i\nu-1})\}, \tag{3.12}$$

where

$$a_{10}=h_0d_{10},\quad a_{20}=h_0d_{20},\quad a_{30}=h_0d_{30},\ a_{1y}=2h_0d_{10},\quad a_{2y}=2h_0d_{20},\ a_{3y}=2h_0d_{30},\ a_{1x}=h_1d_{10}+h_0d_{ix},$$
$$a_{2x}=h_1d_{20}+h_0d_{2x},\quad a_{3x}=h_1d_{30}+h_0d_{3x},\quad a_{1\,x\ln T}=h_0d_{1\,x\ln T},\ a_{2x}=h_0d_{2\,x\ln T},\quad a_{3x}=h_0d_{3\,x\ln T}, \tag{3.13}$$

and

$$f(\mathbf p)=c\tilde\phi_f^*(\mathbf p)s^{-i\nu}\kappa^{2i\nu-2}K^{-2}. \tag{3.14}$$

Consequently (3.12) is of the general form

$$T_{\rm SPB}=\sum_{k=1}^{3}\sum_{j=0,\,y,\,x,\,x\ln T}A_{kj}, \tag{3.15}$$

where, for general final states in (3.14)

$$A_{kj}=\int d^3pf(\mathbf p)a_{kj}T^{t-n_1}(\mathbf p\cdot\hat v)^{n_2}y^{n_3}x^{n_4}(\ln T)^{n_5}, \tag{3.16}$$

in which $n_i=0$ or 1 depending on k and $j(i=1,2,3,4,5)$ as given in (3.12).

It may be shown, using techniques in Sec. III C, that (3.16) may be evaluated in closed form as a sum of hypergeometric functions when ϕ_f represents a hydrogenic wave function with arbitrary quantum numbers nlm.

C. Evaluation for 1*s*–1*s* capture

For 1*s*–*ns* transitions, $f(\mathbf p)=f(p)$ in (3.14) and from (3.15) and (3.16) $T_{\rm SPB}$ is a sum of terms,

$$T_{\rm SPB}=\sum_{k=1}^{3}\sum_{j=0,\,y,\,x,\,x\ln T}A_{kj}, \tag{3.17}$$

where

$$A_{kj}=\int_0^\infty dp\,p^2f(p)a_{kj}\int_{-1}^{1}d(\cos\theta)$$
$$\times\int_0^{2\pi}d\phi\,T^{t-n_1}(\mathbf p\cdot\hat v)^{n_2}y^{n_3}x^{n_4}(\ln T)^{n_5}$$
$$=\int_0^\infty dp\,p^2f(p)a_{kj}2\pi\mathscr I_{kj}, \tag{3.18}$$

where the terms $2\pi\mathscr I_{kj}$ come from the angular integration and the various terms a_{kj} in the integrand are defined by (3.1) – (3.6) and (3.10)-(3.14). In this section we explicitly evaluate the $\mathscr I_{kj}$ terms for 1*s*–1*s* transitions, where

$$\tilde\phi_f(\mathbf p)=2\sqrt{2}/\pi p_0^{5/2}(p^2+p_0^2)^{-2}, \tag{3.19}$$

and we discuss evaluation of A_{kj}. In the next section we show how to reduce the A_{kj} to linear sums of hypergeometric functions for general n,l,m final states.

There are 12 $\mathscr I_{kj}$ terms in (3.17), as defined in (3.12). The first term is $\mathscr I_{10}$. The other terms will be generated by taking derivatives of this first term. Hence let us first consider

$$\mathscr I_{10}=\int_{-1}^{1}d(\cos\theta)\frac{1}{2\pi}\int_0^{2\pi}d\phi\,T^t=\int_{-1}^{1}dz\,T^t$$
$$=\int_{-1}^{1}dz(A+2Lpz)^t$$
$$=\frac{1}{2Lp(t+1)}(A+2Lpz)^{t+1}\Big|_{z=-1}^{z=+1}$$
$$=\frac{1}{2Lp(t+1)}[(A+2Lp)^{t+1}-(A-2Lp)^{t+1}]. \tag{3.20}$$

It is convenient to define

$$F(t,\pm)\equiv[(A+2Lp)^t\pm(A-2Lp)^t]. \tag{3.21}$$

Hence our basic case is

$$\mathscr I_{10}=\int_{-1}^{1}dz\,T^t=\frac{1}{2Lp(t+1)}F(t+1,-), \tag{3.22}$$

where $t=-i\nu$. Similarly,

$$\mathscr I_{20}=\int_{-1}^{1}dz\,T^{t-1}=\frac{1}{2Lpt}F(t,-). \tag{3.23}$$

Next note that a number of terms are of the form

$$\mathscr I_{jk}=\int_{-1}^{1}dz\frac{1}{2\pi}\int_0^{2\pi}d\phi\,(\mathbf p\cdot\mathbf V)T^{t-1}$$
$$=\int_{-1}^{1}dz\frac{1}{2\pi}\int_0^{2\pi}d\phi\,p_lV_l(A+2Lpz)^{t-1}$$
$$=V_l\int_{-1}^{1}dz\frac{1}{2\pi}\int_0^{2\pi}d\phi\,p_l(A+2L_kp_k)^{t-1}$$

$$= \frac{V_l}{2t}\int_{-1}^{1} dz \frac{1}{2\pi}\int_0^{2\pi} d\phi \frac{\partial}{\partial L_l}(A + 2L_k p_k)^t$$

$$= \frac{V_l}{2t}\frac{\partial}{\partial L_l}\int_{-1}^{1} dz \frac{1}{2\pi}\int_0^{2\pi} d\phi (A + 2L_k p_k)^t$$

$$= \frac{V_l}{2t}\frac{L_l}{L}\frac{\partial}{\partial L}\int_{-1}^{1} dz \frac{1}{2\pi}\int_0^{2\pi} d\phi (A + 2Lpz)^t$$

$$= \frac{\mathbf{V}\cdot\mathbf{L}}{2t}\frac{1}{L}\frac{d}{dL}\mathscr{I}_{10}, \tag{3.24}$$

where $2L\,dL = 2L_l\,dL_l$ was used, and $\mathbf{V}$ is an arbitrary vector. Thus terms with $\mathbf{p}\cdot\mathbf{V}$ may be evaluated by differentiating terms without $\mathbf{p}\cdot\mathbf{V}$.

Consequently,

$$\mathscr{I}_{30} = \int_{-1}^{1} dz \frac{1}{2\pi}\int_0^{2\pi} d\phi\, \mathbf{p}\cdot\hat{v}T^{t-1} = \frac{\hat{v}\cdot\mathbf{L}}{2tL}\frac{\partial}{\partial L}\Big\{\frac{1}{2Lp(t+1)}$$

$$\times[(A+2Lp)^{t+1} - (A-2Lp)^{t+1}]\Big\}$$

$$= \frac{\hat{v}\cdot\mathbf{L}}{4t(t+1)L}\Big\{-\frac{1}{L^2p}[(A+2Lp)^{t+1}$$

$$-(A-2Lp)^{t+1}] + \frac{1}{Lp}(t+1)(2p)$$

$$\times\Big[(A+2Lp)^t + (A-2Lp)^t\Big]\Big\} = \frac{\hat{v}\cdot\mathbf{L}}{4L^3t(t+1)}$$

$$\times\{-(1/p)F(t+1,\,-) + 2L(t+1)F(t,\,+)\}, \tag{3.25}$$

where we note that

$$\frac{\partial}{\partial L}F(t,\,\mp) = 2ptF(t-1,\,\pm). \tag{3.26}$$

Increasing t by 1 and changing $\hat{v}$ to $\mathbf{v}/v^2$ and $\mathbf{K}/K^2$ gives $\mathscr{I}_{1x}$ and $\mathscr{I}_{1y}$, namely

$$\mathscr{I}_{1x} = \frac{\mathbf{L}\cdot\mathbf{v}}{4L^3v^2(t+1)(t+2)}\{-(1/p)F(t+2,\,-)$$

$$+ 2L(t+2)F(t+1,\,+)\} \tag{3.27}$$

and

$$\mathscr{I}_{1y} = \frac{\mathbf{L}\cdot\mathbf{K}}{4L^3K^2(t+1)(t+2)}\{-(1/p)F(t+2,\,-)$$

$$+ 2L(t+2)F(t+1,\,+)\}. \tag{3.28}$$

Decreasing t by 1 now gives $\mathscr{I}_{2x}$ and $\mathscr{I}_{2y}$, namely

$$\mathscr{I}_{2x} = \frac{\mathbf{L}\cdot\mathbf{v}}{4L^3v^2t(t+1)}\{-(1/p)F(t+1,\,-)$$

$$+ 2L(t+1)F(t,\,+)\} \tag{3.29}$$

and

$$\mathscr{I}_{2y} = \frac{\mathbf{L}\cdot\mathbf{K}}{4L^3K^2t(t+1)}\{-(1/p)F(t+1,\,-)$$

$$+ 2L(t+1)F(t,\,+)\}. \tag{3.30}$$

To evaluate a term involving $(\mathbf{p}\cdot\mathbf{V})(\mathbf{p}\cdot\mathbf{v})$ the above procedure is applied repeatedly. Thus

$$\mathscr{I}_{3y} = \int_{-1}^{1} dz \frac{1}{2\pi}\int_0^{2\pi} d\phi \left(\frac{\mathbf{p}\cdot\mathbf{K}}{K^2}\right)(\mathbf{p}\cdot\mathbf{v})T^{t-1}$$

$$= \frac{K_l\hat{v}_k}{4t(t+1)K^2}\frac{\partial^2}{\partial L_l\partial L_k}\int_{-1}^{1} dz(A + 2L_k p_k)^{t+1}$$

$$= \frac{K_l\hat{v}_k}{4t(t+1)K^2}\Big(\frac{\delta_{kl}}{L}\frac{\partial}{\partial L} - \frac{L_kL_l}{L^3}\frac{\partial}{\partial L}$$

$$+ \frac{L_kL_l}{L^2}\frac{\partial^2}{\partial L^2}\Big)$$

$$\times\Big\{\frac{1}{2Lp(t+2)}F(t+2,\,-)\Big\}$$

$$= (8K^2t(t+1)(t+2))^{-1}\Big\{\Big(\frac{\mathbf{K}\cdot\hat{v}}{L^3} - \frac{(\mathbf{K}\cdot\mathbf{L})(\hat{v}\cdot\mathbf{L})}{L^5}\Big)$$

$$\times\Big(-\frac{1}{p}F(t+2,\,-) + 2L(t+2)F(t+1,\,+)\Big)$$

$$+ \frac{(\mathbf{L}\cdot\mathbf{K})(\mathbf{L}\cdot\hat{v})}{L^5}\Big(\frac{2}{p}F(t+2,\,-)$$

$$-4L(t+2)F(t+1,\,+)$$

$$+ 4L^2(t+1)(t+2)p\,F(t,\,-)\Big)\Big\}. \tag{3.31}$$

Replacing $\mathbf{K}$ by $\mathbf{v}$, one has

$$\mathscr{I}_{3x} = (8v^2t(t+1)(t+2))^{-1}\Big\{\Big(\frac{v}{L^3} - \frac{(\mathbf{v}\cdot\mathbf{L})(\hat{v}\cdot\mathbf{L})}{L^5}\Big)$$

$$\times\Big(-\frac{1}{p}F(t+2,\,-) + 2L(t+2)F(t+1,\,+)\Big)$$

$$+ \frac{(\mathbf{L}\cdot\mathbf{v})(\mathbf{L}\cdot\hat{v})}{L^5}\Big(\frac{2}{p}F(t+2,\,-)$$

$$-4L(t+2)F(t+1,\,+)$$

$$+ 4L^2(t+1)(t+2)p\,F(t,\,-)\Big)\Big\}. \tag{3.32}$$

There now remain three terms, namely $\mathscr{I}_{kx\ln T}$, which include an $\ln T$ factor in the integrand in (3.17) and (3.18). Noting that

$$\frac{\partial}{\partial t}g^t = \frac{\partial}{\partial t}e^{t\ln g} = \ln g e^{t\ln g} = g^t\ln g, \tag{3.33}$$

we have

$$\mathscr{I}_{1x\ln T} = \int_{-1}^{1} dz \frac{1}{2\pi}\int_0^{2\pi} d\phi \left(\frac{\mathbf{p}\cdot\mathbf{v}}{v^2}\right)T^t\ln T$$

$$= \frac{\partial}{\partial t}\int_{-1}^{1} dz \frac{1}{2\pi}\int_0^{2\pi} d\phi \left(\frac{\mathbf{p}\cdot\mathbf{v}}{v^2}\right)T^t$$

$$= \frac{\partial}{\partial t}\mathscr{I}_{1x}. \tag{3.34}$$

Defining

$$F\ln(t,\,\pm) = \frac{\partial}{\partial t}F(t,\,\pm)$$

$$= [(A+2Lp)^t\ln(A+2Lp)$$

$$\pm (A-2Lp)^t\ln(A-2Lp)], \tag{3.35}$$

we have

$$\mathscr{I}_{1x\ln T} = \frac{\mathbf{L}\cdot\mathbf{v}}{4L^3v^2(t+1)(t+2)}\Big\{\frac{2t+3}{p(t+1)(t+2)}F(t+2,\,-) - \frac{1}{p}F\ln(t+2,\,-) - \frac{2L(t+2)}{(t+1)}F(t+1,\,+) + 2L(t+2)F\ln(t+1,\,+)\Big\}. \tag{3.36}$$

Decreasing t by 1 we have

$$\mathscr{I}_{2x\ln T} = \frac{\partial}{\partial t}\mathscr{I}_{2x} = \frac{\mathbf{L}\cdot\mathbf{v}}{4L^3v^2t(t+1)}\Big\{\frac{2t+1}{pt(t+1)}F(t+1,\,-) - \frac{1}{p}F\ln(t+1,\,-) - \frac{2L(t+1)}{t}F(t,\,+) + 2L(t+1)F\ln(t,\,+)\Big\}. \tag{3.37}$$

Similarly, the last term is given by

$$\begin{aligned}\mathscr{I}_{3x\ln T} = \frac{\partial}{\partial t}\mathscr{I}_{3x} = &-\Big(\frac{(t+1)(t+2)+t(t+2)+t(t+1)}{8v^2t^2(t+1)^2(t+2)^2}\Big)\Big\{\Big(\frac{v}{L^3} - \frac{(\mathbf{v}\cdot\mathbf{L})(\hat{v}\cdot\mathbf{L})}{L^5}\Big)\\ &\times\Big(-\frac{1}{p}F(t+2,\,-) + 2L(t+2)F(t+1,\,+)\Big) + \frac{(\mathbf{L}\cdot\mathbf{v})(\mathbf{L}\cdot\hat{v})}{L^5}\Big(\frac{2}{p}F(t+2,\,-) - 4L(t+2)F(t+1,\,+)\\ &+ 4L^2p(t+2)(t+1)F(t,\,-)\Big)\Big\} + (8v^2t(t+1)(t+2))^{-1}\cdot\Big\{\Big(\frac{v}{L^3} - \frac{(\mathbf{v}\cdot\mathbf{L})(\hat{v}\cdot\mathbf{L})}{L^5}\Big)\\ &\times\Big(-\frac{1}{p}F\ln(t+2,\,-) + 2LF(t+1,\,+) + 2L(t+2)F\ln(t+1,\,+)\Big) + \frac{(\mathbf{L}\cdot\mathbf{v})(\mathbf{L}\cdot\hat{v})}{L^5}\\ &\times\Big(\frac{2}{p}F\ln(t+2,\,-) - 4LF(t+1,\,+) - 4L(t+2)F\ln(t+1,\,+)\\ &+ 4L^2p(t+1)F(t,\,-) + 4L^2p(t+2)F(t,\,-)\\ &+ 4L^2p(t+1)(t+2)F\ln(t,\,-)\Big)\Big\}.\end{aligned} \tag{3.38}$$

This now gives explicit expressions for the $\mathscr{I}_{kj}$ in (3.18) for evaluation of the differential cross section for $1s$–$1s$ electron capture in the SPB aproximation. Capture for $1s\to ns$ may be evaluated by simply taking $p_0\to Z_p/n$ and modifying $\tilde{\phi}_f(p)$ in Eq. (4.3). In order to evaluate $T_{\rm SPB}$ the integration over p in (3.18) may be done numerically. At large p the integrand is cut off by $\tilde{\phi}_f(p)\sim p^{-4}$ so that the integrand varies as p^{-2} at large p, and the cutoff error is proportional to $p_{\max}^{-1}$. In actual evaluation[16] of this integral, integration intervals were chosen as 2^nZ_p with n running from -1 to 11. Convergence at large p may be improved and evaluation of higher-order contributions in $(p/v)^2$ may be obtained by retaining the p^2 terms in the denomination of the $\{\ \}^{-\tau}$ term in (2.28) and in (3.4).

In the evaluation of the p integration of (3.18) there naturally occur two regions, namely $|2Lp|<|A|$ and $|2Lp|\gtrsim|A|$. The latter region is important for evaluation of the Thomas peak at high velocities for systems symmetric (or nearly symmetric) in Z_p and Z_T. In this latter region, $|2Lp|\gtrsim|A|$ the above expressions for the $\mathscr{I}_{kj}$ are useful. The first region, $|2Lp|<|A|$, is important for forward angle scattering, in general, and for systems when $Z_p \ll Z_T$. Here we have encountered numerical roundoff errors in evaluation of the $\mathscr{I}_{kj}$, and in this region we have used a Taylor expansion of the $\mathscr{I}_{kj}$ in powers of $|2Lp|/|A|$ given by

$$\begin{aligned}\mathscr{I}_{10} &= 2(A^t + [t(t-1)/6]4L^2p^2A^{t-2}),\\ \mathscr{I}_{20} &= 2(A^{t-1} + [(t-1)(t-2)/6]4L^2p^2A^{t-3}),\\ \mathscr{I}_{30} &= \tfrac{4}{3}\mathbf{L}\cdot\hat{v}(t-1)p^2A^{t-2},\\ \mathscr{I}_{1x} &= \tfrac{4}{3}(\mathbf{L}\cdot\mathbf{v}/2)\,tp^2A^{t-1},\\ \mathscr{I}_{1y} &= \tfrac{4}{3}(\mathbf{L}\cdot\mathbf{K}/K^2)tp^2A^{t-1},\\ \mathscr{I}_{2x} &= \tfrac{4}{3}(\mathbf{L}\cdot\mathbf{v}/v^2)(t-1)p^2A^{t-2},\\ \mathscr{I}_{2y} &= \tfrac{4}{3}(\mathbf{L}\cdot\mathbf{K}/K^2)(t-1)p^2A^{t-2},\\ \mathscr{I}_{3x} &= \tfrac{2}{3}v^{-1}p^2A^{t-1},\\ \mathscr{I}_{3y} &= \tfrac{2}{3}(\mathbf{K}\cdot\hat{v}/K^2)p^2A^{t-1},\\ \mathscr{I}_{1x\ln T} &= \tfrac{4}{3}(\mathbf{L}\cdot\mathbf{v}/v^2)p^2A^{t-1}(1+t\ln A),\\ \mathscr{I}_{2x\ln T} &= \tfrac{4}{3}(\mathbf{L}\cdot\mathbf{v}/v^2)p^2A^{t-2}(1+(t-1)\ln A),\\ \mathscr{I}_{3x\ln T} &= \tfrac{2}{3}v^{-1}p^2A^{t-1}\ln A.\end{aligned} \tag{3.39}$$

This expansion for $|2Lp|<|A|$ may be used to cross check the numerical values of the full $\mathscr{I}_{kj}$ given above at small $|2Lp|/|A|$.

The full peaking (FP) approximation of Macek and Alston[3] is recovered by retaining the first term in the power series expansions of $\mathscr{I}_{10}$ and $\mathscr{I}_{20}$ given above in (3.39). The transverse peaking (TP) approximation of Alston[9] may be also related to our expressions by setting to zero transverse components of $\mathbf{K}$ and $\mathbf{p}$ and terms $\sim p/Z_T$ in (3.20)-(3.38) and (3.13), except that the phase ν is not approximated in the TP approximation. Both the FP and TP approximations lead to analytic expressions that are simpler and consequently easier to use than our expressions. However, both the FP and TP approximations introduce errors of order Z_p/Z_T, which we avoid. Hence, unlike the FP and TP approximations, our method is applicable for electron capture in systems symmetric in Z_T and Z_p, e.g., $p+H$.

D. Reduction of general case to closed form

The general form for capture from 1s to nlm hydrogenic states is given in the SPB approximation (3.15) and (3.16), namely,

$$T_{\rm SPB}=\sum_{k=1}^{3}\sum_{j=1}^{4}A_{kj},$$

where

$$A_{kj}=c\int d^3p\,\widetilde{\phi}^*_f(\mathbf{p})s(p)^{-i\nu}\kappa^{2i\nu-2}K^{-2}a_{kj} \times T^{t-n_1}(\mathbf{p}\cdot\hat{v})^{n_2}y^{n_3}x^{n_4}(\ln T)^{n_5}. \quad (3.40)$$

The only dependence on $\hat{p}$ in (3.40) not previously considered occurs in $\widetilde{\phi}^*_f(\mathbf{p})$. Choosing unit vectors $\hat{e}_x$, $\hat{e}_y$, and $\hat{e}_z$, one may find the Cartesian tensor components of $\widetilde{\phi}^*_f(\mathbf{p})$ by taking inner products of the $\hat{e}_m$ ($m=x,y,z$) with $\widetilde{\phi}^*_f(\mathbf{p})$. For a final state quantum number $\mathscr{F}$, this yields polynomials in $(\hat{e}_m\cdot\mathbf{p})$ of order l. (The m components of T^{if} may be expressed as linear combinations of the Cartesian components of $T_{\rm SPB}$ thus formed.) Hence the general form of the A_{kj} in (3.40) is

$$A_{kj}=\int d^3p\widetilde{f}(p)(\hat{e}_x\cdot\mathbf{p})^{l_x}(\hat{e}_y\cdot\mathbf{p})^{l_y}(\hat{e}_z\cdot\mathbf{p})^{l_z} \times T^{t-n_1}(\mathbf{p}\cdot\hat{v})^{n_2}x^{n_3}y^{n_4}(\ln T)^{n_5}, \quad (3.41)$$

where

$$l_x+l_y+l_z\leqslant l.$$

Using the methods illustrated in the previous section this expression may be expressed in terms of derivatives of the expression

$$A_b=\int_0^\infty dp\,p^2\widetilde{f}(p)a_{10}\int_{-1}^{1}dz\int_0^{2\pi}d\phi\,(A+2Lpz)^t = \int_0^\infty dp\,p^2\widetilde{f}(p)a_{10}\frac{2\pi}{2Lp(t+1)} \times[(A+2Lp)^{t+1}-(A-2Lp)^{t+1}]. \quad (3.42)$$

One may, of course, apply the methods of the previous section and evaluate integrals over p numerically.

However, one may also reduce the above expression to a linear combination of hypergeometric functions. This may be done by first noting that $\widetilde{f}(p)\sim\Sigma_{n'}c_{n'}p^{n'}/(p^2+p_0^2)^{n'}$, where n' is a positive integer. Each of these terms may be generated by differentiation of the basic expression (taking $t\rightarrow i\nu$)

$$A_{bc}=\int_0^\infty p\,dp(p^2+p_0^2)^{-m-i\nu_1} \times[(A+2Lp)^{-i\nu_2-n}-(A-2Lp)^{-i\nu_2-n}] = \frac{1}{2(-i\nu_2-n+1)}\frac{\partial}{\partial L}\int_{-\infty}^{\infty}dp(p^2+p_0^2)^{-m-i\nu_1} \times(A+2Lp)^{-i\nu_2-n+1}. \quad (3.43)$$

Setting $x=p/p_0$ and $\alpha=A/2Lp_0$, the above integral is proportional to

$$F=\int_{-\infty}^{\infty}dx(x^2+1)^{-m-i\nu_1}(x+\alpha)^{-n-i\nu_2+1}. \quad (3.44)$$

Now setting $(x+\alpha)=(\alpha+i)y$, one may show

$$F=[-(\alpha+i)]^{t_1}(\alpha+i)^{t_2+1}[-(\alpha-i)]^{t_1} \times\int_c dy\,y^{t_2}(1-y)^{t_1}(1-yz)^{t_1}, \quad (3.45)$$

where $z=\alpha+i/\alpha-i$ and c is a suitable contour in the complex y plane. One may show that the contour integral satisfies a hypergeometric equation so that F can be expressed as a linear combination of an independent pair of two hypergeometric functions satisfying the above-mentioned hypergeometric equation. Thus one may write

$$F=[-(\alpha+i)]^t(\alpha+i)^{t_2+1}[-(\alpha-1)^{t_1}] \times\Big[C_2F_1\Big(-t_1,t_2+1;-2t_1;-\frac{2i}{\alpha-i}\Big) +D\Big(\frac{-2i}{\alpha-i}\Big)^{2t_1+1}{}_2F_1\Big(2t_1+t_2+2,t_1+1; 2(t_1+1);\frac{-2i}{\alpha-i}\Big)\Big] \quad (3.46)$$

The coefficients C and D may be determined by taking $\alpha\rightarrow 0$ and ∞, and evaluating (3.44) directly.

Thus the SPB T matrix may be expressed as a sum of integrals as given by (3.40) and (3.41). These integrals may be generated by parametric differentiation of the basic integral in (3.43). Thus the SPB T matrix may be expressed as a sum of derivations of hypergeometric functions.

IV. SUMMARY

The electron capture problem is characterized by two different asymptotic potentials, V_p, the interaction of the electron with the projectile, and V_T, the interaction to the electron with the target. In the SPB approximation the T matrix is expanded to first order in the weak potential, identified here as V_p, retaining all order in the strong potential V_T. Errors in this approximation are (i) order $(\int V_p\,dt)^2$, i.e., order $(Z_p/nv)^2$, where Z_p is the projectile charge, and v is the collision velocity, and n is the principal quantum number of the final state; and (ii) order $(1/M)$ where $M\sim 10^3$ for heavy projectiles.

Taking V_T as a Coulomb potential (rigorously speaking, a screened Coulomb potential in the limit of zero screening), one may serve an integral expression (2.18) for $T_{\rm SPB}$ in terms of the off-energy shell Coulomb wavefunction, which has a phase divergence in the on-energy shell limit. Ignoring terms of order $(Z_p/nv)^2$, Macek's integral expression (2.28) may be obtained. At this point we introduce a further approximation by expanding various terms to order $(Z_p/nv)^2$. Our technique, described in Sec. III, is to expand all terms in powers of (p/v) (where $p<Z_p/n$ is the integration variable), except the term arising from the Green's function operator, which gives rise to the Thomas peak. The result for 1s–1s transitions may be expressed as

$$T_{\mathrm{SPB}} = \sum_{k=1}^{3} \sum_{j=1}^{4} A_{kj}$$
$$= \Sigma_{kj}\, 2\pi \int_0^\infty dp\, p^2 f(p) a_{kj}(p) \mathscr{I}_{kj}(p), \tag{4.1}$$

where

$$f(p) = c\widetilde{\phi}_f^{\,*} s^{-i\nu} \kappa^{2i\nu-2} K^{-2}, \tag{4.2}$$

with

$$\begin{aligned} &c = (Z_T/2)^{3/2} 16\pi Z_p, \quad \widetilde{\phi}_f^{\,*} = (2\sqrt{2}/\pi) p_0^{5/2} (p^2 + p_0^2)^{-2}, \\ &p_0 = Z_p/n, \quad s = (p^2 + p_0^2)/4v^2, \\ &\nu = Z_T/v = it, \quad K^2 = K_\perp^2 + K_z^2, \\ &K_\perp = (M_p M_T/(M_p + M_T)) v\theta_{cm}, \\ &K_z^2 = (1/4v^2)(v^2 + 2(\epsilon_i - \epsilon_f))^2 + Z_p^2/n^2, \\ &\mathbf{J} + \mathbf{K} + m\mathbf{v} = 0, \quad \kappa^2 = \mu^2 + J^2, \quad \mu = Z_T, \\ &\mathbf{L} = \mathbf{J} - i\mu\hat{v}, \quad L^2 = \mathbf{L}\cdot\mathbf{L}, \quad L = (L^2)^{1/2}, \\ &A = (\mu - iv)^2 + K^2. \end{aligned} \tag{4.3}$$

The coefficients a_{kj} are given by (3.13), (3.10), and (3.6) and the 12 $\mathscr{I}_{jk}$ terms are given by (3.22)–(3.38), with $F(t, \pm)$ defined by (3.21) and $F\ln(t, \pm)$ by (3.35).

Using this integral expression, (4.1), one may generate higher-order corrections in (p/v) by replacing in (4.2) [in accord with (2.28)] K^{-2} by $(K^2 + p^2)^{-1}$ and modifying by replacing v^{-2} by $(v^2 + p^2)^{-1}$ so that the integrand in (4.1) always converges at large p. This also tends to improve the calculations for $v \sim Z_T$.

The above expressions (4.1)–(4.3) may also be generalized in a straightforward way to evaluate cross sections for capture to final states with arbitrary quantum numbers n, l, m. This is discussed in Sec. III D, where it is also shown that these SPB amplitudes may be reduced to a linear combination of derivatives of hypergeometric functions.

ACKNOWLEDGMENT

This work was supported by a grant from the Division of Chemical Sciences, United States Department of Energy.

[1] J. S. Briggs, J. Phys. B **10**, 3075 (1977).

[2] J. H. Macek and K. Taulbjerg, Phys. Rev. Lett. **46**, 170 (1980).

[3] J. H. Macek and S. Alston, Phys. Rev. A **26**, 250 (1982).

[4] J. S. Briggs, J. H. Macek, and K. Taulbjerg, Comm. At. Mol. Phys. **12**, 1 (1982).

[5] R. Shakeshaft and L. Spruch, Rev. Mod. Phys. **51**, 369 (1979).

[6] J. H. McGuire, J. Eichler, and P. R. Simony, Phys. Rev. A **28**, 2104 (1983).

[7] K. Taulbjerg and J. S. Briggs, Abstracts of Contributed Papers, Thirteenth International Conference on the Physics of Electronic and Atomic Collisions, 1983 (unpublished), p. 496; cf. also Ref. 4.

[8] K. Taulbjerg and J. Macek (private communication). However, it is not now clear whether the SPB approximation contains additional errors which are small at high velocity, e.g., order $(Z_p/v)\,(Z_p/Z_T)$.

[9] S. Alston, Ph.D. thesis, University of Nebraska, 1982; Phys. Rev. A **27**, 2342 (1983).

[10] J. H. Macek and R. Shakeshaft, Phys. Rev. A **22**, 1441 (1980).

[11] J. H. McGuire, Phys. Rev. A **28**, 2518 (1983).

[12] R. A. Mapleton, *Theory of Charge Exchange* (Wiley, New York, 1972).

[13] E. J. Kelsey and J. Macek, J. Math. Phys. **17**, 1182 (1976).

[14] Unlike Eq. (4.17) of Ref. 3, our Eq. (2.28) contains a factor of $+p^2$ in the $(v^2 + 2\mathbf{p}\cdot\mathbf{v} + p^2)$ expression. Although this may be dropped to order $(p/v)^2$, retaining this term enables one to generate higher-order corrections in p/v.

[15] E. Horsdal-Pedersen, C. L. Cocke, and M. Stockli, Phys. Rev. Lett. **50**, 1910 (1983).

[16] J. H. McGuire and N. C. Sil, Phys. Rev. A **28**, 3679 (1983).

[17] E. Horsdal-Pedersen, C. L. Cocke, and M. Stockli, Abstracts of Contributed Papers, Thirteenth International Conference on the Physics of Electronic and Atomic Collisions, Berlin, 1983 (Ref. 7), p. 478; (private communication).

Scattering by a parabolic cylinder—A uniform asymptotic expansion

R. H. Ott
RF Systems Research Office, Electronics Research Laboratory, The Aerospace Corporation, P. O. Box 92957, Los Angeles, California 90009

(Received 30 April 1984; accepted for publication 1 November 1984)

A uniform asymptotic expansion in the variable, determining the location of the observer for the fields scattered by a perfectly conducting parabolic cylinder, is derived. This expansion can be applied to scattering by arbitrary smooth convex surfaces of variable curvature. The accuracy of numerical results is examined.

I. INTRODUCTION

In this paper, the fields on the surface of a parabolic cylinder derived by Fock,[1] Jones,[2] Rice,[3] Hong,[4] and Ott[5] are used together with the divergence theorem in two dimensions to derive a uniform asymptotic expansion for the fields scattered by a parabolic cylinder. The expansion is "uniform" in the variable locating the observers coordinates as the observation point moves from the lit region through the penumbra and into the shade. Higher-order terms in the asymptotic expansion were retained near the light-shadow boundary.

This expansion can be applied to scattering by arbitrary smooth convex surfaces of variable curvature by using the curvature of the surface in the asymptotic expansion derived in this paper. The solution is valid in the short-wavelength region.

The accuracy of the numerical results computed using a 96-point Gaussian quadrature algorithm for the transition function are examined as a function of the integration interval length. Contour plots of the magnitude of the total field in the region around the parabolic cylinder are presented for two angles of incidence. The asymptotics for the transition function are also presented.

II. ANALYSIS

The geometry of the incident plane wave and parabolic cylinder is shown in Fig. 1. The equations defining the transformation from rectangular coordinates to parabolic cylinder coordinates are

$$x = \tfrac{1}{2}(\xi^2 - \eta^2)\,, \tag{1a}$$

$$y = \xi\eta\,. \tag{1b}$$

The equations defining the transformation from parabolic cylinder coordinates to cylindrical coordinates are

$$\xi = \sqrt{2r}\cos(\phi/2)\,, \tag{2a}$$

$$\eta = \xi\tan(\phi/2)\,. \tag{2b}$$

A unit tangent vector to the η curve is

$$\mathbf{e}_T = \mathbf{e}_x\,\sin(\phi/2) - \mathbf{e}_y\,\cos(\phi/2)\,, \tag{3}$$

and a unit normal to the surface of the parabolic cylinder $\xi = \xi_0$ is

$$\mathbf{e}_n = \mathbf{e}_x\,\cos(\phi/2) + \mathbf{e}_y\,\sin(\phi/2)\,. \tag{4}$$

A unit vector in the direction of the incident wave is

$$\mathbf{e}_I = \mathbf{e}_x\,\cos\phi_0 + \mathbf{e}_y\,\sin\phi_0\,. \tag{5}$$

From Eqs. (4) and (5) the incident wave grazes the cylinder, $\xi = \xi_0$ when

$$\mathbf{e}_I \cdot \mathbf{e}_n = 0 \tag{6}$$

or

$$\phi = 2\phi_0 - \pi\,,$$

and from Eq. (2b), $\eta = -\xi_0\cos\phi_0$.

From the divergence theorem in two dimensions and

$$\frac{\partial\Phi}{\partial n} = 0\,, \quad \text{for } \xi = \xi_0\,, \tag{7}$$

we have, for the scattered field,

$$\Phi_s(x,y) = \int_{\xi_0} \Phi(\xi_0,\eta)\frac{\partial G}{\partial n}\,d\eta\,, \tag{8}$$

where $\Phi(\xi_0,\eta)$ is the total field on the cylinder and the two-dimensional Green's function is ($e^{i\omega t}$ time convention)

$$G(x,y|x',y') = (i/4)H_0^{(2)}\{k\,[(x-x')^2 + (y-y')^2]^{1/2}\}\,, \tag{9}$$

with x', y' coordinates on the surface of the cylinder $\xi = \xi_0$. The field on the surface of the cylinder is given by[5]

$$\Phi(\xi_0,\eta) = \frac{-1}{2\pi}\int_{-\infty}^{\infty} d\mu\,\Gamma\Big(i\mu + \frac{1}{2}\Big)\frac{(\tan\frac{1}{2}\phi_0)^{i\mu - 1/2}}{\cos(\phi_0/2)} \times D_{-i\mu-1/2}(h\eta)/D'_{i\mu-1/2}(h\xi_0)\,, \tag{10}$$

where $h = \sqrt{2ik}$ and $D_{-i\mu-1/2}(h\eta)$ is the parabolic cylinder function of complex order. The integral in Eq. (10) has a saddle point near $\mu = \mu_{sp} = \frac{1}{2}k\xi_0^2$, which coalesces with the point $\eta = -\xi_0\cos\phi_0$ near grazing. Ott[5] has shown that

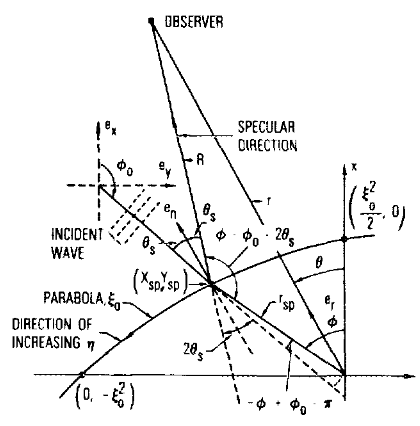


FIG. 1. Geometry for incident plane wave, stationary phase point (x_{sp}, y_{sp}) and observers coordinates.

 0022-2488/85/040854-07$02.50

near the saddle point Eq. (10) has the following asymptotic expansion:

$$\Phi(\xi_0, y) = (e^{i5\pi/6}/2\pi)(\tfrac{1}{2}k\xi_0^2)^{-1/3} \times \int_{-\infty}^{\infty} d\mu \, \frac{\exp[-i\mu(\pm\xi_2 - \ln\tan(\phi_0/2))]}{\mathrm{Ai}'(e^{i2\pi/3}\xi_1\mu^{2/3})}, \tag{11}$$

where the plus sign applies for $\eta > 0$, the minus sign for $\eta < 0$, and

$$\frac{2}{3}\xi_1^{3/2} = \begin{cases} 2\int_1^{(\mu_{sp}/\mu)^{1/2}} \sqrt{s^2-1}\, ds, & -\pi < \arg\mu < 0, \\ 2\int_{(\mu_{sp}/\mu)^{1/2}}^{1} \sqrt{s^2-1}\, ds, & -2\pi < \arg\mu < -\pi, \end{cases} \tag{12}$$

and

$$\xi_2 = \left(\frac{k\eta^2}{2\mu}\right)^{1/2}\left(1+\frac{k\eta^2}{2\mu}\right)^{1/2} + \ln\left[\left(\frac{k\eta^2}{2\mu}\right)^{1/2} + \left(1+\frac{k\eta^2}{2\mu}\right)^{1/2}\right]. \tag{13}$$

In the neighborhood of the saddle point we expand μ as

$$\mu = \mu_{sp} + \mu_{sp}^{1/3}\tau, \quad |\tau| < 1. \tag{14}$$

Making the substitution

$$s = \sqrt{v}, \quad ds = (2s)^{-1}\, dv, \tag{15}$$

and noting that $1 < |s| < 1 + \epsilon$, with ϵ a small number,

$$\frac{2}{3}\xi_1^{3/2} = \begin{cases} \int_1^{(\mu_{sp}/\mu)} \sqrt{v-1}\, dv, & -\pi < \arg\mu < 0, \\ \int_{(\mu_{sp}/\mu)}^{1} \sqrt{v-1}\, dv, & -2\pi < \arg\mu < -\pi, \end{cases}$$

$$= \begin{cases} (v-1)^{3/2}/\tfrac{3}{2}\big|_1^{(\mu_{sp}/\mu)}, \\ (v-1)^{3/2}/\tfrac{3}{2}\big|_{(\mu_{sp}/\mu)}^{1}, \end{cases} \tag{16}$$

$$\xi_1 = \begin{cases} \mu_{sp}/\mu - 1, & -\pi < \arg\mu < 0, \\ 1 - \mu_{sp}/\mu, & -2\pi < \arg\mu < -\pi. \end{cases} \tag{17}$$

Substituting Eq. (14) into Eq. (17)

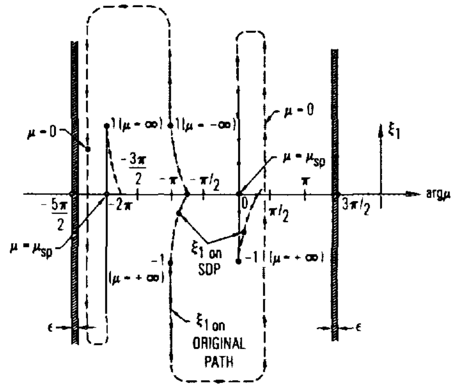


FIG. 2. Region of validity for asymptotic expansions for parabolic cylinder functions and the variable ξ, vs $\arg\mu$.

$$\xi_1 = \begin{cases} -\mu_{sp}^{-2/3}\tau, & -\pi < \arg\mu < 0, \\ \mu_{sp}^{-2/3}\tau, & -2\pi < \arg\mu < -\pi. \end{cases} \tag{18}$$

Figure 2 shows the variation of ξ_1 vs $\arg\mu$ along the original path and versus $\arg\tau$ along the steepest descent path. Also shown is the region of validity of the asymptotic expansions for the parabolic cylinder functions, $D_{i\mu-1/2}(h\eta)$. The connection between the branches is

$$\arg\{\tfrac{2}{3}[\xi_1((\mu_{sp}/\mu)e^{i\pi})]^{3/2}\} = -\arg\{\tfrac{2}{3}[\xi_1(\mu_{sp}/\mu)]^{3/2}\} + \pi. \tag{19}$$

Ott[5] has shown that near the saddle point, the exponent of the exponential function in the integrand of Eq. (11) has a Taylor series expansion about the grazing point, $\eta = \xi_0 \cot\phi_0$, given by

$$-i\mu(\pm\xi_2 - \ln\tan(\phi_0/2)) = -i\left\{\frac{k\xi_0\omega}{\sin^2\phi_0}\left(1 - \frac{\omega\cos\phi_0}{2\xi_0}\right) + \frac{k\xi_0^2}{2}\left(\frac{-\cos\phi_0}{\sin^2\phi_0} + \frac{\omega^3}{3\xi_0^2}\right) + \left(\frac{k}{2\xi_0}\right)^{1/3}\omega\tau\left(1 + \frac{\omega\cos\phi_0}{2\xi_0}\right)\right\}, \tag{20}$$

where

$$\omega = \eta\sin\phi_0 + \xi_0\cos\phi_0, \tag{21}$$

with τ defined in Eq. (14).

From Eq. (9)

$$\nabla G = (-ik/4)[(x-x')\mathbf{e}_x + (y-y')\mathbf{e}_y]\,\frac{H_1^{(2)}\{k[(x-x')^2+(y-y')^2]^{1/2}\}}{[(x-x')^2+(y-y')^2]^{1/2}}. \tag{22}$$

Now

$$\nabla G\cdot\mathbf{e}_n = \frac{\partial G}{\partial n}, \tag{23}$$

and substituting Eqs. (22) and (4) into (23) and using the asymptotic form for $H_1^{(2)}$ we have

$$\frac{\partial G}{\partial n}\Big|_{\substack{y'=\xi_0\eta \\ x'=\frac{1}{2}(\xi_0^2-\eta^2)}} \sim (-i/4)(2k/\pi)^{1/2}e^{i3\pi/4}\{\xi_0[x\tfrac{1}{2}(\xi_0^2-\eta^2)] + \eta(y-\xi_0\eta)\} \cdot \frac{\exp\{-ik[(x-\tfrac{1}{2}(\xi_0^2-\eta^2))^2+(y-\xi_0\eta)^2]^{1/2}\}}{[(x-\tfrac{1}{2}(\xi_0^2-\eta^2))^2+(y-\xi_0\eta)^2]^{3/4}}, \tag{24}$$

where

$$\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y , \tag{25}$$

$$\xi_0/\cos(\phi/2) = \sqrt{\xi_0^2 + \eta^2} = \xi_0\sqrt{1 + \tan^2\phi/2} . \tag{26}$$

Substituting Eqs. (11), (20), and (24) into Eq. (8), we find

$$\begin{aligned}\Phi_s(x,y) = {} & \frac{e^{i13\pi/12}}{8\pi}\left(\frac{2k}{\pi}\right)^{1/2}\left(\frac{k\xi_0^2}{2}\right)^{-1/3}\int_{-\infty}^{\infty} d\eta \int_{-\infty}^{\infty} d\mu \, \frac{[\xi_0(x - \frac{1}{2}(\xi_0^2 - \eta_0^2)) + \eta(y - \xi_0\eta)]}{R^{3/2}\mathrm{Ai}'(e^{i\pi/3}\xi_1\mu^{2/3})} \\ & \cdot \exp\left\{ -i\left[\frac{k\xi_0\omega}{\sin^2\phi_0}\left(1 - \frac{\omega\cos\phi_0}{2\xi_0}\right) + \frac{k\xi_0^2}{2}\left(-\frac{\cos\phi_0}{\sin^2\phi_0} + \frac{\omega^3}{3\xi_0^3}\right)\right.\right. \\ & \left.\left. + \left(\frac{k}{2\xi_0}\right)^{1/3}\omega\tau\left(1 + \frac{\omega\cos\phi_0}{2\xi_0}\right) + k\,[(x - \tfrac{1}{2}(\xi_0^2 - \eta^2))^2 + (y - \xi_0\eta)^2]^{1/2}\right]\right\} . \end{aligned} \tag{27}$$

The result in Eq. (27) is exact in the far field for the given boundary condition. We assume $k\xi_0^2$ is a large parameter and evaluate the η integration in Eq. (27) asymptotically. The "rapidly" varying portion of the exponent in the exponential function in Eq. (27) is

$$f(\mu,\eta) = \frac{k\xi_0\omega}{\sin^2\phi_0} - \frac{k\omega^2\cos\phi_0}{2\sin^2\phi_0} - \frac{k\xi_0^2\cos\phi_0}{2\sin^2\phi_0} + k\,\{[(x - \tfrac{1}{2}(\xi_0^2 - \eta^2))^2 + (y - \xi_0\eta)^2]^{1/2}\} , \tag{28}$$

where $k\xi_0^2\cos\phi_0/2\sin^2\phi_0$ is the incident wave. The "slowly"varying portion of the integrand for the η integration is

$$H(\eta) = (\exp\{-i(k/2\xi_0)^{1/3}\omega\tau - ik\omega^3/6\xi_0\}/R^{3/2})[\xi_0(x - \tfrac{1}{2}(\xi_0^2 - \eta^2)) + \eta(y - \xi_0\eta)] . \tag{29}$$

We expand this in a Taylor's series about the stationary phase point, $\mu_{\rm sp}$, as

$$H(\eta) = H(\eta_{\rm sp}) + \frac{1}{2}\left.\frac{\partial^2 H}{\partial\eta^2}\right|_{\eta = \eta_{\rm sp}}(\eta - \eta_{\rm sp})^2 + \cdots , \tag{30}$$

where the odd term $(\eta - \eta_{\rm sp})$ integrates to zero when multiplied by $\exp\{-(|f''|/2)(\eta - \eta_{\rm sp})^2\}$ over the infinite range of integration on η.

After a great deal of algebra

$$\begin{aligned}\frac{\partial^2 H}{\partial\eta^2} = {} & \frac{\exp\{-i(k/2\xi_0)^{1/3}\omega\tau - ik\omega^3/6\xi_0\}}{R^{3/2}}\left\{\left[-i\left(\frac{k}{2\xi_0}\right)^{1/3}\tau - i\,\frac{kK^2\cos^2\theta_s}{2\xi_0}\right]^2 \cdot KR\sin^2\phi_0\cos\theta_s\right. \\ & + (y - \xi_0\eta)\left[-i\left(\frac{k}{2\xi_0}\right)^{1/3}\tau - i\,\frac{kK^2\cos^2\theta_s}{2\xi_0}\right]\sin\phi_0 + \frac{3}{2}K^2\sin\theta_s\cos\theta_s\sin\phi_0\left[-i\left(\frac{k}{2\xi_0}\right)^{1/3}\tau\right. \\ & \left. - ik\,\frac{K^2\cos^2\theta_s}{2\xi_0}\right] - i\,\frac{kR}{\xi_0}K^2\cos\theta_s\sin^2\phi_0 - \xi_0 - \frac{3}{2}K\cos\theta_s\,\frac{[K^2 + x - \frac{1}{2}(\xi_0^2 - \eta^2)]}{R} \\ & - \frac{3}{2}K\cos\theta_s\,\frac{(y - \xi_0\eta)}{R} - \frac{3K^3\sin^3\theta_s}{R} - \frac{3}{2}K\sin\theta_s\,\frac{(y - \xi_0\eta)}{R} \\ & \left. - \frac{3}{2}K^2\cos\theta_s\sin\theta_s\sin\phi_0\left[-i\left(\frac{k}{2\xi_0}\right)^{1/3}\tau - i\,\frac{kK^2\cos^2\theta_s}{2\xi_0}\right] + \frac{9}{4}\frac{K^3\cos\theta_s\sin^2\theta_s}{R}\right\} \end{aligned} \tag{31}$$

where the curvature is

$$K = \xi_0/\cos\phi/2 . \tag{32}$$

To find the stationary phase point, we compute from Eq. (28)

$$\begin{aligned}\frac{\partial f}{\partial\eta} &= \frac{k\xi_0}{\sin\phi_0} - \frac{k\omega\cos\phi_0}{\sin\phi_0} \\ &\quad + k\,\{\eta[x - \tfrac{1}{2}(\xi_0^2 - \eta^2)] - \xi_0(y - \xi_0\eta)\}/R \\ &= k\xi_0/\sin\phi_0 - [k(\eta\sin\phi_0 + \xi_0\cos\phi_0)\cos\phi_0]/\sin\phi_0 \\ &\quad - k(\eta^2 + \xi_0^2)^{1/2}|\mathbf{e}_n \times \mathbf{e}_R| \\ &= k\xi_0(\cos(\phi/2)\sin\phi_0 - \sin(\phi/2)\cos\phi_0) - k\xi_0\sin\theta_s\end{aligned}$$

and

$$\frac{\partial f}{\partial\eta} = 0 = \cos(\phi/2)\sin\phi_0 - \sin(\phi/2)\cos\phi_0 - \sin\theta_s ,$$

or one solution in the lit region is

$$\sin[\pi - (\phi_0 - \phi/2)] = \sin\theta_s , \quad \theta_s = \pi - (\phi_0 - \phi/2) , \tag{33}$$

and a second solution valid in the shade

$$\sin\theta_s = \sin(\phi_0 - \phi/2) , \quad \theta_s = \phi_0 - \phi/2 . \tag{34}$$

These two solutions are shown in Fig. 3 for one incidence angle ϕ_0. The solution in Eq. (34) removes the incident wave in the shadow region. After a great deal of algebra, from Eq. (28) we obtain

$$\begin{aligned}\frac{\partial^2 f}{\partial\eta^2} &= -k\cos\phi_0 + \frac{k\xi_0^2}{R}\cos^2\theta_s\left(1 + \tan^2\frac{\phi}{2}\right) \\ &\quad + (k/R)[x - \tfrac{1}{2}(\xi_0^2 - \eta^2)] \\ &= (k/R)\{(x - x_{\rm sp}) - R\cos\phi_0 + (K\cos\theta_s)^2\} \\ &\cong (k/R)\{2x + (K\cos\theta_s)^2\} ,\end{aligned} \tag{35}$$

with

$$x_{\rm sp} = \tfrac{1}{2}(\xi_0^2 - \eta_{\rm sp}^2) , \quad y_{\rm sp} = \xi_0\eta_{\rm sp} . \tag{36}$$

From Eqs. (28), (33), and (35) we have

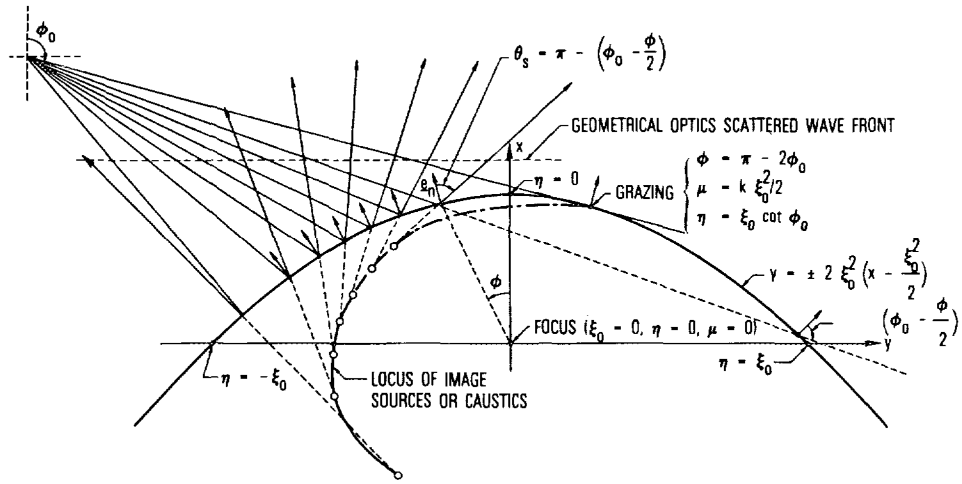


FIG. 3. Definition of stationary phase points in lit and shadow regions plus focus of caustic points inside parabolic cylinder.

$$f(\mu,\eta) = k\,[R + (y_{sp}\sin\phi_0 + x_{sp}\cos\phi_0)] + (k/R)[2x + (K\cos\theta_s)^2]\,[(\eta-\eta_{sp})^2/2] + \cdots. \tag{37}$$

From Eqs. (30), (31), and (32) we have

$$\begin{aligned} H(\eta) &= \exp\left\{-i\left(\frac{k}{2\xi_0}\right)^{1/3} K\cos\theta_s\tau - i\frac{kK^3}{6\xi_0}\cos^3\theta_s\right\} \\ &\times\left\{\frac{\xi_0\cos\theta_s}{R^{1/2}\cos(\theta/2)} - \left[\xi_0 + i(y-\xi_0\eta)\right.\right. \\ &\times\left.\left.\tau\left(\frac{k}{2\xi_0}\right)^{1/3}\sin\phi_0\right]\frac{(\eta-\eta_{sp})^2}{2R^{3/2}}\right\} \\ &= \frac{1}{R^{1/2}}\exp\left\{-i\left(\frac{k}{2\xi_0}\right)^{1/3}K\cos\theta_s\tau - i\frac{kK}{6\xi_0}\cos\theta_s\right\} \\ &\times\left\{K\cos\theta_s - \left[\xi_0 + i(y-\xi_0\eta)\tau\left(\frac{k}{2\xi_0}\right)^{1/3}\sin\phi_0\right]\right. \\ &\times\left.\frac{(\eta-\eta_{sp})^2}{2R}\right\}. \end{aligned} \tag{38}$$

The η integration is, from Eqs. (37) and (38)

$$\begin{aligned} &\int_{-\infty}^{\infty} H(\eta)e^{-if(\mu,\eta)}\,d\eta \\ &= \exp\left\{-i\left(\frac{k}{2\xi_0}\right)^{1/3}K\cos\theta_s\tau - i\frac{kK^3\cos^3\theta_s}{6\xi_0}\right\}e^{-i\pi/4} \\ &\cdot\left\{\frac{K\cos\theta_s}{R^{1/2}}\sqrt{\frac{2\pi}{|f''|}}\begin{Bmatrix}1\\ i\end{Bmatrix} - i\left[\xi_0 + i(y-\xi_0\eta)\right.\right. \\ &\times\left.\left.\tau\left(\frac{k}{2\xi_0}\right)^{1/3}\sin\phi_0\right]\frac{1}{2R^{3/2}}\sqrt{\frac{2\pi}{|f''|^3}}\begin{Bmatrix}1\\ -i\end{Bmatrix}\right\}, \end{aligned} \tag{39}$$

where $f''(\eta_s) > 0$ corresponds to one and $f''(\eta_s) < 0$ corresponds to i,

$$\begin{aligned} &= \sqrt{\frac{2\pi}{k}}\exp\left\{-i\left(\frac{k}{2\xi_0}\right)^{1/3}K\cos\theta_s\tau\right. \\ &\quad\left.- i\frac{kK^3}{6\xi_0}\cos^3\theta_s\right\}e^{-i\pi/4} \\ &\quad\cdot\left\{\left[\frac{(K\cos\theta_s)^2}{2x+(K\cos\theta_s)^2}\right]^{1/2}\begin{Bmatrix}1\\ i\end{Bmatrix}\right. \\ &\quad\left.+ i\frac{[(k\xi_0^2)^{1/2} + iy\tau\sqrt{k}(k/2\xi_0)^{1/3}\sin\phi_0]}{2\{k\,[2|x| + (K\cos\theta_s)^2]\}^{3/2}}\begin{Bmatrix}1\\ -i\end{Bmatrix}\right\}. \end{aligned} \tag{40}$$

The μ integration in Eq. (27) is transformed to the τ plane by Eq. (14) together with

$$\begin{aligned} d\mu &= {\mu_{sp}}^{1/3}\,d\tau, \\ e^{i\pi/3}\xi_1\mu^{2/3} &= e^{-i2\pi/3}\tau. \end{aligned} \tag{41}$$

Substituting Eqs. (37), (40), and (41) into Eq. (27) gives

$$\begin{aligned} \Phi_s(x,y) = \frac{e^{i5\pi/6}}{4\pi}&\exp\{-ik\,[R + (x_{sp}\cos\phi_0 + y_{sp}\sin\phi_0)]\}\left(\left\{\left[\frac{(K\cos\theta_s)^2}{2x+(K\cos\theta_s)^2}\right]^{1/2}\begin{Bmatrix}1\\ i\end{Bmatrix}\right.\right. \\ &+ i\frac{(k\xi_0^2/4)^{1/2}}{\{k\,[2x+(K\cos\theta_s)^2]\}^{3/2}}\left.\begin{bmatrix}1\\ -i\end{bmatrix}\right\}\exp\left\{-i\frac{kK^3\cos^3\theta_s}{6\xi_0}\right\}\int_c d\tau\,\frac{\exp\{i(k/2\xi_0)^{1/3}K\cos\theta_s\tau\}}{\mathrm{Ai}'(e^{-i2\pi/3}\tau)} \\ &- \frac{\sqrt{k}(k/2\xi_0)^{1/3}\sin\phi_0}{2\{k\,[2|x| + (k\cos\theta_s)^2]\}^{3/2}}\exp\left\{-ik\frac{K^3\cos^3\theta_s}{6\xi_0}\right\}\int_c \tau\,d\tau\,\frac{\exp\{i(k/2\xi_0)^{1/3}K\cos\theta_s\tau\}}{\mathrm{Ai}'(e^{-i2\pi/3}\tau)}\left.\begin{Bmatrix}1\\ -i\end{Bmatrix}\right), \end{aligned} \tag{42}$$

where the contour c is yet to be specified. We select the path in the complex τ plane where $\mathrm{Ai}'(e^{-i2\pi/3}\tau)$ in the denominator of Eq. (42) is exponentially large. This occurs along the rays $\infty e^{i2\pi/3}$ to the origin and the negative real axis. The zeros of $\mathrm{Ai}'(e^{-i2\pi/3}\tau)$ in the complex τ plane lie along the ray $\arg(\tau) = \pi/3$. Traveling waves in the complex τ plane lie along the ray $\arg(\tau) = -\pi$. By the Cauchy–Goursat theorem

$$\oint_c \frac{e^{-i\alpha\tau}}{\mathrm{Ai}'(e^{-i2\pi/3}\tau)}\, d\tau = 0\,, \tag{43}$$

where

$$\alpha = -(k/2\xi_0)^{1/3} K \cos\theta_s\,, \tag{44}$$

and c consists of the ray from $-\infty$ to the origin ($\arg\tau = -\pi$), from the origin to $\infty e^{i2\pi/3}$ ($\arg\tau = 2\pi/3$), and along the arc at ∞ so as to close the path. The contribution along the arc at ∞ goes to zero and Eq. (43) becomes

$$\left\{\int_{-\infty}^{0} + \int_0^{\infty e^{i2\pi/3}}\right\} d\tau \frac{e^{-i\alpha\tau}}{\mathrm{Ai}'(e^{-i2\pi/3}\tau)} = 0 \tag{45}$$

or

$$\int_{-\infty}^{0} \frac{e^{-i\alpha\tau}}{\mathrm{Ai}'(e^{-2\pi/3}\tau)}\, d\tau = \int_{\infty e^{i2\pi/3}}^{0} \frac{e^{-i\alpha\tau}}{\mathrm{Ai}'(e^{-i2\pi/3}\tau)}\, d\tau\,. \tag{46}$$

Thus, c in Eq. (42) is

$$\left\{\int_{\infty e^{i2\pi/3}}^{0} + \int_0^{\infty}\right\} \frac{e^{-i\alpha\tau}}{\mathrm{Ai}'(e^{-i2\pi/3}\tau)}\, d\tau\,. \tag{47}$$

Now introduce the rotation

$$t = e^{-i2\pi/3}\tau\,, \quad dt = e^{-i2\pi/3}\, d\tau\,, \tag{48}$$

and the contour becomes the one shown in Fig. 4 and our uniform asymptotic expansion is

$$\begin{aligned}\Phi_s(x,y) = {} & \frac{-i}{4\pi} \exp\{-ik\,[R + (x_{\rm sp}\cos\phi_0 + y_{\rm sp}\sin\phi_0)]\} \\ & \times\left[\frac{(K\cos\theta_s)^2}{2x + (K\cos\theta_s)^2}\right]^{1/2} \begin{Bmatrix}1\\ i\end{Bmatrix} \\ & + i\frac{(k\xi_0^2/4)^{1/2}}{\{k\,[2|x| + (K\cos\theta_s)^2]\}^{3/2}} \\ & \times\begin{Bmatrix}1\\ -i\end{Bmatrix} e^{-i\alpha^3/3} G(\alpha) - \frac{1}{2^{4/3}} \\ & \times\frac{ky}{\{k\,[2|x| + (K\cos\theta_s)^2]\}^{3/2}} \\ & \times\frac{\sin\phi_0}{(k\xi_0^2)^{1/6}} e^{-i\alpha^3/3} \frac{\partial G}{\partial\alpha} \begin{Bmatrix}1\\ -i\end{Bmatrix}, \end{aligned} \tag{49}$$

with

$$G(\alpha) = \int_{\infty e^{-i2\pi/3}}^{\infty e^{i2\pi/3}} \frac{e^{\alpha e^{i\pi/6} z}}{\mathrm{Ai}'(z)}\, dz \tag{50}$$

and

$$\alpha = -(k\xi_0^2/2)^{1/3}[\cos\theta_s/\cos(\phi/2)]\,, \tag{51}$$

which is valid in a uniform sense with respect to the variable θ_s or the observers location. The θ_s varies from 0 to $\pi/2$ in the lit region to $\pi/2$ at grazing from $\pi/2$ to π in the shade. For large values of α we have

$$\mathrm{Ai}'(z) \sim (iz^{1/4}/2\sqrt{\pi})e^{2/3 z^{3/2}}\,, \quad \pi/3 < \arg z < 5\pi/3\,, \tag{52}$$

and Eq. (50) becomes

$$G(\alpha) \sim \int_{\infty e^{-i2\pi/3}}^{\infty e^{i2\pi/3}} -i2\sqrt{\pi}\exp\frac{\{\alpha z e^{i\pi/6} - \frac{2}{3}z^{3/2}\}}{z^{1/4}}\, dz\,. \tag{53}$$

Now

$$f(z) = \alpha z e^{i\pi/6} - \tfrac{2}{3}z^{3/2}$$

has a stationary phase point when

$$f'(z) = 0 = \alpha e^{i\pi/6} - z^{1/2}$$

or

$$z_{\rm sp} = \alpha^2 e^{i\pi/3}\,.$$

Also

$$f''(z) = -\tfrac{1}{2}z^{-1/2}\,.$$

Thus,

$$\begin{aligned} G(\alpha) &= -i2\sqrt{\pi}(1/z_{\rm sp}^{1/4})\sqrt{2\pi}\sqrt{2}z_{\rm sp}^{1/4}e^{i\alpha^3/3} \\ &\sim -i4\pi e^{i\alpha^3/3}\,, \end{aligned} \tag{54}$$

and the scattered field in the lit region becomes

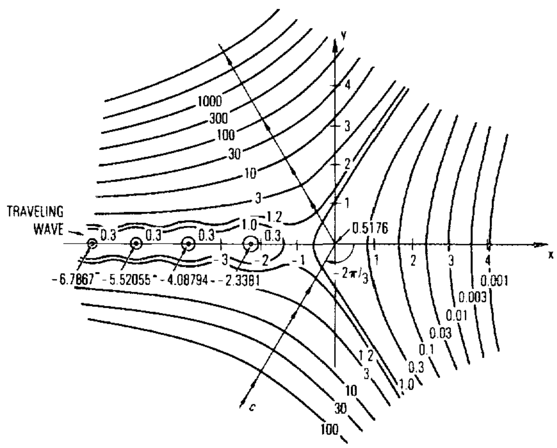


FIG. 4. Magnitude of 2 $\mathrm{Ai}'(z)$ in complex $z = x + iy$ plane. Also, integration contour c in evaluating transition function.

TABLE I. $G(\alpha)$ in Eq. (50) vs α.

α	$\lvert G(\alpha)\rvert$	$\arg G(\alpha)$ (rad)
-3.0	1.997 616 4	2.004 535 1
-2.5	1.993 570 6	-0.481 236 61
-2.0	1.981 733 5	2.070 852 8
-1.5	1.947 577 2	-2.653 338 8
-1.0	1.860 646 7	-1.840 348 7
-0.5	1.681 886 6	-1.544 396 7
0	1.399 375 7	-1.570 796 327
0.5	1.059 087 8	-1.769 061 9
1.0	0.738 222 56	-2.035 778 6
1.5	0.488 137 90	-2.313 634 1
2.0	0.315 335 19	-2.582 729 6
2.5	0.202 482 86	-2.843 095 4
3.0	0.130 022 47	-3.099 222 8

TABLE II. Convergence of $G(\alpha)$ with ϵ.

ϵ	Re $G(1.0)$	Im $G(1.0)$
0.05	3.592 306 7	$+i$ 2.011 229 7
0.10	3.592 306 7	$+i$ 2.011 229 7
0.2	3.592 306 8	$+i$ 2.011 229 7
1	3.592 310 4	$+i$ 2.011 229 8
2	3.592 768 0	$+i$ 2.010 690 3

$$\phi_s = -\exp\{-ik[R + (x_{sp}\cos\phi_0 + y_{sp}\sin\phi_0)]\} \cdot [(K\cos\theta_s)^2/(2x + (K\cos\theta_0)^2)]^{1/2}. \tag{55}$$

The factor

$$\text{Divergence} = \left[\frac{(K\cos\theta_s)^2}{2x + (K\cos\theta_s)^2}\right]^{1/2} \tag{56}$$

is the classical divergence factor and the minus sign in Eq. (56) is the reflection coefficient.

Figure 3 shows the locus of image points at a distance $K\cos\theta_s$ from the reflector surface.

III. NUMERICAL RESULTS

Table I gives a tabulation of the function $G(\alpha)$ in Eq. (50). The results in Table I were obtained on a CDC Cyber 170-835 with approximately a 56 ns cycle time. The integral $G(\alpha)$ was computed using 96 Gaussian quadrature points per interval of length ϵ. The range of floating point numbers on the CDC 170-835 is

$$10^{-293} \leqslant x \leqslant 10^{322},$$

with 60 bits per word (about 14 significant figures). The float-

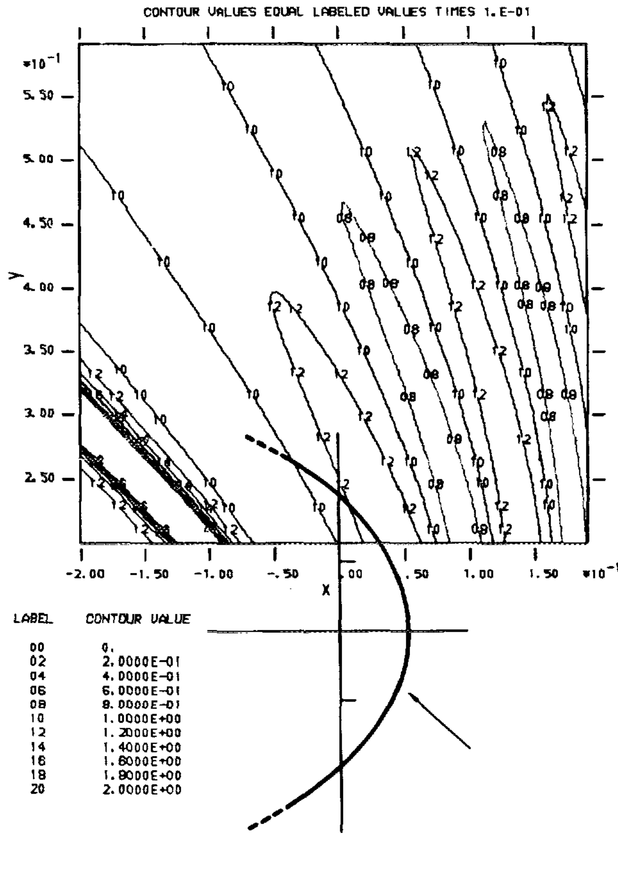


FIG. 6. Contour plots of magnitude of total field for $\phi_0 = 3\pi/4$ for $-2\lambda < x < 2\lambda$, $2\lambda < y < 6\lambda$.

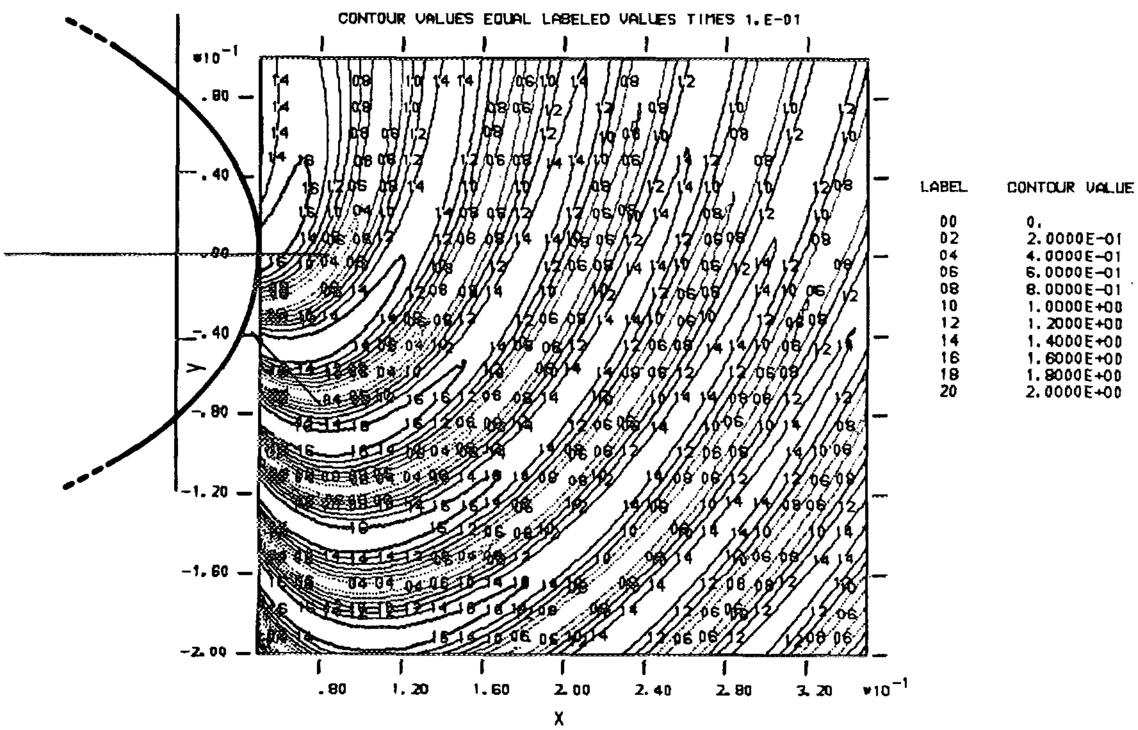


FIG. 5. Contour plots of magnitude of total field for $\phi_0 = 3\pi/4$ for $-0.2\lambda < y < \lambda$, $0 < x < 3.2\lambda$.

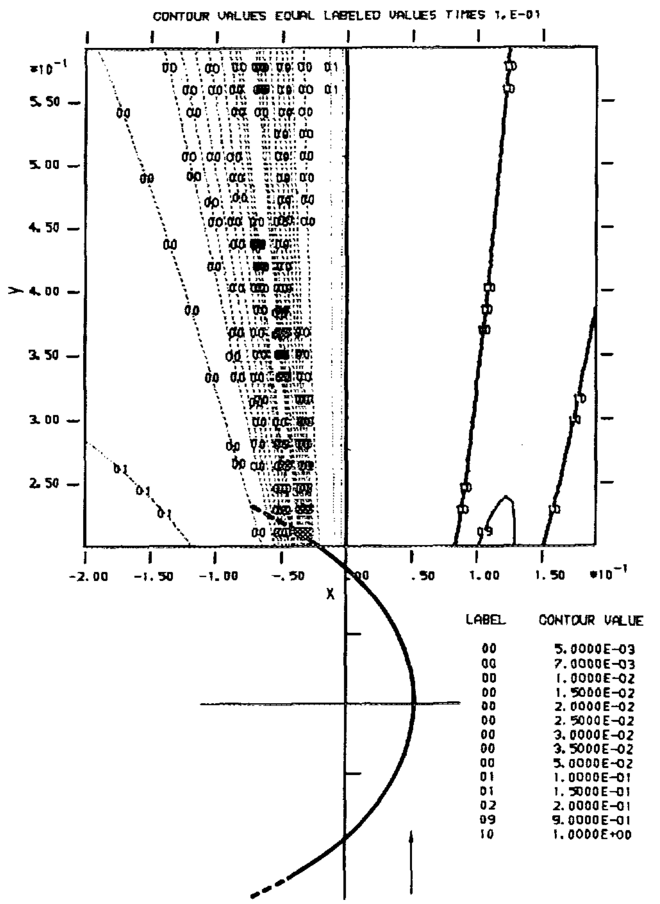


FIG. 7. Contour plots of magnitude of total field for $\phi_0 = \pi/2$ for $-2\lambda \leqslant x \leqslant 2\lambda$, $2\lambda \leqslant y \leqslant 6\lambda$.

ing point range limited the approximation of the infinite limits in Eq. (50) to around 107 [i.e., $\exp(+\frac{2}{3}(107)^{3/2}) \cong 10^{322}$]. Table II shows the variation in $G(\alpha)$ for $\alpha = 1.0$ for different values of the interval length ϵ. From Table II, $\epsilon \cong 0.15$ will yield about 8 significant figure accuracy so $\epsilon = 0.14925$ was used in the algorithm.

In Fig. 5 we show a contour plot of the magnitude of the total field

$$\Phi(x, y) = e^{-ikr\cos(\phi - \phi_0)} + \Phi_s(x, y), \tag{57}$$

with $\Phi_s(x, y)$ calculated from Eq. (49). The parameters for the plot are $\lambda = \xi_0^2$ and $\phi_0 = 3\pi/4$. The modal pattern in Fig. 5 is the result of the interference of the incident plane wave and the scattered plane wave. The magnitude of Φ in Eq. (57) varies from about 0.4 to 1.7 for $0 \leqslant x \leqslant 4\lambda$ and $-2\lambda \leqslant y \leqslant \lambda$. The magnitude does not vary from 0 to 2 in Eq. (57) because the divergence factor is less than unity.

When $\theta_s > \pi/2$ the total field in the shade is given by

$$\Phi(x, y) = \Phi_s(x, y). \tag{58}$$

From Fig. 1,

$$r \cong R + r_{\rm sp}\cos(\phi - \theta) \tag{59}$$

or

$$r \cong R - r_{\rm sp}\cos(\phi - \phi_0 - 2\theta_s), \tag{60}$$

and from Eqs. (59) and (60)

$$-2\theta_s + \phi - \phi_0 = \phi - \theta - \pi$$

or

$$\theta_s = \pi/2 + (\theta - \phi_0)/2, \tag{61}$$

and $\theta_s = \pi/2$ when $\theta = \phi_0$.

In Fig. 6 contour plots for the magnitude of the total field are shown for $-2\lambda \leqslant x \leqslant 2\lambda$ and $2\lambda \leqslant y \leqslant 6\lambda$. Generally, the magnitude of the field in this region varied from about 0.8 to 1.2. The light–shadow boundary intersects the lower left-hand corner of Fig. 6 at the point $x = 2\lambda$, $y = 2\lambda$. In this region the scattered field is traveling in a direction closer to the direction of the incident wave than for the contour plot in Fig. 5.

In Fig. 7 contour plots of the magnitude of the total field are shown for $-2\lambda \leqslant x \leqslant 2\lambda$ and $2\lambda \leqslant y \leqslant 6\lambda$ for $\phi_0 = \pi/2$. The lines of constant field magnitude are nearly parallel to the incident wave front.

IV. CONCLUDING REMARKS

A uniform asymptotic expansion for the fields scattered by a parabolic cylinder is derived. This expansion involves a transition function that allows for the expansion to remain valid as the observer moves from the lit region through the penumbra and into the shade. The solution is applicable to scattering by arbitrarily curved convex surfaces by replacing the curvature of the parabolic cylinder with the curvature of the surface in question.

ACKNOWLEDGMENTS

The author wishes to thank Dr. Robert Dybdal of the RF Systems Research Office at The Aerospace Corporation for many interesting and stimulating discussions regarding this problem.

This work was supported by the Air Force under Contract No. FO4701-83-C-0084.

[1]V. A. Fock, *International Series of Monographs on Electromagnetic Waves* (Pergamon, New York, 1965), Chaps. 1, 2, and 5.
[2]D. S. Jones, *The Theory of Electromagnetism* (Macmillan, New York, 1964), pp. 467–478.
[3]S. O. Rice, Bell Syst. Tech. J. **33**, 417 (1953).
[4]S. Hong, J. Math. Phys. **8**, 1123 (1967).
[5]R. H. Ott, J. Math. Phys. **11**, 590 (1970).

Exact solutions of the wave equation with complex source locations

Richard W. Ziolkowski
Lawrence Livermore National Laboratory, P.O. Box 808, L-156, Livermore, California 94550

(Received 6 August 1984; accepted for publication 9 November 1984)

New exact solutions of the homogeneous, free-space wave equation are obtained. They originate from complex source points moving at a constant rate parallel to the real axis of propagation and, therefore, they maintain a Gaussian profile as they propagate. Finite energy pulses can be constructed from these Gaussian pulses by superposition.

I. INTRODUCTION

A recent article by Brittingham[1] has indicated the existence of a new type of solution of the homogeneous, free-space wave equation

$$\Box \Phi(\mathbf{r},t) = \{\Delta - \partial_{ct}^2\}\Phi(\mathbf{r},t) = 0, \tag{1}$$

which was termed a focus wave mode (FWM) for its alleged solitonlike properties. It has been found that this FWM is but one of a class of solutions of (1). In particular, assuming the desired direction of propagation is along the z axis, a solution of the form

$$\Phi(\mathbf{r},t) = e^{ik(z+ct)}F_k(x,y,z-ct) \tag{2}$$

reduces (1) to a Schrödinger equation; i.e.,

$$e^{-ik(z+ct)}\Box\, e^{ik(z+ct)}F_k = \{\Delta_\perp + 4ik\,\partial_\tau\}F_k = 0, \tag{3}$$

where the transverse Laplacian is $\Delta_\perp = \partial_x^2 + \partial_y^2$ and the characteristic variables are $(\tau,\sigma) = (z-ct, z+ct)$. Equation (2) has a symmetric solution ($\rho^2 = x^2 + y^2$)

$$F_k(x,y,\tau) = \exp[-k\rho^2/(z_0 + i\tau)]/4\pi i(z_0 + i\tau), \tag{4}$$

that originates at the complex source location $(\rho,\tau) = (0, iz_0)$,

$$\{\Delta_\perp + 4ik\,\partial_\tau\}F_k(x,y,\tau) = \delta(\boldsymbol{\rho})\delta(\tau - iz_0), \tag{5}$$

the right-hand side being identically zero for a point in real space-time.

Clearly, the source location $z = iz_0 + ct$ moves parallel to the real z axis. Moreover, defining the complex variance $V = z_0 + i\tau$ so that

$$1/V = 1/A - i/R, \tag{6}$$

it is recognized immediately that (4) represents a moving Gaussian beam with beam spread $A = z_0^2 + (\tau^2/z_0)$, phase front curvature $R = \tau + (z_0^2/\tau)$, and normalized beam waist $(A/k)^{1/2}$. Consequently, combining (2) and (4), the fundamental solution

$$\Phi(\mathbf{r},t) = \phi_k(\rho,\tau,\sigma) = \frac{\exp[ik\sigma - k\rho^2/(z_0 + i\tau)]}{4\pi i(z_0 + i\tau)} \tag{7}$$

is a modulated, moving Gaussian pulse.

II. SOLUTION BEHAVIOR

A short time history of $\mathrm{Re}\,\Phi$ with $f = kc/2\pi = 3.0\times 10^9$ and $z_0 = 1.0$ is shown in Fig. 1. The second subplots are contour plots of the three-dimensional surface plots given in the first subplot. The Gaussian profile of the pulse is apparent. Notice that this profile is maintained during propagation with only local variations. The latter occur primarily near the profile center $(\rho,\tau) = (0,0)$. The variation in the shape of Φ with k is illustrated in Fig. 2. The pulse is concentrated near the ρ axis for small k and becomes more concentrated along the z axis for large k. The unusual features of the plots in Fig. 2(c) such as the jagged peaks and the ragged contours are artifacts of the coarseness of the computational grid. It has also been demonstrated that the pulse amplitude decreases as z_0 increases.

The FWM solution of Maxwell's equations in Ref. 1 is readily obtained from (7) with a Hertzian potential formula-

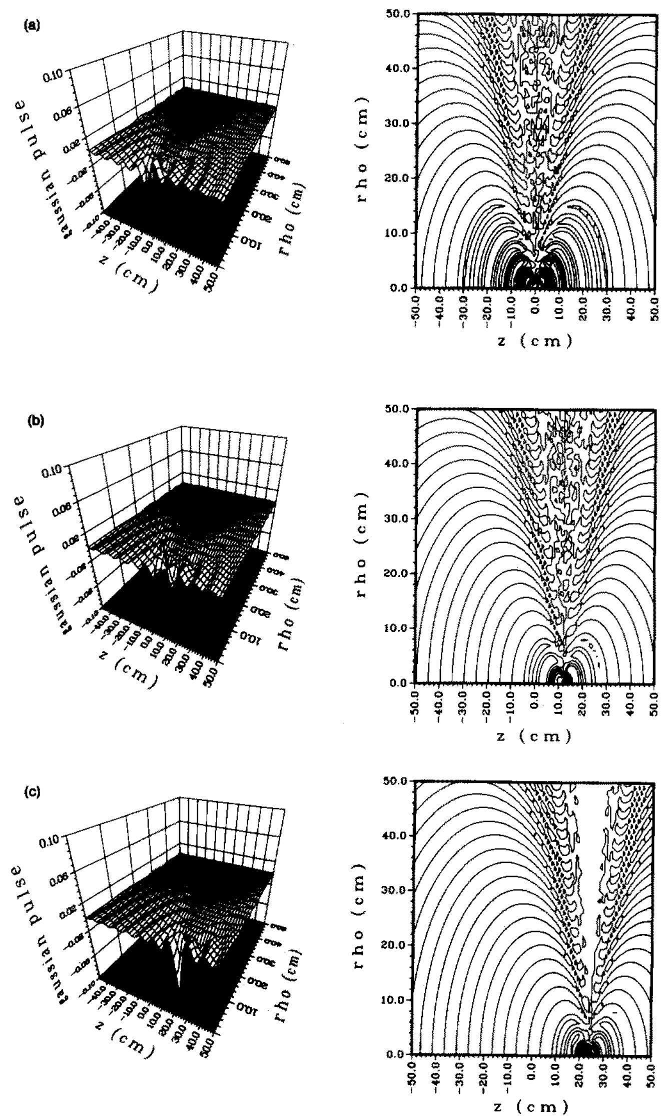


FIG. 1. A time sequence of the fundamental solution (7) for $f = kc/2\pi = 3.0\times 10^9$ and $z_0 = 1.0$ demonstrates that it is a modulated moving Gaussian pulse: (a) $t = 0.0$, (b) $t = 4.0\times 10^{-10}$, (c) $t = 8.0\times 10^{-10}$.

 0022-2488/85/040861-03$02.50

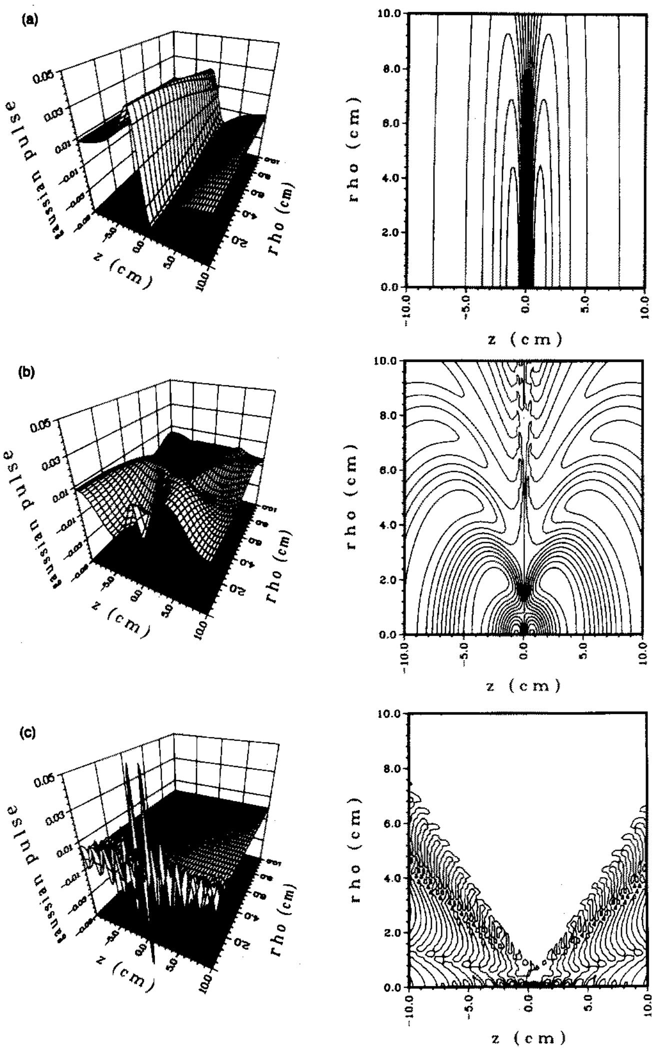


FIG. 2. As k increases, the Gaussian pulse profile becomes more concentrated along the z axis than along the ρ axis: (a) $f = 3.0\times 10^7$, (b) $f = 3.0\times 10^9$, (c) $f = 3.0\times 10^{11}$.

tion. It is the zeroth-order mode in a sequence of multipoles that can be generated in a cylindrical (rectangular) geometry by applying Laguerre (Hermite) polynomial operators to the fundamental Gaussian mode (7). Dr. Brittingham has brought to my attention that Bélanger[2] also recognized this point. However, contrary to the original article[1] and to Ref. 2, Fig. 1 demonstrates that the solution is neither focused nor packetlike nor a boost solution (translationally invariant). Moreover, recognizing that these pulses originate from complex source locations connects these results to a large body of literature. In particular, the concept that a Gaussian beam is *equivalent paraxially* to a spherical wave with a center at a (stationary) complex location was introduced by Deschamps[3] and was later used extensively by Felsen (for example, see Ref. 4) to model the propagation and scattering of Gaussian beams. In contrast to those beam descriptions, (7) is an *exact* solution of (1). On the other hand, the fundamental Gaussian pulse satisfies all of the properties associated with Gaussian beams. For example, its propagation through an optical system will be described by the *ABCD* transformation law (see Ref. 5, Sec. 6.7).

The approach that led from (1) to (7) can also be used to define Gaussian pulse solutions of related equations of import. Consider the Klein–Gordon equation

$$\Box\Psi - \mu^2\Psi = 0, \tag{8}$$

where $\mu = mc/\hbar$. It has the exact axially symmetric solution

$$\Psi(r,t) = \exp(-i\mu^2\tau/4k)\Phi(r,t) = \exp(i\mu^2 ct\,2k)e^{ik_{\rm eff}\sigma}F_k, \tag{9}$$

where the effective modulation frequency

$$\omega_{\rm eff} = k_{\rm eff}c = (k - \mu^2/4k)c = [1 - (mc/2\hbar k)^2]kc \tag{10}$$

has been modified by the mass of the particle. Note that this modulation frequency disappears when $\hbar k = mc/2$ leaving only the time harmonic portion $\exp(i\mu^2 ct/2k) = \exp[i(mc^2)t/\hbar]$. Similarly, the wave equation in a transverse quadratic medium

$$\Box\Psi - (\epsilon_0 + \epsilon_x x^2 + \epsilon_y y^2)\Psi = 0 \tag{11a}$$

reduces to a harmonic oscillator Schrödinger equation

$$4ik\,\partial_\tau F_k = -\Delta_\perp F_k + (\epsilon_0 + \epsilon_x x^2 + \epsilon_y y^2)F_k. \tag{11b}$$

Restricting the problem to one transverse spatial dimension, one can obtain an exact solution from path integral literature (e.g., see Ref. 6, Chap. 6) that is readily converted to one originating from a complex source location. A modified Gaussian pulse is obtained. In fact, when the transverse medium coefficients are small, the results reduce to those discussed above. One should then be able to modify standard quantum electronic results (e.g., see Ref. 5, Chap. 6) to apply directly to these exact complex center pulse solutions.

A strong objection to the results in Ref. 1 has been raised essentially because the solution (7) has infinite energy. This is not a drawback *per se*. Plane wave solutions of (1) also share the infinite energy property and are commonly employed in constructing physical signals. The Gaussian pulse solutions offer a new set of modes that can be used to construct finite energy solutions of (1). In particular, the function

$$\begin{aligned} f(\mathbf{r},t) \equiv h(\rho,\tau,\sigma) &= \int_0^\infty dk\, F(k)\,\phi_k(\rho,\tau,\sigma) \\ &= [4\pi i(z_0 + i\tau)]^{-1}\int_0^\infty dk\, F(k)e^{-ks(\rho,\tau,\sigma)}, \end{aligned} \tag{12}$$

where $s(\rho,\tau,\sigma) = -i\sigma + \rho^2/(z_0 + i\tau)$ satisfies (1) in real space-time. The wave number has been restricted to nonnegative values in this expansion (as well as assuming that $z_0 > 0$) to guarantee the finiteness of the kernel ϕ_k. The resulting Laplace transform expression yields a rich class of possible solutions. An inversion formula corresponding to the Gaussian pulse expansion (12) is

$$F(k) = \int_{-\infty}^{\infty} d\sigma \int_{-\infty}^{\infty} d\tau \int_0^\infty \rho\, d\rho\, \psi_k(\rho,\tau,\sigma)\, h(\rho,\tau,\sigma), \tag{13}$$

where the kernel is

$$\psi_k(\rho,\tau,\sigma) = 8\pi^{1/2}e^{-(\tau/4kz_0)^2}\phi_k^*(\rho,\tau,\sigma), \tag{14}$$

ϕ_k^* being the complex conjugate of ϕ_k. Equivalently, the completeness relation

$$\int_{-\infty}^{\infty} d\sigma \int_{-\infty}^{\infty} d\tau \int_0^\infty \rho\, d\rho\, \psi_k(\rho,\tau,\sigma)\,\phi_{k'}^*(\rho,\tau,\sigma) = \delta(k - k') \tag{15}$$

is satisfied by ψ_k and ϕ_k. The density $\exp[-(\tau/4kz_0)^2]d\tau$ represents a Gaussian measure over τ with real variance $8(kz_0)^2$, kz_0 being the source phase distance, which guarantees the finiteness of the τ integration.

Consider, as an example, the spectrum $F(k) = \exp(-ak)$. Equation (12) gives the pulse

$$f(\mathbf{r},t) = [4\pi i(z_0 + i\tau)]^{-1}\,[s(\rho,\tau,\sigma) + a]^{-1}. \tag{16}$$

Setting $f_+(\mathbf{r},t) = f(\mathbf{r},t)$, $f_-(\mathbf{r},t) = f_+(\mathbf{r},-t)$, and $a = z_0 = \gamma$, the composite pulse

$$\Psi(\mathbf{r},t) = f_+(\mathbf{r},t) - f_-(\mathbf{r},t) = \frac{1}{2\pi}\,\frac{\gamma(ct)}{\{[\rho^2 + (z-ct)(z+ct) + \gamma^2]^2 + 4\gamma^2(ct)^2\}} \tag{17}$$

is a real, exact solution of (1). A time sequence of a pulse (17) with $\gamma = 1.0$ is given in Fig. 3. The pulse has zero amplitude at $t = 0$ and its maxima occur at $\rho = z = 0$, for $0 < t \ll \gamma/c$ and lie on the sphere $\rho^2 + z^2 = R^2 = (ct)^2 - \gamma^2$, for $t > \gamma/c$. Its amplitude on that sphere $[8\pi\gamma(ct)]^{-1}$ decreases essentially as R^{-1} for $ct \gg \gamma$. The likeness of these figures to those describing a pebble dropped in a pond precipitated the name "splash pulse." As the figures illustrate, the support of the splash pulse is localized in space and separates space into two regions of null field for $t \gg \gamma/c$, the pulse layer being relatively thin. The apparent spikes in the surface subplot in Fig. 3(c) again are due to the coarseness of the computational grid employed in the graphics routine.

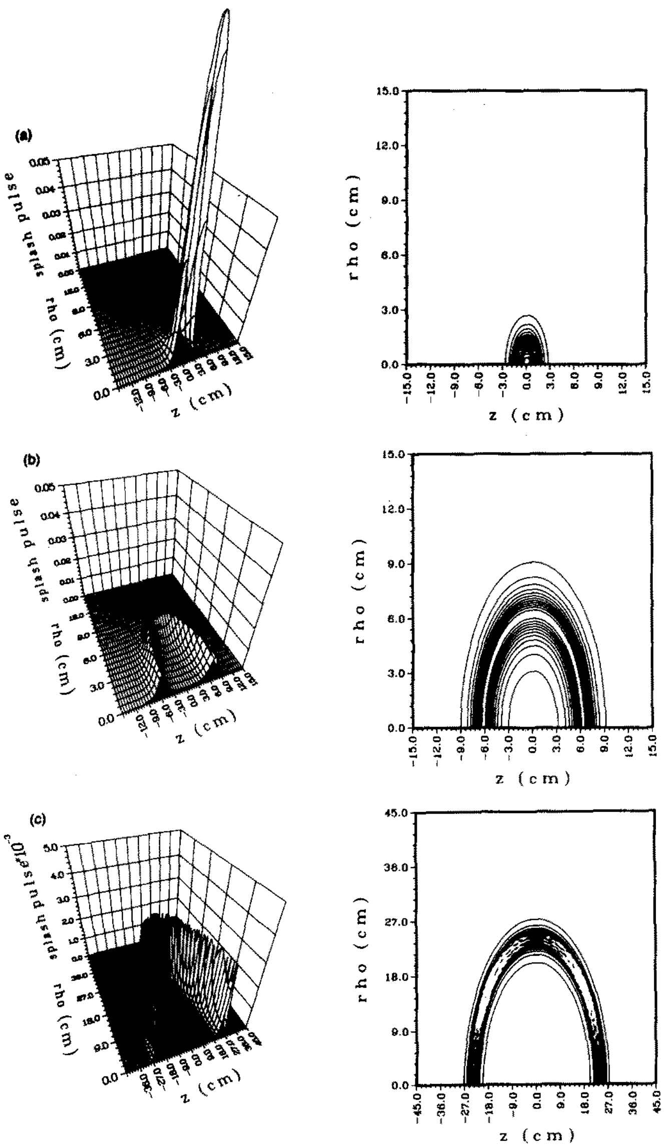


FIG. 3. Time sequence of the splash pulse (17) with $\gamma = 1.0$: (a) $t = 8.0\times 10^{-11}$, (b) $t = 2.1\times 10^{-10}$, (c) $t = 8.0\times 10^{-10}$.

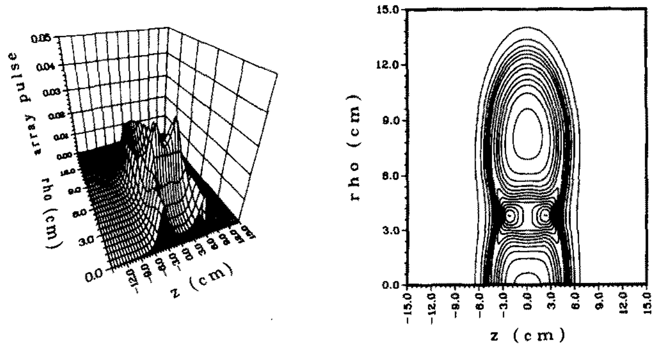


FIG. 4. The interaction of two splash pulses ($\gamma = 1.0$) confirms the linearity of the problem. The splash centers are $(\rho,z) = (0,0)$ and $(\rho,z) = (7.5,0)$.

The interaction of two splash pulses is depicted in Fig. 4. The apparent splash centers are the point $(\rho,z) = (0,0)$ and the ring $(\rho,z) = (7.5,0)$. The linear nature of the problem is reflected in the simple superposition in the overlap region and the decoupled propagation of each splash pulse.

III. CONCLUSION

Several issues remain outstanding and are currently under investigation. Foremost is the possible launchability of pulses derived from the fundamental Gaussian pulses. The physical connection between resonating structures and Gaussian beams (stationary complex center descriptions) leads one to speculate that such pulses may be associated with some special type of resonator cavity. In addition to the indicated k-superposition/transform pair, another class of solutions, those constructed by superposition of the complex source location z_0, may lead to other physically interesting pulses. Finally, with (2), nonlinear wave equations reduce immediately to the corresponding nonlinear Schrödinger equations. For instance, the cubic wave equation

$$\Box\Phi - \alpha|\Phi|^2\Phi = 0 \tag{18}$$

reduces to the cubic Schrödinger equation

$$4ik\,\partial_\tau F_k = -\Delta_\perp F_k + \alpha|F_k|^2 F_k. \tag{19}$$

At least for one transverse dimension, (19) has known soliton solutions (see Ref. 7, Sec. 5.3). Extensions of these solitons to ones associated with complex source locations may yield other physically interesting wave equation solutions.

ACKNOWLEDGMENTS

The author would like to thank Dr. James N. Brittingham for several interesting discussions on his FWM results.

This work was performed by the Lawrence Livermore National Laboratory under the auspices of the U.S. Department of Energy under Contract No.W-7405-ENG-48.

[1]J. N. Brittingham, J. Appl. Phys. **54**, 1179 (1983).
[2]P. A. Bélanger, J. Opt. Soc. Am. A **1**, 723 (1984).
[3]G. A. Deschamps, Electron. Lett. **7**, 23 (1971).
[4]L. B. Felsen, Symp. Math. **18**, 39 (1976).
[5]A. Yariv, *Quantum Electronics* (Wiley, New York, 1975), 2nd ed.
[6]L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
[7]G. L. Lamb, Jr., *Elements of Soliton Theory* (Wiley, New York, 1980).

Torsion of certain beams

Evelyn E. Obaid
Department of Computer Studies, North Carolina State University, Raleigh, North Carolina 27695
S. A. Obaid
Department of Mathematics, Pennsylvania State University, York, Pennsylvania 17403

(Received 9 April 1984; accepted for publication 10 September 1984)

Complex variable methods are introduced to derive exact and closed expressions for the stress functions, torsional rigidities, and peripheral shearing stresses of certain isotropic cylinders under torsion. Numerical results in some special cases are presented in tabular and graphic form. The equations for the boundaries of the cross sections for these cylinders have the polar forms $r^2 = a^2(2\cos 2\theta - 1)$, $a>0$ ($|\theta|\leqslant\pi/6$), $r^4 = 2na^4/(n-1+\cos 4\theta)(n+1+\cos 4\theta)$, $a>0$ ($|\theta|\leqslant\pi$, $2<n<\infty$), and $r^{m-2} = a^{m-2}(\cos^2\theta - \cos^2\delta)/\sin^2\delta\cos m\theta$, $a>0$ ($m>2$, $0<\delta<\pi/2m$, $|\theta|\leqslant\delta$).

I. INTRODUCTION

The classical Saint–Venant torsion problem has long been a favorite of engineers and applied mathematicians. It has been solved for many boundary forms by using various methods. References concerning this subject are to be found in textbooks by Muskhelishvili,[1] Sokolnikoff,[2] and Timoshenko and Goodier.[3] Comprehensive reviews of exact and approximate methods for solving the problem have been given by Higgins.[4,5]

Complex variable methods have been applied successfully to obtain the solutions for the problem corresponding to various curves, most notably the ellipse, cardioid, lemniscate of Bernoulli, elliptic limacon, lemniscate of Booth, semicircle, and semicardioid. These solutions and others can be found in the papers by Abbassi,[6] Bassali,[7] Deutsch,[8] Ghosh,[9] Milne-Thomson,[10] Morris,[11] Bassali and Obaid,[12,13] and Stevenson.[14]

We shall use a generalization of Stevenson's method (see Ref. 14) to solve the torsion problem for isotropic cylinders with cross sections bounded by curvilinear edges whose polar equations are

$$r^2 = a^2(2\cos 2\theta - 1), \quad a>0 \quad (|\theta|\leqslant\pi/6), \tag{1.1}$$

$$r^4 = 2na^4/(n-1+\cos 4\theta)(n+1+\cos 4\theta), \quad a>0 \quad (|\theta|\leqslant\pi,\ 2<n<\infty), \tag{1.2}$$

$$r^{m-2} = a^{m-2}(\cos^2\theta - \cos^2\delta)/\sin^2\delta\cos m\theta, \quad a>0 \quad (m>2,\ 0<\delta\leqslant\pi/2m,\ |\theta|\leqslant\delta). \tag{1.3}$$

We denote these boundaries by Γ, Γ_n, and Γ_m^δ, respectively. Figures are sketched showing the shapes of the cross sections corresponding to various values of the parameters n, m, and δ. The equation of the boundary Γ is similar to that of the lemniscate of Bernoulli, with the exception of the additional constant a^2. Solving the torsion problem for the cross section bounded by Γ determines the influence of the additional constant a^2 on the solution. Each member of the family of curves Γ_n has four axes of symmetry, namely, $\theta = 0, \pi/4, \pi/2$, and $3\pi/4$. For large n, this family is approximately a family of concentric circles. The two-parameter family Γ_m^δ is symmetric with respect to the initial line; the two tangent lines to the curve Γ_m^β at the origin contain an angle 2β. Moreover, each member of the family $\Gamma_m^{\pi/2m}$ is a sector formed by the two straight lines $\theta = \pm\pi/2m$ and a certain curve. For example, the cross section bounded by $\Gamma_3^{\pi/6}$ is an equilateral triangle, and the cross section bounded by $\Gamma_4^{\pi/8}$ is a sector with a hyperbolic base. Closed and exact expressions are established for the torsional rigidities and shearing stresses. The variation of certain torsional rigidities with the parameters involved are illustrated in tabular form. The distribution of shearing stress on the boundary is investigated in certain cases, and graphs illustrating its variation are sketched.

II. FUNDAMENTAL EQUATIONS

In the torsion problem for an elastic isotropic cylinder of uniform simply connected cross section S bounded by a simple closed curve C in the z plane ($z = x+iy$), we assume, with the usual notation, that u, v, and w are the displacements, where w is parallel to the Z axis and generators of the cylinders. It is known that

$$u + iv = i\tau zZ, \quad w = \tau\phi(x,y) = \tau[\Omega(z) + \bar{\Omega}(\bar{z})]/2, \tag{2.1}$$

where τ is the constant angle of twist per unit length of the cylinder, $\Omega(z)$ is the complex torsion function which is analytic throughout the region S, and $\bar{\Omega}(\bar{z})$ is the complex conjugate of the function $\Omega(z)$. It is also known that the imaginary part of $\Omega(z)$, $\psi(x,y)$, satisfies the boundary condition

$$\psi(x,y) = \tfrac{1}{2}(x^2+y^2) \quad \text{on } C. \tag{2.2}$$

In order to solve the torsion problem for a particular bar, it is necessary to find the function $\psi(x,y)$ or, equivalently, the stress function

$$\Psi(x,y) = \psi(x,y) - \tfrac{1}{2}(x^2+y^2), \tag{2.3}$$

which satisfies Poisson's equation

$$\nabla^2\Psi = -2, \tag{2.4}$$

throughout the region S, and the simple boundary condition

$$\Psi(x,y) = 0 \quad \text{on } C. \tag{2.5}$$

In polar coordinates (r,θ) the stresses are given by

$$\widehat{rZ} = \frac{\mu\tau}{r}\frac{\partial\Psi}{\partial\theta}, \quad \widehat{\theta Z} = -\mu\tau\frac{\partial\Psi}{\partial r}, \tag{2.6}$$

where μ is the rigidity of the material for the cylinder.

 0022-2488/85/040864-06$02.50

The solution of the torsion problem is not complete until the torsional rigidity D has been determined (see Ref. 14). The torsional rigidity is given by

$$D = 2\mu \iint_S \Psi(r,\theta)\, r\, dr\, d\theta \tag{2.7}$$

or

$$D = \mu(I + J), \tag{2.8}$$

where

$$I = \iint_S r^2\, dS, \tag{2.9}$$

$$J = \mathrm{Re} \int_C \frac{1}{2} r^2 \frac{d\Omega}{dz}\, dz, \tag{2.10}$$

and Re denotes the real part.

We now consider the Cartesian equation for the closed boundary C of the cross section of the cylinder, in the general form

$$\mathrm{Re}\, F(z) - (\alpha x^2 + \beta y^2 + \gamma) = 0, \tag{2.11}$$

where $F(z)$ is an arbitrary function of z, which is analytic in the domain S enclosed by C; and α, β, γ are real constants subject to the condition $\alpha + \beta \neq 0$. Let us assume the stress function $\Psi(x,y)$ has the form

$$\Psi(x,y) = A\, [\mathrm{Re}\, F(z) - (\alpha x^2 + \beta y^2 + \gamma)],$$

where A is a constant to be determined by Poisson's equation (2.4). It is obvious that Ψ satisfies the boundary condition (2.5) and that $A = 1/(\alpha + \beta)$. Thus the complex torsion function $\Omega(z)$ and the stress function $\Psi(x,y)$ are given, respectively, by the following fomulas:

$$\Omega(z) = [i/(\alpha + \beta)]\,[F(z) + \tfrac{1}{2}(\beta - \alpha) z^2 - \gamma], \tag{2.12}$$

$$\Psi(x,y) = [1/(\alpha + \beta)]\,[\mathrm{Re}\, F(z) - (\alpha x^2 + \beta y^2 + \gamma)]. \tag{2.13}$$

The value of $\Psi(x,y)$ is obviously proportional to the expression in Eq. (2.11) for the boundary C of the cross section. This result clearly generalizes those of Leibenson,[15] who gave only some examples satisfying this property, and generalizes the results of Stevenson (see Ref. 14).

III. CROSS SECTION BOUNDED BY A LOOP OF A HIPPOPEDE

Taking

$$F(z) = 2az^2/(z + a), \quad \alpha = \beta = 1, \quad \gamma = 0, \tag{3.1}$$

in Eq. (2.11) yields the closed curve with polar equation

$$r^2 = a^2(2 \cos 2\theta - 1). \tag{3.2}$$

Imposing the condition $|\theta| < \pi/6$ in order to obtain one loop, Eq. (3.2) takes the following form:

$$r^2 = a^2 \cos 3\theta / \cos \theta \quad (|\theta| < \pi/6). \tag{3.3}$$

Thus the cross section, illustrated in Fig. 1, is bounded by a loop Γ. The Cartesian equation for the quartic curve Γ is given by

$$x^4 + y^4 + 2x^2y^2 + a^2(3y^2 - 2x^2) = 0 \quad (x > 0). \tag{3.4}$$

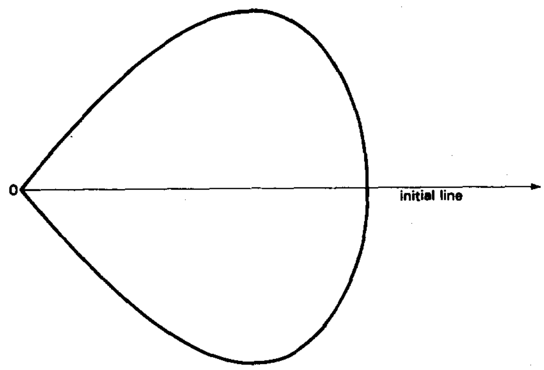


FIG. 1. Cross section bounded by the curve Γ.

Substituting the values of $F(z)$, α, β, and γ from Eq. (3.1) in Eqs. (2.12) and (2.13), we obtain, respectively, the complex torsion function and the stress function corresponding to the curve Γ

$$\Omega(z) = iaz^2/(z + a), \tag{3.5}$$

$$\Psi(r,\theta) = \frac{r^2}{2}\left(\frac{2a^2 \cos 2\theta - a^2 - r^2}{r^2 + a^2 + 2ar\cos\theta}\right). \tag{3.6}$$

Obviously, $\Psi(r,\theta)$ vanishes on Γ.

Using Eqs. (2.9) and (3.2) we find

$$I = a^4(2\pi - 3\sqrt{3})/8 = 0.1359 a^4. \tag{3.7}$$

After some calculations, the complex torsion function (3.5) together with Eq. (2.10) yield

$$J = (a^4/2)[L + 12I_0 - 174I_2 + 704I_4 - 1248I_6 + 768I_8 - 16K_0 + 152K_2 - 384K_4 + 384K_6], \tag{3.8}$$

where

$$L = \int_0^{\pi/6} \left[\frac{1 - \sqrt{(1 - s^2)(1 - 4s^2)}}{s^2}\right] d\theta = \pi - \sqrt{3} \quad (s = \sin\theta), \tag{3.9}$$

$$K_{2l} = \frac{1}{2^{2l+1}} \int_0^1 u^{2l} \sqrt{1 - u^2}\, du = 2l!\pi/(l + 1)2^{4l+3}(l!)^2, \tag{3.10}$$

$$I_{2l} = \int_0^{\pi/6} \sin^{2l}\theta\, d\theta, \tag{3.11}$$

$$I_0 = \pi/6, \quad I_2 = (2\pi - 3\sqrt{3})/24,$$
$$I_4 = (4\pi - 7\sqrt{3})/64, \quad I_6 = (5\pi - 9\sqrt{3})/96, \tag{3.12}$$
$$I_8 = (280\pi - 507\sqrt{3})/6144.$$

The results in Eqs. (3.9), (3.10), and (3.12) provide the required value for J

$$J = a^4(351\pi - 672\sqrt{3})/512 = -0.1196 a^4. \tag{3.13}$$

The torsional rigidity D for the beam is given by

$$D = \mu a^4(479\pi - 864\sqrt{3})/512 = 0.0163 \mu a^4. \tag{3.14}$$

The torsional rigidity D_L, for a loop of the lemniscate of Bernoulli

$$r^2 = 2a^2 \cos 2\theta,$$

is given by (see Ref. 14, p. 210)

$$D_L = \mu a^4(\pi^2 - 8)/4\pi = 0.1488\,\mu a^4.$$

Comparing the above result with (3.14) we obtain

$$D_L/D = 9.1275.$$

We now examine the distribution of peripheral shear for the cross section bounded by Γ. Substituting from Eq. (3.6) in Eq. (2.6) yields, after some algebraic manipulation, the following expressions for the boundary values of the shearing stresses:

$$\widehat{rZ} = \mu\tau r^3/(r^2 + a^2 + 2ar\cos\theta), \tag{3.15}$$

$$\widehat{\theta Z} = -2\mu\tau a^2 r\sin\theta/(r^2 + a^2 + 2ar\cos\theta), \tag{3.16}$$

$$\sigma(r,\theta) = \sqrt{(\widehat{rZ})^2 + (\widehat{\theta Z})^2} = \mu\tau a^2 r\sqrt{1 + 8\sin^2\theta}/(r^2 + a^2 + 2ar\cos\theta). \tag{3.17}$$

The resultant shearing stress at any point P of Γ is along the tangent to Γ at P. At $(a,0)$ we have $\sigma(a,0) = \mu\tau a/4$, which is the maximum value for $\sigma(r,\theta)$. Also, at $(0,\pi/6)$ we have $\sigma(0,\pi/6) = 0$, which is the minimum value for $\sigma(r,\theta)$.

IV. CROSS SECTION BOUNDED BY MEMBERS OF THE FAMILY Γ_n

Taking

$$F(z) = \sqrt{a^4 - z^4}, \quad \alpha = \beta = \sqrt{n/2}, \quad \gamma = 0, \tag{4.1}$$

in Eq. (2.11), and applying the following result, which is valid for real functions u, v:

$$\mathrm{Re}\sqrt{u + iv} = [\tfrac{1}{2}(u + \sqrt{u^2 + v^2})]^{1/2}, \tag{4.2}$$

we obtain the family of closed curves Γ_n, with polar equation

$$r^4 = 2na^4/(n - 1 + \cos 4\theta)(n + 1 + \cos 4\theta) \quad (2 < n < \infty). \tag{4.3}$$

The condition $2 < n < \infty$ is imposed in order to obtain closed curves. Each member of the above family has four axes of symmetry. Examining Fig. 2, we notice that, for values of n close to 6, the closed curves are approximately squares with curvilinear corners; and for large values of n, the curves Γ_n are approximately circles. We will further investigate the two limiting cases later in this section.

Substituting the values of $F(z)$, α, β, and γ from Eq. (4.1)

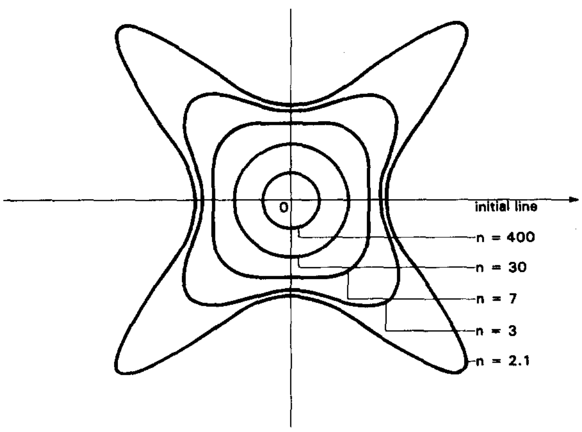


FIG. 2. Cross sections bounded by the family of curves Γ_n.

into Eqs. (2.12) and (2.13), we obtain the complex torsion function and the stress function, respectively,

$$\Omega(z) = i\sqrt{a^4 - z^4}/\sqrt{2n}, \tag{4.4}$$

$$\Psi(r,\theta) = (\mathrm{Re}\sqrt{a^4 - z^4} - r^2\sqrt{n/2})/\sqrt{2n} = \big(\sqrt{a^4 - r^4\cos 4\theta + (a^8 + r^8 + 2a^4r^4\cos 4\theta)^{1/2}} - \sqrt{n}\,r^2\big)/2\sqrt{n}. \tag{4.5}$$

We now evaluate the torsional rigidity D_n for the beam with cross section bounded by Γ_n. The polar moment of inertia I_n for the cross section takes the form

$$I_n = \iint_S r^2\,dS = na^4\int_0^\pi \frac{d\phi}{(n - 1 + \cos\phi)(n + 1 + \cos\phi)} = \frac{na^4}{2}\int_0^\pi\left(\frac{1}{n - 1 + \cos\phi} - \frac{1}{n + 1 + \cos\phi}\right)d\phi$$
$$= (na^4\pi/2)([\sqrt{n(n-2)}]^{-1/2} - [\sqrt{n(n+2)}]^{-1/2}). \tag{4.6}$$

Applying formula (2.10), the line integral J_n becomes

$$J_n = -\frac{1}{\sqrt{2n}}\,\mathrm{Re}\int_C \frac{ir^2z^3}{(a^4 - z^4)^{1/2}}\,dz. \tag{4.7}$$

The boundary values of dz and $\sqrt{a^4 - z^4}$ on Γ_n can be written in the form

$$dz = \frac{r_n e^{i\theta}}{na^4}\,[r_n^4\sin 4\theta(n + \cos 4\theta) + ina^4]\,d\theta, \tag{4.8}$$

$$\sqrt{a^4 - z^4} = r_n^2(n - i\sin 4\theta)/\sqrt{2n}, \tag{4.9}$$

where r_n is the boundary value of r on Γ_n. After some calculations and application of Eqs. (4.8) and (4.9), the integral J_n reduces to

$$J_n = 4na^4\int_0^\pi \frac{\cos^2\phi + n\cos\phi + 1}{(n - 1 + \cos\phi)^2(n + 1 + \cos\phi)^2}\,d\phi = na^4\int_0^\pi\left[\frac{n+2}{(n + 1 + \cos\phi)^2} - \frac{n-2}{(n - 1 + \cos\phi)^2}\right]d\phi.$$

In view of the formula (see Ref. 16)

$$\int\frac{dx}{(c + b\cos x)^2} = \frac{1}{(c^2 - b^2)}\left[\frac{2c}{(c^2 - b^2)^{1/2}}\tan^{-1}\right.$$
$$\left.\times\left(\frac{(c^2 - b^2)^{1/2}\tan(x/2)}{c + b}\right) - \frac{b\sin x}{c + b}\right], \quad c^2 > b^2,$$

we obtain the following value for J_n

$$J_n = a^4\pi\left[\frac{n+1}{[n(n+2)]^{1/2}} - \frac{n-1}{[n(n-2)]^{1/2}}\right]. \tag{4.10}$$

The exact and closed expression for the torsional rigidity D_n is given by

$$D_n = \mu(I_n + J_n) = \mu a^4\pi(\sqrt{n+2} - \sqrt{n-2})/2\sqrt{n}. \tag{4.11}$$

One can easily show that D_n is a strictly decreasing function of n.

It is desirable to have the points of intersection of the family Γ_n with the initial line coincide. To do so, we replace a by $\sqrt[4]{(n+2)/2}\,b$ in Eq. (4.3) thus obtaining

$$D_n = \mu b^4 \pi (n+2)(\sqrt{n+2} - \sqrt{n-2})/4\sqrt{n}, \quad (4.12)$$

as the torsional rigidity of the beam corresponding to the cross section bounded by

$$r^4 = n(n+2)b^4/(n-1+\cos 4\theta)(n+1+\cos\theta). \quad (4.13)$$

Setting $n=6$ in Eq. (4.12) yields

$$D_6 = 2\sqrt{6}(\sqrt{2}-1)\pi\mu b^4/3 = 2.1250\,\mu b^4.$$

Comparing this with the value

$$D_s = \left[\frac{16}{3} - \frac{2^{10}}{\pi^5}\left(\frac{e^\pi - 1}{e^\pi + 1}\right)\right]\mu b^4 = 2.2644\mu b^4,$$

for the torsional rigidity of a beam whose cross section is a square with side $2b$ (see Ref. 2, p. 132), we find $D_6/D_s = 0.9384$.

We now consider the family of closed curves (4.13) for large values of n. In this case, $\cos 4\theta \ll n$, and the number $n(n+2)$ is of the same order as $(n-1)(n+1)$. Thus the family reduces to the circle $r=b$. The expression for the torsional rigidity (4.12) can be rewritten in the form

$$D_n = \frac{\mu b^4 \pi}{4}(n+2)(\sqrt{1+2/n} - \sqrt{1-2/n})$$
$$= \frac{\mu b^4 \pi}{4}(n+2)[(1+1/n+O(1/n^2)) - (1-1/n+O(1/n^2))],$$

or

$$D_n \cong \mu b^4 \pi (n+2)/2n.$$

Thus for large n, D_n is approximately $\mu b^4\pi/2$, which equals the torsional rigidity of a circular cylinder.

Substituting from Eq. (4.5) into Eq. (2.6) yields, after some algebraic manipulation, the following expressions for the boundary values of the shearing stresses:

$$rZ = 2\mu\tau r \sin 4\theta\,(n+\cos 4\theta)/(n^2+\sin^2 4\theta), \quad (4.14)$$

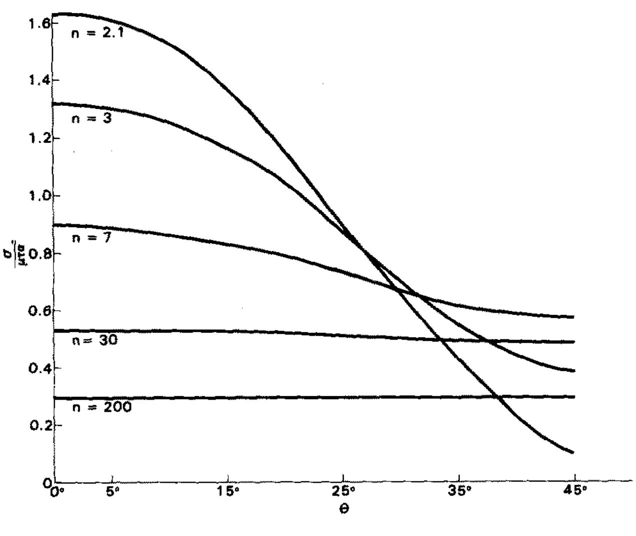


FIG. 3. Variation of peripheral shearing stress along the boundary of Γ_n.

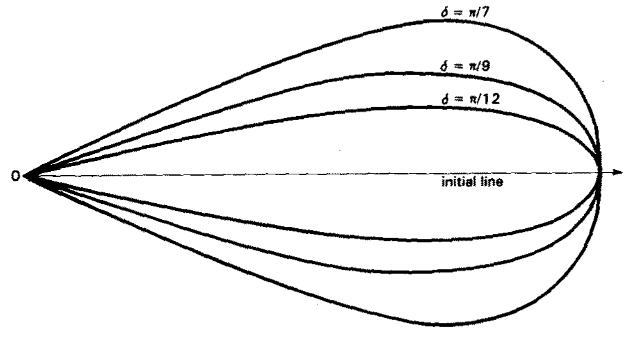


FIG. 4. Cross section bounded by the family of curves Γ_3^δ.

$$\theta Z = \mu\tau r(n-1+\cos 4\theta)(n+1+\cos 4\theta)/(n^2+\sin^2 4\theta). \quad (4.15)$$

The variation of peripheral shearing stress along the boundaries in Fig. 2 is shown in Fig. 3. We remark that, for large n, $\sigma(r,\theta) = \sqrt{(rZ)^2 + (\theta Z)^2}$ obviously reduces to a constant, as expected.

Finally, we notice that if $F(z)$ is replaced by $\sqrt{a^4+z^4}$ with $\alpha = \beta = \sqrt{n/2}$ and $\gamma = 0$, we obtain a family of curves having the same shapes as Γ_n, rotated by an angle of $\pi/4$.

V. CROSS SECTIONS BOUNDED BY THE FAMILY OF CURVES Γ_m^δ

Setting

$$F(z) = z^m, \quad \alpha = a^{m-2}, \quad \beta = -a^{m-2}\cot^2\delta, \quad \gamma = 0 \quad (5.1)$$

in Eq. (2.11) we obtain

$$r^{m-2} = a^{m-2}(\cos^2\theta - \cos^2\delta)/\sin^2\delta\cos m\theta. \quad (5.2)$$

Imposing the conditions $|\theta| \leqslant \delta$, $m>2$, and $0<\delta\leqslant\pi/2m$, Eq. (5.2) reduces to the family of closed curves Γ_m^δ. Each curve is symmetric with respect to the initial line $\theta = 0$. We also notice that the curve Γ_3^δ is a cubic curve. Figure 4 illustrates the family Γ_3^δ, for $\delta = \pi/12, \pi/9, \pi/7$.

Setting $\delta = \pi/2m$ in Eq. (5.3) we obtain

$$\left[\cos 2\theta - \cos\left(\frac{\pi}{m}\right)\right]\left[\frac{r^{m-2}\cos m\theta}{\cos 2\theta - \cos(\pi/m)} - \frac{a^{m-2}}{2\sin^2(\pi/2m)}\right] = 0. \quad (5.3)$$

The factor $[\cos 2\theta - \cos(\pi/m)]$ results from writing $\cos m\theta$, according to m even or odd, respectively, as follows (see Jolley[17]):

TABLE I. Values of $D_m^\delta \times 10^6/\mu a^4$.

m	5°	10°	15°	20°	25°
2.1	2.266 90	19.2561	72.2089	200.963	496.040
2.3	10.794 7	91.8896	345.960	969.474	2 421.95
2.5	21.394 8	182.506	689.929	1947.41	4 931.23
2.7	32.793 8	280.336	1064.17	3027.02	7 785.01
2.9	44.418 0	380.512	1450.65	4160.70	10 899.3
3.1	55.964 5	480.454	1839.78	5324.66	14 268.7
3.3	67.265 6	578.724	2226.28	6507.71	17 967.3
3.5	78.228 7	674.518	2607.24	7706.70	22 244.9
3.7	88.804 9	767.407	2981.19	8924.84	26 430.2

$$\cos 2n\theta = 2^{n-1} \prod_{\nu=1,3,\ldots}^{2n-1} [\cos 2\theta - \cos(\nu\pi/2n)]$$

$$(n = 2,3,\ldots), \tag{5.4}$$

$$\cos(2n+1)\theta = 2^n \cos\theta \prod_{\nu=1,3,\ldots}^{2n-1} [\cos 2\theta - \cos(\nu\pi/2n+1)]$$

$$(n = 1,2,\ldots). \tag{5.5}$$

Equation (5.3) represents a sector $\Gamma_m^{\pi/2m}$, bounded by the two straight lines $\theta = \pm\pi/2m$, with base given by Eq. (5.2) itself. For $m = 3$, Eq. (5.3) simplifies to the equation of an equilateral triangle with side $2a/\sqrt{3}$.

The stress function $\Psi(r,\theta)$ corresponding to the curve (5.2) is given by

$$\Psi(r,\theta) = [-2r^m \cos m\theta \sin^2\delta + a^{m-2}r^2 \cos 2\theta - a^{m-2}r^2 \cos 2\delta\,]/2a^{m-2}\cos 2\delta, \tag{5.6}$$

and the torsional rigidity D_m^δ takes the closed form

$$D_m^\delta = K_m^\delta \int_0^\delta (\cos^2\theta - \cos^2\delta)^{(m+2)/(m-2)} \times (\cos m\theta)^{4/(2-m)}\, d\theta, \tag{5.7}$$

where

$$K_m^\delta = \mu a^4 \sec 2\delta\, (\sin^2\delta)^{4/(2-m)}(m-2)/(m+2). \tag{5.8}$$

Evaluation of the integral in Eq. (5.7) is, in general, relatively difficult and requires the use of computers for special cases. We shall find the exact expression for D_3^δ, where

$$D_3^\delta = \left(\frac{K_3^\delta}{256}\right) \int_0^\delta \frac{(\cos^2\theta - \cos^2\delta)^5}{\cos^4\theta\,(\cos^2\theta - \frac{3}{4})^4}\, d\theta. \tag{5.9}$$

Using partial fractions, integration, and some formulas from Ref. 16, we arrive at

$$D_3^\delta = \frac{K_3^\delta}{32}\left[A_0 \tan\delta + A_1 \tan^3\delta + A_2 \ln\left|\frac{1-\sqrt{3}\tan\delta}{1+\sqrt{3}\tan\delta}\right| + \frac{A_3 \cot\delta\,(1-\cot^2\delta)}{(3-\cot^2\delta)^2} + A_4 \sin 2\delta\right], \tag{5.10}$$

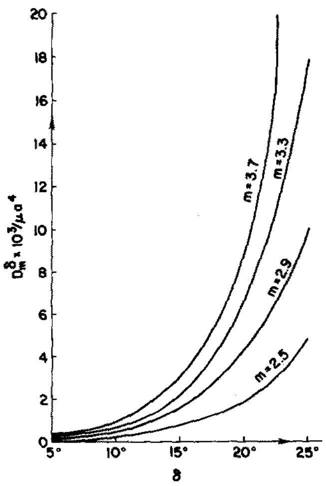


FIG. 5. Variations of the torsional rigidities D_m^δ.

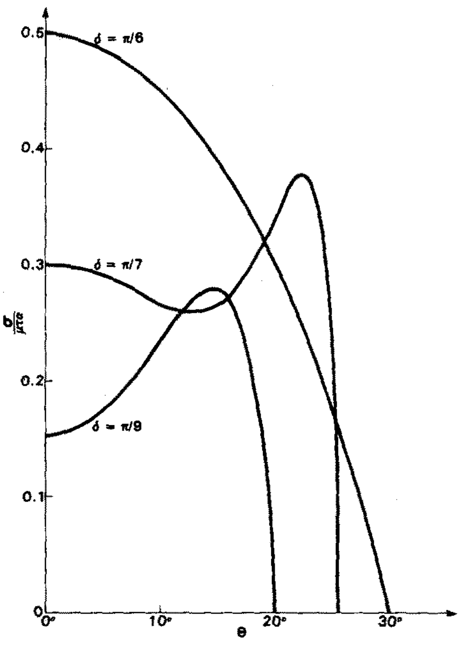


FIG. 6. Distribution of shearing stress on the boundaries of Γ_3^δ.

where

$$A_0 = (32\lambda^8/243)(15 - 19\lambda^2),$$

$$A_1 = -32\lambda^{10}/243,$$

$$A_2 = (\sqrt{3}/729)(4672\lambda^6 - 7824\lambda^4 + 4428\lambda^2 - 27)(\lambda^2 - \tfrac{3}{4})^2,$$

$$A_3 = -(768/243)(8\lambda^2 + 9)(\lambda^2 - \tfrac{3}{4})^4,$$

$$A_4 = (2/243)(304\lambda^4 - 856\lambda^2 + 121)(\lambda^2 - \tfrac{3}{4})^2,$$

with $\lambda = \cos\delta$. For $\delta = \pi/6$, we notice that

$$K_3^{\pi/6} = 512\mu a^4/5, \quad A_0 = \tfrac{1}{32},$$

$$A_1 = -\tfrac{1}{32}, \quad A_2 = A_3 = A_4 = 0.$$

Thus we have

$$D_3^{\pi/6} = \sqrt{3}\mu a^4/45, \tag{5.11}$$

which is in agreement with the value given on p. 125 of Ref. 2 for the torsional rigidity corresponding to the equilateral triangle.

Numerical values for the torsional rigidity D_m^δ given by Eq. (5.7) are presented in Table I, and graphically plotted in Fig. 5.

Substituting from Eq. (5.6) in Eq. (2.6) yields the following expressions for the stresses at any point (r,θ) of the cross section

$$\widehat{rZ} = \mu\tau r[m(r/a)^{m-2} \sin^2\delta \sin m\theta - \sin 2\theta\,]/(\cos 2\delta), \tag{5.12}$$

$$\widehat{\theta Z} = \mu\tau r[m(r/a)^{m-2} \sin^2\delta \cos m\theta + \cos 2\delta - \cos 2\theta\,]/(\cos 2\delta). \tag{5.13}$$

Thus, the resultant shearing stress σ_m^δ at any point (r,θ) of the cross section is given by

$$\sigma_m^\delta = (\mu\tau r/\cos 2\delta)\{1 + (\cos^2 2\delta) - (2\cos 2\delta)\cos(2\theta) - 2m\sin^2\delta\,(r/a)^{m-2}[\cos(m-2)\theta - (\cos 2\delta)\cos m\theta] + m^2\sin^4\delta\,(r/a)^{2m-4}\}^{1/2}. \quad (5.14)$$

On the boundary (5.2), the above expression reduces to

$$\sigma_m^\delta = (\mu\tau r/\cos 2\delta)\{\sin^2 2\theta + [m^2\tan^2 m\theta + (m-2)^2](\cos^2\theta - \cos^2\delta) - 2m\sin 2\theta\tan m\theta\,(\cos^2\theta - \cos^2\delta)\}^{1/2}. \quad (5.15)$$

The distribution of shearing stress on the boundary Γ_3^δ is illustrated in Fig. 6 for $\delta = \pi/9$, $\pi/7$, $\pi/6$.

The stresses for the equilateral triangle can be easily derived by substituting $m = 3$ and $\delta = \pi/6$ in Eqs. (5.12), (5.13), and (5.15).

[1]N. I. Muskhelishvili, *Some Basic Problems of the Mathematical Theory of Elasticity* (Noordhoff, Groningen, Holland, 1954).
[2]I. S. Sokolnikoff, *Mathematical Theory of Elasticity* (McGraw–Hill, New York, 1956).
[3]S. Timoshenko and J. N. Goodier, *Theory of Elasticity* (McGraw–Hill, New York, 1970).
[4]T. J. Higgins, Am. J. Phys. **10**, 248 (1942).
[5]T. J. Higgins, J. Appl. Phys. **14**, 469 (1943).
[6]M. M. Abbassi, Q. J. Appl. Math. **1**, 75 (1956).
[7]W. A. Bassali, J. Mech. Phys. Solids **7**, 272 (1959).
[8]E. Deutsch, Q. J. Mech. Appl. Math. **15**, 303 (1962).
[9]S. Ghosh, Bull. Calcutta Math. Soc. **39**, 1 (1947).
[10]L. M. Milne-Thomson, Trans. Am. Math. Soc. **90**, 143 (1959).
[11]R. M. Morris, Math. Anna. **116**, 374 (1939).
[12]W. A. Bassali and S. A. Obaid, J. Univ. Kuwait (Sci) **8**, 33 (1981).
[13]W. A. Bassali and S. A. Obaid, J. Appl. Math. Mech. **61**, 639 (1981).
[14]A. C. Stevenson, Philos. Trans. R. Soc. London **237**, 161 (1938).
[15]L. Leibenson, Wiss. Ber. Mosk. Univ. **2**, 99 (1924).
[16]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals* (Academic, New York, 1965).
[17]L. B. W. Jolley, *Summation of Series* (Dover, New York, 1961).

Immiscible two-phase flow in a porous medium: Utilization of a Laplace transform boost

Michael J. King
SOHIO Research and Development, Cleveland, Ohio 44128

(Received 17 July 1984; accepted for publication 26 October 1984)

A boost operator is defined for the Laplace transform. It is used in conjunction with the Laplace transform to solve the diffusion equation on a domain with multiple moving boundaries. This problem arises in the study of immiscible multiphase flow in a finite medium, including end effects. Extensions of this method to the solution of other linear differential equations are conceivable.

I. INTRODUCTION

This paper is concerned with the solution of a problem which arises in the study of the flow of two immiscible fluids through a finite porous medium. To this end, a boost operator is introduced which, in conjunction with the Laplace transform, allows a linear differential equation of flow, with multiple moving boundaries, to be solved.

The general multiphase flow problem is intrinsically nonlinear. The effective diffusivity and convective speed are functions of the fluid saturation, i.e., the fraction of pore space occupied by each fluid.[1,2] These equations depend upon phenomenological relative permeability functions for each fluid, and a capillary pressure function, which characterize the pore size distribution of the medium and the rock–fluid interaction.

The corresponding class of nonlinear evolution equations has been analyzed[3] previously using the symmetry approach of Fokas.[4] The symmetry analysis is related to that of inverse scattering theory.[5] It leads to a choice of relative permeability and capillary pressure functions for which the resulting partial differential equation can be linearized. This linearization is not a mathematical approximation to the nonlinear problem. Instead, it corresponds to a "fine tuning" of physical parameters so that the resulting nonlinear evolution equation has sufficient symmetries for a Lie–Bäcklund transformation to exist. The resulting relative permeability and capillary pressure functions are similar to empirically determined functions, although of a more restricted form. Related evolution equations[6] have arisen in the study of solid crystalline molecular hydrogen,[7] nonlinear heat conduction,[8] and in plasma physics.[9]

The nonlinear problem can be mapped to the ordinary diffusion equation. The inlet and outlet boundaries are mapped to moving boundaries under the same transformation. These boundary conditions cannot be solved by conventional use of the Laplace transform because there is no coordinate system in which both boundaries are at rest.

The boost operator allows multiple moving boundaries to be treated. The action of this operator on the general solution of the diffusion equation is examined in detail. It is conceivable that this same approach may be utilized in the solution of other linear differential equations with more complicated boundary motions. However, such extensions are beyond the scope of this paper.

The linearized evolution equation for two-phase flow was solved[3] on a semi-infinite domain by Yortsos and Fokas. This neglected the capillary end effects at the outlet of a finite medium due to the preferential wettability of one fluid with respect to the medium. Typical laboratory work involves the simultaneous flow of oil and water through a water-wet rock, leading to a transient period of flow in which only oil is produced from the rock. The time to water breakthrough and the distribution of fluids through the rock are measured in the laboratory. It is important to understand how the capillary "holdback" of water delays breakthrough and leads to a banking up of wetting phase at the outlet of the medium. This capillary end effect changes the problem from one on an infinite domain to one on a finite domain. Difficulties arise because in the linearized version of the problem, each boundary is moving.

This paper is organized into five sections. After this Introduction, the linearization of Yortsos and Fokas[3] is summarized. In Sec. III the boost operator is derived and several properties are proven. The problem is solved without capillary end effect in Sec. IV. Finally, the problem of a finite medium with end effect is solved until water breakthrough occurs.

II. EVOLUTION EQUATIONS

The evolution equation for the two-phase flow of fluids (water and oil) in a one-dimensional porous medium follows from Darcy's law and mass conservation equations. For the injection of water at a constant rate into a porous medium,

$$0 = \frac{\partial s}{\partial t} + \frac{\partial f_w}{\partial x}, \tag{2.1}$$

$$f_w = F(s) - G(s)\frac{\partial s}{\partial x}. \tag{2.2}$$

This is a second-order, nonlinear, parabolic, partial differential equation.

Here, s is the saturation of water; the saturation of oil is given by $1-s$. The fractional flow of water, $f_w(x,t)$ is defined as the volumetric flow of water relative to the total volumetric flow of water and oil. In (2.1) and (2.2), x and t are dimensionless distance and time. Here $F(s)$ only depends on the relative permeabilities, and $G(s)$ depends on relative permeabilities and capillary pressure, vanishing when the latter is negligible.

For the linearization of Yortsos and Fokas,[3] one chooses

$$F(s) = Ms/[1+(M-1)s], \tag{2.3}$$

$$G(s) = \epsilon^2/[1+(M-1)s]^2. \tag{2.4}$$

 0022-2488/85/040870-08$02.50

M is the mobility of water relative to oil and ϵ characterizes the relative strength of capillary to viscous forces. The linearization takes advantage of (2.1) to introduce a change of variable:

$$d\bar{x} \equiv \epsilon^{-1}[1 + (M-1)s]dx + \epsilon^{-1}[1 + (M-1)(1-f_w)]dt \tag{2.5a}$$

$$= [G(s)]^{-1/2}dx + \epsilon^{-1}[1 + (M-1)(1-f_w)]dt, \tag{2.5b}$$

$$d\bar{t} \equiv dt. \tag{2.6}$$

The integrability of (2.5) is assured by (2.1).

The evolution equation becomes

$$0 = \frac{\partial}{\partial \bar{t}}[\epsilon(1 + (M-1)s)^{-1}] + \frac{\partial}{\partial \bar{x}}[(1 + (M-1)s)^{-1}(1 + (M-1)(1-f_w))]. \tag{2.7}$$

It is linearized in terms of $\phi(\bar{x},\bar{t})$

$$[1 + (M-1)s]^{-1} \equiv -(\epsilon/M)(\phi_{\bar{x}}/\phi). \tag{2.8}$$

Notice that ϕ is determined up to a gauge transformation,

$$\phi(\bar{x},\bar{t}) \rightarrow \phi(\bar{x},\bar{t})\Lambda(\bar{t}). \tag{2.9}$$

One obtains

$$0 = \frac{\partial}{\partial \bar{x}}\left(\frac{-\phi_{\bar{t}} + \phi_{\bar{x}\bar{x}}}{\phi}\right).$$

Without loss of generality,

$$\phi_{\bar{t}} - \phi_{\bar{x}\bar{x}} = 0. \tag{2.10}$$

The heat equation results.

Equation (2.1), without convective terms, was studied by Rogers and Shadwick.[6] It was previously known that the special case of $G(s)$, (2.4), is integrable.[10]

The fractional water flow takes a simple form:

$$f_w = [M + \epsilon\phi_{\bar{t}}/\phi_{\bar{x}}]/(M-1). \tag{2.11}$$

Typically, f_w is fixed at a boundary. A linear equation in ϕ,

$$0 = [1 + (M-1)(1-f_w)]\phi_{\bar{x}} + \epsilon\phi_{\bar{t}}, \tag{2.12}$$

results. At the inlet, for waterflood,

$$x = 0, \quad t \geqslant 0, \quad f_w = 1. \tag{2.13}$$

With this boundary condition one obtains the transformed independent variables,

$$\bar{x} = \epsilon^{-1}\left[x + t + (M-1)\int_0^x s(x',t)dx'\right], \tag{2.14}$$

$$\bar{t} = t. \tag{2.15}$$

The inlet boundary condition becomes

$$\bar{x} = \epsilon^{-1}\bar{t}, \quad \bar{t} \geqslant 0, \quad 0 = \phi_{\bar{x}} + \epsilon\phi_{\bar{t}}. \tag{2.16}$$

Therefore, in a comoving coordinate system, ϕ is stationary,

$$\phi(\epsilon^{-1}\bar{t},\bar{t}) = \phi(0,0).$$

Since ϕ can be rescaled,

$$\phi(0,0) = 1, \tag{2.17}$$

$$\phi(\epsilon^{-1}\bar{t},\bar{t}) = 1, \quad \bar{t} \geqslant 0, \quad \epsilon \neq 0. \tag{2.18}$$

The initial condition can also be stated in the new variables. Initially, the porous medium is fully saturated with oil,

$$x \geqslant 0, \quad t = 0, \quad s = 0. \tag{2.19}$$

From (2.8) and (2.17)

$$\phi(\bar{x},0) = \varphi(\bar{x}), \quad \bar{x} \geqslant 0, \tag{2.20}$$

$$\varphi(\bar{x}) \equiv \exp(-M\bar{x}/\epsilon). \tag{2.21}$$

Alternative choices of the second boundary condition necessary to solve (2.10) will be stated in Secs. IV and V.

The new independent variable $\bar{x}$ and dependent variable ϕ have simple interpretations. Aside from a trivial time translation and rescaling, $\bar{x}$, (2.14), differs from x by $\int_0^x s(x',t)dx'$. The integral is the dimensionless volume of water under the saturation profile up to the distance x. Through the core, $\bar{x}$ is advanced from $\epsilon^{-1}(x+t)$ by an amount proportional to this volume.

At the inlet, this integral vanishes and $\bar{x}$ moves uniformly with t. At the outlet, before breakthrough, all the injected water is still in the core and this integral is given by t. The outlet will also move uniformly with t.

After breakthrough, the volume of water in the rock is less than t. For large t, the saturation profile reaches a steady state and the volume integral becomes constant. From (2.14), this implies that the outlet and inlet boundaries are asymptotically parallel. Asymptotically, one expects the water to occupy all of the available pore space,

$$\int_0^L s(x',t)dx' \rightarrow L.$$

Therefore, $\bar{x} \rightarrow \epsilon^{-1}(LM + t)$, and the inlet and outlet boundary curves will never cross. Typical domains are shown in Fig. 1. Notice that for $M < 1$, the no-breakthrough outlet boundary leads to a domain which only exists for a finite time. Consequently, the solution is singular. The physical solution is not singular since breakthrough will occur before the domain lines will cross.

The interpretation of the dependent variable follows by reexpressing ϕ in terms of the original independent variables

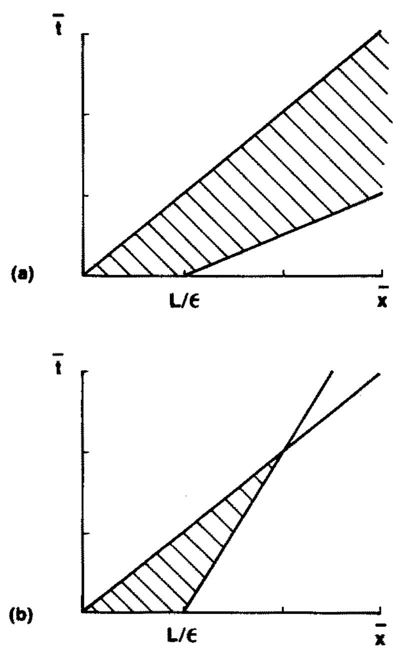


FIG. 1. Domain of solution, finite medium before breakthrough: (a) $M > 1$, (b) $M < 1$.

x,t. From (2.5), (2.8), and (2.12) one obtains

$$\partial_x \phi = -(M/\epsilon^2)\phi, \tag{2.22}$$

$$\partial_t \phi = 0. \tag{2.23}$$

Utilizing the inlet boundary condition (2.18) gives

$$\phi = \exp(-Mx/\epsilon^2). \tag{2.24}$$

Therefore, ϕ is simply related to x and drops off exponentially through the rock.

III. LAPLACE TRANSFORM BOOST

The purpose of introducing a Laplace transform boost is to treat directly a moving boundary problem. This is necessary when multiple moving boundaries are considered. It may also be convenient to use the boost operator on problems with single moving boundaries.

The boost operator is obtained by taking the Laplace transform of the operator of spatial translation. Consider a function $\phi(x,t)$, analytic in x, which solves a given partial differential equation. Let $\psi(x-\delta t, t)$ be the corresponding solution in a moving coordinate system

$$\phi(x,t) = \psi(x-\delta t, t) \tag{3.1}$$

$$= \sum_{n=0}^{\infty} \frac{(-\delta t)^n}{n!} \frac{\partial^n}{\partial x^n} \psi(x,t) \tag{3.2}$$

$$= \exp\left(-\delta t \frac{\partial}{\partial x}\right)\psi(x,t). \tag{3.3}$$

The series (3.2) is assumed to converge uniformly in x to $\phi(x,t)$.

The Laplace transform $\tilde{\phi}(x,p)$ can be related to the transform $\tilde{\psi}(x,p)$. Substituting the power series (3.2) into the definition

$$\tilde{\phi}(x,p) \equiv \int_0^{\infty} dt\, e^{-pt}\phi(x,t), \tag{3.4}$$

leads to the power series

$$\tilde{\phi}(x,p) = \sum_{n=0}^{\infty} \frac{\delta^n}{n!} \frac{\partial^n}{\partial x^n} \frac{\partial^n}{\partial p^n} \int_0^{\infty} dt\, e^{-pt}\psi(x,t). \tag{3.5}$$

This defines the boost operator for uniform translational motion

$$\tilde{\phi}(x,p) \equiv \exp\left(\delta \frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)\tilde{\psi}(x,p). \tag{3.6}$$

The interchanges of summation, integration, and differentiation are justified because of the uniform convergence of the series.

The boost operator is useful because its action can be simply stated. The following two theorems provide the tools needed to solve the problems of Secs. IV and V.

Theorem 1: Consider a function $\tilde{\phi}(x,p)$ of complex x, p whose only singularities throughout the finite portion of the plane are poles, i.e., no branch cuts or limit points (meromorphic function). Then

$$\exp\left(\delta \frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)\tilde{\phi}(x,p) = \tilde{\phi}\left(x + \delta\frac{\partial}{\partial p}, p + \delta\frac{\partial}{\partial x}\right). \tag{3.7}$$

Differentiation is understood to act to the right, only.

Proof: Because $\tilde{\phi}$ is meromorphic it can be represented by a uniformly convergent power series away from a pole. The domain of convergence is determined by the location of the nearest pole. Without loss of generality, expand around $x=0, p=0$:

$$\tilde{\phi}(x,p) = \sum_{k,l=0}^{\infty} \phi_{kl} x^k p^l.$$

Then

$$\exp\left(\delta \frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)\tilde{\phi}(x,p) = \sum_{n=0}^{\infty} \frac{\delta^n}{n!}\left(\frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)^n \sum_{k,l=0}^{\infty} \phi_{kl}x^k p^l = \sum_{k,l=0}^{\infty} \phi_{kl}\left[\sum_{n=0}^{\infty} \frac{\delta^n}{n!}\frac{k!}{(k-n!)}\frac{l!}{(l-n)!}x^{k-n}p^{l-n}\right].$$

The sum over n gives an operator product

$$= \sum_{k,l=0}^{\infty} \phi_{kl}\left(x + \delta\frac{\partial}{\partial p}\right)^k\left(p + \delta\frac{\partial}{\partial x}\right)^l$$

$$= \tilde{\phi}\left(x + \delta\frac{\partial}{\partial p}, p + \delta\frac{\partial}{\partial x}\right).$$

The product of operators and the shifted form of ϕ is unambiguous because the operators commute,

$$\left[x + \delta\frac{\partial}{\partial p}, p + \delta\frac{\partial}{\partial x}\right] = 0.$$

The result describes the boost in terms of an operator valued shift in the arguments of $\tilde{\phi}(x,p)$.

The following theorem is useful for problems involving the heat equation. Theorem 1 is not applicable to functions of the form (3.8) because of the branch cut in $\sqrt{p}$.

Theorem 2: If

$$\tilde{\phi}(x,p) = A(\sqrt{p})\exp(-x\sqrt{p}), \tag{3.8}$$

$\tilde{\phi}(x,p)$ vanishes sufficiently fast for large p, and $A(\sqrt{p})$ is analytic with at most a countable number of singularities in the complex $\sqrt{p}$ plane, then

$$\exp\left(\delta \frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)A(\sqrt{\phi})\exp(-x\sqrt{p})$$

$$= \left[1 - \tfrac{1}{2}\delta/\sqrt{\tfrac{1}{4}\delta^2 + p}\,\right]A\left(-\delta/2 + \sqrt{\delta^2/4 + p}\right)$$

$$\times \exp\left[-\left(-\delta/2 + \sqrt{\delta^2/4 + p}\right)x\right]. \tag{3.9}$$

The proof will be given in the Appendix.

IV. SOLUTION WITHOUT CAPILLARY END EFFECT

Consider the linearized problem of flow in a semi-infinite porous medium. The partial differential equation (PDE) is given by (2.10) with initial condition (2.20) and inlet boundary condition (BC1) (2.18). On the semi-infinite domain, the boundary condition at infinity follows from mass conservation. The dimensionless volume of water which enters the system is given by

$$\int_0^\infty s(x,t)dx = t. \tag{4.1}$$

Therefore, for large x at fixed t, $s(x,t)\to 0$. In terms of $\bar{x}, \bar{t}, \phi(\bar{x},\bar{t})$ one obtains BC2:

$$\bar{x}\to\infty,\quad \bar{t}\geqslant 0,\quad \phi(\bar{x},\bar{t})\to 0. \tag{4.2}$$

For the remainder of this paper, the symbols x, t will be given in place of $\bar{x}, \bar{t}$ unless otherwise noted.

The semi-infinite domain follows from Fig. 1 as $L\to\infty$. No distinction need be made between the cases $M>1$, $M<1$. Also, as $L\to\infty$, (4.2) follows from (2.24).

The Laplace transform of (2.10) leads to the differential equation for $\tilde{\phi}(x,p)$:

$$0 + \tilde{\phi}_{xx}(x,p) - p\tilde{\phi}(x,p) + \varphi(x). \tag{4.3}$$

BC1 cannot be treated, while BC2 leads to

$$x\to\infty,\quad \tilde{\phi}(x,p)\to 0. \tag{4.4}$$

The general solution to (4.3) is

$$\tilde{\phi}(x,p) = A_p\exp(-x\sqrt{p}) + B_p\exp(x\sqrt{p}) + (p-(M/\epsilon)^2)^{-1}\exp(-Mx/\epsilon). \tag{4.5}$$

From (4.4)

$$B_p = 0, \tag{4.6}$$

but A_p cannot be obtained.

To implement BC1, work in a coordinate system comoving with the inlet:

$$y \equiv x - \epsilon^{-1}t, \tag{4.7}$$

$$\phi(x,t)\equiv\psi(y,t). \tag{4.8}$$

One obtains the differential equation

$$0 = \psi_{yy}(y,t) + \epsilon^{-1}\psi_y(y,t) - \psi_t(y,t) \tag{4.9}$$

and boundary conditions BC1 and BC2

$$y=0,\quad t\geqslant 0,\quad \psi(y,t)=1, \tag{4.10}$$

$$y\to\infty,\quad t\geqslant 0,\quad \psi(y,t)\to 0. \tag{4.11}$$

The Laplace transform of this system is

$$0 = \tilde{\psi}_{yy}(y,p) + \epsilon^{-1}\tilde{\psi}_y(y,p) - p\tilde{\psi}_y(y,p) + \varphi(y), \tag{4.12}$$

with

$$y=0,\quad \tilde{\psi}(y,p) = p^{-1}, \tag{4.13}$$

$$y\to\infty,\quad \tilde{\psi}(y,p)\to 0. \tag{4.14}$$

The general solution to (4.12) is

$$\begin{aligned}\tilde{\psi}(y,p) &= C_p\exp[-(1/2\epsilon + \sqrt{1/4\epsilon^2+p})y] \\ &\quad + D_p\exp[(-1/2\epsilon + \sqrt{1/4\epsilon^2+p})y] \\ &\quad + (p - M(M-1)/\epsilon^2)^{-1}\exp(-My/\epsilon).\end{aligned} \tag{4.15}$$

The boundary conditions imply

$$D_p = 0, \tag{4.16}$$

$$C_p = p^{-1} - (p - M(M-1)/\epsilon^2)^{-1}. \tag{4.17}$$

One may obtain $\phi(x,t)$ by inverting (4.15)–(4.17) for $\psi(y,t)$ and substituting (4.7) and (4.8). However, instead of this usual approach, $\tilde{\phi}(x,p)$ will be obtained from $\tilde{\psi}(y,p)$ by utilizing the boost operator. The inverse Laplace transform will then be performed to obtain $\phi(x,t)$.

From (3.6),

$$\tilde{\phi}(x,p) = \exp\left(\epsilon^{-1}\frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)\tilde{\psi}(x,p)$$

or

$$\tilde{\psi}(y,p) = \exp\left(-\epsilon^{-1}\frac{\partial}{\partial y}\frac{\partial}{\partial p}\right)\tilde{\phi}(y,p).$$

The boost of the first term of $\tilde{\phi}(x,p)$ is given by Theorem 2:

$$\begin{aligned}&\exp\left(-\epsilon^{-1}\frac{\partial}{\partial y}\frac{\partial}{\partial p}\right)A[\sqrt{\phi}]\exp(-y\sqrt{\phi}) \\ &\quad = \left[1 + \frac{1/2\epsilon}{\sqrt{1/4\epsilon^2+p}}\right]A\left[\frac{1}{2\epsilon} + \sqrt{\frac{1}{4\epsilon^2+p}}\right] \\ &\quad\quad \times\exp\left[-\left(\frac{1}{2\epsilon} + \sqrt{\frac{1}{4\epsilon^2+p}}\right)y\right].\end{aligned}$$

Comparison with (4.15) and (4.17) gives

$$\left[1 + \frac{1/2\epsilon}{\sqrt{1/4\epsilon^2+p}}\right]A\left[\frac{1}{2\epsilon} + \sqrt{\frac{1}{4\epsilon^2}+p}\right] = p^{-1} - (p - M(M-1)/\epsilon^2)^{-1}.$$

Therefore,

$$\begin{aligned}2\sqrt{p}A[\sqrt{p}] &= 1/\sqrt{p} + 1/(\sqrt{p}-1/\epsilon) \\ &\quad - 1/(\sqrt{p}+(M-1)/\epsilon) - 1/(\sqrt{p}-M/\epsilon)\end{aligned} \tag{4.18}$$

$$\equiv \sum_{j=1}^{4} s_j/(\sqrt{p}-b_j), \tag{4.19}$$

where

$$s_j = +1,\ +1,\ -1,\ -1, \tag{4.20a}$$

$$b_j = 0,\ 1/\epsilon,\ -(M-1)/\epsilon,\ M/\epsilon. \tag{4.20b}$$

The inhomogeneous term may be boosted utilizing Theorem 1

$$\begin{aligned}&\exp\left(-\epsilon^{-1}\frac{\partial}{\partial y}\frac{\partial}{\partial p}\right)\left(p - \left(\frac{M}{\epsilon}\right)^2\right)^{-1}\exp\left(-\frac{My}{\epsilon}\right) \\ &\quad = \left(\left(p - \epsilon^{-1}\frac{\partial}{\partial y}\right) - \left(\frac{M}{\epsilon}\right)^2\right)^{-1}\exp\left(\frac{-My}{\epsilon}\right) \\ &\quad = (p - M(M-1)/\epsilon^2)^{-1}\exp(-My/\epsilon).\end{aligned}$$

This reproduces the inhomogeneous term of $\tilde{\psi}(y,p)$ (4.15).

Equation (4.3) and BC1 are solved by

$$\begin{aligned}\tilde{\phi}(x,p) &= \frac{1}{2}\sum_{j=1}^{4} s_j\frac{1}{\sqrt{p}}\frac{1}{\sqrt{p}-b_j}\exp(-x\sqrt{p}) \\ &\quad + (p-(M/\epsilon)^2)^{-1}\exp(-Mx/\epsilon),\end{aligned} \tag{4.21}$$

which can be inverted for $\phi(x,t)$[11] provided $x\geqslant 0$

$$\begin{aligned}\phi(x,t) &= \frac{1}{2}\sum_{j=1}^{4} s_j\exp[-b_j(x-b_jt)] \\ &\quad \times\operatorname{erfc}[(\tfrac{1}{2}x - b_jt)/\sqrt{t}] \\ &\quad + \exp[-(M/\epsilon)(x-Mt/\epsilon)].\end{aligned} \tag{4.22}$$

This is the solution of Yortsos and Fokas. The form of (4.22) is more compact than that of Yortsos and Fokas. They obtained separate expressions for $M>1$ and $M<1$. The form of the solution (4.22) is similar to that obtained by Wei.[12]

The solution $\phi(\bar{x},\bar{t})$ decreases with $\bar{x}$ while increasing with $\bar{t}$. At each time the saturation is determined by (2.8) knowing ϕ and $\phi_{\bar{x}}$. The location of this saturation x follows from (2.24). The resulting saturation profile decreases monotonically through the porous medium (Fig. 2).

V. SOLUTION WITH CAPILLARY END EFFECT

The linearized problem will now be solved with the inclusion of a capillary end effect. The end effect occurs because the rock is preferentially water-wet. Consequently, water cannot flow from the outlet until the water pressure inside the rock is greater than the oil pressure beyond the outlet. Before breakthrough, f_w vanishes at the outlet; water can flow to the outlet but not beyond. During this transient period, water is banking up at the outlet while oil continues to flow.

As in the previous section, the PDE is given by (2.10), the initial condition by (2.20), and BC1 by (2.18). In terms of the original variables, the outlet boundary condition (BC2) is

$$x = L, \quad t \geqslant 0, \quad f_w = 0. \tag{5.1}$$

L is the dimensionless length of the medium. Since there is no water flow past $x = L$,

$$\int_0^L s(x, t)dx = t$$

and

$$\bar{x} = \epsilon^{-1}(L + M\bar{t}). \tag{5.2}$$

Reverting to the convention of unbarred variables for $\bar{x}, \bar{t}$, and using (2.12) gives

$$0 = M\phi_x + \epsilon\phi_t. \tag{5.3}$$

In a comoving coordinate system,

$$\frac{d}{dt}\phi(\epsilon^{-1}(L + Mt), t) = 0.$$

Utilizing the initial condition gives BC2,

$$\phi(\epsilon^{-1}(L + Mt), t) = \exp(-ML/\epsilon^2), \quad t \geqslant 0. \tag{5.4}$$

It is necessary that x lie between the inlet and the outlet boundaries:

$$\epsilon^{-1}t \leqslant x \leqslant \epsilon^{-1}(L + Mt). \tag{5.5}$$

This gives a necessary condition on t

$$(1 - M)t \leqslant L.$$

If $M > 1$ then the inequality will be strictly true for all t. Both

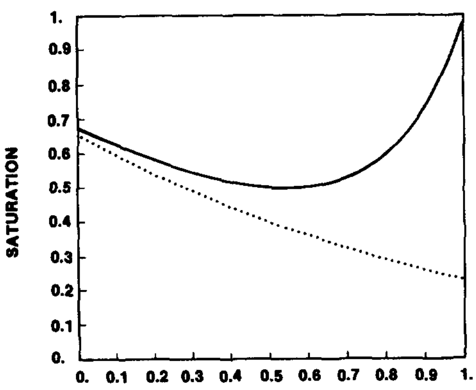


FIG. 2. Saturation profiles at breakthrough: $M = 2.0, L = 1.0, \epsilon = 1.0$. Solid curve with end effect, dotted curve without end effect.

$\tilde{\phi}(x,p)$ and $\phi(x, t)$ exist in this case. On the other hand, $M < 1$ requires that t be restricted,

$$t \leqslant L/(1 - M), \quad M < 1. \tag{5.6}$$

A singularity is expected in the solution for $M < 1$ as t approaches $L/(1 - M)$. In this case the domain of solution exists for only a finite time and the Laplace transform cannot exist. However, although the series for $\tilde{\phi}(x,p)$ is formally divergent, the resulting series for $\phi(x,t)$ is convergent.

Although $\phi(x, t)$ will exist, the no-breakthrough boundary condition is not reasonable. There is a finite pore volume in the rock and once this is completely saturated with water, water must flow from the rock. In practice, breakthrough occurs much earlier than this argument would indicate. Water will bank up until a saturation s^* determined by the capillary pressure function is reached. As will be shown, there is still a substantial amount of oil in the rock when breakthrough occurs and multiphase flow from the outlet begins. Without loss of generality, s^* will be set to 1.

As before, the Laplace transform of (2.10) leads to (4.3) and the general solution (4.5). Neither A_p nor B_p can be determined. In the moving coordinate system (4.7) one obtains (4.12) and BC1, (4.13). With the general solution (4.15), BC1 implies

$$C_p + D_p = p^{-1} - (p - M(M-1)/\epsilon^2)^{-1}. \tag{5.7}$$

To implement BC2 (5.4) introduce a comoving coordinate

$$z \equiv x - Mt/\epsilon, \tag{5.8}$$

$$\phi(x,t) \equiv A(z, t). \tag{5.9}$$

The resulting PDE is

$$0 = A_{zz}(z, t) + M\epsilon^{-1}A_z(z, t) - A_t(z, t), \tag{5.10}$$

with boundary conditions BC1 and BC2:

$$z = \epsilon^{-1}(1 - M)t, \quad A(z, t) = 1, \tag{5.11}$$

$$z = \epsilon^{-1}L, \quad A(z, t) = \exp(-ML/\epsilon^2). \tag{5.12}$$

The Laplace transform of (5.10) is

$$0 = \tilde{A}_{zz}(z,p) + M\epsilon^{-1}\tilde{A}_z(x,p) - p\tilde{A}(z,p) + \varphi(z), \tag{5.13}$$

and BC2 becomes

$$p\tilde{A}(\epsilon^{-1}L, p) = \exp(-ML/\epsilon^2). \tag{5.14}$$

The general solution to (5.13) is

$$\begin{aligned}\tilde{A}(z,p) = {} & E_p \exp[-(M/2\epsilon + \sqrt{M^2/4\epsilon^2 + p})z] \\ & + F_p \exp[(-M/2\epsilon + \sqrt{M^2/4\epsilon^2 + p})z] \\ & + p^{-1}\exp(-Mz/\epsilon).\end{aligned} \tag{5.15}$$

Then BC2 implies

$$\begin{aligned}0 = {} & E_p \exp\left[-\frac{L}{\epsilon}\left(\frac{M}{2\epsilon} + \sqrt{\frac{M^2}{4\epsilon^2} + p}\right)\right] \\ & + F_p \exp\left[\frac{L}{\epsilon}\left(\frac{M}{2\epsilon} + \sqrt{\frac{M^2}{4\epsilon^2} + \phi}\right)\right].\end{aligned} \tag{5.16}$$

The general solution in the original coordinates, (4.5), is boosted into the moving coordinates (4.7) and (5.8) to satisfy (5.7) and (5.16).

Consider the boost of the homogeneous terms of (4.5) into the coordinate y, (4.7):

$$\exp\left(-\epsilon^{-1}\frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)\{A[\sqrt{p}]\exp(-x\sqrt{p})+B[\sqrt{p}]\exp(x\sqrt{p})\}$$
$$=\left[1+\frac{1}{2\epsilon}\Big/\sqrt{\frac{1}{4\epsilon^2}+p}\right]A\left[\frac{1}{2\epsilon}+\sqrt{\frac{1}{4\epsilon^2}+p}\right]\exp\left[-x\left(\frac{1}{2\epsilon}+\sqrt{\frac{1}{4\epsilon^2}+p}\right)\right]+\left[1-\frac{1}{2\epsilon}\Big/\sqrt{\frac{1}{4\epsilon^2}+p}\right]$$
$$\times B\left[-\frac{1}{2\epsilon}+\sqrt{\frac{1}{4\epsilon^2}+p}\right]\exp\left[x\left(-\frac{1}{2\epsilon}+\sqrt{\frac{1}{4\epsilon^2}+p}\right)\right].$$

Therefore,

$$C_p=\left[1+\frac{1}{2\epsilon}\Big/\sqrt{\frac{1}{4\epsilon^2}+p}\right]A\left[\frac{1}{2\epsilon}+\sqrt{\frac{1}{4\epsilon^2}p}\right],$$
$$D_p=\left[1-\frac{1}{2\epsilon}\Big/\sqrt{\frac{1}{4\epsilon^2}+p}\right]B\left[-\frac{1}{2\epsilon}+\sqrt{\frac{1}{4\epsilon^2}+p}\right],$$

and BC1, (5.7), becomes

$$\sqrt{p}A[\sqrt{p}]+\left(\sqrt{p}-\frac{1}{\epsilon}\right)B\left[\sqrt{p}-\frac{1}{\epsilon}\right]=\frac{1}{2}\sum_{j=1}^{4}\frac{s_j}{\sqrt{p}-b_j}. \tag{5.17}$$

Similarly, boosting (4.5) into the coordinate system (5.8) leads to

$$0=\sqrt{p}A[\sqrt{p}]+(\sqrt{p}-M/\epsilon)B[\sqrt{p}-M/\epsilon]\times\exp[(L/\epsilon)(2\sqrt{p}-M/\epsilon)]. \tag{5.18}$$

To solve the functional equations (5.17) and (5.18), it is convenient to define

$$\xi\equiv\sqrt{p}, \tag{5.19}$$

$$H(\xi)\equiv\sqrt{p}A[\sqrt{p}]. \tag{5.20}$$

Equation (5.18) gives

$$A(\xi)=\xi^{-1}H(\xi), \tag{5.21}$$

$$B(\xi)=-\xi^{-1}H(\xi+(M/\epsilon))\times\exp[-(L/\epsilon)(2\xi+M/\epsilon)]. \tag{5.22}$$

Substituting into (5.17) gives the functional equation

$$H(\xi)=\frac{1}{2}\sum_{j=1}^{4}\frac{s_j}{\xi-b_j}+H\left(\xi+\frac{M-1}{\epsilon}\right)\times\exp[-(L/\epsilon)(2\xi+(M-2)/\epsilon)]. \tag{5.23}$$

It may be solved with the ansatz

$$H(\xi)=\frac{1}{2}\sum_{j=1}^{4}s_j\sum_{n=0}^{\infty}\left(\xi+\frac{n(M-1)}{\epsilon}-b_j\right)^{-1}\times\exp[-(L/\epsilon)(c_n\xi+d_n)].$$

Substituting into (5.23) gives

$$H(\xi)=\frac{1}{2}\sum_{j=1}^{4}s_j\sum_{n=0}^{\infty}\left(\xi+\frac{n(M-1)}{\epsilon}-b_j\right)^{-1}\times\exp[-(nL/\epsilon)(2\xi+n(M-1)/\epsilon-1/\epsilon)]. \tag{5.24}$$

The convergence of (5.24) is dominated by the exponential $\exp[-n^2L(M-1)/\epsilon^2]$. For $M>1$ the sum converges rapidly. For $M<1$ the series diverges, as expected. For $M=1$ the sum converges and may be performed explicitly.

Consider the inverse Laplace transform[11] of the homogeneous term $A_p\exp(-x\sqrt{p})$ (4.5).

$$A_p\exp(-x\sqrt{p})=\frac{1}{2}\sum_{j=1}^{4}s_j\sum_{n=0}^{\infty}\exp\left[-\frac{nL}{\epsilon^2}(n(M-1)-1)\right]\times(\sqrt{p})^{-1}(\sqrt{p}+n(M-1)/\epsilon-b_j)^{-1}\times\exp[-(x+2nL/\epsilon)\sqrt{p}]$$
$$\to\frac{1}{2}\sum_{j=1}^{4}s_j\sum_{n=0}^{\infty}\exp[n^2\alpha+n\beta_j+\gamma_j]\times\operatorname{erfc}[n\delta+\sigma_j], \tag{5.25}$$

where

$$\alpha\equiv(M-1)(L+(M-1)t)/\epsilon^2, \tag{5.26a}$$

$$\beta_j\equiv\epsilon^{-1}[x(M-1)+L/\epsilon-2b_j(L+(M-1)t)], \tag{5.26b}$$

$$\gamma_j\equiv-b_j(x-b_jt), \tag{5.26c}$$

$$\delta\equiv\epsilon^{-1}(L+(M-1)t)/\sqrt{t}, \tag{5.26d}$$

$$\sigma_j\equiv(\tfrac{1}{2}x-b_jt)/\sqrt{t}. \tag{5.26e}$$

The second homogeneous term transforms as

$$B_p\exp(x\sqrt{p})=-\frac{1}{2}\sum_{j=1}^{4}s_j\sum_{n=1}^{\infty}\exp\left[-\frac{nL}{\epsilon^2}(n(M-1)+1)\right]\times(\sqrt{p})^{-1}(\sqrt{p}+n(M-1)/\epsilon+b_j)^{-1}\times\exp[-(2nL/\epsilon-x)\sqrt{p}]$$
$$\to-\frac{1}{2}\sum_{j=1}^{4}s_j\sum_{n=1}^{\infty}\exp[n^2\alpha-n\beta_j+\gamma_j]\times\operatorname{erfc}[n\delta-\sigma_j]. \tag{5.27}$$

The erfc terms assure convergence for all M. For large n,[13] the terms are dominated by

$$\exp[n^2(M-1)(L+(M-1)t)/\epsilon^2]\times\operatorname{erfc}[n\epsilon^{-1}(L+(M-1)t)/\sqrt{t}]$$
$$\sim(\epsilon\sqrt{t}/n\sqrt{\pi})(L+(M-1)t)^{-1}\times\exp[-n^2(L/\epsilon^2t)(L+(M-1)t)],$$

which vanishes rapidly because of the positivity of (5.5). The argument can only vanish if $M<1$ and $t=L/(1-M)$. In this case the summation may be performed explicitly. For any M, the large n cancellation between (5.26) and (5.27) further enhances convergence.

The solution to the flow problem, with capillary holdback, before breakthrough, is given by

$$\phi(x,t) = \exp\left[-\left(\frac{M}{\epsilon}\right)\left(x - \frac{Mt}{\epsilon}\right)\right] + \frac{1}{2}\sum_{j=1}^{4} s_j \exp(\gamma_j)$$
$$\times \operatorname{erfc}(\sigma_j) + \frac{1}{2}\sum_{j=1}^{4} s_j \sum_{n=1}^{\infty} \exp(n^2\alpha + \gamma_j)$$
$$\times \{\exp(n\beta_j)\operatorname{erfc}(n\delta + \sigma_j)$$
$$- \exp(-n\beta_j)\operatorname{erfc}(n\delta - \sigma_j)\}. \quad (5.28)$$

As $L \to \infty$, the $n \geqslant 1$ terms vanish and the solution without end effect (4.22) results.

The resulting saturation profile is shown in Fig. 3. For early times, the solutions (4.22) and (5.28) are indistinguishable. However, as water reaches the outlet but does not flow from the core, it will bank up. Finally, at breakthrough ($s = 1$ at outlet), the end effect is quite substantial (Fig. 2).

VI. DISCUSSION

The effects of capillarity for two-phase flow in porous media have been described for a "finely tuned" porous medium. This medium, characterized by the functions $F(s)$, $G(s)$ [(2.3) and (2.4)] correspond to physically reasonable choices of relative permeability and capillary pressure functions. It is found that the capillary holdback of the wetting phase can be quite substantial. Substantial changes in production data and saturation distributions arise. The problem has been solved up to the moment of water breakthrough.

Beyond breakthrough, the problem is intrinsically nonlinear. The location of the outlet $x(t)$ depends on the total saturation of water in the rock. To solve for this saturation couples again to $\bar{x}(t)$, leading to the nonlinearity.

To determine the saturation before breakthrough a boost operator has been constructed to deal with multiple moving boundaries. It allows boundary conditions in a moving frame to be stated in a fixed coordinate system. The differential equation and boundary conditions can then be treated using the Laplace transform.

The utility of the boost operator follows from Theorems 1 and 2. The second theorem is specialized to solutions of the diffusion equation and follows from examining the general solution in shifted coordinates. It should be possible to prove analogous theorems for other differential equations and thereby extend the utility of the boost operator.

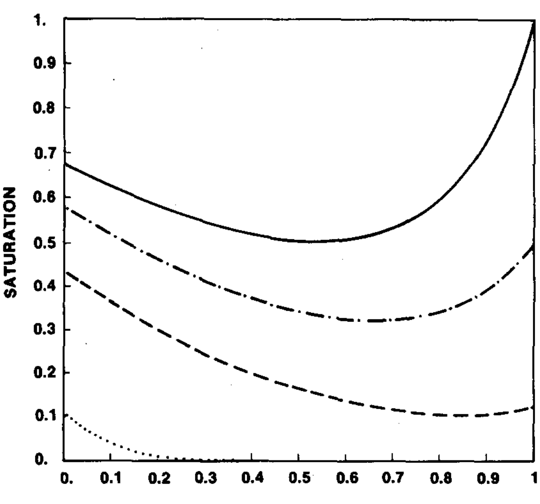


FIG. 3. Saturation profile at successive times: $M = 2.0$, $L = 1.0$, $\epsilon = 1.0$; $t = 0.01$ (dot), 0.20 (dash), 0.40 (dot–dash), 0.59 (solid).

ACKNOWLEDGMENTS

The author would like to thank his colleagues at SOHIO and F. Rohrlich for useful discussions on the subject of this paper.

In addition, the author thanks SOHIO Petroleum Company for permission to publish this work.

APPENDIX: PROOF OF BOOST ACTION

We prove Theorem 2 [(3.8) and (3.9)].

Consider the diffusion equation (4.3) with general homogeneous solution

$$\tilde{\phi}_H(x,p) = A[\sqrt{p}]\exp(-x\sqrt{p}) + B[\sqrt{p}]\exp(x\sqrt{p}). \quad (A1)$$

In a boosted coordinate system

$$y \equiv x - \delta t, \quad (A2)$$

$$\phi(x,t) \equiv \psi(y,t), \quad (A3)$$

one obtains the diffusion equation

$$\psi_{yy}(y,t) + \delta\psi_y(y,t) - \psi_t(y,t) = 0, \quad (A4)$$

and transformed equation

$$\tilde{\psi}_{yy}(y,p) + \delta\tilde{\psi}_y(y,p) - p\tilde{\psi}(y,p) + \varphi(y) = 0. \quad (A5)$$

The general form of the homogeneous solution is given by

$$\tilde{\psi}_H(y,p) = C[\sqrt{p}]\exp[-(\delta/2 + \sqrt{\delta^2/4 + p})y] + D[\sqrt{p}]\exp[(-\delta/2 + \sqrt{\delta^2/4 + p})y]. \quad (A6)$$

The solutions (A1) and (A6) are related by the boost (3.6). From the differing dependence on x of each term,

$$\exp\left(-\delta\frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)A[\sqrt{p}]\exp(-x\sqrt{p})$$
$$= C[\sqrt{p}]\exp[-x(\delta/2 + \sqrt{\delta^2/4 + p})], \quad (A7)$$

$$\exp\left(-\delta\frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)B[\sqrt{p}]\exp(x\sqrt{p})$$
$$= D[\sqrt{p}]\exp[x(-\delta/2 + \sqrt{\delta^2/4 + p})]. \quad (A8)$$

Consider (A7) for $A[\sqrt{p}] = 1$,

$$C[\sqrt{p}] = \exp\left[x\left(\frac{\delta}{2} + \sqrt{\frac{\delta^2}{4} + p}\right)\right]$$
$$\times \exp\left[-\delta\frac{\partial}{\partial x}\frac{\partial}{\partial p}\right]\exp(-x\sqrt{p})$$
$$= \sum_{jkl=0}^{\infty} \frac{x^j}{j!}\left(\frac{\delta}{2} + \sqrt{\frac{\delta^2}{4} + p}\right)^j$$
$$\times \frac{(-\delta)^k}{k!}\left(\frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)^k \frac{(-)^l}{l!}x^l(\sqrt{p})^l$$
$$= \sum_{jkl=0}^{\infty} (-)^{k+l}\frac{x^{j+l-k}\delta^k}{j!k!(l-k)!}$$
$$\times\left(\frac{\delta}{2} + \sqrt{\frac{\delta^2}{4} + p}\right)^j\left(\frac{\partial}{\partial p}\right)^k p^{l/2}.$$

Because $C[\sqrt{p}]$ is independent of x, only the $j+l-k=0$ term contributes. Further, only terms for which $l \geqslant k$ are nonvanishing. Therefore,

$$C[\sqrt{p}] = \sum_{l=0}^{\infty} \frac{\delta^l}{l!}\left(\frac{\partial}{\partial p}\right)^l p^{l/2}$$
$$= 1 + \sum_{l=1}^{\infty} \frac{\delta^l}{l!} \frac{(l/2)!}{(-l/2)!} p^{-l/2}.$$

Only odd l terms survive

$$C[\sqrt{p}] = 1 + (\delta/\sqrt{p}) \sum_{m=0}^{\infty} \left(\frac{\delta^2}{p}\right)^m \frac{1}{\Gamma(2m+2)} \frac{\Gamma(\frac{3}{2}+m)}{\Gamma(\frac{1}{2}-m)}.$$

Utilizing known properties[13] of $\Gamma(z)$,

$$C[\sqrt{p}] = 1 + \left(\frac{\delta}{2\sqrt{p}}\right) \sum_{m=0}^{\infty} \binom{-\frac{1}{2}}{m} \left(\frac{\delta^2}{4p}\right)^m$$
$$= 1 + (\delta/2)(\delta^2/4+p)^{-1/2}.$$

Therefore,

$$\exp\left(-\delta \frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)\exp(-x\sqrt{p})$$
$$= \left[1 + \frac{\delta}{2}\left(\frac{\delta^2}{4}+p\right)^{-1/2}\right]$$
$$\times \exp\left[-x\left(\frac{\delta}{2} + \sqrt{\frac{\delta^2}{4}+p}\right)\right]. \tag{A9}$$

One may act $(-\partial/\partial x)^k$ on (A9) to obtain

$$\exp\left(-\delta \frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)(\sqrt{p})^k \exp(-x\sqrt{p})$$
$$= \left[1 + \frac{\delta}{2}\left(\frac{\delta^2}{4}+p\right)^{-1/2}\right]\left(\frac{\delta}{2} + \sqrt{\frac{\delta^2}{4}+p}\right)^k$$
$$\times \exp[-x(\delta/2 + \sqrt{d^2/4+p})]. \tag{A10}$$

Similarly the k th integrated integral gives

$$\exp\left(-\delta \frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)(\sqrt{p})^k \exp(-x\sqrt{p})$$
$$= \left[1 + \frac{\delta}{2}\left(\frac{\delta^2}{4}+p\right)^{-1/2}\right]\left(\frac{\delta}{2} + \sqrt{\frac{\delta^2}{4}+p}\right)^{-k}$$
$$\times \exp\left[-x\left(\frac{\delta}{2} + \sqrt{\frac{\delta^2}{4}+p}\right)\right]. \tag{A11}$$

This result may be extended using the techniques of the fractional calculus[14] for any real number r:

$$\exp\left(-\delta \frac{\partial}{\partial x}\frac{\partial}{\partial p}\right)(\sqrt{p})^r \exp(-x\sqrt{p})$$
$$= \left[1 + \frac{\delta}{2}\left(\frac{\delta^2}{4}+p\right)^{-1/2}\right]\left(\frac{\delta}{2} + \sqrt{\frac{\delta^2}{4}+p}\right)^r$$
$$\times \exp\left[-x\left(\frac{\delta}{2} + \sqrt{\frac{\delta^2}{4}+p}\right)\right]. \tag{A12}$$

Further, if $A[\sqrt{p}]$ falls off sufficiently fast for large $\sqrt{p}$, and if $A[\sqrt{p}]$ is an analytic function of $\sqrt{p}$ with at most a countable number of poles, then (3.9) is obtained.

[1] H. C. Brinkman, Appl. Sci. Reš. A 1, 333 (1948).
[2] F. J. Fayers, and J. W. Sheldon, Pet. Trans. AIME **216**, 147 (1959).
[3] A. S. Fokas, and Y. C. Yortsos, SIAM J. Appl. Math. **42**, 318 (1982).
[4] A. S., Fokas, J. Math. Phys. **21**, 1318 (1980).
[5] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[6] C. Rogers and W. F. Shadwick, *Bäcklund Transformations and Their Applications* (Academic, New York, 1982).
[7] G. Rosen, Phys. Rev. B **19**, 2398 (1979).
[8] G. W. Bluman and S. Kumei, J. Math. Phys. **21**, 1019 (1980).
[9] J. G. Berryman and C. J. Holland, J. Math. Phys. **19**, 2476 (1978).
[10] W. F. Ames, *Nonlinear Partial Differential Equations in Engineering* (Academic, New York, 1965), Vols. I and II.
[11] A. V. Luikov, *Analytical Heat Diffusion Theory* (Academic, New York, 1968).
[12] M. S. Wei (personal communication).
[13] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
[14] S. Izen (personal communication).

Isotropic homogeneous universe with viscous fluid

N. O. Santos[a)]
Institut Henri Poincaré, 11, rue Pierre et Marie Curie, Paris, France
R. S. Dias
Instituto de Fisica, Universidade Federal de Rio de Janeiro, Rio de Janeiro, Brazil
A. Banerjee
Department of Physics, Jadavpur University, Calcutta, India

(Received 8 May 1984; accepted for publication 31 August 1984)

Exact solutions are obtained for the isotropic homogeneous cosmological model with viscous fluid. The fluid has only bulk viscosity and the viscosity coefficient is taken to be a power function of the mass density. The equation of state assumed obeys a linear relation between mass density and pressure. The models satisfying Hawking's energy conditions are discussed. Murphy's model is only a special case of this general set of solutions and it is shown that Murphy's conclusion that the introduciton of bulk viscosity can avoid the occurrence of space-time singularity at finite past is not, in general, valid.

I. INTRODUCTION

The investigation of cosmological solutions in Einstein's theory usually has the energy momentum tensor of matter as that due to a perfect fluid. If we want to consider more realistic matter sources we should take into account dissipative processes due to viscosity. Belinskii and Khalatnikov[12] have previously made qualitative analyses of such models with bulk as well as shear viscosity. It is obviously of some interest to have available examples of exact solutions. Murphy[3] constructed a homogeneous isotropic cosmological model with a fluid containing bulk viscosity. The model obtained by Murphy possessed an interesting feature that the big bang type singularity of infinite space-time curvature does not occur at finite past. But the relationship assumed by Murphy between the viscosity coefficient and the matter density is not acceptable at large density. In the present paper we have used a more reasonable relation to describe the characteristics of the models near high density. The asymptotic forms of the viscosity coefficients for small and large values of the energy limit for some simple cases are given approximately by some power function of the mass density such as $\eta = \alpha\rho^{\nu}$. For large values of ρ, ν is quite small and one would expect in this the condition $0<\nu<\frac{1}{2}$. For small density, ν may even be equal to unity as used in Murphy's work for simplicity. Exact solutions are here obtained for different ranges of ν, that is, for $0<\nu<\frac{1}{2}$, $\nu=\frac{1}{2}$, and $\nu>\frac{1}{2}$. The equation of state assumed $p=\lambda\rho$. In the models constructed for $0<\nu<\frac{1}{2}$ it has been observed in some simple choices of ν for radiation universes that the big bang singularity of infinite density occurs at finite past and thus the conclusions in this case differ from those obtained by Murphy. Although these results are observed for some special choices of parameters like λ and ν, it appears that it is true perhaps for other choices, too. The result, however, that the collapse is unavoidable for an irrotational fluid with Hawking's energy condition[4] being satisfied, can be shown from Raychaudhuri's[5] equation below:

$$\dot\theta = -2\sigma^2 - \tfrac{1}{3}\theta^2 + R_{\mu\nu}v^{\mu}v^{\nu},$$

where θ and σ have the usual significances of the expansion and shear scalars, respectively. If the energy condition is satisfied, $R_{\mu\nu}v^{\mu}v^{\nu}<0$ and, consequently, $\dot\theta<0$. The meaning of this result is simple. A contracting system will collapse to a point and an expanding one will gradually slow down the speed. The effect of viscosity is more prominent at the beginning where θ, the expansion scalar, and ρ, the mass density, are quite large. At later stages the viscosity may play only an insignificant role. The value of λ, of course, is not constant throughout the period of evolution for a realistic model. It will change, as pointed out by Murphy, as the proportions of nonrelativistic particles, radiation, and other particles change.

II. SOLUTIONS OF EINSTEIN'S EQUATIONS FOR A VISCOUS FLUID ($\kappa = 0$ CASE)

The most general line element for a spatially isotropic universe is given by the Robertson–Walker metric in the following form:

$$ds^2 = -dt^2 + R^2(t)[dr^2/(1-\kappa r^2) + r^2\, d\Omega^2]. \tag{2.1}$$

In the above metric $R(t)$ is the well-known scale factor and κ stands for the spatial curvature constant; κ may be $+1, 0$, or -1. We assume that the fluid has bulk viscosity, so that its energy momentum tensor can be written as

$$T_{\mu\nu} = (\rho + \bar p)v_\mu v_\nu + \bar p g_{\mu\nu},$$

where $\bar p = p - \eta v^{\mu}{}_{;\mu}\, \rho$, v_μ, and p have the usual meanings; and η stands for the coefficient of bulk viscosity. The two independent equations are

$$3\kappa/R^2 + 3(\dot R/R)^2 = 8\pi\rho, \tag{2.2}$$

$$\kappa + 2R\ddot R + \dot R^2 = -8\pi\bar p R^2. \tag{2.3}$$

Combining (2.2) and (2.3) one can immediately obtain

$$\frac{\kappa}{R^2} + \frac{2\ddot R}{R} + \left(\frac{\dot R}{R}\right)^2 = -8\pi\left(\lambda\rho - 3\alpha\rho^{\nu}\frac{\dot R}{R}\right). \tag{2.4}$$

In (2.4) we have utilized the equation of state $p=\lambda\rho (0<\lambda<1)$ along with an assumption like

$$\eta = \alpha\rho^{\nu} \quad (0<\nu<\tfrac{1}{2}),$$

[a] On leave from the Instituto de Fisica, Universidade Federal de Rio de Janeiro, Rio de Janeiro, Brazil.

 0022-2488/85/040878-04$02.50

where α, ν, λ are constants. Equation (2.4) can now, in view of (2.2), be written in the form

$$2R\ddot{R} = -(3\lambda + 1)(\dot{R}^2 + \kappa) + (8\pi)^{(1-\nu)}3^{(1+\nu)}\alpha \times (R\ddot{R}/R^{2\nu})(\dot{R}^2 + \kappa)^\nu. \quad (2.5)$$

Writing $A = (3\lambda + 1)$ and $B = -(8\pi)^{(1-\nu)}3^{(1+\nu)}\alpha$, Eq. (2.5) becomes

$$2R\ddot{R} + A(\dot{R}^2 + \kappa) + B(R\ddot{R}/R^{2\nu})(\dot{R}^2 + \kappa)^\nu = 0. \quad (2.6)$$

The differential equation (2.6) involves only R and its time derivatives. The general solution of this equation is not easy to obtain. We may try for solutions with the assumption that the time derivative of the scale factor, that is $\dot{R}$, is itself a function of R. In that case, writing y for $(\dot{R}^2 + \kappa)$ the equation (2.6) can be written in the form

$$R\frac{dy}{dR} + Ay + BR^{(1-2\nu)}y^\nu(y-\kappa)^{1/2} = 0. \quad (2.7)$$

At this stage we integrate (2.7) for $\kappa = 0$, that is, for zero spatial curvature, and later we do it for more complicated situations $\kappa = \pm 1$.

With $\kappa = 0$ and defining $\mu = R^4 y$, Eq. (2.7) becomes

$$\mu^{-(\nu+1/2)}\frac{d\mu}{dR} + BR^{[(1/2)A - \nu(A+2)]} = 0, \quad (2.8)$$

and the integration of which yields for $\nu \neq \frac{1}{2}$,

$$\frac{1}{\frac{1}{2}-\nu}\mu^{(1/2-\nu)} + B\frac{R^{(1/2)A + 1 - \nu(A+2)}}{\frac{1}{2}A + 1 - \nu(A+2)} = C_1,$$

where C_1 is an integration constant. Going back to the original notation, we have

$$\dot{R}/R = [CR^{-(1/2+1)(1-2\nu)} - B/(A+2)]^{1/(1-2\nu)}, \quad (2.9)$$

with C being a constant quantity.

The integration of (2.8) for $\nu = \frac{1}{2}$ yields

$$R^{[(1/2)(A+B)+1]} = [\tfrac{1}{2}(A+B)+1](ct+\alpha), \quad (2.10)$$

where c and d are integration constants.

We now proceed to study the behavior of the models given in (2.9) and (2.10). Three different cases, (I) $2\nu > 1$, (II) $2\nu < 1$, and (III) $2\nu = 1$, are studied separately in the following. The case treated by Murphy $\nu = 1$ may be valid at low density. Near the big bang, $0 \leqslant \nu \leqslant \frac{1}{2}$ is a more appropriate assumption (Ref. 1).

Case I: $2\nu > 1$. In this case we have

$$\dot{R}/R = 1/[b^2 + CR^{[(1/2)A+1](2\nu-1)}]^{1/(2\nu-1)}, \quad (2.11)$$

where b^2 is written for $|B|/(A+2)$. In Murphy's case $\nu = 1$ and $b^2 = 3\alpha/\gamma$, where γ in his paper is related to our λ by $\lambda = (\gamma - 1)$. Here as $R \to 0$, $\dot{R}/R$ approaches a finite magnitude showing a steady-state characteristic and the motion is very slow ($\dot{R} \to 0$ at this limit). Murphy discussed three cases. In all these three cases the models have $R = 0$ singularity with finite mass density ρ at infinite time past. For other values of ν in the range $2\nu > 1$ we get similar characteristics. For example, if the fluid content is radiation, that is, $(A/2 + 1) = 2$, and further $2\nu = 3/2$, we get explicitly after integration

$$b^2 \ln R + \tfrac{1}{2}CR^2 = t + t_0,$$

t_0 being the integration constant. Then, $R \to 0$ corresponds to $t \to -\infty$. The model starts from infinite past at zero proper volume.

It is worth noting that when $R \to 0$, $\dot{R}$, $\ddot{R}$, and all other higher time derivatives vanish at the same time.

For $C < 0$, the universe may expand from a finite density at $R = 0$ and increase to an infinite density developing a singularity $\dot{R}/R \to \infty$, or the universe may explode from this singularity itself. Murphy describes that at this epoch the observers "run out of time."

Case II: $2\nu < 1$. In this case the relation (2.11) can be written as

$$\dot{R}/R = [b^2 + CR^{(1/2A+1)(1-2\nu)}]^{1/(1-2\nu)}. \quad (2.12)$$

Here the model with $C > 0$ explodes from $R = 0$, where $\dot{R}$ is infinitely large and the density ρ is also infinitely large.

For $\nu = 0$ one can integrate (2.12) and get

$$b^2 R^{[(1/2)A+1]} = \exp[\tfrac{1}{2}(t + t_0)|B|] - C,$$

t_0 being an integration constant. For $2\nu = \frac{1}{2}$ the integration yields

$$\frac{1}{b^4}\ln(b^2R + C) = t + t_0 + \frac{1}{b^2}\frac{R}{(b^2R + C)}.$$

It is easy to conclude that for the above two special cases the singularity $R = 0$ occurs at finite past, which is a distinct deviation from Murphy's case ($\nu = 1$) or other value of ν with $2\nu > 1$.

For $C = 0$, we have $\dot{R}/R = \text{const}$ and represents a steady-state model with $(R - t)$ curve as an exponential one.

For $C < 0$ one notes that $\dot{R} = 0$ for a finite magnitude $R = R_0$, where

$$|C|R^{-[(1/2)A+1](1-2\nu)} = b^2.$$

The density is zero at this stage and gradually increases as R increases till the density attains a finite maximum when $R(t)$ becomes indefinitely large. Thus in this case the universe evolves from a finite dimension.

Case III: $2\nu = 1$. The solution for $R(t)$ in this case can be explicitly given. It is

$$R^q = ct + d, \quad (2.13)$$

where $q = ((A+B)/2 + 1)$ and c, d are arbitrary constants. Since

$$\dot{R}/R = (c/q)R^{-1},$$

one observes that $(c/q) > 0$ leads to an expansion and $(c/q) < 0$ to a contraction. The universe is a monotonically expanding or contracting one. For collapse the point singularity is reached in finite time. This is the singularity of $R \to 0$, $\dot{R}/R \to -\infty$, and $\rho \to \infty$. On the other hand, the model may explode from a pointlike singularity and increase in size in course of time with decrease of density. The behavior is somewhat similar to the Robertson–Walker model with zero spatial curvature $\kappa = 0$.

Hawking–Penrose energy conditions. By using Einstein's field equations one finds that the Hawking–Penrose energy condition $R_{\mu\nu}v^\mu v^\nu \leqslant 0$ leads to the relation $(\rho + 3\bar{p}) \geqslant 0$, which again in turn implies, in view of field equations, $\ddot{R} \leqslant 0$. The other energy condition mentioned in Murphy's paper, for example, $(\rho + \bar{p}) \geqslant 0$, is then automatically satisfied. Equation (2.6) for $\kappa = 0$ is now

$$2R\ddot{R}+A\dot{R}^2-|B|(\dot{R}^{(2\nu+1)}/R^{(2\nu-1)})=0. \tag{2.14}$$

If we demand $\ddot{R}\leqslant 0$, the relation (2.14) yields

$$A(\dot{R}/R)^2 \geqslant |B|(\dot{R}/R)^{(2\nu+1)}.$$

Denoting the so-called Hubble's constant $(\dot{R}/R)$ by the symbol H, we write the above condition as

$$H^{(1-2\nu)} \geqslant |B|/A. \tag{2.15}$$

It is evident from (2.11) and (2.12) that for both $2\nu>1$ and $2\nu<1$ the condition (2.15) is not satisfied when the constant $C\leqslant 0$. It means that except for C positive the energy conditions are violated throughout the evolution. But for $C>0$ and $2\nu<1$ the energy condition which is equivalent to (2.15) is satisfied at the initial stage of expansion and violated at later stages, whereas for $C>0$ and $2\nu>1$ the energy condition is violated at the beginning and satisfied at later stages of evolution. When $2\nu=1$, the solution is (2.13), and $\ddot{R}\leqslant 0$ needs the condition $A\geqslant|B|$ to be satisfied, which merely restricts the magnitudes of parameters λ and α.

III. SOLUTIONS FOR $\kappa=\pm 1$

It is more difficult to solve in general the differential equation (2.7) for nonzero κ, that is for $\kappa=\pm 1$. However, we are able to integrate it in the case $\nu=\frac{1}{2}$. The integration yields

$$\begin{aligned}\ln(C_1R) = &-(A/F)\ln(yF+G)\\ &+(B/F)\ln[2(y^2-\kappa y)^{1/2}+(2y-\kappa)]\\ &+(A/F)\ln(yF+G)+(A/F)\ln[2\kappa AB\\ &\times(y^2-\kappa y)^{1/2}+\kappa^2B^2-(\kappa F+2\kappa B^2)y],\end{aligned} \tag{3.1}$$

where

$$y=(\dot{R}^2+\kappa),\quad F=(A^2-B^1)\neq 0,\quad G=\kappa B^2.$$

For those expressions of y, F, and Q we can write (3.1) in the form

$$DR^{(A^2-B^2)/A}=\frac{2\kappa|B|A[\dot{R}^2(\dot{R}^2+\kappa)]^{1/2}-(A^2+B^2)\kappa\dot{R}^2-A^2\kappa^2}{[\dot{R}^2(A^2-B^2)+\kappa A^2]^2\{2[\dot{R}^2(\dot{R}^2+\kappa)]^{1/2}+2\dot{R}^2+\kappa\}^{|B|/A}}, \tag{3.2}$$

with D being a constant. It is quite difficult to perform the second integration. However, it is possible to study the motion to some extent from Eq. (3.2). From the field equations (2.2) and (2.3) we have

$$8\pi(\rho+3\bar{p})=-6\ddot{R}/R, \tag{3.3}$$

$$8\bar{\eta}(\rho+\bar{p})=2\frac{\kappa}{R^2}+2\left(\frac{\dot{R}}{R}\right)^2-2\left(\frac{\ddot{R}}{R}\right), \tag{3.4}$$

and the mass density ρ is given by

$$8\bar{\eta}\rho=3(\kappa+\dot{R}^2)/R^2. \tag{3.5}$$

If we restrict to positive mass density the energy conditions $(\rho+3\bar{p})\geqslant 0$ and $(\rho+\bar{p})\geqslant 0$ are both satisfied independent of the sign of κ provided $\ddot{R}\leqslant 0$.

Properties of the solutions. The relation (3.2) can be written, after a little manipulation, as

$$\begin{aligned}DR^{(A^2-B^2)/A} &= -\kappa[A(\dot{R}^2+\kappa)^{1/2}+|B|\dot{R}]^{-2}\\ &\quad\times[(\dot{R}^2+\kappa)^{1/2}+\dot{R}]^{-2|B|/A};\end{aligned} \tag{3.6}$$

$|B|$ stands for the magnitude of B, where $B=-(8\pi)^{1/2}(27)^{1/2}\alpha$ and $A=(3\lambda+1)$. In view of (3.6) it is evident that D is greater, equal to, or less than zero accordingly as κ is less than, equal to, or greater than zero. We consider below four different situations.

Case I: $A^2>B^2$, $\kappa=+1$ and so $D<0$. Writing $-\alpha^2$ for D (3.6) may be written as

$$\begin{aligned}a^2R^{(A^2-B^2)/A} = &[A(\dot{R}^2+1)^{1/2}\\ &+|B|\dot{R}]^{-2}[(\dot{R}^2+1)^{1/2}+\dot{R}]^{-2|B|/A},\end{aligned} \tag{3.7}$$

and from (2.6),

$$-2R\ddot{R}=A(\dot{R}^2+1)-|B|\dot{R}(\dot{R}^2+1)^{1/2}. \tag{3.8}$$

Here $\ddot{R}<0$ and hence energy conditions are satisifed. There is a big bang type singularity, where the explosion takes place with $\dot{R}\to\infty$ and expansion continues up to a finite volume, where $\dot{R}=0$. This corresponds to the point of maximum expansion analogous to the case of the Friedman universe.

Case II: $A^2>B^2$, $\kappa=-1$ and therefore $D>0$. Writing now $D=a^2$, we have

$$\begin{aligned}a^2R^{(A^2-B^2)/A} = &[A(\dot{R}^2-1)^{1/2}+|B|\dot{R}]^{-2}\\ &\times[(\dot{R}^2-1)^{1/2}+\dot{R}]^{-2|B|/A}.\end{aligned} \tag{3.9}$$

From (2.6) with $\nu=\frac{1}{2}$ and $\kappa=-1$ one has

$$-2R\ddot{R}=|B|\dot{R}(\dot{R}^2-1)^{1/2}-A(\dot{R}^2-1). \tag{3.10}$$

We have $\ddot{R}<0$ provided $\dot{R}^2\geqslant A^2/(A^2-B^2)$. In this case, therefore, the expansion starts from $R\approx 0$ and continues till $\dot{R}^2=A^2/(A^2-B^2)$, when $\ddot{R}$, $\dot{R}$, and higher time derivatives vanish. Further, in future R continues to increase, with the expansion rate $\dot{R}$ becoming slower till $\dot{R}=1$, where we have $\rho=0$. Our solution (3.9) is not valid for $\dot{R}^2<1$.

Case III: $A^2<B^2$, $\kappa=+1$ and so $D<0$. It is not difficult to see that in this model R and $\dot{R}$ simultaneously become very large and the density increases indefinitely. For $A^2<B^2$ and $\kappa=-1$ the energy conditions are violated. These two models are not physically meaningful.

Case IV: $A^2=B^2$, $\kappa=+1$. For an expanding model, that is, for $\dot{R}>0$ only when $\kappa=+1$, one has $\ddot{R}<0$ and the energy conditions are satisfied. When $\kappa=-1$, though the energy conditions are violated in an expanding model, they are satisfied for the collapsing situation, where $\dot{R}$ is negative. Now if one looks for the solution, one finds that (3.2) is not valid in the case $A^2=B^2$. Integration of (2.7) for $\nu=\frac{1}{2}$ and $\kappa=+1$ yields

$$\begin{aligned}CR^A = &[(\dot{R}^2+1)^{1/2}+\dot{R}]\\ &\times\exp\{(\dot{R}^2+1)^{1/2}[(\dot{R}^2+1)^{1/2}+\dot{R}]\}.\end{aligned} \tag{3.11}$$

From (3.11) it is obvious that one can construct an exploding model from a big bang type singularity ($\dot{R}\to\infty$) which slows down in course of time to a finite dimension. Similarly there can be a collapsing model to a singularity with the beginning at a finite dimension.

IV. CONCLUSION

It may be more realistic to assume the presence of anisotropy in the early universe and this anisotropy might have been smoothed out by the dissipative mechanism such as viscosity in course of evolution (Misner,[6] 1969; Weinberg,[7] 1972). The model discussed here is isotropic and homogeneous and, in view of the assumption of isotropy, the shear viscosity cannot exist. The effect of bulk viscosity changes the behavior from that of a perfect fluid model. Murphy introduced bulk viscosity and observed that the big bang type singularity may be avoided in the finite past. The conclusion arrived at by Murphy is not, in general, true. It is shown explicitly in the above that for certain choices of ν in the range $2\nu \leqslant 1$ the model expand from zero proper volume at finite past. There is no problem with the Hawking–Penrose energy conditions for $\kappa = 0$ and $\nu = \frac{1}{2}$, because it is satisfied throughout the evolution for certain ranges of λ and α. But the energy conditions are satisfied for part of this period either at initial stages or at later stages according as $2\nu < 1$ or $2\nu > 1$.

For models with $\kappa = \pm 1$ the behavior has some similarities with the corresponding Robertson–Walker models. For example, with $\kappa = +1$ and $A^2 > B^2$ the model explodes from a big bang type space-time.

ACKNOWLEDGMENT

One of us (N.O.S.) is grateful to CAPES of Brazil for financial support.

[1]U. A. Belinskii and I. M. Khalatnikov, Sov. Phys. JETP **42**, 205 (1976).
[2]U. A. Belinskii and I. M. Khalatnikov, Sov. Phys. JETP **45**, 1 (1977).
[3]G. L. Murphy, Phys. Rev. D **8**, 4231 (1973).
[4]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space Time* (Cambridge U. P., Cambridge, 1973).
[5]A. K. Raychaudhuri, *Theoretical Cosmology* (Clarendon, Oxford, 1979).
[6]C. Misner, Phys. Rev. Lett. **22**, 1071 (1969).
[7]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).

ERRATA

Erratum: On some general properties of the point spectrum of three particles moving in one dimension [J. Math. Phys. 25, 2589 (1984)]

J. Fernando Perez, C. P. Malta, and F. A. B. Coutinho
Departmento de Física Matemática, Instituto de Física da USP, São Paulo, Brazil

(Received 30 November 1984; accepted for publication 14 December 1984)

We found some misprints and mistakes after expression (25) in Sec. IV of our work. The corrections are the following.

(1) The assumption on the energies ϵ_0, ϵ_1, and ϵ_2 is that $\epsilon_0 < \epsilon_1 < \epsilon_2$ such that $\epsilon_0 + \epsilon_2 < 2\epsilon_1$ (instead of $\epsilon_0 + \epsilon_1 < 2\epsilon_2$ as printed).

(2) The paragraph after result R2 should be replaced by the following.

The Hall theorem also tells us that $(\epsilon_0 + \epsilon_2)$ is a lower bound both for $\psi_A^{(+)}$ and for an excited state of type $\psi_S^{(+)}$ and $2\epsilon_1$ is a lower bound for an excited state of type $\psi_S^{(+)}$. Due to the assumption $\epsilon_0 + \epsilon_2 < 2\epsilon_1$ it follows that $\epsilon_0 + \epsilon_2$ is a lower bound for the first excited state of type $\psi_S^{(+)}$. Therefore we have the result R3.

(3) In result R3 the assumption $2\epsilon_2 - E_{2B}^0 > 0$ should be replaced by $\epsilon_0 + \epsilon_2 - E_{2B}^0 > 0$.

We have now a new result concerning interception of the bands, if the two-body energies $\epsilon_0 < \epsilon_1 < \epsilon_2$ are such that $\epsilon_0 + \epsilon_2 > 2\epsilon_1$. In this case, the Hall theorem tells us that $2\epsilon_1$ is the lower bound for the first excited state of type $\psi_S^{(+)}$. Therefore, if $\epsilon_0 + \epsilon_2 - E_{2B}^1 > 0$ and $2\epsilon_1 - E_{2B}^0 < 0$, $\psi_A^{(+)}$ is unbound while the first excited state of type $\psi_S^{(+)}$ might be bound and the excited state band will intercept the ground state band.

Erratum: A rule for the total number of topologically distinct Feynman diagrams [J. Math. Phys. 25, 3489 (1984)]

Franco Battaglia and Thomas F. George
Department of Chemistry, University of Rochester, Rochester, New York 14627

(Received 19 December 1984; accepted for publication 4 January 1985)

In Table I, the value of $F(n)$ for $n = 7$ should be 1 708 394 instead of 1 708 391. We thank Dr. A. E. Jacobs for pointing out the misprint and his paper,[1] which gives the same results with a different method.

[1] A. E. Jacobs, Phys. Rev. D 23, 1760 (1981). See also Ref. 2 quoted therein.

0022-2488/85/040882-01$02.50