# Variational nature of eigenvalue problems with a parameter

James L. Schwarzmeier

*Courant Institute of Mathematical Sciences, New York University, New York, New York 10012*

Keith R. Symon

*Department of Physics, University of Wisconsin, Madison, Wisconsin 53706*

(Received 21 August 1978)

The following theorem is proved. Let $D(\omega)$ be an operator with eigenvalues and eigenfunctions $\{d_k(\omega), \nu_k(\omega)\}$, where $\omega$ is a complex parameter. Given a complex number $d_{k0}$, let $\omega_0$ be such that $d_k(\omega_0) = \langle\tilde{\nu}_k(\omega_0)|D(\omega_0)|\nu_k(\omega_0)\rangle = d_{k0}$, where $\tilde{\nu}_k(\omega_0)$ is the dual eigenfunction to $\nu_k(\omega_0)$. Suppose $\psi$ and $\tilde{\psi}$ approximate $\nu_k(\omega_0)$ and $\tilde{\nu}_k(\omega_0)$, respectively, to order $\epsilon$. Then, if $D(\omega)$ is analytic in $\omega$ in the neighborhood of $\omega_0$, and if $\omega'$ is such that $\langle\tilde{\psi}|D(\omega')|\psi\rangle = d_{k0}$, $\omega'$ usually will approximate $\omega_0$ to order $\epsilon^2$. By applying this theorem it is shown that roots of the inhomogeneous plasma dispersion relation usually will be accurate to second order if the associated normal modes and their duals are known merely to first order. The theorem can also be applied to solutions of the dispersion relation in a truncated function space.

## INTRODUCTION

A standard result from operator theory is that the eigenvalues of a Hermitian operator can be calculated to second order accuracy if the trial eigenfunctions used are accurate merely to first order. Here we consider a related but less obvious result. We consider an eigenvalue problem in which a complex parameter $\omega$ appears, and we prove a theorem about how accurately the parameter $\omega$ can be calculated such that an approximate eigenvalue (which is a function of $\omega$) attains a specified value. Then we apply the theorem to the problem of determining the roots of the inhomogeneous plasma dispersion relation. A formula is derived for the value of $\omega$ for which an approximate eigenvalue of the dispersion matrix vanishes; we show that this procedure is usually equivalent to solving approximately the dispersion relation. We refer to a procedure (given in Ref. 1) for determining *a priori* an approximate normal mode, and, for the case of Hamiltonian systems, show that an approximate dual to the normal mode follows automatically. Finally, we prove a theorem relating roots of the dispersion relation for a small truncation of the dispersion matrix to roots of the exact dispersion relation.

## FORMULATION OF THE PROBLEM

Let $D(\omega)$ be an operator on vectors in a Hilbert space, where the complex number $\omega$ appears as a parameter. The eigenvalues and eigenvectors of $D(\omega)$ satisfy

$$D(\omega)\nu_k(\omega) = d_k(\omega)\nu_k(\omega). \tag{1}$$

$D(\omega)$ may not be Hermitian so we cannot, in general, guarantee that its eigenfunctions are orthogonal. [We note that the eigenfunctions may also not be complete, that is, two or more of them may become degenerate, if two or more eigenvalues $d_k(\omega)$ are degenerate. Typically, i.e., if $D(\omega)$ depends in a nontrivial way on $\omega$, this will happen only at certain discrete values of $\omega$. This possibility does not invalidate our present argument.] Thus we introduce the dual set $\{\tilde{\nu}_k(\omega)\}$ (a tilde will always denote adjoint quantities), belonging to the adjoint operator $[D(\omega)]^\dagger$:

$$[D(\omega)]^\dagger\tilde{\nu}_k(\omega) = \tilde{d}_k(\omega)\tilde{\nu}_k(\omega). \tag{2}$$

The following relationships are satisfied (we use Dirac's bracket notation for inner products):

$$\langle\tilde{\nu}_k|\nu_{k'}\rangle = \delta_{kk'} \tag{3}$$

and

$$\tilde{d}_k(\omega) = [d_k(\omega)]^*, \tag{4}$$

where an asterisk denotes complex conjugation. From Eqs. (1) and (3) we obtain the eigenvalue $d_k(\omega)$ as

$$d_k(\omega) = \frac{\langle\tilde{\nu}_k|D(\omega)|\nu_k\rangle}{\langle\tilde{\nu}_k|\nu_k\rangle} = \langle\tilde{\nu}_k|D(\omega)|\nu_k\rangle. \tag{5}$$

Let us now focus on a particular eigenvalue $d_k(\omega)$ and henceforth drop the subscript $k$. Given an arbitrary complex number $d_0$, let $\omega_0$ be such that

$$d(\omega_0) = \langle\tilde{\nu}_0|D(\omega_0)|\nu_0\rangle = d_0, \tag{6}$$

where we have introduced the notation $\nu_0 \equiv \nu(\omega_0)$ and $\tilde{\nu}_0 \equiv \tilde{\nu}(\omega_0)$. Suppose that $\nu_0$ and $\tilde{\nu}_0$ are not known exactly but only to order $\epsilon$:

$$\psi = \nu_0 + \epsilon\nu_1 \tag{7a}$$

$$\tilde{\psi} = \tilde{\nu}_0 + \epsilon\tilde{\nu}_1, \tag{7b}$$

where $\epsilon \ll 1$. The eigenvectors $\nu_0$ and $\tilde{\nu}_0$ have been normalized to one, and normalizing $\psi$ and $\tilde{\psi}$ similarly, $\langle\tilde{\psi}|\psi\rangle = 1$, implies that

$$\langle\tilde{\nu}_0|\nu_1\rangle + \langle\tilde{\nu}_1|\nu_0\rangle = -\epsilon\langle\tilde{\nu}_1|\nu_1\rangle. \tag{8}$$

Now let us solve an equation similar to Eq. (6); namely, let us solve for the value of $\omega'$ near $\omega_0$, for which the approximate eigenvalue of $D(\omega')$ equals the number $d_0$:

$$\langle\tilde{\psi}|D(\omega')|\psi\rangle = d_0. \tag{9}$$

 0022-2488/79/060979-05$01.00  

We have the following theorem.

*Theorem I*: Let $\psi$ and $\tilde{\psi}$ approximate to order $\epsilon$ a particular eigenfunction and its dual, respectively, of the operator $D(\omega)$, for a particular value $\omega_0$ of the complex parameter $\omega$. Let $\omega_0$ be such that the corresponding eigenvalue of $D(\omega)$ evaluated at $\omega_0$ equals a given complex number $d_0$, $d(\omega_0) = d_0$. If $D(\omega)$ is an analytic function of $\omega$ in the neighborhood of $\omega_0$, then the value $\omega'$ which satisfies Eq. (9), will approximate $\omega_0$ to order $\epsilon^2$, except in special circumstances. (The special circumstances will be exhibited in the proof.)

*Proof*: If $D(\omega)$ is an analytic functionof $\omega$ in the neighborhood of $\omega_0$, we have the expansion

$$D(\omega) = D(\omega_0) + (\omega - \omega_0)\left(\frac{\partial D}{\partial \omega}\right)_{\omega_0} + \tfrac{1}{2}(\omega - \omega_0)^2\left(\frac{\partial^2 D}{\partial \omega^2}\right)_{\omega_0} + \cdots$$
$$= D_0 + (\omega - \omega_0)D'_0 + \tfrac{1}{2}(\omega - \omega_0)^2 D''_0 + \cdots. \quad (10)$$

Substitute Eqs. (7) into Eq. (9) to obtain

$$d_0 = \langle \tilde{\nu}_0 + \epsilon\tilde{\nu}_1 | D(\omega') | \nu_0 + \epsilon\nu_1 \rangle$$
$$= \langle \tilde{\nu}_0 | D(\omega') | \nu_0 \rangle + \epsilon(\langle \tilde{\nu}_1 | D(\omega') | \nu_0 \rangle + \langle \tilde{\nu}_0 | D(\omega') | \nu_1 \rangle) + \epsilon^2 \langle \tilde{\nu}_1 | D(\omega') | \nu_1 \rangle. \quad (11)$$

In the last equation, for $D(\omega')$ substitute the expansion (10). Using Eqs. (4), (6), and (8) and the definition of the adjoint, we obtain

$$0 = (\omega' - \omega_0)\langle \tilde{\nu}_0 | D'_0 | \nu_0 \rangle + \tfrac{1}{2}(\omega' - \omega_0)^2 \langle \tilde{\nu}_0 | D''_0 | \nu_0 \rangle + \epsilon(\omega' - \omega_0)(\langle \tilde{\nu}_1 | D'_0 | \nu_0 \rangle + \langle \tilde{\nu}_0 | D'_0 | \nu_1 \rangle) + \epsilon^2 \langle \tilde{\nu}_1 | D_0 - d_0 | \nu_1 \rangle + \cdots. \quad (12)$$

We now assume that

$$\langle \tilde{\nu}_0 | D'_0 | \nu_0 \rangle \neq 0, \quad (13)$$

and that $\epsilon$ and $\omega' - \omega_0$ are sufficiently small so that

$$\epsilon |\langle \tilde{\nu}_1 | D'_0 | \nu_0 \rangle + \langle \tilde{\nu}_0 | D'_0 | \nu_1 \rangle| \ll |\langle \tilde{\nu}_0 | D'_0 | \nu_0 \rangle| \quad (14)$$

and

$$|(\omega' - \omega_0)\langle \tilde{\nu}_0 | D''_0 | \nu_0 \rangle| \ll |\langle \tilde{\nu}_0 | D'_0 | \nu_0 \rangle|. \quad (15)$$

According to the Hellman–Feynman theorem,[2,3] condition (13) will fail if $\omega_0$ is a multiple root of Eq. (6). Condition (14) requires that $\omega_0$ not be, even approximately, a multiple root of Eq. (6). Failure of this condition is the special circumstance referred to in the theorem. If conditions (14) and (15) hold, we may neglect all but the first and fourth terms in Eq. (12) to obtain

$$\omega' = \omega_0 + \epsilon^2 \frac{\langle \tilde{\nu}_1 | d_0 - D_0 | \nu_1 \rangle}{\langle \tilde{\nu}_0 | D'_0 | \nu_0 \rangle} + O(\epsilon^3). \quad (16)$$

The result (16) shows that condition (15) will hold if condition (14) holds and $\epsilon$ is small enough. Condition (13) is sufficient to guarantee that condition (14) holds if $\epsilon$ is sufficiently small.

## APPLICATION TO THE PLASMA DISPERSION RELATION

Now we apply the theorem just proved to the solution of the inhomogeneous plasma dispersion relation. After we establish the connection between the theorem and the solution of the dipersion relation, we show that roots of the dispersion relation usually will be accurate to second order if the associated normal modes and their duals are known merely to first order. This is what is meant by the phrase "variational nature" of solutions of the dispersion relation. The result to be proved here was first discovered[4,5] numerically while studying the stability of Bernstein–Greene–Kruskal[6] equilibria using the recent stability formalism of Lewis and Symon.[1] Quantities associated with the approximate normal mode appear in several places in the dispersion relation. A note of caution here,[4,5] from a computational point of view, is that the variational nature of solutions of the dispersion relation holds only if the *same* approximation to the normal mode is used *everywhere* in the dispersion relation.

If the perturbation distribution function is eliminated from the linearized Vlasov and field equations, the self-consistent, Laplace transformed perturbation potential, $\hat{\phi}_1(x,\omega)$, is determined from the equation[1]

$$D(x,\omega)\hat{\phi}_1(x,\omega) = I(x,\omega), \quad (17)$$

where $D(x,\omega)$ is the dispersion operator, which operates on functions of $x$, $I(x,\omega)$ is the initial value term for the perturbation, $\omega$ is the Laplace transform variable, and a caret over a variable denotes the Laplace transform of the variable. In general $x$ stands for the vector $\vec{x}$, $\hat{\phi}_1(x,\omega)$ may be a column vector containing the perturbation scalar and vector potentials as components, $D(x,\omega)$ is a dispersion operator matrix, and $I(x,\omega)$ is the vector of initial perturbations.

The solution of Eq. (17) is

$$\hat{\phi}(x,\omega) = D^{-1}(x,\omega)I(x,\omega). \quad (18)$$

The normal modes of $\phi(x,t)$ arise from the poles of $\hat{\phi}(x,\omega)$. If we imagine evaluating the right member of Eq. (18) by expanding in eigenfunctions of $D(x,\omega)$, we see that the poles occur at frequencies $\omega_0$, where $D(x,\omega_0)$ has a zero eigenvalue $d_k(\omega_0) = 0$, and that the form of the normal mode is given by the corresponding eigenfunction $\nu_0(x)$ of $D(x,\omega_0)$. If we know an approximation $\psi(x)$ to $\nu_0(x)$, and a corresponding approximation $\tilde{\psi}(x)$ to the dual $\tilde{\nu}_0(x)$, good to order $\epsilon$, then our Theorem I shows that a solution $\omega'$ of the approximate dispersion relation

$$\langle \tilde{\psi}(x) | D(x,\omega') | \psi(x) \rangle = \int dx [\tilde{\psi}(x)]^* D(x,\omega')\psi(x) = 0 \quad (19)$$

will approximate $\omega_0$ to order $\epsilon^2$, provided $\omega_0$ is a simple zero of $d_k(\omega)$.

All that remains before applying Theorem I to Eq. (19) is to establish that $D(\omega)$ has the required analytic properties in $\omega$. In the work of Lewis and Symon it is shown that the dispersion operator $D(\omega)$ is an analytic function of $\omega$ for complex $\omega$, and that $D(\omega)$ has discrete branch points along the real $\omega$ axis. $D(\omega)$ is defined originally with its branch cut along the real $\omega$ axis, but $D(\omega)$ can be analytically continuted from the upper half $\omega$-plane into the lower half $\omega$-plane by redefining the branch cuts, as long as the equilibrium distribution function and initial data are analytic as required. Therefore, even if $\omega_0$ is a real frequency which is not a branch point of $D(\omega)$ (we do not consider such a possibility), the expansion (10) is still valid, but then it must be understood that it is the analytic continuation of the dispersion operator that is used.

Now the theorem can be applied, so from Eq. (16) (with $d_0 = 0$) the final formula for the approximate eigenfrequency $\omega'$ is

$$\omega' = \omega_0 - \epsilon^2 \frac{\langle \tilde{v}_1 | D(x,\omega_0) | v_1 \rangle}{\langle \tilde{v}_0 | D'(x,\omega_0) | v_0 \rangle}. \tag{20}$$

This result holds if $\omega_0$ is a simple zero of a nondegenerate eigenvalue $d(\omega_0)$. [It is easy to show that $\omega_0$ is a simple zero of $d(\omega)$ if and only if $\langle \tilde{v}_0 | D_0' | v_0 \rangle \neq 0$].[2,3]

In practice, we may solve Eq. (17) for $\hat{\phi}(x,\omega)$ by expanding the same in some set of basis functions $\{\eta_n(x)\}$. In the stability analysis of Lewis and Symon[1] for inhomogeneous Vlasov equilibria, the freedom to choose appropriate basis functions is provided by introducing a certain transformation on the particle distribution function. An operator $\Lambda(x)$ is obtained whose eigenfunctions, $\{\eta_n(x)\}$, are used in expanding the perturbation potential. Since $\Lambda(x)$ may not be a Hermitian operator, we introduce the dual set, $\{\tilde{\eta}_n(x)\}$, which are eigenfunctions of the adjoint $[\Lambda]^\dagger$, satisfying the orthogonality condition

$$\langle \tilde{\eta}_n | \eta_{n'} \rangle = \int_R dx [\tilde{\eta}_n(x)]^* \eta_{n'}(x) = \delta_{nn'}. \tag{21}$$

The integral in the last equation is to be performed over all configuration space consistent with whatever boundary conditions exist on $\hat{\phi}_1(x,\omega)$. It is often possible to choose $\Lambda$ so that one of its eigenfunctions, $\eta_1$, say, is a good approximation to an eigenfunction of $D(x,\omega_0)$ so that the above theorem can be applied.[1,4,5]

Let us assume the perturbation potentials of interest can be represented by a finite expansion

$$\hat{\phi}_1(x,\omega) = \sum_{n=1}^{N} \hat{\alpha}_n(\omega) \eta_n(x), \tag{22}$$

where $N$ may be arbitrarily large (though finite). An infinite expansion leads to an infinite-dimensional dispersion matrix whose determinant may not converge. Rather than burden ourselves with determining under what conditions the determinant will be guaranteed to exist, we simply consider a finite (perhaps large) $N \times N$ dispersion matrix and limit the present argument to situations where this approximates well enough the exact case. The Laplace transformed expansion coefficients $\hat{\alpha}_n(\omega)$ are obtained by substituting Eq. (22) into Eq. (17) and inverting:

$$\hat{\alpha}_n(\omega) = \sum_{n'=1}^{N} D_{nn'}^{-1}(\omega) I_{n'}(\omega), \tag{23}$$

where

$$D_{nn'}(\omega) = \langle \tilde{\eta}_n | D(\omega) | \eta_{n'} \rangle = \int_R dx [\tilde{\eta}_n(x)]^* D(x,\omega) \eta_{n'}(x) \tag{24}$$

is the dispersion matrix (in the basis $\{\eta_n, \tilde{\eta}_n\}$), and $I_n(\omega) = \langle \tilde{\eta}_n | I(x,\omega) \rangle$. We will use $D(\omega)$ to denote the dispersion matrix.

The poles of $\hat{\alpha}_n(\omega)$ give rise to normal modes of $\phi_1(x,t)$, so the roots of the dispersion relation or "eigenfrequencies" $\omega_0$ are solutions of the equation

$$\det D(\omega_0) = 0. \tag{25}$$

Since the determinant of a matrix equals the product of its eigenvalues, if $\omega_0$ satisfies the dispersion relation (25), then

$$d_k(\omega_0) = 0. \tag{26}$$

for some $k$. Thus condition (26) is equivalent for a finite-dimensional matrix $D(\omega)$ to the dispersion relation (25). It is possible for $\omega_0$ to be a simple zero of the dispersion relation and yet have Eq. (26) satisfied for *two* $k$'s; in this case, if $\omega_0$ is not real it can be shown[1] that the two corresponding eigenvectors of $D(\omega_0)$ and the two dual eigenvectors of $[D(\omega_0)]^\dagger$ become degenerate (i.e., linearly dependent). We will not consider this special case any further.

The extent to which the dispersion matrix is diagonalized at the frequency $\omega$ is a measure of how well the basis function $\{\eta_n, \tilde{\eta}_n\}$ approximate the eigenfunctions $\{v_k(x,\omega),\ \tilde{v}_k(x,\omega)\}$. Therefore, one useful way to choose a set of basis functions which converges rapidly to a particular normal mode at frequency $\omega_0$ is to choose a basis which will diagonalize approximately the dispersion matrix in the neighborhood of $\omega_0$. In other words, choose a set of basis functions such that one member of the set is a first order approximation to the normal mode, $\psi = \eta_n(x) = v_0 + O(\epsilon)$.

It is shown in Ref. 1 that for a completely Hamiltonian system (particle and field equations derivable from a single variational principle), the operator $D(x,\omega)$ is a Hermitian function of $\omega$, that is, its adjoint is given by

$$[D(x,\omega)]^\dagger = D(x,\omega^*). \tag{27}$$

It follows that if we have an analytic procedure for determining an approximation $\psi(x,\omega)$ to an eigenfunction of $D(x,\omega)$, the same procedure should lead to an approximation to its dual

$$\tilde{\psi}(x,\omega) = \psi(x,\omega^*). \tag{28}$$

This result simplifies the application of Eq. (19) to such systems [with $\psi(x) = \psi(x,\omega_0)$]. This procedure has been demonstrated[4,5] for a class of Bernstein–Greene–Kruskal equilibria, when a crude guess, $\Omega_0$, of the exact eigenfrequency $\omega_0$ is known *a priori*. Thus it is possible to determine *a priori* an approximate normal mode, $\psi(x)$:

$$\psi(x) = v_0 + \epsilon v_1(x), \tag{29}$$

where $\epsilon \ll 1$. $\psi$ may depend parametrically on $\Omega_0$. For the Bernstein–Greene–Kruskal equilibria, the approximate normal mode $\psi$ is independent of $\Omega_0$ to lowest order. The next nonvanishing order involves $\Omega_0^2$, and since for this problem $\Omega_0$ (and indeed $\omega_0$) are purely imaginary, $\Omega_0^{*2} = \Omega_0^2$, so by Eq. (27) the basis functions are dual to themselves.

## SOLUTION IN A TRUNCATED FUNCTION SPACE

Generally the most that one can hope to derive *a priori* is a field operator $\Lambda(x)$, one of whose eigenfunctions is close to the normal mode. If an even better approximation to the normal mode is desired than that furnished by the one eigenfunction alone, or if it is not possible to derive *a priori* an appropriate $\Lambda$ operator, one might hope that an $M$-term expansion [like Eq. (22), but with $M$ small] would yield an accurate approximation to the normal mode. In practice what is done is to assume that an $M$-term expansion (typically $M \lesssim 10$) of the normal mode is adequate, and then an $M \times M$ dispersion matrix is constructed, yielding an approximate eigenfrequency $\omega'$. If the remaining terms $(n > M)$ in the expansion (22) are at most of order $\epsilon$, we would like to be able to claim from Theorem I that $\omega'$ is within order $\epsilon^2$ of $\omega_0$. To do this, we must establish that the exact and the approximate expansion coefficients for $1 \leqslant n \leqslant M$ differ by at most order $\epsilon$. This is the subject of the following theorem.

*Theorem II.* Let $\omega_0$ yield a zero eigenvalue of the operator $D(\omega)$; and let $\phi$ be the corresponding eigenvector:

$$D(\omega_0)\phi = 0. \tag{30}$$

Let $S$ be a subspace within which $\phi$ can be approximated to order $\epsilon$:

$$\phi = P\phi + \epsilon\xi, \tag{31}$$

where $P$ is a projection operator onto $S$, and $\phi$, $\xi$ are suitably normalized vectors. Assume the dual $\tilde{\phi}$ can likewise be approximated to order $\epsilon$ in $S$. Let $\underline{D}(\omega)$ be the operator $D(\omega)$ restricted to the subspace $S$:

$$\underline{D}(\omega) = PD(\omega)P. \tag{32}$$

Let $\omega'$ yield a zero eigenvalue of $\underline{D}(\omega)$:

$$\underline{D}(\omega')\underline{\phi} = 0, \tag{33}$$

where $\underline{\phi}$ lies in $S$:

$$P\underline{\phi} = \underline{\phi}. \tag{34}$$

Then, if $\omega_0$ is a simple root of Eq. (30), Eq. (33) has a root $\omega'$ which approximates $\omega_0$ to order $\epsilon^2$. The corresponding eigenvector $\underline{\phi}$ approximates $\phi$ to order $\epsilon$.

*Proof*: Expand $D(\omega')$ in a Taylor series:

$$D(\omega') = D(\omega_0) + \frac{1}{k!}(\omega' - \omega_0)^k D^{(k)}(\omega_0) + \cdots, \tag{35}$$

where

$$D^{(k)}(\omega_0) = \left.\frac{\partial^k D(\omega)}{\partial\omega^k}\right)_{\omega=\omega_0} \tag{36}$$

is the lowest order derivative of $D(\omega)$ for which

$$D^{(k)}(\omega_0)\phi \neq 0. \tag{37}$$

If $\omega_0$ is a simple root of Eq. (30) (for fixed $\phi$), $k = 1$. We assume that for $\epsilon$ sufficiently small, Eq. (33) has a root which approximates $\omega_0$ in the sense that

$$\omega' - \omega_0 \sim \epsilon^\alpha, \quad \alpha > 0. \tag{38}$$

Project Eq. (30) onto the subspace $S$, and use Eqs. (31), (32) and (35):

$$PD(\omega_0)\phi = \underline{D}(\omega')\phi + \epsilon PD(\omega_0)\xi - \frac{(\omega' - \omega_0)^k}{k!} PD^{(k)}(\omega_0)P\phi + \cdots = 0. \tag{39}$$

Subtract Eq. (39) from Eq. (33):

$$D(\omega')(\underline{\phi} - \phi) = \epsilon PD(\omega_0)\xi - \frac{(\omega' - \omega_0)^k}{k!} PD^{(k)}(\omega_0)P\phi + \cdots, \tag{40}$$

where dots represent higher order terms in $(\omega' - \omega_0)$.

In view of the definition (32) and since $P^2 = P$, we can replace $\phi$ by $P\phi$ on the left in Eq. (40). Let $N$ be a projection operator onto the null space of $\underline{D}(\omega')$. Since Eq. (33) allows us to add to $\underline{\phi}$ any vector in the null space of $\underline{D}(\omega')$, we can arrange that

$$(1 - N)(\underline{\phi} - P\phi) = \underline{\phi} - P\phi, \tag{41}$$

where 1 is the unit operator in $S$. Define the operator

$$D_N = (1 - N)\underline{D}(\omega')(1 - \nu). \tag{42}$$

Within the subspace orthogonal to the null space of $\underline{D}(\omega')$, $D_N$ has an inverse $D_N^{-1}$. Project Eq. (40) onto that subspace, use Eq. (41), and solve for $\underline{\phi} - P\phi$. In view of Eq. (31), we can write the solution in the form

$$\underline{\phi} = \phi - \epsilon\xi + D_N^{-1}(1 - N) \times \left[\epsilon PD(\omega_0)\xi - \frac{(\omega' - \omega_0)^k}{k!} PD^{(k)}(\omega_0)P\phi + \cdots\right]. \tag{43}$$

Now assume that

$$k\alpha \geqslant 1. \tag{44}$$

Then we have

$$\underline{\phi} - \phi \sim \epsilon, \tag{45}$$

which is the second conclusion of the theorem. A similar argument shows that $\tilde{\underline{\phi}}$ approximates $\tilde{\phi}$ to order $\epsilon$. In concluding the result (45) from Eq. (43), we are assuming that $\epsilon$ is small enough that there are no other roots of the dispersion relation in the neighborhood of $\omega_0$ that is specified by Eq. (38). Since $\underline{\phi}$ and $\tilde{\underline{\phi}}$ lie in $S$,

$$\langle \tilde{\underline{\phi}} \underline{D}(\omega) | \underline{\phi} \rangle = \langle \tilde{\underline{\phi}} | D(\omega) | \underline{\phi} \rangle, \tag{46}$$

and we can apply Theorem I to deduce the first conclusion of Theorem II. (In this case, $k = 1$.)

A simple extension of the argument leading to Eq. (16) shows that for $k > 1$,

$$\omega' = \omega_0 + O(\epsilon^{2/k}). \tag{47}$$

Comparison of Eqs. (47) and (38) yields

$$\alpha = 2/k, \tag{48}$$

which is compatible with assumption (44). A similar argument from the contrary assumption ($k\alpha < 1$) leads to an inconsistency.

## REMARKS

(1) In a fixed subspace $S$, the order $\epsilon$ of the error involved in projecting $\phi$ onto the space $S$ is fixed. Statements like (45) strictly apply only as $\epsilon \to 0$, i.e., as the subspace $S$ increases to approach the Hilbert space containing $\phi$. In that case, the projection operator $P$ will depend on $\epsilon$. This does not invalidate our argument.

(2) In applications where $S$ is a fixed subspace, usually of small dimensionality, Theorem II will guarantee that $\omega'$ approximates $\omega_0$ with a relative error of order of magnitude $\epsilon^2$ only if there is no other root of Eq. (30) near $\omega_0$, i.e., only if the term $k = 1$ in Eq. (35) dominates the later terms.

(2) When $S$ is finite dimensional, solving the dispersion relation

$$|D(\omega')| = 0 \tag{49}$$

yields a root $\omega'$ for which Eq. (33) has a solution $\phi$.

(4) Equation (33) has in general many solutions $\phi$ and $\omega'$. Likewise Eq. (30) has at least as many, usually many more solutions $\phi$, $\omega_0$. Therorem II does not guarantee that every solution of Eq. (33) approximates a solution of Eq. (30), but only that, for each solution of Eq. (30) satisfying the specified conditions there is a solution of Eq.(33) which approximates it. In many cases which arise in practice, Eq. (33) has other solutions which lie far from any solution of Eq. (30).

(5) The practical significance of Theorem II is that it may be computationally more efficient to expend additional effort generating a carefully chosen set of basis functions with which to expand a normal mode of interest (so that only a small number of terms are needed in the expansion), than it is to follow the standard procedure of expanding the normal mode in eigenfunctions of the Laplacian (which may require a large number of terms).

(6) Theorem II refers only to zero eigenvalues of $D(\omega)$ because that is the important case for the applications we have in mind. At a cost of a slight additional algebraic complexity, the proof can be generalized as in Theorem I to the case of an eigenvalue of $D(\omega)$ having any fixed value $d_0$.

## ACKNOWLEDGMENTS

We wish to thank H. Weitzner for reading the manuscript and offering some valuable suggestions. This research was supported by the Department of Energy, Contracts No. EY-76-5-02-2387 and No. EY-76-C-02-3077.

[1]H.R. Lewis and K.R. Symon, J. Math. Phys., **20**, 413 (1979).
[2]G. Gel'man, *Einfuhrung in die Quantenchemie* (Deuticke, Leipzig, 1937), p. 285.
[3]R.P. Feynman, Phys. Rev. **56**, 340 (1939).
[4]J.L. Schwarzmeier, H.R. Lewis, B. Abraham-Schrauner, and K.R. Symon, Phys. Fluids (to be published).
[5]J.L. Schwarzmeier, Ph.D. thesis, University of Wisconsin (1977).
[6]I.B. Bernstein, J.M. Greene, and M.D. Kruskal, Phys. Rev. **108**, 546 (1957).

# Extension of KMS states and angular momentum

G. Loupias

*Département de Mathématiques, Université des Sciences et Techniques du Languedoc, 34060-Montpellier Cedex, France*

M. Mebkhout

*Université d'Aix-Marseille II, UER Scientifique de Luminy, 70, Route Léon Lachamp, 13288—Marseille Cedex 2, France*

(Received 13 July 1978)

We present a description of $\beta$-KMS equilibrium states with respect to some one-parameter subgroup of $(T\times R^3)\rtimes K$, over an asymptotically Abelian algebra of observables (here $T\times R^3$ is the group of time and space translations and $K$ a compact group acting on $R^3$). In the case where $K=\mathrm{SO}(3)$, the group of space rotations, we get a characterization of the angular momentum of such states as a parameter occurring in their description, similar to the chemical potential, and classify all of them according to their invariance group, a closed subgroup of $T\times R\times \mathrm{SO}(2)$.

## INTRODUCTION

In their work (quoted in Ref. 1) describing the algebraic background of the chemical potential, Araki *et al.* are using a $C^*$-system $\{\mathscr{F},T\times G,\alpha\times\gamma\}$, where $\mathscr{F}$ is the field algebra, $T\times G$ the direct product of the group of time translations by a general compact gauge group, $\alpha\times\gamma$ a continuous representation of $T\times G$ by *-automorphisms of $\mathscr{F}$, the $G$-fixed point subalgebra $\mathfrak{A}$ of $\mathscr{F}$ being the algebra of observables.

In the present work, we give an analogous description of the angular momentum as a parameter occurring in the characterization of equilibrium states through the KMS condition with respect to a one-parameter subgroup of the group $(T\times R^3)\square\mathrm{SO}(3)$. [Here $R^3$ is the group of space translations and SO(3) the group of space rotation, $\square$ being the semidirect product.]

In the first half of this work, we discuss, more generally, the case of a compact group $K$ [instead of SO(3)] acting on $R^3$. As in the work of Araki *et al.*, we have two algebras: $\mathfrak{A}$, the algebra of observables, and $\mathfrak{A}_K$, the fixed points under the action of $K$. In Secs. III and V, we show that a state on $\mathfrak{A}_K$, of "KMS type at zero energy" with respect to time and space (this condition being deduced from some stability requirement under perturbations by $K$-fixed observables as shown in Sec. IV) can be extended, under some hypothesis (some clustering property of the state and some asymptotic Abelianness condition over $\mathfrak{A}$ and $\mathfrak{A}_K$), as a state on $\mathfrak{A}$, KMS with respect to time, space, and $K$ transformations.

After a study of the representations generated by such states (Sec. VI), we return to the case of a state invariant by some closed subgroup of $T\times R^3\square\mathrm{SO}(3)$ (the structure and construction of such subgroups being worked out in Sec. II). We show in Sec. VII that states KMS (in a nontrivial way) with respect to time, space, and rotations can be invariant by closed subgroups of $T\times R\times\mathrm{SO}(2)$ only, with SO(2) (the group of plane rotations) acting trivially on $R$, and we give a complete description of all of them. We close this part by giving another description of the angular momentum in terms of Radon–Nikodym derivative. As our situation contains the Araki *et al.* one, and as the case $K$ Abelian is especially interesting, we take this opportunity to present some specific proofs adapted to this case.

## I. NOTATION AND DEFINITION

We consider the following locally compact symmetry group:

$$\mathscr{G}=T\times(R^3\square K)=(T\times R^3)\square K=\{(t,x,k)\},\qquad(1)$$

where

$t\in T$: time evolution group, isomorphic to the reals $R$,

$x\in R^3$: space translations group,

$k\in K$: compact group of (eventually trivial) continuous actions on $R^3$.

In what follows, $K$ will be SO(3), the group of space rotations acting on $R^3$, but it can also represent any compact gauge group as in Ref. 1, commuting with $T\times R^3$ and without any action on it.

The sign $\square$ denotes the semidirect product (reducing to the direct product in the case of trivial actions)

$$(t,x,k)(t',x',k')=(t+t',x+kx',kk').\qquad(2)$$

We will prove, in the next part, that all closed subgroups $\mathscr{H}$ of $\mathscr{G}$ are extensions (denoted $\square^\lambda$, where $\lambda$ means some factor system),

$$\mathscr{H}=\mathscr{A}\square^\lambda\mathscr{K},\qquad(3)$$

where $\mathscr{A}$ is a closed subgroup of $T\times R^3$ and

$$\mathscr{K}=\{\mu(k),k\},$$

where

$$\mu(k)\in T\times R^3/\mathscr{A}\quad\text{with}\ \mu(k_1k_2)=\mu(k_1)+k_1\mu(k_2),\qquad(4)$$

is a closed group of "screwing operations," i.e., actions $k$ followed by translations $\mu(k)$ in the quotient space $T\times R^3/\mathscr{A}$. If $\mu=0$, these "screwing operations" are "pure operations" and $\mathscr{K}$ is then a closed subgroup of $K$.

$\mathfrak{A}$ will be a $C^*$-algebra of quasilocal observables (or of fields as Ref. 1) acted upon by $\mathscr{G}$, i.e., there is a faithful homomorphism

 0022-2488/79/060984-15$01.00  

$$(t,x,k)\to\alpha_{(t,x,k)}=\alpha_t\circ\alpha_x\circ\alpha_k, \quad (5)$$

from $\mathscr{G}$ to the group of *-automorphisms of $\mathfrak{A}$ such that $(t,x,k)\to\alpha_{(t,x,k)}(a)$ is norm continuous for any $a\in\mathfrak{A}$.

We will assume the existence of an automorphism $\tau$ of $\mathfrak{A}$ commuting with $\mathscr{G}$ (for instance $\tau=\alpha_t$ for some $t$) with respect to which $\mathfrak{A}$ is weakly asymptotically Abelian, i.e.,

$$M_n\{\omega(a[\tau^n b,c]d)\} = \lim_{N\to\infty}\frac{1}{N}\sum_{n=-N}^{N}\omega(a[\tau^n b,c]d)=0, \quad a,b,c,d\in\mathfrak{A}, \quad (6)$$

where [ , ] denotes the commutator.

Let us mention that the group $\{\tau^n, n\in Z\}$ could be replaced by any amenable group of * automorphisms of $\mathfrak{A}$ (for instance $T$), $M_n$ being replaced by any invariant mean on it.

$\mathfrak{A}_K$ will be the fixed point sub-$C^*$ algebra of $\mathfrak{A}$ under the action of $K$ (the group of all pure operations).

If $K$ is Abelian [for instance, $K=\mathrm{SO}(2)$], let $\widehat{K}$ be its (discrete) dual group. The sets

$$\mathfrak{A}^{(p)}=\{a\in\mathfrak{A},\ \alpha_k a=\langle p,k\rangle a,\ k\in K\},\quad p\in\widehat{K}, \quad (7)$$

are all nonvoid, $\tau$ and $\alpha_t$ $(t\in T)$ invariant, fulfill

$$\mathfrak{A}^{(p)}.\mathfrak{A}^{(q)}=\mathfrak{A}^{(pq)},\quad p,q,\in\widehat{K},$$
$$\mathfrak{A}^{(p)*}=\mathfrak{A}^{(p^{-1})}, \quad (8)$$
$$\mathfrak{A}^{(0)}=\mathfrak{A}_K,$$

and have a linear closure dense in $\mathfrak{A}$ for the weak topology.

Let $\omega$ be a state on $\mathfrak{A}$, invariant with respect to some closed subgroup of $\mathscr{G}$,

$$\mathscr{H}_\omega=\mathscr{A}_\omega\Box\mathscr{K}_\omega. \quad (9)$$

This state will be said to be weakly $\tau$ clustering if

$$\omega(\tau a)=\omega(a),\quad a\in\mathfrak{A},$$
$$M_n\{\omega(a\tau^n b)\}=\omega(a)\omega(b),\quad a,b,\in\mathfrak{A}, \quad (10)$$

where $M_n$ has been defined in (6). It is known[2] that, if $\mathfrak{A}$ is weakly asymptoticaly Abelian, (10) is equivalent to the extremal $\tau$ invariance of $\omega$. The remark following (6) being valid in (10), if $\{\tau^n, n\in Z\}$ is replaced by $T$, then (10) means that $\omega$ represents a "pure phase."

We call $\mathfrak{A}_\omega$ the fixed point sub-$C^*$ algebra of $\mathfrak{A}$ under the action of the subgroup $K_\omega$ of all pure operations in $\mathscr{H}_\omega$. Of course,

$$\mathfrak{A}_K\subset\mathfrak{A}_\omega\subset\mathfrak{A}, \quad (11)$$

but if $\mathscr{H}_\omega$ is such that $\mu\neq0$, then $K_\omega=\{1\}$ and $\mathfrak{A}_\omega=\mathfrak{A}$; on the other hand, if $\mu=0$, then $K_\omega=\mathscr{K}_\omega$ and $\mathfrak{A}_\omega\neq\mathfrak{A}$ unless $\mathscr{K}_\omega=\{1\}$.

At this point, it is interesting to notice a phenomenon that will play an important role in the sequel. If $\omega$ is invariant with respect to (9), the state $\omega$ restricted to $\mathfrak{A}_K$ and denoted $\omega_K$ can be invariant with respect to some subgroup $\mathscr{A}_{\omega_K}$ such that

$$\mathscr{A}_\omega\subset\mathscr{A}_{\omega_K}\{((t,x),\mu(k));(t,x)\in\mathscr{A}_\omega,(\mu(k),k)\in\mathscr{K}_\omega\}. \quad (12)$$

This is due to the fact that if

$$\omega(\alpha_{(t,x)}\alpha_{(\mu(k),k)}a)=\omega(a),$$
$$(t,x)\in\mathscr{A}_\omega,\quad(\mu(k),k)\in\mathscr{K}_\omega,\quad a\in\mathfrak{A},$$

then

$$\omega(\alpha_{((t,x),\mu(k))}A)=\omega(A),\quad A\in\mathfrak{A}_K,$$

and so

$$\mathscr{A}_\omega\subset\{((t,x),\mu(k))\}\subset\mathscr{A}_{\omega_K}.$$

But, conversely, if

$$\omega_K(\alpha_{(u,y)}A)=\omega_K(A),\quad(u,y)\in\mathscr{A}_{\omega_K},\quad A\in\mathfrak{A}_K,$$

then Lemma III,2(1) (see Sec. III) will prove the existence of some $k\in K$, depending on $(u,y)$, such that

$$\omega(\alpha_{(u,y)}\alpha_k a)=\omega(a),\quad a\in\mathfrak{A},$$

i.e., $(u,y,k)\in\mathscr{H}_\omega$, and is such that

$$(u,y,k)=((t,x),(\mu(k),k)),$$
$$(t,x)\in\mathscr{A}_\omega,\quad(\mu(k),k)\in\mathscr{K}_\omega,\quad\mu(k)=T\times R^3/\mathscr{A}_\omega.$$

Hence $(u,y)=((t,x),\mu(k))$, which proves (12), and we can notice that $\mathscr{A}_\omega=\mathscr{A}_{\omega_K}$ if an only if $\mu=0$ or $\mathscr{K}_\omega$ is trivial.

Our last definition will be the so-called KMS condition, a mathematical formalization of the notion of a "limit Gibbs state." Let us first say that a pair of tempered distributions on $R$ is a $\beta$-KMS pair $(\beta\in R)$ if

$$\langle\widehat{S},\varphi\rangle=\langle e^{\beta E}\widehat{T},\varphi\rangle=\langle\widehat{T},e^{\beta E}\varphi\rangle,\quad\varphi\in\mathscr{S}(R),\ E\in R, \quad (13)$$

where $\widehat{\ }$ means the Fourier transform. Then, if $\sigma_t$ is a continuous one-parameter group of automorphisms of $\mathfrak{A}$, $\omega$ is said to be $\beta$-KMS with respect to $\sigma_t$ whenever the functions

$$f_{ab}(t)=\omega(b\cdot\sigma_t a),\quad g_{ab}(t)=\omega(\sigma_t a\cdot b), \quad (14)$$

from a $\beta$-KMS pair of distributions.

It is known that a $\beta$-KMS state $\omega$ is automatically invariant. Therefore, the functions (14) are linear combinations of positive type functions and their Fourier transforms are bounded Radon measures. The $\beta$-KMS condition can also be stated as the fact that

$$\int\hat{v}(t)f_{ab}(t)\,dt=\int\widehat{w}(t)\,g_{ab}(t)\,dt, \quad (15)$$

where $v,w\in\mathscr{D}(R)$ are such that $w(E)=e^{\beta E}v(E)$, $E\in R$, or that there exists a function $u(z)$ holomorphic in the strip $0<\mathrm{Im}z<\beta$, continuous and bounded on its closure and such that

$$f_{ab}(t)=u(t),\quad g_{ab}(t)=u(t+i\beta). \quad (16)$$

It is also known that the left kernel $M_\omega=\{a\in\mathfrak{A},\omega(a^*a)=0\}$ is a two-sided ideal of $\mathfrak{A}$ such that $M_\omega=\mathrm{Ker}\pi_\omega$ if $\pi_\omega$ is the cyclic representation of $\mathfrak{A}$ deduced from $\omega$ through the GNS construction. Hence the faithfulness of $\omega$ and $\pi_\omega$ are then equivalent, and automatic if $\mathfrak{A}$ is simple. On the other hand, there exists a unique faithful normal state $\widetilde{\omega}$ on the von Neumann algebra $\pi_\omega(\mathfrak{A})''$ such that

$$\widetilde{\omega}\circ\pi_\omega=\omega,\quad\tilde{\sigma}_t\circ\pi_\omega=\pi_\omega\circ\sigma_{-\beta t},\quad\text{on }\mathfrak{A}, \quad (17)$$

where $\tilde{\sigma}_t$ is the modular automorphism of $\widetilde{\omega}$. Moreover, if $\rho$ is an automorphism of $\mathfrak{A}$ such that $\omega\circ\rho=\omega$ and if $\omega$ is faithful (or $\mathfrak{A}$ is simple), then

$$\sigma_t\circ\rho=\rho\circ\sigma_t. \quad (18)$$

## II. STRUCTURE AND CONSTRUCTION OF CLOSED SUBGROUPS OF $\mathscr{G}$

We use here techniques already introduced in Refs. 3 and 4.

Let $\mathscr{H}$ be a closed subgroup of $\mathscr{G}$ and

$$\mathscr{A} = (T\times R^3)\cap\mathscr{H}. \tag{19}$$

$\mathscr{A}$ is a closed normal subgroup of $\mathscr{H}$, and it is possible to define the quotient group

$$\mathscr{K} = \mathscr{H}/\mathscr{A}. \tag{20}$$

Let $N(\mathscr{A})$ be the normalizer of $\mathscr{A}$ into $\mathscr{G}$: it contains $\mathscr{A}$ and $T\times R^3$ as normal subgroups and $\mathscr{H}$ as a closed subgroup. Hence we can define the quotient groups

$$Q = N(\mathscr{A})/(T\times R^3) \quad\text{and}\quad \mathscr{N} = N(\mathscr{A})/\mathscr{A}. \tag{21}$$

It is easy to check that $Q$ is the closed subgroup of $K$ keeping $\mathscr{A}$ globally invariant and that

$$N(\mathscr{A}) = (T\times R^3)\Box Q. \tag{22}$$

On the other hand, the maps

$$\mathscr{H}\to\mathscr{K}$$

and

$$\mathscr{H}\to N(\mathscr{A})\to Q,$$

having the same kernel (i.e., $\mathscr{A}$), there exists a continuous, injective, nonnecessarily closed mapping $\varphi$ from $\mathscr{K}$ into $Q$ such that $\varphi(K)$ is a (nonnecessarily closed) subgroup of $Q$.

Let us now notice that it is possible to write

$$T\times R^3 = \mathscr{A}\Box^\lambda T\times R^3/\mathscr{A}, \tag{23}$$

where $\lambda$ means some central extension. Hence, if $(t,x,k)\in\mathscr{H}$, we can write

$$(t,x,k) = (\{(t,x)_{\mathscr{A}},\mu\},k) = ((t,x)_{\mathscr{A}},1)(\{0,\mu\},k)$$
$$= \{(t,x)_{\mathscr{A}},(\mu,k)\}, \tag{24}$$

where $(t,x)_{\mathscr{A}}\in\mathscr{A}$, $\mu\in T\times R^3/\mathscr{A}$, $(\mu;k)\in\mathscr{K}$.

As $\varphi((\mu,k)) = k$ and $\varphi$ is injective, then

$$\mu = \mu(k) \text{ with } \mu(k\cdot k') = \mu(k) + k\mu(k'), \tag{25}$$

where $k$ acts on $T\times R^3/\mathscr{A}$ through the quotient action because $k\in\varphi(\mathscr{K})\subset Q$ stabilizes $\mathscr{A}$. Moreover, using this quotient action,

$$\mathscr{N} = N(\mathscr{A})/\mathscr{A} = \frac{(T\times R^3)\Box Q}{\mathscr{A}} = \frac{T\times R^3}{\mathscr{A}}\Box Q. \tag{26}$$

Hence, $\mathscr{H}$ being closed into $N(\mathscr{A})$, $\mathscr{K}$ is also closed into $\mathscr{N}$, and because the mapping

$$(\mu(k),k)\in\mathscr{K}\to\mu(k)\in T\times R^3/\mathscr{A} \tag{27}$$

is a projection parallel to a compact, the set

$$\{\mu(k),k\in\varphi(\mathscr{K})\} \tag{28}$$

is closed into $T\times R^3/\mathscr{A}$. All these considerations allow us to write

$$N(\mathscr{A}) = (T\times R^3)\Box Q = \left(\mathscr{A}\Box^\lambda \frac{T\times R^3}{\mathscr{A}}\right)\Box Q$$
$$= \mathscr{A}\Box^{\lambda'}\left\{\frac{T\times R^3}{\mathscr{A}}\Box Q\right\} = \mathscr{A}\Box^{\lambda'}\mathscr{N}, \tag{29}$$

where the extension $\lambda'$ can easily be deduced from the extension $\lambda$. Consequently,

$$\mathscr{H} = \mathscr{A}\Box^{\lambda'}\mathscr{K}, \tag{30}$$

where $\mathscr{A}$ is some closed subgroup of $T\times R^3$ and $\mathscr{K}$ some closed subgroup of $\mathscr{N}$ made of "screwing operations" $(\mu(k),k)$, where $\mu$ is some cocycle (if $\mu=0$, then $\mathscr{K}$ is a closed subgroup of "pure operations" in $K$), the extension $\lambda'$ being deduced from the extension (23).

In the special case where $K = \mathrm{SO}(3)$ [or SO(2)] it can be considered as the connected component of O(3) [or O(2)], the extension of SO(3) [or SO(2)] by the operation $(-I)$ defined according to $(-I)x = -x$. If the closed subgroup $\mathscr{K}$ can also be extended by some operation $(\bar\mu(-I),(-I))$, then $\mu$ is the restriction to $\varphi(\mathscr{K})$ of some cocycle $\bar\mu$ on $\varphi(\mathscr{K})\times(-I)$ such that $\bar\mu(-I)\in T\times R^3/\mathscr{A}$ and

$$\bar\mu((-I)k) = \bar\mu(k\cdot(-I)) = \bar\mu(-I) + (-I)\mu(k)$$
$$= \mu(k) + k\bar\mu(-I)$$

or (31)

$$\mu(k) - (-I)\mu(k) = (I-k)\bar\mu(-I).$$

Calling $\bar\mu_\perp(k)$ the projection of $\bar\mu(k)$ on the "space" part of $T\times R^3/\mathscr{A}$, we get

$$\mu_\perp(k) + \mu_\perp(k) = (I-k)\bar\mu_\perp(-I). \tag{32}$$

So if $\mu_0$ is an element of $T\times R^3/\mathscr{A}$ such that $\mu_0 + \mu_0 = \bar\mu(-I)$, then

$$\mu_\perp(k) = (I-k)\mu_{0\perp}, \tag{33}$$

which means that $\mu_\perp$ is a coboundary and shows that, in the case when $\varphi(\mathscr{K})$ acts trivially on $T\times R^3/\mathscr{A}$, such an extension of $\mathscr{K}$ by $(\bar\mu(-I),(-I))$ is possible only if $\bar\mu(-I) = 0$.

Using the language of exact sequences, we can represent this whole construction in Fig. 1.

## III. EXTENSION OF INVARIANT STATES

The main theorem of this part relies on two successive lemmas that we recall without proof because they are identical to (Ref. 1, Lemma II.1 and Theorem II.1):

*Lemma III.*1: If $\omega$ is weakly $\tau$ clustering on $\mathfrak{A}$, the norm closure $\overline{C^1_\omega(K)}$ of

$$C^1_\omega(K) = \{k\in K\to f^\omega_a(k) = \omega(\alpha_k a), a\in\mathfrak{A}\}\subset\mathscr{C}(K) \tag{34}$$

coincides with $\mathscr{C}(K_\omega\backslash K)$.

*Lemma III.*2: (1) If $\omega_1$ and $\omega_2$ are two weakly $\tau$ clustering states on $\mathfrak{A}$ (nonnecessarily asymptotically Abelian with respect to $\tau$) and if their restrictions to $\mathfrak{A}_K$ are equal, then there exists a $k\in K$ such that $\omega_2 = \omega_1\circ\alpha_K$ with $K_{\omega_2} = k^{-1}K_{\omega_1}k$ ($k$ is unique up to an element of $K_{\omega_1}$).

(2) If $\mathfrak{A}$ is weakly asymptotically Abelian relative to $\tau$ and if the state $\omega_K$ on $\mathfrak{A}_K$ is weakly $\tau$ clustering, then there exists an extension $\omega$ of $\omega_K$ to a weakly $\tau$ clustering state of $\mathfrak{A}$.

This lemma proves that either $\omega_K$ has a unique $K$-invariant extension or $\omega_K$ has a plurality of non-$K$-invariant extensions labeled by the elements of $K$.

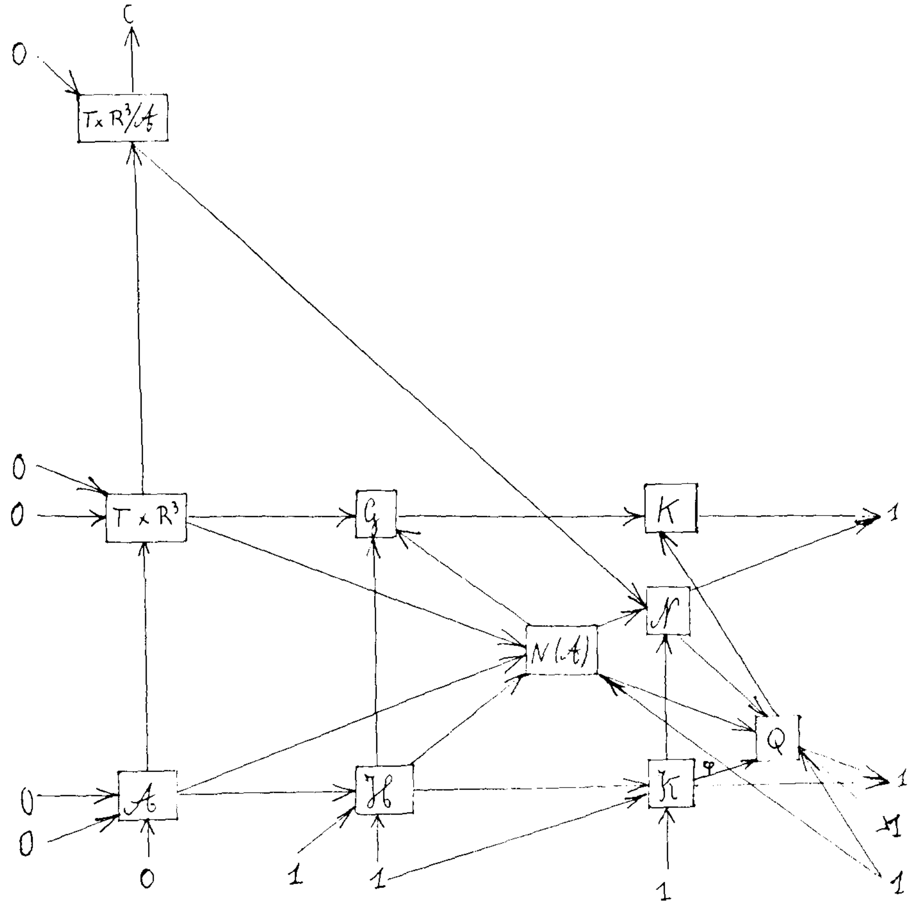

FIG. 1.

As the case $K$ Abelian will be one of the most interesting ones in what follows, we will give a new proof of Lemma III.2(1) adapted to this situation. It depends on the following lemma:

*Lemma III.3*: If $\omega$ is weakly $\tau$ clustering, and $K$ Abelian, the set

$$\Sigma_\omega = \{ p\in\widehat{K}: \text{there exists } a\in\mathfrak{A}^{(p)} \text{ with } \omega(a)\neq 0\} \qquad (35)$$

is a subgroup of $\widehat{K}$.

*Proof*: Let $p,q\in\Sigma_\omega$, $a\in\mathfrak{A}^{(p)}$, $b\in\mathfrak{A}^{(q)}$ with $\omega(a)\neq 0$ and $\omega(b)\neq 0$. As $M_n\{\omega(a\tau^n b)\} = \omega(a)\omega(b)\neq 0$, there exists some $n\in Z$ such that $\omega(a\tau^n b)\neq 0$. As $a\tau^n b\in\mathfrak{A}^{(pq)}$, then $pq\in\Sigma_\omega$. On the other hand, as $a^*\in\mathfrak{A}^{(p^{-1})}$ and $\omega(a^*) = \overline{\omega(a)}\neq 0$, $p^{-1}\in\Sigma_\omega$.

*Proof of Lemma III.2(1) in the case K Abelian*: Let $a\in\mathfrak{A}^{(p)}$, $b\in\mathfrak{A}^{(p^{-1})}$, $p\in\widehat{K}$, $n\in Z$. Then $a\tau^n b\in\mathfrak{A}^{(0)} = \mathfrak{A}_K$ and so $\omega_1(a\tau^n b) = \omega_2(a\tau^n b)$. Applying $M_n$ to both sides, we get $\omega_1(a)\omega_1(b) = \omega_2(a)\omega_2(b)$ and in particular (if $b = a^*$) $|\omega_1(a)| = |\omega_2(a)|$ which means that $\Sigma_{\omega_1} = \Sigma_{\omega_2}$. If $p\in\Sigma_{\omega_1}$ (of $\Sigma_{\omega_2}$) we can, thanks to Lemma III.3, choose $b$ such that $\omega_2(b)\neq 0$ and so $\omega_2(a) = [\omega_1(b)/\omega_2(b)]\omega_1(a) = c_p\,\omega_1(a)$, where $c_p$ is a constant depending on $p$, of modulus 1 and such that $c_{p^{-1}} = \bar{c}_p = c_p^{-1}$. If $p\notin\Sigma_{\omega_1}$, we have the same (but now trivial) relation with $c_p$ arbitrary. Moreover, if $p = qr$ ($q,r\in\Sigma_{\omega_1} = \Sigma_{\omega_2}$) and if we replace $b$ by $c\tau^m d$, where $c\in\mathfrak{A}^{(q^{-1})}$ and $d\in\mathfrak{A}^{(r^{-1})}$, then $\omega_1(a)\omega_1(c)\omega_1(d) = \omega_2(a)\omega_2(c)\omega_2(d)$ and so $c_{qr} = c_q c_r$. Hence $c_p$ is a character on the (discrete) group $\Sigma_{\omega_1} = \Sigma_{\omega_2}\subset\widehat{K}$, which can be extended in an arbitrary way to a character on $\widehat{\widehat{K}}$. Hence there is some $k\in K$ such that $\omega_2(a) = \langle p,k\rangle\omega_1(a) = \omega_1(\langle p,k\rangle a) = \omega_1(\alpha_K a)$, from which we deduce $\omega_2 = \omega_1\circ\alpha_K$ over $\mathfrak{A}$ by linearity and density.

*Theorem III.3*: If $\omega$ is a weakly $\tau$ clustering state on $\mathfrak{A}$ (nonnecessarily asymptotically Abelian with respect to $\tau$) and if $\omega_K = \omega\mid_{\mathfrak{A}_K}$ is "invariant" under some continuous one-parameter subgroup $\alpha_I\alpha_{x_t}$ of $\mathscr{A}_{\omega_K}\subset T\times R^3$ (i.e., $\omega\circ\alpha_I\alpha_{x_t} = \omega$ on $\mathfrak{A}_K$, in spite on the fact that $\alpha_I\alpha_{x_t}$ is not an automorphism of $\mathfrak{A}_K$), then there exists a unique continuous one-parameter family $k_t\in K$ with

$$\begin{aligned} k_{(t+t')} &= (-x_t + k_{t'}x_{t'}k_{t'}k_t) = (-x_t,I)(0,k_{t'})(x_t,I)(0,k_t) \\ &= (-x_{t'} + k_t x_{t'}, k_t k_{t'}) \\ &= (-x_{t'},I)(0,k_t)(x_{t'},I)(0,k_{t'}) \qquad (36) \end{aligned}$$

such that $\omega$ is invariant under the continuous one-parameter subgroup of $\mathscr{H}_\omega$: $\{\alpha_I\alpha_{x_t}\alpha_{k_t}\}$. If $k_t x_{t'} = x_{t'}$, the family $k_t$ is a continuous one-parameter subgroup of $K$. The family $k_t$ is nontrivial if and only if $\mu\neq 0$ and $\mathscr{H}_\omega$ is nontrivial.

*Proof*: Let $\omega_t$ be the state on $\mathfrak{A}$ defined according to $\omega_t(a) = \omega(\alpha_I\alpha_{x_t}a)$, $a\in\mathfrak{A}$: $\omega_t$ is a weakly $\tau$ clustering state on $\mathfrak{A}$, which is different from $\omega$ if and only if $\alpha_I\alpha_{x_t}\notin\mathscr{A}_\omega$ (i.e., if and only if $\mu\neq 0$ and $\mathscr{H}_\omega$ is nontrivial), and whose restriction to $\mathfrak{A}_K$ coincides with $\omega$ by hypothesis. By Lemma III.2(1), there exists $k_t^{-1}\in K$, nontrivial if and only if $\mu\neq 0$ and $\mathscr{H}_\omega$ is

nontrivial, and hence unique, such that
$\omega_t(a) = \omega(\alpha_t\alpha_{x_t}a) = \omega(k_t^{-1}a)$, $a\in\mathfrak{A}$, or else
$\omega(a) = \omega(\alpha_t\alpha_{x_t}\alpha_{k_t}a)$, $a\in\mathfrak{A}$.

Because $\omega(\alpha_t\alpha_{x_t}a) = \omega(\alpha_{k_t^{-1}}a)$ and
$\omega(\alpha_{t'}x_{t'}a) = \omega(\alpha_{k_{t'}^{-1}}a)$, then

$$\omega(\alpha_{t+t'}\alpha_{x_{t+t'}}a) = \omega(\alpha_{k_{t+t'}^{-1}}a) = \omega(\alpha_t\alpha_{x_t}\alpha_{t'}\alpha_{x_{t'}}a)$$

$$= \omega(\alpha_{k_t^{-1}}\alpha_{t'}\alpha_{x_{t'}}a) = \omega(\alpha_{t'}\alpha_{k_t^{-1}x_{t'}}\alpha_{k_t^{-1}}a)$$

$$= \omega(\alpha_{t'}\alpha_{x_{t'}}\alpha_{-x_{t'}+k_t^{-1}x_{t'}}\alpha_{k_t^{-1}}a)$$

$$= \omega(\alpha_{k_{t'}^{-1}}\alpha_{-x_{t'}+k_t^{-1}x_{t'}}\alpha_{k_t^{-1}}a)$$

$$= \omega(\alpha_{-k_{t'}^{-1}x_{t'}+k_{t'}^{-1}k_t^{-1}x_{t'}}\alpha_{k_{t'}^{-1}k_t^{-1}}a),$$

and consequently,
$$k_{t+t'}^{-1} = (-k_{t'}^{-1}x_{t'} + k_{t'}^{-1}k_t^{-1}x_{t'}, k_{t'}^{-1}k_t^{-1}),$$
or else

$$k_{t+t'} = (0,k_tk_{t'})(k_{t'}^{-1}x_{t'} - k_{t'}^{-1}k_t^{-1}x_{t'}, I)$$

$$= (k_tx_{t'} - x_{t'}, k_tk_{t'}).$$

Then

$$(t,x_t,k_t)(t',x_{t'},k_{t'})$$

$$= (t+t', x_t + k_tx_{t'}, k_tk_{t'})$$

$$= (t+t', x_t + x_{t'} - x_{t'} + k_tx_{t'}, k_tk_{t'})$$

$$= (t+t', x_{t+t'}, k_{t+t'})$$

and so $\alpha_t\alpha_{x_t}\alpha_{k_t}$ is a continuous one-parameter group.

## IV. A STABILITY REQUIREMENT

In what follows, we will assume that $\omega$ verifies, on $\mathfrak{A}_K$, a "KMS-type" relation at zero energy, i.e.,

$$\int_{-\infty}^{+\infty} \omega(A\cdot\alpha_t\alpha_{x_t}B)\,dt = \int_{-\infty}^{+\infty} \omega(\alpha_t\alpha_{x_t}B\cdot A)\,dt, \quad A,B\in\mathfrak{A}_K \tag{37}$$

(remember that $\alpha_t\alpha_{x_t}$ is not an automorphism of $\mathfrak{A}_K$). We are first going to show that this condition can be deduced from some stability requirement as follows. Let us consider the two automorphisms on $\mathfrak{A}$: $\alpha_t\alpha_{x_t}$ and $\alpha_t\alpha_{x_t}\alpha_{k_t}$, and let $h\in\mathfrak{A}_K^{\text{s.a.}}$ be some self-adjoint element of $\mathfrak{A}_K$. Because $\alpha_{k_t}h = h$, the unitary cocycle $P_t^h$ defined equivalently by

$$i\frac{d}{dt}P_t^h = P_t^h\alpha_t\alpha_{x_t}(h), \quad P_0^h = I$$

or (38)

$$i\frac{d}{dt}P_t^h = P_t^h\alpha_t\alpha_{x_t}\alpha_{k_t}(h), \quad P_0^h = I$$

allows us to define the two perturbed automorphisms on $\mathfrak{A}$ [5]:

$$(\alpha_t\alpha_{x_t})^h(a) = P_t^h\alpha_t\alpha_{x_t}(a)P_t^{h*}, \tag{39}$$

$$(\alpha_t\alpha_{x_t}\alpha_{k_t})^h(a) = P_t^h\alpha_t\alpha_{x_t}\alpha_{k_t}(a)P_t^{h*}$$
$$= (\alpha_t\alpha_{x_t})^h\circ\alpha_{k_t}(a), \tag{40}$$

which coincide on $\mathfrak{A}_K$:

$$(\alpha_t\alpha_{x_t})^h(A) = (\alpha_t\alpha_{x_t}\alpha_{k_t})^h(A), \quad A\in\mathfrak{A}_K. \tag{41}$$

Consequently, if $\omega$ is weakly $\tau$ clustering, the element $k_t^h\in K$ which, according to Lemma III. 2, is such that

$$\omega((\alpha_t\alpha_{x_t}\alpha_{k_t})^h(a)) = \omega((\alpha_t\alpha_{x_t})^h\alpha_{k_t^h}(a)), \quad a\in\mathfrak{A},$$

can be chosen equal to $\alpha_{k_t}$, hence independent of $h$. As it is known that any KMS state on $\mathfrak{A}$ is stable under local perturbations of the dynamic,[6] it is natural, conversely, to say that $\omega$ is stable on $\mathfrak{A}_K$ for some perturbation by a self-adjoint $h\in\mathfrak{A}_K^{\text{s.a.}}$ if there exists some application $h\to\omega^h$ from some absorbing neighborhood $\mathscr{V}$ of 0 in $\mathfrak{A}_K^{\text{s.a.}}$ into the set of weakly $\tau$ clustering states of $\mathfrak{A}$ such that

$$\omega^h((\alpha_t\alpha_{x_t})^h(A)) = \omega^h(A), \quad A\in\mathfrak{A}_K, \tag{42a}$$

$$\omega^{\lambda h}(A) \xrightarrow{\lambda\to 0} \omega(A), \quad A\in\mathfrak{A}_K, \tag{42b}$$

$$\omega^h(\alpha_t\alpha_{x_t}(A)) \underset{t\to\pm\infty}{\to} \omega(A), \quad A\in\mathfrak{A}_K. \tag{42c}$$

Moreover, condition (42a) implies, according to Theorem III.3, the existence of $k_t^h$ such that

$$\omega^h((\alpha_t\alpha_{x_t})^h\alpha_{k_t^h}(a)) = \omega^h(a), \quad a\in\mathfrak{A}. \tag{43}$$

We will assume moreover (as is already the case for $\omega$) that, for any $h$ and $\omega^h$,

$$k_t^h = k_t. \tag{44}$$

We then have the following theorem.

*Theorem IV.*1: Let $\omega$ weakly $\tau$ clustering on $\mathfrak{A}$ such that $\omega_K$ is "invariant" with respect to $\alpha_t\alpha_{x_t}$. Let us assume the existence of a dense set $\mathfrak{A}_K^\circ$ in $\mathfrak{A}_K$ such that ($L^1$-Abelian asymptotism):

$$\int_{-\infty}^{+\infty} \|[B,\alpha_t\alpha_{x_t}(A)]\|dt < \infty, \quad A,B\in\mathfrak{A}_K^\circ. \tag{45}$$

Then, if $\omega$ is stable on $\mathfrak{A}_K$ under perturbations $h\in\mathfrak{A}_K^{\text{s.a.}}$, then

$$\int_{-\infty}^{+\infty} \omega(A\cdot\alpha_t\alpha_{x_t}B)dt = \int_{-\infty}^{+\infty} \omega(\alpha_t\alpha_{x_t}B\cdot A)dt, \quad A,B\in\mathfrak{A}_K. \tag{46}$$

*Proof*: According to (42a), (44), and Theorem III.3,

$$\omega^h((\alpha_t\alpha_{x_t}\alpha_{k_t^h})^h(a)) = \omega^h(a), \quad a\in\mathfrak{A},$$

and, by differentiation for any $a$ in the dense set of differentiable elements,

$$\omega^h\left(\frac{d}{dt}\bigg|_{t=0}(\alpha_t\alpha_{x_t}\alpha_{k_t})^h(a)\right) = 0, \quad a\in\mathfrak{A},$$

or else, according to (38) and (40),

$$\omega^h\left(\frac{d}{dt}\bigg|_{t=0}\alpha_t\alpha_{x_t}\alpha_{k_t}(a)\right) = i\omega^h([h,a]), \quad a\in\mathfrak{A}.$$

Let

$$a = \int_S^T \alpha_t \alpha_{x_t}(A)\, dt = \int_S^T \alpha_t \alpha_{x_t} \alpha_{k_t}(A)\, dt, \quad A \in \mathfrak{A}_K^\circ.$$

Then

$$\frac{d}{dt}\Big|_{t=0} \alpha_t \alpha_{x_t} \alpha_{k_t}(a) = \alpha_T \alpha_{x_T}(A) - \alpha_S \alpha_{x_S}(A)$$

and so

$$\omega^h(\alpha_T \alpha_{x_T}(A) - \alpha_S \alpha_{x_S}(A)) = i\omega^h\Big(\Big[h, \int_S^T \alpha_t \alpha_{x_t}(A)\, dt\Big]\Big).$$

If $S \to -\infty$ and $T \to +\infty$, we get, thanks to (42c) and (45),

$$0 = i\omega^h\Big(\int_{-\infty}^{+\infty} [h, \alpha_t \alpha_{x_t}(A)]\, dt\Big), \quad h \in \mathfrak{A}_K^\circ \cap \mathscr{Y}, \; A \in \mathfrak{A}_K^\circ$$

and finally, thanks to (42b),

$$0 = \int_{-\infty}^{+\infty} \omega([h, \alpha_t \alpha_{x_t}(A)])dt, \quad A \in \mathfrak{A}_K^\circ, \; h \in \mathfrak{A}_K^\circ \cap \mathscr{Y},$$

which leads to (46) by linearity and density.

*Remark*: Following the same procedure as in Ref. 7, end of Sec. II, we might get a stronger theorem by replacing $\mathfrak{A}_K$ by the smallest $C^*$ algebra of fixed points under the action of $\{\alpha_x \alpha_k, x \in R^3, k \in K\}$ and asking to $\omega$ to be invariant under $\alpha_t$ and stable under perturbations $h$ such that $\alpha_x \alpha_k h = h$ for any $x \in R^3$ and $k \in K$.

## V. EXTENSION OF KMS STATES

In this part, we will prove that if (37) is true, then $\omega$ is KMS on $\mathfrak{A}$ with respect to some one-paremeter group $\alpha_t \alpha_{x_t} \alpha_{k_t}$. The strategy of the proof is the following one. Starting from a "KMS-type" condition on $\mathfrak{A}_K$, (37), with respect to $\alpha_t \alpha_{x_t}$, we first get a true KMS condition at zero energy on $\mathfrak{A}_\omega$ with respect to $\alpha_t \alpha_{\tilde{x}_t}$ (where $\tilde{x}_t$ is some averaged direction) if $\mu = 0$ and with respect to $\alpha_t \alpha_{x_t} \alpha_{k_t}$ if $\mu \neq 0$ (and then $\mathfrak{A}_\omega = \mathfrak{A}$). We then transform this KMS condition at zero energy into a KMS condition for any energy, and so we are done if $\mu \neq 0$. If $\mu = 0$, we have to move this KMS condition on $\mathfrak{A}_\omega$ with respect to $\alpha_t \alpha_{\tilde{x}_t}$ up to a KMS condition on $\mathfrak{A}$ with respect to some one-parameter subgroup $\alpha_t \alpha_{\tilde{x}_t} \alpha_{k_t}$: This can be done only up to some "asymmetry subgroup" of $K_\omega$.

The first step is contained in

*Theorem V.1*: Let $\omega$, weakly $\tau$ clustering on $\mathfrak{A}$ weakly $\tau$ asymptotically Abelian, such that (37) is true and $\omega_K$ is "invariant" with respect to $\alpha_t \alpha_{x_t}$. Then

$$\int_{-\infty}^{+\infty} \omega(a \cdot \alpha_t \alpha_{x_t} \alpha_{k_t} b)dt = \int_{-\infty}^{+\infty} \omega(\alpha_t \alpha_{x_t} \alpha_{k_t} b \cdot a)dt, \quad a, b \in \mathfrak{A}_\omega, \tag{47}$$

where $k_t$ is given by Theorem III.3. Since the proof is identical to that in Ref. 1, Lemma II.5 where $k(E) = \delta_0(E)$, we do not reproduce it.

*Corollary V.2*: If $\mu = 0$, $k_t = I$ and we get

$$\int_{-\infty}^{+\infty} \omega(a \cdot \alpha_t \alpha_{x_t} b)dt = \int_{-\infty}^{+\infty} \omega(\alpha_t \alpha_{x_t} b \cdot a)dt, \quad a, b \in \mathfrak{A}_\omega. \tag{48}$$

But $\mathfrak{A}_\omega$ and $\omega$ being respectively fixed and invariant under $K_\omega$, this can be written

$$\int_{-\infty}^{+\infty} \omega(a \cdot \alpha_t \alpha_{\tilde{x}_t} b)dt = \int_{-\infty}^{+\infty} \omega(\alpha_t \alpha_{\tilde{x}_t} b \cdot a)dt, \quad a, b \in \mathfrak{A}_\omega, \tag{49}$$

where

$$\tilde{x}_t = \int_{K_\omega} k x_t\, dt \quad \text{and} \quad K_\omega \tilde{x}_t = \tilde{x}_t \tag{50}$$

so that $\alpha_t \alpha_{\tilde{x}_t}$ is now an automorphism of $\mathfrak{A}_\omega$ and (49) is a true KMS condition on $\mathfrak{A}_\omega$ at zero energy with respect to $\alpha_t \alpha_{\tilde{x}_t}$.

If $\mu \neq 0$, $\mathfrak{A}_\omega = \mathfrak{A}$ and (47) is a true KMS condition on $\mathfrak{A}$ at zero energy with respect to $\alpha_t \alpha_{x_t} \alpha_{k_t}$.

The second step is now given by

*Theorem V.3*: Let $\omega$ weakly $\tau$ clustering on $\mathfrak{A}$ such that (37) is true and $\omega_K$ is "invariant" with respect to $\alpha_t \alpha_{x_t}$, with $x_t = \tilde{x}_t$ if $\mu = 0$. Let us assume that $\mathfrak{A}_\omega$ is simple and weakly asymptotically Abelian with respect to $\tau$ and $\alpha_t \alpha_{x_t} \alpha_{k_t}$. Then, if the spectrum of $\alpha_t \alpha_{x_t} \alpha_t$ is not one-sided,

$$d\,[\omega(a \cdot \alpha_t \alpha_{x_t} \alpha_{k_t} b)]\hat{}\,(E) = e^{\beta E} d\,[\omega(\alpha_t \alpha_{x_t} \alpha_{k_t} b \cdot a)]\hat{}\,(E), \quad a, b \in \mathfrak{A}_\omega, \; \beta \in R. \tag{51}$$

*Proof*: Let us rewrite (47) with $a = a_1 \tau^n a_2$ and $b = b_1 \tau^n b_2$, $a_1, a_2, b_1, b_2 \in \mathfrak{A}_\omega$.

$$\int_{-\infty}^{+\infty} \omega(a_1 \tau^n a_2 \cdot \alpha_t \alpha_{x_t} \alpha_{k_t}(b_1 \tau^n b_2))dt = \int_{-\infty}^{+\infty} \omega(\alpha_t \alpha_{x_t} \alpha_{k_t}(b_1 \tau^n b_2) \cdot a_1 \tau^n a_2)dt.$$

By $\tau$-Abelian asymptotism [see Lemma V.7, (56)], this equality will be equivalent, in the mean $M_n$, to

$$\int_{-\infty}^{+\infty} \omega(a_1 \cdot \alpha_t \alpha_{x_t} \alpha_{k_t} b_1 \cdot \tau^n(a_2 \cdot \alpha_t \alpha_{x_t} \alpha_{k_t} b_2))dt = \int_{-\infty}^{+\infty} \omega(\alpha_t \alpha_{x_t} \alpha_{k_t} b_1 \cdot a_1 \tau^n(\alpha_t \alpha_{x_t} \alpha_{k_t} b_2 \cdot a_2))dt,$$

whose mean is, by weak $\tau$ clustering,

$$\int_{-\infty}^{+\infty} \omega(a_1 \cdot \alpha_t \alpha_{x_t} \alpha_{k_t} b_1)\omega(a_2 \cdot \alpha_t \alpha_{x_t} \alpha_{k_t} b_2)dt = \int_{-\infty}^{+\infty} \omega(\alpha_t \alpha_{x_t} \alpha_{k_t} b_1 \cdot a_1)\omega(\alpha_t \alpha_{x_t} \alpha_{k_t} b_2 \cdot a_2)dt,$$

and the proof follows in the same way as in Refs. 5 or 8 thanks to the results of (Ref. 9, final remark), which we reproduce here for completeness.

*Lemma V.4*: Let $\omega$ be weakly $\tau$ clustering on $\mathfrak{A}_\omega$ and $\omega_K$ "invariant" on $\mathfrak{A}_K$ with respect to $\alpha_t \alpha_{x_t}$. Let us assume that $\mathfrak{A}_\omega$ is simple and weakly $\tau$ and $\alpha_t \alpha_{x_t} \alpha_{k_t}$-asymptotically Abelian. Let $(\pi_\omega, U_\omega)$ the representation of $(\mathfrak{A}_\omega, \alpha_t \alpha_{x_t} \alpha_{k_t})$ generated by $\omega$. Then we have the following alternative:

(i) The spectrum of $U$ is one sided (contained in $R^+$ or $R^-$).

(ii) The spectrum of $U$ is the whole of $R$.

We are now left with the last step ($\mu = 0$). For that purpose we first need the following lemma, whose proof can be found in Ref. 3, Lemma II.3 and II.4:

*Lemma V.5*: Let $\omega$ be weakly $\tau$ clustering on $\mathfrak{A}_\omega$ weakly $\tau$ asymptotically Abelian. There exists a closed normal subgroup $G_\omega$ of $K_\omega$ such that

$$\overline{C_\omega(K_\omega)} \cap \overline{C_\omega(K_\omega)}^* = \mathscr{C}(K_\omega/G_\omega) \tag{52}$$

and

$$\overline{C_{\omega_1}(K_\omega)} = \overline{C_{\omega_1}(K_\omega)}^* = \mathscr{C}(K_\omega/G_\omega), \tag{53}$$

where

$$C_\omega(K_\omega) = \{k\in K_\omega \to f^\omega_{a,b}(k) = \omega(a\cdot\alpha_k b),\ a,b\in\mathfrak{A}\} \tag{54}$$

and $\omega_1 = \omega|_{\mathfrak{A}_{G_\omega}}$. This subgroup is defined by

$$G_\omega = \{k\in K_\omega : \omega(a\cdot\alpha_k b) = \omega(ab),\ a,b\in\mathfrak{A}\} \tag{55}$$

and is trivial if $\omega$ is faithful. This lemma leads to the following theorem as in Ref. 3, Theorem II.4:

*Theorem V.6*: With the hypothesis of Theorem V.3 with $\mu = 0$, there exists a continuous one-parameter subgroup $k_t$ of $K_\omega$ commuting with $G_\omega$ such that

(i) The restriction of $\omega$ to $\mathfrak{A}_{G_\omega}$ is KMS with respect to $\alpha_t\,\alpha_{\dot{x}_t}\alpha_{k_t}$;

(ii) The image $\dot{k}_t = k_t G_\omega$ of $k_t$ into $K_\omega/G_\omega$ is in the center of $K_\omega/G_\omega$;

(iii) the $G_\omega$ spectrum of $\omega$ (i.e., the set of all irreducible representations $p$ of $G_\omega$ contained in $U_\omega|G_\omega$) is one-sided (i.e., if $p_1$ and $p_2$ are in the $G_\omega$ spectrum, $p_1\otimes p_2$ is also in it, and if both $p$ and its conjugate $\bar{p}$ are in it, then $p = 1$).

Here too, we are going to give a new proof of this theorem in the case where $K$ is Abelian. For that purpose, we need the following lemmas.

*Lemma V.7*: Let $\omega$ be weakly $\tau$ clustering on $\mathfrak{A}$ weakly $\tau$ asymptotically Abelian. Then

(i) if $a_1,a_2,b_1,b_2\in\mathfrak{A}$, $a_n = a_1\tau^n a_2$, $b_n = b_1\tau^n b_2$, $n\in Z$,

$$M_n\{\omega(a_n\cdot\alpha_t\,\alpha_{x_t}b_n)\} = \omega(a_1\cdot\alpha_t\,\alpha_{x_t}b_1)\omega(a_2\cdot\alpha_t\,\alpha_{x_t}b_2), \tag{56}$$

$$M_n\{\omega(\alpha_t\,\alpha_{x_t}b_n\cdot a_n)\} = \omega(\alpha_t\,\alpha_{x_t}b_1\cdot a_1)\omega(\alpha_t\,\alpha_{x_t}b_2\cdot a_2),$$

(ii) the set

$$\Theta_\omega = \{(p,q)\in\hat{K}_\omega\times\hat{K}_\omega : \text{there exists } a\in\mathfrak{A}^{(p)}, b\in\mathfrak{A}^{(q)},$$

$$\text{with } \omega(ab)\neq 0\} \tag{57}$$

[or in an equivalent way $\omega(a\cdot\alpha_t\alpha_{\dot{x}_t}b)\not\equiv 0$ or $\omega(\alpha_t\alpha_{\dot{x}_t}b\cdot a)\not\equiv 0$] is stable for the composition law induced by the one of $\hat{K}_\omega$. If $\omega|_{\mathfrak{A}_\omega}$ is faithful (in particular if $\mathfrak{A}_\omega$ is simple) $\Theta_\omega$ is a group containing the antidiagonal $\Delta = \{(p,p^{-1}), p\in\hat{K}\}$. If $\Theta_\omega = \Delta$, $\omega$ is $K_\omega$ invariant, and non-$K_\omega$ invariant otherwise.

*Proof*: (i) It is sufficient to prove the first line of (56).

$$M_n\{\omega(a_n\cdot\alpha_t\alpha_{x_t}b_n)\} = M_n\{\omega(a_1\cdot\alpha_t\alpha_{x_t}b_1\cdot\tau^n(a_2\cdot\alpha_t\alpha_{x_t}b_2))\}$$

by $\tau$-Abelian asymptotism and is also equal to $\omega(a_1\cdot\alpha_t\alpha_{x_t}b_1)\times\omega(a_2\cdot\alpha_t\alpha_{x_t}b_2)$ by weak $\tau$ clustering.

(ii) Let $(p,q)$ and $(p',q')\in\Theta_\omega$ and $a_1\in\mathfrak{A}^{(p)}$, $a_2\in\mathfrak{A}^{(p')}$, $b_1\in\mathfrak{A}^{(q)}$, $b_2\in\mathfrak{A}^{(q')}$ such that $\omega(a_1\cdot\alpha_t\alpha_{x_t}b_1)\not\equiv 0$ and $\omega(a_2\cdot\alpha_t\alpha_{x_t}b_2)\not\equiv 0$. By (56) there exists some $n\in Z$ such that $\omega(a_n\cdot\alpha_t\alpha_{x_t}b_n)\not\equiv 0$. But $a_n\in\mathfrak{A}^{(pp')}$ and $b_n\in\mathfrak{A}^{(qq')}$ and $(pp',qq')\in\Theta_\omega$. If $\omega|_{\mathfrak{A}_\omega}$ is faithful, and $a\in\mathfrak{A}^{(p)}$ such that $a\neq 0$, then $a^*\in\mathfrak{A}^{(p^{-1})}$, $aa^*\in\mathfrak{A}^{(1)} = \mathfrak{A}_\omega$, $\omega(aa^*)\neq 0$, and so $(p,p^{-1})\in\Theta_\omega$ for any $p\in\hat{K}_\omega$. If $\Theta_\omega = \Delta$, $\Theta_\omega$ is a group and $\omega$ is zero on all $\mathfrak{A}^{(p)}$, $p\neq 1$ which means that $\omega$ is $K_\omega$ invariant.

If $\Delta\subset\Theta_\omega$, let $(p,q)\in\Theta_\omega$ with $q\neq p^{-1}$. Then $(p,q)(q,q^{-1}) = (pq,1)\in\Theta_\omega$ and $(p^{-1},q)(p,q) = (1,pq)\in\Theta_\omega$; hence $(pqr,r^{-1})\in\Theta_\omega$ and $(r^{-1},pqr)\in\Theta_\omega$ for any $r\in\hat{K}_\omega$. On the other hand, if $a\in\mathfrak{A}^{(p)}$, $b\in\mathfrak{A}^{(q)}$ with $\omega(ab)\neq 0$, then $ab\in\mathfrak{A}^{(pq)}$, $b^*a^*\in\mathfrak{A}^{(p^{-1}q^{-1})}$, and $\omega(b^*a^*) = \overline{\omega(ab)}\neq 0$ which means that $(p^{-1},q^{-1})\in\Theta_\omega$. This proves that $\Theta_\omega$ is a group made of an array of manifolds "parallel" to $\Delta$. Moreover, $\omega(\alpha_k(ab)) = \langle pq,k\rangle\omega(ab)\neq\omega(ab)$ and $\omega$ is non-$K_\omega$ invariant.

*Lemma V.8*: (i) Let $f_n$ and $g_n$ be a $\beta$-KMS pair of linear combinations of positive type functionals such that $f_n(t)\underset{n\to\infty}{\longrightarrow} f(t)$ and $g_n(t)\underset{n\to\infty}{\longrightarrow} g(t)$ pointwise. Then $f$ and $g$ are a $\beta$-KMS pair.

(ii) Let $f_n$ and $g_n$ be a $\beta$-KMS pair of linear combinations of positive type functionals. Then $f_N(t) = 1/N\,\Sigma_{n=1}^N f_n(t)$ and $g_N(t) = 1/N\,\Sigma_{n=1}^N g_n(t)$ are a $\beta$-KMS pair.

(iii) If $a_m\in\mathfrak{A}^{(p_m)}$, $b_n\in\mathfrak{A}^{(q_n)}$, $m$ and $n = 1,2,\cdots,k$, $p_m\in\hat{K}_\omega$, $q_n\in\hat{K}_\omega$, $\Pi_{m=1}^k p_m = \Pi_{n=1}^k q_n = 1$, and if $\omega$ is $\beta$-KMS on $\mathfrak{A}^{(0)} = \mathfrak{A}_\omega$ with respect to $\alpha_t\alpha_{\dot{x}_t}$, then $\Pi_{m=1}^k\omega(a_m\cdot\alpha_t\alpha_{\dot{x}_t}b_m)$ and $\Pi_{m=1}^k\omega(\alpha_t\alpha_{\dot{x}_t}b_m\cdot a_m)$ are a $\beta$-KMS pair.

*Proof*: If $\int\hat{v}(t)f_n(t)\,dt = \int\hat{w}(t)g_n(t)\,dt$ as in (15), then $\int\hat{v}(t)f(t)\,dt = \int\hat{w}(t)g(t)\,dt$ by Lebesgue's dominated convergence theorem. In the same way, (ii) is straightforward by linearity of the integrals. Note that if $\{\tau^n, n\in Z\}$ is replaced by an amenable group $G$ and $M_n$ by some invariant mean:

$$\lim_{n\to\infty}\frac{1}{m(G_n)}\int_{G_n}\cdot\,dm(h),$$

where $dm$ is the Haar measure of $G$ and $G_n$ an increasing sequence of neighborhoods of the unit in $G$, the same conclusion holds for function $f_h(t)$ and $g_h(t)$ continuous in $h\in G$ and $t\in R$, and

$$f_N(t) = \frac{1}{m(G_N)}\int_{G_N} f_h(t)\,dm(h)$$

and

$$g_N(t) = \frac{1}{m(G_N)}\int_{G_N} g_h(t)\,dm(h)$$

thanks to Fubini's theorem.

Let us prove (iii) first when $k = 2$. Then $p_1\cdot p_2 = 1$, $q_1\cdot q_2 = 1$, $a_1\tau^n a_2\in\mathfrak{A}_\omega$, $b_1\tau^n b_2\in\mathfrak{A}_\omega$, and the pair $\omega(a_1\tau^n a_2\cdot\alpha_t\alpha_{\dot{x}_t}(b_1\tau^n b_2))$, $\omega(\alpha_t\alpha_{\dot{x}_t}(b_1\tau^n b_2)\cdot a_1\tau^n a_2)$, is a $\beta$-KMS pair by hypothesis. The conclusion comes from Lemma V.7.(i) and points (i) and (ii) of the present lemma. It is then sufficient to proceed by induction replacing $p_k$ by $p_k\cdot p_{k+1}$, $q_k$ by $q_k\cdot q_{k+1}$, $a_k$ by $a_k\tau^n a_{k+1}$, $b_k$ by $b_k\tau^n b_{k+1}$, applying $M_n$ and using the same trick.

*Proof of Theorem V.6 when K is Abelian and ω faithful on $\mathfrak{A}_\omega$ (for instance $\mathfrak{A}_\omega$ simple)*: Let us start from the β-KMS pair (see Lemma V.8) $\Pi_{m=1}^k \omega(a_m \cdot \alpha_t \rho_{tx_t} b_m)$ and $\Pi_{m=1}^k \omega(\alpha_t \rho_{\bar{x}_t} b_m \cdot a_m)$ with $\Pi_{m=1}^k a_m \in \mathfrak{A}_\omega$, $\Pi_{m=1}^k b_m \in \mathfrak{A}_\omega$. Writing (15) in Fourier transform, we get

$$\int v(E)\Big[d \prod_{m=1}^{k} \omega(a_m \cdot \alpha_t \rho_{\bar{x}_t} b_m)\Big]^\wedge(E)$$
$$= \int w(E)\Big[d \prod_{m=1}^{k} \omega(\alpha_t \rho_{\bar{x}_t} b_m \cdot a_m)\Big]^\wedge(E),$$

where $v$ and $w$ can now be continuous bounded functions such that $w(E) = e^{\beta E} v(E)$. Using the fact that Fourier transform turns products into convolution, we get

$$\int v(E_1 + E_2 + \cdots + E_k) \prod_{m=1}^{k} [d\omega(a_m \cdot \alpha_t \rho_{\bar{x}_t} b_m)]^\wedge(E_m)$$
$$= \int w(E_1 + E_2 + \cdots + E_k) \prod_{m=1}^{k} [d\omega(\alpha_t \rho_{\bar{x}_t} b_m \cdot a_m)]^\wedge(E_m).$$

If we replace $b_m$ by $\alpha_{t_m} \alpha_{\bar{x}_{t_m}} b_m$ (which changes nothing to the hypothesis as $\alpha_k$ and $\alpha_t \rho_{\bar{x}_t}$ commute) we get

$$\int v(E_1 + E_2 + \cdots + E_k) \prod_{m=1}^{k} e^{-t_m E_m} [d\omega(a_m \cdot \alpha_t \rho_{\bar{x}_t} b_m)]^\wedge(E_m)$$
$$= \int w(E_1 + E_2 + \cdots + E_k) \prod_{m=1}^{k} e^{-t_m E_m}$$
$$\times [d\omega(\alpha_t \rho_{\bar{x}_t} b_m \cdot a_m)]^\wedge(E_m)$$

which proves that, if we take $v = 1$,

$$\prod_{m=1}^{k} [d\omega(a_m \cdot x_t \rho_{\bar{x}_t} b_m)]^\wedge(E_m)$$
$$= \prod_{m=1}^{k} e^{\beta E_m} [d\omega(\alpha_t \rho_{\bar{x}_t} b_m \cdot a_m)]^\wedge(E_m)$$

If $k = 2$ and $(p,q) \in \Theta_\omega$, then $(p^{-1}, q^{-1}) \in \Theta_\omega$ and we can choose $\omega(a_2 \cdot \alpha_t \rho_{\bar{x}_t} b_2) \not\equiv 0$. Hence, for some convenient function $f(E_2)$, we have

$$[d\omega(a_1 \cdot \alpha_t \rho_{\bar{x}_t} b_1)]^\wedge(E_1)$$
$$= \frac{\int e^{\beta E_2} f(E_2)[d\omega(\alpha_t \rho_{\bar{x}_t} b_2 \cdot a_2)]^\wedge(E_2)}{\int f(E_2)[d\omega(a_2 \cdot \alpha_t \rho_{\bar{x}_t} b_2)]^\wedge(E_2)}$$
$$\times e^{\beta E_1} [d\omega(\alpha_t \rho_{\bar{x}_t} b_1 \cdot a_1)]^\wedge(E_1).$$

As this relation is trivial if $(p,q) \notin \Theta_\omega$, there exists some constant $c_{p,q}$ [arbitrary if $(p,q) \notin \Theta_\omega$] such that

$$[d\omega(a \cdot \alpha_t \rho_{\bar{x}_t} b)]^\wedge(E) = c_{p,q} e^{\beta E} [d\omega(\alpha_t \rho_{\bar{x}_t} b \cdot a)]^\wedge(E),$$

for $a \in \mathfrak{A}^{(p)}$, $b \in \mathfrak{A}^{(q)}$, $p,q \in \hat{K}_\omega$. Thanks to the above product formula we get

$$\prod_{m=1}^{k} c_{p_m q_m} = 1 \text{ if } \prod_{m=1}^{k} p_m = \prod_{m=1}^{k} q_m = 1,$$
$$c_{p^{-1}, q^{-1}} = c_{p,q}^{-1},$$
$$c_{pp', qq'} = c_{p,p'} \cdot c_{q,q'}.$$

On the other hand, if we remark that

$$[d\omega(a^* \cdot \alpha_t \rho_{\bar{x}_t} b^*)]^\wedge(E) = [d\omega(a \cdot \alpha_t \rho_{\bar{x}_t} b)]^\wedge(E)$$

and

$$[d\omega(a \cdot \alpha_t \rho_{\bar{x}_t} b)]^\wedge(E) = [d\omega(\alpha_t \rho_{\bar{x}_t} a \cdot b)]^\wedge(-E),$$

we get

$$\bar{c}_{p,q} = c_{q^{-1}, p^{-1}} \text{ and } c_{p,q} = c_{q,p}^{-1},$$

which, combined with the preceding formulas, gives

$$c_{p,q} = c_{q^{-1}, p^{-1}}.$$

Hence we can take $c_{p,q}$ real and $c_{p,p^{-1}} > 0$, as $\omega(a \cdot \alpha_t \rho_{\bar{x}_t} a^*)$ and $\omega(\alpha_t \rho_{\bar{x}_t} a^* \cdot a)$ are both of positive type. As $\omega$ is by hypothesis $K_\omega$ invariant, $\Theta_\omega = \Delta$ which means that the only nontrivial relation is

$$[d\omega(a \cdot \alpha_t \rho_{\bar{x}_t} b)]^\wedge(E)$$
$$= c_{p,p^{-1}} e^{\beta E} [d\omega(\alpha_t \rho_{\bar{x}_t} b \cdot a)]^\wedge(E), \quad a \in \mathfrak{A}^{(p)}, b \in \mathfrak{A}^{(p^{-1})}.$$

Writing

$$c_{p,p^{-1}} = c_{p,1} \cdot c_{1,p^{-1}} = c_{p,1}^2 = e^{\beta d(p)},$$

where

$$d(p,q) = d(p) + d(q) \quad \text{and} \quad d(p^{-1}) = -d(p),$$

we get

$$[d\omega(a \cdot \alpha_t \rho_{\bar{x}_t} b)]^\wedge(E - d(p)) = e^{\beta E} [d\omega(\alpha_t \rho_{\bar{x}_t} b \cdot a)]^\wedge$$
$$\times (E - d(p)),$$

but

$$[d\omega(a \cdot \alpha_t \rho_{\bar{x}_t} b)]^\wedge(E - d(p)) = d[e^{-id(p)t} \omega(a \cdot \alpha_t \rho_{\bar{x}_t} b)]^\wedge(E).$$

As $e^{-id(p)t}$ is, for each $t$, a character on $\hat{K}_\omega$, there exists some $k_t \in K_\omega$ such that $e^{-id(p)t} = \langle p, k_t \rangle$ and $k_t$ is a one-parameter group, and we get

$$d[e^{-id(p)t} \omega(a \cdot \alpha_t \rho_{\bar{x}_t} b)]^\wedge(E) = [d\omega(a \cdot \alpha_t \rho_{\bar{x}_t} \alpha_{k_t} b)]^\wedge(E)$$

and

$$[d\omega(a \cdot \alpha_t \rho_{\bar{x}_t} \alpha_{k_t} b)]^\wedge(E) = e^{\beta E} [d\omega(\alpha_t \rho_{\bar{x}_t} \alpha_{k_t} b \cdot a)]^\wedge(E),$$

which proves the theorem.

*Remark* 1: If $K$ is Abelian, the faithfulness of $\omega$ on $\mathfrak{A}_\omega$ is sufficient to ensure the triviality of $G_\omega$. If $a \in \mathfrak{A}^{(p)}$, $a \neq 0$, and $p \in \hat{K}_\omega$, then $\omega(a^* \alpha_k a) = \langle p, k \rangle \omega(a^* a)$. If $\omega$ is faithful, $\langle p, k \rangle \omega(a^* a) = \omega(a^* a)$ implies that $k = I$.

*Remark* 2: If $K_\omega = \mathrm{SO}(2) = \mathbb{T}$ (the one-dimensional torus), then $\hat{K}_\omega = Z$. Hence $p$ is some relative integer and there exists a real constant $\nu$ such that $d(n) = \nu n$, and $c_{p,p^{-1}} = e^{\beta \nu n}$: $\nu$ is nothing but the angular momentum of $\omega$.

## VI. ANALYSIS OF REPRESENTATIONS AND VON NEUMANN ALGEBRAS

In this part, we just want to recall that all results of Ref. 1 about the analysis of representations $\pi_\omega$ and their associated von Neumann algebras, and the restoration of the broken $K$ invariance remain valid here (Ref. 1, Theorems III.2.1, III.2.3, III.3.2, III.3.3, Corollaries III.3.5 and III.3.6).

More precisely, let us call

$$\bar{\omega} = \int_K \omega_k \circ \alpha_k \, dk \tag{58}$$

the unique $K$-invariant extension of $\omega_K = \omega|_{\mathfrak{A}_K}$ to $\mathfrak{A}$ and let us consider the following von Neumann algebras:

$$\mathcal{M} = \pi_\omega(\mathfrak{A})'', \tag{59}$$

$$\mathcal{M}_{K_\omega} = \pi_\omega(\mathfrak{A}_\omega)''$$

$$= \{a \in \mathcal{M} : \alpha_k(a) = U_\omega(k) a U_\omega(k)^{-1} = a, k \in K_\omega\}, \tag{60}$$

$$\tilde{\mathcal{M}} = \pi_{\bar{\omega}}(\mathfrak{A})'', \tag{61}$$

$$\mathcal{N} = \pi_\omega(\mathfrak{A}_K)'' = \mathcal{M}_K$$

$$= \{a \in \mathcal{M} : \alpha_k(a) = U_\omega(k) a U_\omega(k)^{-1} = a, k \in K\}, \tag{62}$$

$$\mathcal{N}_K = \pi_{\omega_K}(\mathfrak{A}_K)''. \tag{63}$$

We have the following equalities:

*Theorem VI.1*:

(i) $\bar{\omega} = \int_{K_\omega \backslash K} \omega \circ d\dot{k}$, where $d\dot{k}$ is the measure on the quotient space $K_\omega \backslash K$,

(ii) $(\pi_{\bar{\omega}}, U_{\bar{\omega}}) = \operatorname{Ind}_{K_\omega \uparrow K} (\pi_\omega, U_\omega)$ in the sense of Ref. 10,

(iii) $(\tilde{\mathcal{M}}, K, \alpha_k) = \operatorname{Ind}_{K_\omega \uparrow K} (\mathfrak{m}, K_\omega, \alpha_k)$ in the sense of Ref. 11,

(iv) $\tilde{\mathcal{M}} = \mathcal{M} \bar{\otimes} \mathcal{L}^\infty(K_\omega \backslash K)$,

(v) $\mathcal{M}_{K_\omega} = \tilde{\mathcal{M}}_K$,

(vi) $\mathcal{N} = \mathcal{M}_{K_\omega}$,

(vii) If $\omega$ is seperating, $\mathcal{N}_K = \mathcal{N}$.

*Theorem* VI.2: (i) If $G_\omega$ is trivial, any automorphism $\nu$ of $\mathcal{M}$ leaving $\mathcal{N}$ pointwise invariant and commuting with $\tau$ is of the type $\nu = \alpha_{k_\nu}$ with $k_\nu \in K_\omega$.

(ii) $\mathcal{M}$ is the dual crossed product of $\mathcal{N}$ by some dual action $\hat{\gamma}$ of $K_\omega$ on $\mathcal{N}$ and the action $\gamma$ of $K_\omega$ on $\mathcal{M}$ is the dual of $\hat{\gamma}$.

*Corollary* VI.3: Theorem VI.2 provides an alternative proof of Theorem V.6(i) and (ii).

## VII. THE CASE OF THE ROTATION GROUP AND THE ANGULAR MOMENTUM

From now on, we will assume that $K = \mathrm{SO}(3)$, the connected real three-dimensional orthogonal group. Its closed proper subgroups are the finite ones, cyclic of order $n$:$C_n$ (rotations of $2\pi k/n$ around an axis), dihedral $C_{nh}$ (add the rotations of $\pi$ around the perpendicular axis) and those related to the regular polyhedrons: tetrahedral, octahedral, and icosahedral; the nonfinite ones: SO(2) (all rotations around an axis) and O(2) (add the rotations of $\pi$ around a perpendicular axis).

According to (18) $\alpha_t \alpha_{x_t} \alpha_{k_t}$ has to be in the center of $\mathcal{C}_\omega = \mathcal{A}_\omega \square^\lambda \mathcal{H}_\omega$. Hence, if

$$(t, t_x, k_t) = [(t, y_t), (\mu(k_t), k_t)] \in \mathcal{A}_\omega \square^\lambda \mathcal{H}_\omega,$$

and

$$[(u, y), (\mu(k), k)] \in \mathcal{A}_\omega \square^\lambda \mathcal{H}_\omega,$$

we get

$$[(t + u, y_t + k_t y), (\mu(k_t k), k_t k)]$$
$$= [(u + t, y + k y_t), (u(k k_t), k k_t)],$$

which implies that

$$y_t + k_t y = y + k y_t, \quad k_t k = k k_t, \quad y \in \mathcal{A}_\omega, (\mu(k), k) \in \mathcal{H}_\omega, \tag{64}$$

or, equivalently,

$$y_t = k y_t, \quad y = k_t y, \quad k_t k = k k_t. \tag{64'}$$

Let $\varphi(\mathcal{H}_\omega) \subset Q_\omega$ be the (nonnecessarily closed) projection of $\mathcal{H}_\omega$ into $Q_\omega$. We can distinguish three cases.

*First case*: $\overline{\varphi(\mathcal{H}_\omega)} = \mathrm{SO}(3)$ or a finite group $\not\subset \mathrm{O}(2)$. In such a case $\varphi(\mathcal{H}_\omega)$ has a trivial center, which implies that $k_t = I$, and no fixed directions, which implies that $y_t = 0$. Hence $\omega$ is only an $\alpha_t$-KMS state and $\mathcal{G}_\omega$ is a closed subgroup of $(T \times R^3) \square \mathrm{SO}(3)$ not included into $(T \times R) \square \mathrm{O}(2)$ as a subgroup.

*Second case*: $\overline{\varphi(\mathcal{H}_\omega)} = \mathrm{O}(2)$ or SO(2) or a finite group $\subset \mathrm{O}(2)$. In such a case $\varphi(\mathcal{H}_\omega)$ is Abelian, and $k k_t = k_t k$ is always true, while $y$ and $y_t$ have to be in the fixed axis of O(2). Hence $\mathcal{A}_\omega \subset R$ and $\omega$ is an $\alpha_t \alpha_{x_t} \alpha_{(\mu(k_t), k_t)}$-KMS state. If the state is moving uniformly, $k_t = \nu t$ and $y_t = \upsilon t$ where $\nu$ is the angular momentum and $\upsilon$ the velocity. Moreover, $\mathcal{G}_\omega$ is a closed subgroup of $(T \times R) \square \mathrm{O}(2)$.

*Third case*: $\overline{\varphi(\mathcal{H}_\omega)} = I$. All required conditions (64) are satisfied and $\omega$ is an $\alpha_t \alpha_{x_t}$-KMS state while $\mathcal{G}_\omega$ is a closed subgroup of $T \times R^3$.

It would be interesting to list all closed subgroup of $(T \times R^3) \square \mathrm{SO}(3)$. This can be done using the techniques of Sec. II. The first step consists in the construction of all closed subgroups of $T \times R^3$ and their corresponding $Q$.

### A. Construction of the closed subgroups of $T \times R^3$

As it is well known, the closed subgroups of $T \times R^3 \approx R^4$ are all isomorphic to:

$$O, Z, R, Z^2, Z \oplus R, R^2, Z^3, Z^2 \oplus R, Z \oplus R^2, R^3,$$
$$Z^4, Z^3 \oplus R, Z^2 \oplus R^2, Z \oplus R^3, R^4 \tag{65}$$

But this classification is not detailed enough for us because $\mathcal{H}$ is acting of the space component of $\mathcal{A}$ only. Hence we have to be more precise and specify the relative position of these subgroups with respect to time and space. For this purpose, let us construct the closed subgroups of $T \times R^3$ using the techniques of Sec. II. The procedure is here much more simple because $R^3$ does not act onto $T$. We get three kinds of situations according to the type of "pure time translations."

(1) *The pure time translations are the whole of T*

In such a case we get the direct product of $T$ by a closed subgroup of $R^3$, that is to say

$$T\,[Q=\mathrm{SO}(3)],\ T\times Z\,[Q=\mathrm{O}(2)],\ T\times R\,[Q=\mathrm{O}(2)],\ T\times Z^2\ (Q \text{ finite}),$$
$$T\times(Z\oplus R)\ (Q\text{ finite}), T\times R^2\,[Q=\mathrm{O}(2)],\ T\times Z^3\ (Q\text{ finite}), \tag{66}$$
$$T\times(Z^2\oplus R)\ (Q\text{ finite}),\ T\times(Z\oplus R^2)\,[Q=\mathrm{O}(2)],\ T\times R^3\,[Q=\mathrm{SO}(3)].$$

A more precise description of these subgroups can be given by the choice of a basis $(\mathbf{a}_i)$ $i=1,2,3$ of $R^3$: their generic element can then be written

$$\left(t,\ \sum_{i=1}^{p} r^i\mathbf{a}^i+\sum_{j=p+1}^{r} n^j\mathbf{a}_j\right),\quad t\in R, r^i\in R, n^j\in Z, \tag{67}$$

where $p$ and $r$ are integers such that $0\leqslant p\leqslant r\leqslant 3$ called, respectively, the dimension and the rank of the space component of these groups.[12]

(2) *The pure time translations are a subgroup Z of R*

The structure of these subgroups is

$$Z\square^{\lambda}(\chi(\mathbf{x}),\mathbf{x}), \tag{68}$$

where $\mathbf{x}$ describes some closed subgroup of $R^3$, $\chi$ some continuous additive (hence linear) map from such closed subgroup of $R^3$ into the one-dimensional torus and $\lambda$ the natural extension of $Z$ by this torus, that is to say

$$Z\ [Q=\mathrm{SO}(3)],\ Z\square^{\lambda}(\chi,Z)[Q=\mathrm{O}(2)],\ Z\square^{\lambda}(\chi,R)\ [Q=\mathrm{O}(2)],$$
$$Z\square^{\lambda}(\chi,Z^2)\ (Q\text{ finite if }\chi=0,\text{ trivial if }\chi\neq0),$$
$$Z\square^{\lambda}(\chi,Z\oplus R)\ (Q\text{ finite if }\chi=0,\text{ trivial if }\chi\neq0),$$
$$Z\square^{\lambda}(\chi,R^2)\ [Q=\mathrm{O}(2)\text{ if }\chi=0,\text{ trivial if }\chi\neq0], \tag{69}$$
$$Z\square^{\lambda}(\chi,Z^3)\ (Q\text{ finite if }\chi=0,\text{ trivial if }\chi\neq0),$$
$$Z\square^{\lambda}(\chi,Z^2\oplus R)\quad (Q\text{ finite if }\chi|_{Z^2}=0,\text{ trivial if }\chi|_{Z^2}\neq0),$$
$$Z\square^{\lambda}(\chi,Z\oplus R^2)\quad [Q=\mathrm{O}(2)\text{ if }\chi|_{Z^2}=0,\text{ trivial if }\chi|_{Z^2}\neq0],$$
$$Z\square^{\lambda}(\chi,R^3)\ [Q=\mathrm{SO}(3)\text{ if }\chi=0,\text{ trivial if }\chi\neq0].$$

If $a$ is the fundamental length of $Z$ and $(a_i)_{i=1,2,3}$ some basis of $R^3$, the generic element of these subgroups can be written [with the same notation in (67)]

$$\left(na+\sum_{i=1}^{p} r^i\chi(\mathbf{a}_i)+\sum_{j=p+1}^{r} n^j\chi(\mathbf{a}_j),\ \sum_{i=1}^{p} r^i\mathbf{a}_i+\sum_{j=p+1}^{r} n^j\mathbf{a}_j\right). \tag{70}$$

(3) *The pure time translations reduce to* $\{0\}$

The closed subgroups are then the closed graphs of continuous linear applications $(\chi(\mathbf{x}),\mathbf{x})$ from some (nonnecessarily closed) subgroup of $T\times R^3$ to $R$. Being closed subgroups of $R^4$, there exists a basis $\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3,\mathbf{e}_4$ of $R^4=T\times R^3$ such that

$$(\chi(\mathbf{x}),\mathbf{x})=\sum_{i=1}^{p} r^i\mathbf{e}_i+\sum_{j=p+1}^{r} n^j\mathbf{e}_j,$$
$$r^i\in R,\ n^j\in Z,\ 0\leqslant p\leqslant r\leqslant 4,\ p\leqslant 3 \tag{71}$$

(the condition $p\leqslant 3$ coming from the fact that all pure time translations are zero). In the canonical basis $\boldsymbol{\epsilon}_0\in T$, $(\boldsymbol{\epsilon}_1,\boldsymbol{\epsilon}_2,\boldsymbol{\epsilon}_3)\in R^3$ of $T\times R^3$,

$$\mathbf{e}_i=\alpha_i^0\boldsymbol{\epsilon}_0+\sum_{l=1}^{3}\alpha_i^l\boldsymbol{\epsilon}_l=\alpha_i^0\boldsymbol{\epsilon}_0+\mathbf{a}_i$$
$$=\chi(\mathbf{a}_i)\boldsymbol{\epsilon}_0+\mathbf{a}_i=(\chi(\mathbf{a}_i),\mathbf{a}_i), \tag{72}$$

with $\mathbf{a}_i\neq0$, i.e., for any $i$ there is some $l$ such that $\alpha_i^l\neq0$, and so

$$(\chi(\mathbf{x}),\mathbf{x})=\left(\sum_{i=1}^{p} r^i\chi(\mathbf{a}_i)+\sum_{j=p+1}^{r} n^j\chi(\mathbf{a}_j),\ \sum_{i=1}^{p} r^i\mathbf{a}_i+\sum_{j=p+1}^{r} n^j\mathbf{a}_j\right) \tag{73}$$

but now the $(\mathbf{a}_i)_{i=1,\cdots,r}$ are not necessarily linearly independent.

Another way to represent these groups is to write

$$\mathbf{x}=\sum_{i=1}^{p}\sum_{l=1}^{3} r^i\alpha_i^l\boldsymbol{\epsilon}_l+\sum_{j=p+1}^{r}\sum_{l=1}^{3} n^j\alpha_j^l\boldsymbol{\epsilon}_l$$
$$=\sum_{l=1}^{3}\left\{\sum_{i=1}^{p}\alpha_i^l r^i+\sum_{j=p+1}^{l}\alpha_j^l n^j\right\}\boldsymbol{\epsilon}_l=A\mathbf{y}, \tag{74}$$

where $A$ is the $3\times r$ matrix $(\alpha_i^l)$ and $y=(r^1,\cdots r^p,n^{p+1},\cdots,n^r)$. Hence $y$ describes a closed subgroup of $R^4$ and

$$(\chi(\mathbf{x}),\mathbf{x})=(\chi(A\mathbf{y}),A\mathbf{y})=(\mathbf{k}\cdot\mathbf{y},A\mathbf{y}),\quad \mathbf{k}=(k_i)=(\chi(a_i)). \tag{73'}$$

Let us list the different cases, with their corresponding $Q$.

$-r=p=0$, $\mathscr{A}=\{0\}$, $Q=\mathrm{SO}(3)$,

$-r=1$, $p=0$, $\mathscr{A}=n(\chi(\mathbf{a}),\mathbf{a})\sim Z$, $\mathbf{a}\neq0$, $\mathbf{a}\in R^3$, $n\in Z$; $Q=\mathrm{O}(2)$,

$-r=1$, $p=1$, $\mathscr{A}=r(\chi(\mathbf{a}),\mathbf{a})\sim R$, $\mathbf{a}\neq0$, $\mathbf{a}\in R^3$, $r\in R$; $Q=\mathrm{O}(2)$,

$-r=2$, $p=0$, $\mathscr{A}\sim Z^2$, but we have to distinguish different cases.

1*st case*: $\mathbf{a}_1$ and $\mathbf{a}_2$ are linearly independent. Let $\chi(\mathbf{a}_2)=\alpha\chi(\mathbf{a}_1)$. Then $\mathscr{A}=\{((n^1+\alpha n^2)\chi(\mathbf{a}_1),n^1\mathbf{a}_1+n^2\mathbf{a}_2)\}$.

The $R^3$ projection is $Z^2$, the $T$ projection is $Z$ or dense according to whether $\alpha$ is rational or not. $Q$ is finite if $\chi = 0$ and is otherwise trivial.

*2nd case*: $\mathbf{a}_2 = \lambda\mathbf{a}_1 \neq 0, \chi(\mathbf{a}_1) \neq 0$ [$\chi(\mathbf{a}_1)$ and $\chi(\mathbf{a}_2)$ cannot both be zero by independence $\mathbf{e}_1$ and $\mathbf{e}_2$], $\chi(\mathbf{a}_2) = \alpha\chi(\mathbf{a}_1)$, $\alpha \neq \lambda$ (for the same reason). Then $\mathscr{A} = \{((n^1 + \alpha n^2)\chi(\mathbf{a}_1), (n^1 + \lambda n^2)\mathbf{a}_1), \quad \alpha \neq \lambda, \lambda \text{ irrational}\}$, the condition $\lambda$ irrational coming from the nullity of pure time translations. The $R^3$ projection is dense in the one-dimensional subspace generated by $\mathbf{a}_1$, the $T$ projection is $Z$ or dense according to $\alpha$ rational or not. As $\chi \neq 0$, $Q$ is trivial.

$- r = 2, p = 1$, $\mathscr{A} \sim Z \oplus R$, and we should have two cases as before, but the second one is excluded by the fact that $r + \lambda n$ can be zero for nonzero values or $r$ and $n$, giving rise to nonzero time translations. Hence we have $\mathbf{a}_1$ and $\mathbf{a}_2$ linearly independent, $\chi(\mathbf{a}_2) = \alpha\chi(\mathbf{a}_1)$, $\mathscr{A} = \{((r + \alpha n)\chi(\mathbf{a}_1), r\mathbf{a}_1 + n\mathbf{a}_2)\}$. The $R^3$ projection is $Z \oplus T$ and the $T$ projection is $R$. $Q$ is finite if $\chi = 0$ and trivial otherwise.

$- r = 2, p = 2$ $\mathscr{A} \sim R^2$ with only one case as above. With $\mathbf{a}_1$ and $\mathbf{a}_2$ linearly independent, $\chi(\mathbf{a}_2) = \alpha\chi(\mathbf{a}_1)$, $\mathscr{A} = \{((r^1 + \alpha r^2)\chi(\mathbf{a}_1), r^1\mathbf{a}_1 + r^2\mathbf{a}_2)\}$. The $R^3$ projection is $R^2$, the $T$ projection is $R$. $Q = \mathrm{O}(2)$ if $\chi = 0$, trivial otherwise.

$- r = 3, p = 0$, $\mathscr{A} \sim Z^3$, but we have to distinguish different cases.

*1st case*: $\mathbf{a}_1, \mathbf{a}_2$ and $\mathbf{a}_3$ linearly independent. Let $\chi(\mathbf{a}_2) = \alpha\chi(\mathbf{a}_1), \chi(\mathbf{a}_3) = \beta\chi(\mathbf{a}_1)$. Then $\mathscr{A} = \{(n^1 + \alpha n^2 + \beta n^3)\chi(\mathbf{a}_1), n^1\mathbf{a}_1 + n^2\mathbf{a}_2 + n^2\mathbf{a}_3\}$. The $R^3$ projection is $z^3$, the $T$ projection is $Z$ or dense according to $\alpha$ and $\beta$ both rational or not. $Q$ is finite if $\chi = 0$ and trivial otherwise.

*2nd case*: $\mathbf{a}_1 = \lambda\mathbf{a}_2 + \mu\mathbf{a}_3, \chi(a_1) \neq 0$ and $1 - \lambda\alpha - \mu\beta \neq 0$ for the linear independence of the $(e_i)$. Then $\mathscr{A} = ((n^1 + \alpha n^3 + \beta n^3)\chi(\mathbf{a}_1), (n^1\lambda + n^2)\mathbf{a}_2 + (n^1\mu + n^3)\mathbf{a}_3)$ with $\lambda$ or $\mu$ irrational. The $R^3$ projection if made of $Z^2$ (from $n^2\mathbf{a}_2 + n^3\mathbf{a}_3$) and of $Z$ [from $n^1(\lambda\mathbf{a}_2 + \mu\mathbf{a}_3)$], with dense projection on $\mathbf{a}_2$ or $\mathbf{a}_3$ or both, the $T$ projection is $Z$ or dense according to $\alpha$ and $\beta$ both rational or not. As $\chi \neq 0$, $Q$ is trivial.

*3rd case*: $\mathbf{a}_1 = \lambda\mathbf{a}_2 = \mu\mathbf{a}_3$. This case is impossible by the linear independence of the $(e_i)$.

$- r = 3, p = 1$, $\mathscr{A} \sim Z^2 \oplus R$ with three different cases.

*1st case*: $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ linearly independent. With the same notations, $\mathscr{A} = \{r^1 + \alpha n^2 + \beta n^3)\chi(a_1), r^1\mathbf{a}_1 + n^2\mathbf{a}_2 + n^3\mathbf{a}_3\}$. The $R^3$ projection is $Z^2 \oplus R$, the $T$ projection is $R$. $Q$ is finite if $\chi|_{Z^2}$ is zero, (i.e., $\alpha$ and $\beta$ zero) and trivial otherwise.

*2nd case*: $\mathbf{a}_1 = \lambda\mathbf{a}_2 + \mu\mathbf{a}_3, \chi(\mathbf{a}_1) \neq 0, \quad 1 - \alpha\lambda - \beta\mu \neq 0$. Then $\mathscr{A} = \{(r^1 + \alpha n^2 + \beta n^3)\chi(\mathbf{a}_1), (r^1\lambda + n^2)\mathbf{a}_2 + (r^1\mu + n^3)\mathbf{a}_3\}$ with $\lambda$ and $\mu$ not both zero and $\lambda/\mu$ irrational to avoid pure time translations. The $R^3$ projection is made of $Z^2$ and $R$, the $T$ projection is $R$. $Q$ is trivial.

*3rd case*: $\mathbf{a}_1 = \lambda\mathbf{a}_2 = \mu\mathbf{a}_3$ is impossible as above.

$- r = 3, p = 2$, $\mathscr{A} \sim Z \oplus R^2$ with three different cases. cases.

*1st case*: $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ linearly independent, $\chi(\mathbf{a}_1) = \alpha\chi(\mathbf{a}_3)$, $\chi(\mathbf{a}_2) = \beta\chi(\mathbf{a}_3)$. Then $\mathscr{A} = \{(\alpha r^1 + \beta r^2 + n^3)\chi(\mathbf{a}_3), r^1\mathbf{a}_1 + r^2\mathbf{a}_2 + n^3\mathbf{a}_3\}$. The $R^3$ projection is $Z \oplus R^2$, the $T$ projection is $R$. $Q$ is $\mathrm{O}(2)$ if $\chi|_{R^2}$ is zero (i.e., if $\alpha = \beta = 0$) and trivial otherwise.

*2nd case*: $\mathbf{a}_3 = \lambda\mathbf{a}_1 + \mu_2\mathbf{a}_2$ is impossible if we do not want pure time translations.

*3rd case*: $\mathbf{a}_1 = \lambda\mathbf{a}_2 = \mu\mathbf{a}_3$ is impossible as above.

$- r = 3, p = 3$, $\mathscr{A} \sim R^3$ with three cases.

*1st case*: $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ linearly independent, $\chi(\mathbf{a}_2) = \alpha\chi(\mathbf{a}_1)$, $\chi(\mathbf{a}_3) = \beta\chi(\mathbf{a}_1)$. Then $\mathscr{A} = \{(r^1 + \alpha r^2 + \beta r^3)\chi(\mathbf{a}_1), r^1\mathbf{a}_1 + r^2\mathbf{a}_2 + r^3\mathbf{a}_3\}$. The $R^3$ projection is $R^3$, the $T$ projection is $R$. $Q = \mathrm{SO}(3)$ if $\chi = 0$ and trivial otherwise.

*The 2nd and 3rd cases* are excluded for the same reasons as above.

$- r = 4, p = 0$, $\mathscr{A} \sim Z^4$ with four different cases.

*1st case*: $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ linearly independent. This is impossible because $R^3$ is of dimension 3.

*2nd case*: $\mathbf{a}_4 = \lambda\mathbf{a}_1 + \mu\mathbf{a}_2 + \nu\mathbf{a}_3$, $\chi(a_4) \neq 0$, $\chi(\mathbf{a}_1) = \alpha\chi(\mathbf{a}_4)$, $\chi(\mathbf{a}_2) = \beta\chi(\mathbf{a}_4)$, $\chi(\mathbf{a}_3) = \gamma\chi(\mathbf{a}_4)$, $\alpha\lambda + \beta\mu + \gamma\nu - 1 \neq 0$. Then $\mathscr{A} = \{(n^1\alpha + n^2\beta + n^3\gamma + n^4)\chi(\mathbf{a}_4), (n^1 + n^4\lambda)\mathbf{a}_1 + (n^2 + n^4\mu)\mathbf{a}_2 + (n^3 + n^4\nu)\mathbf{a}_3\}$ with $\lambda$ or $\mu$ or $\nu$ irrational. The $R^3$ projection is made of $Z^3$ and $Z$, with projection on $\mathbf{a}_1$ or $\mathbf{a}_2$ or $\mathbf{a}_3$ dense according to the irrationality of $\lambda$ or $\mu$ or $\nu$; the $T$ projection is $Z$ or dense according to $\alpha, \beta$ and $\gamma$ all rational or not. $Q$ is trivial.

*3rd and 4th cases* (i.e., $\mathbf{a}_3 = \lambda\mathbf{a}_1 + \mu\mathbf{a}_2$ and $\mathbf{a}_4 = \nu\mathbf{a}_1 + \delta\mathbf{a}_2$, or $\mathbf{a}_1 = \lambda\mathbf{a}_2 = \mu\mathbf{a}_3 = \nu\mathbf{a}_4$) are impossible by linear independence of the $(e_i)$.

$- r = 4, p = 1$, $\mathscr{A} \sim Z^3 \oplus R$: also four cases, the only realistic one being, with the same notation and conditions as before, $\mathscr{A} = \{(n^1\alpha + n^2\beta + n^3\gamma + r^4)\chi(\mathbf{a}_4), (n^1 + \lambda r^4)\mathbf{a}_1 + (n^2 + \mu r^4)\mathbf{a}_2 + (n^3 + \nu r^4]\mathbf{a}_3\}$ with $\lambda/\mu$ or $\mu/\nu$ or $\nu/\lambda$ irrational. The $R^3$ projection is made of $Z^3$ and $R$, the $T$ projection is $R$. $Q$ is trivial.

$- r = 4, p = 2$, $\mathscr{A} \sim Z^2 \oplus R^2$. Here also, only one case is realistic, $\mathscr{A} = \{(n^1\alpha + n^2\beta + r^3\gamma + r^4)\chi(\mathbf{a}_4), (n^1 + \lambda r^4)\mathbf{a}_1 + (n^2 + \mu r^4)\mathbf{a}_2 + (r^3 + \nu r^4)\mathbf{a}_3\}$ with $\mu/\nu$ irrational. The $R^3$ projection is made of $Z^2 \oplus R$ and $R$. $Q$ is trivial.

$- r = 4, p = 3$ and $r = 4, p = 4$ are all impossible.

## B. Construction of the closed subgroups of $T \times R \,\square\, \mathrm{SO}(2)$

As the sequel of the construction of closed subgroups of $(T \times R^3) \square \mathrm{SO}(3)$ is rather tedious, let us proceed only in the most interesting case (cf. second case) of closed subgroups of $(T \times R) \square \mathrm{SO}(2)$. In such a case, the $\mathscr{A}$'s reduce to the following list, $Q$ being always $\mathrm{O}(2)$, where $\mathbb{T}$ means the one-dimensional torus:

(a)

$$\mathscr{A} = T \times I = \{(t,0), t \in T\}; \frac{T \times R}{\mathscr{A}} = R; \mathscr{N} = R \times \mathrm{SO}(2)$$

(b) $$\mathscr{A} = T \times Z = \{(t,na), t \in T, n \in Z\}; \frac{T \times R}{\mathscr{A}} = \mathbb{T}_R; \quad \mathscr{N} = \mathbb{T}_R \times \mathrm{SO}(2)$$

(c)
$\mathscr{A} = T \times R = \{(t,ra), t\in T, r\in R\}; \dfrac{T\times R}{\mathscr{A}} = I; \mathscr{N} = \mathrm{SO}(2)$

(d) $\mathscr{A} = Z \times I = \{(nt_0,0), n\in Z\}; \dfrac{T\times R}{\mathscr{A}} = \mathbb{T}_T \times R;$
$\mathscr{N} = (\mathbb{T}_T \times R)\times \mathrm{SO}(2)$

(e)
$\mathscr{A} = Z\square^\lambda(\chi,Z) = \{(nt_0 + mt_1, ma), n,m\in Z\};$
$\dfrac{T\times R}{\mathscr{A}} = \mathbb{T}_T \times \mathbb{T}_R; \mathscr{N} = (\mathbb{T}_T\times\mathbb{T}_R)\times\mathrm{SO}(2)$

(f)
$\mathscr{A} = Z\square^\lambda(\chi,R) = \{(nt_0 + rt_1, ra), n\in R, r\in R\};$
$\dfrac{T\times R}{\mathscr{A}} = \mathbb{T}_T; \mathscr{N} = \mathbb{T}_T\times\mathrm{SO}(2)$

(g)
$\mathscr{A} = I = (0,0); \dfrac{T\times R}{\mathscr{A}} = T\times R; \mathscr{N} = (T\times R)\times\mathrm{SO}(2)$

(h) $\mathscr{A} = (\chi,Z) = \{n(t_0,a), n\in Z\}; \dfrac{T\times R}{\mathscr{A}} \sim \mathbb{R}\times\mathbb{T};$
$\mathscr{N}\sim(\mathbb{R}\times\mathbb{T})\times\mathrm{SO}(2)$

(i)
$\mathscr{A} = (\chi,R) = \{r(t_0,a), r\in R\}; \dfrac{T\times R}{\mathscr{A}} \sim \mathbb{R}; \mathscr{N}\sim\mathbb{R}\times\mathrm{SO}(2)$

(j)
$\mathscr{A} = Z^2 = \{n(t_0,a) + m(\alpha t_0, \lambda a), \alpha\neq\lambda, \lambda$ irrational,
$n,m\in Z\};$
$\dfrac{T\times R}{\mathscr{A}} \sim \mathbb{T}\times\mathbb{T}; \mathscr{N}\sim(\mathbb{T}\times\mathbb{T})\times\mathrm{SO}(2)$

(a) We have to determine a closed subgroup $\mathscr{H}$ of $R\times\mathrm{SO}(2)\sim\mathbb{R}\times\mathbb{T}$, whose universal covering is $R^2$,

$$R\times R \xrightarrow{P} R\times\mathrm{SO}(2), \quad \mathrm{Ker}P = Z. \tag{75}$$

Hence $P^{-1}(\mathscr{H})$ is closed subgroup of $R^2$, with closed projection onto the first factor $R$ and intersection with $P^{-1}(R)$ included into Ker$P$ (to avoid pure space translation). So

$$P^{-1}(\mathscr{H}) = \{(nx_0 + \alpha m x_0, (n+\lambda m)a)\} = \{n(x_0,a) + m(\alpha x_0, \lambda a)\}\sim Z^2,$$

where $x_0,\alpha,\lambda,a\in R$, $n,m\in Z$, $x_0\neq0$ and $\lambda$ irrational (to exclude pure space translations), $\alpha$ rational (to get a closed projection onto $R$), $(n+\lambda m)a\neq 2k\pi$ for any nonzero couple of integers $n$ and $m$ if $\alpha\neq0$ or else for any $n\neq0$ and $m$ integers if $\alpha=0$ (to avoid pure space translations); or else

$$P^{-1}(\mathscr{H}) = \{n(x_0,a)\}\sim Z, \quad a\in R, n\in Z,$$

with $na\neq 2k\pi$ for $n\neq0$ if $x_0\neq0$; or else

$$P^{-1}(\mathscr{H}) = \{r,a\}\sim R, \quad r\in R, a\in R.$$

This gives rise to the three following groups:

$$\mathscr{G} = \{(t,(n+\alpha m)x_0, n\theta_1 + m\theta_2)\},$$

where $t,x_0\in R$, $n,m\in Z$, $x_0\neq0$, $\theta_1,\theta_2\in\mathrm{SO}(2)$, $\lambda = (\theta_2/\theta_1)$ irrational, $\alpha$ rational, $n\theta_1 + m\theta_2\neq 2k\pi$ for any nonzero couple of integers $n$ and $m$ if $\alpha\neq0$ or else for any $n\neq0$ and $m$ integers if $\alpha=0$;

$$\mathscr{G} = \{(t,nx_0,n\theta_0)\}$$

where $t,x_0\in R$, $n\in Z$, $\theta_0\in\mathrm{SO}(2)$, $\theta_0/2\pi$ irrational if $x_0\neq0$;

$$\mathscr{G} = \{(t,r\theta_0)\},$$

with $t,r\in R$, $\theta_0\in\mathrm{SO}(2)$. The corresponding states are $\alpha_{t}$-, $\alpha_t$- and $\alpha_t\alpha_{\theta_t}$-KMS, respectively.

(b) By the same kind of discussion, we get

$$\mathscr{G} = \left\{\left(t,(nx_0,mx_1),m\frac{2\pi}{k}\right)\right\},$$

where $t,x_0\in R$, $m,n,k\in Z$, $x_1\in\mathbb{T}_R$, $kx_1 = 0$; $\quad x_0\neq0$ and

$$\mathscr{G} = \{(t,(nx_0,rx_1),r\theta_0)\}$$

where $t,x_0,r\in R$, $n\in Z$, $\theta_0\in\mathrm{SO}(2)$, $2\pi x_1/\theta_0 = 0$, $\quad x_0\neq0$. The corresponding states are, respectively, $\alpha_t$ and $\alpha_t\alpha_{x_t}\alpha_{\theta_t}$-KMS.

(c) We immediately get

$$\mathscr{G} = \{(t,x)\},$$

with $t,x\in R$, or else

$$\mathscr{G} = \left\{\left(t,x,n\frac{2\pi}{k}\right)\right\},$$

with $t,x\in R$, $n$, $k\in Z$, and

$$\mathscr{G} = \{(t,x,r\theta_0)\},$$

with $t,x,r\in R$, $\theta_0\in\mathrm{SO}(2)$. The corresponding states are, respectively, $\alpha_t\alpha_{x_t}$, $\alpha_t\alpha_{x_t}$ and $\alpha_t\alpha_{x_t}\alpha_{\theta_t}$-KMS.

(d) We have to determine a closed subgroup $\mathscr{H}$ of $\mathbb{T}_T\times R\times\mathrm{SO}(2)$ whose projections onto $R\times\mathrm{SO}(2)$, $\mathbb{T}_T\times R$ and $R$ have to be closed. Hence, using the results of (1), we get

$$\mathscr{G} = \{((pt_0,(n+\beta m)t_1),(n+\alpha m)x_0, n\theta_1 + m\theta_2)\},$$

where $t_0,x_0\in R$, $p,n,m\in Z$, $t_1\in\mathbb{T}_T$, $x_0\neq0$, $\theta_1,\theta_2\in\mathrm{SO}(2)$, $\lambda = (\theta_2/\theta_1)$ irrational, $\alpha$ is rational, $n\theta_1 + m\theta_2\neq2k\pi$ for any nonzero couple of integers $n$ and $m$ if $\alpha$ and $\beta$ are not both zero, or else for any $n\neq0$ and $m$ integers if $\alpha$ and $\beta$ are both zero,

$$\mathscr{G} = \{((pt_0,nt_1),nx_0,n\theta_0)\}$$

where $t_0,x_0\in R$, $p,n,\in Z$, $t_1\in\mathbb{T}_T$, $\theta_0\in\mathrm{SO}(2)$, $(\theta_0/2\pi)$ irrational if $x_0\neq0$, $\theta_1 = t_1 = 2\pi/k$ if $x_0 = 0$, $\theta_0 = 2\pi/k$ if $x_0 = t_1 = 0$;

$$\mathscr{G} = \{((pt_0,rt_1),r\theta_0)\},$$

where $t_0\in R$, $t_1\in\mathbb{T}_T$, $p\in Z$, $r\in R$ and $\theta_0\in\mathrm{SO}(2)$. The last one alone gives rise to an $\alpha_t\alpha_{\theta_t}$-KMS state.

(e) Here all projections of $\mathscr{H}$ on any factor have to be closed. Using the results of (b), we get

$$\mathscr{G} = \{((nt_0,mt_1,pt_2),(mx_0,px_1),p\theta_0)\},$$

where $t_0,x_0\in R$, $t_1,t_2\in\mathbb{T}_T$, $x_1\in\mathbb{T}_R$, $\theta_0\in\mathrm{SO}(2)$, $n,m,p\in Z$, $x_0\neq0$, $\theta_0 = 2\pi/k$, $kx_1 = 0$, $kt_2 = t_1$;

$$\mathscr{G} = \{((nt_0,mt_1,rt_2),(mx_0,rx_1),r\theta_0)\},$$

where $t_0,x_0\in R$, $t_1,t_2\in\mathbb{T}_T$, $x_1\in\mathbb{T}_R$, $\theta_0\in\mathrm{SO}(2)$, $n,m\in Z$, $r\in R$, $x_0\neq0$,

$$\frac{2\pi x_1}{\theta_0} = 0, \quad \frac{2\pi t_2}{\theta_0} = t_1.$$

The second group only gives rise to an $\alpha\rho_{x_t}\alpha_{\theta_t}$-KMS state.

(f) Here $\mathscr{H}$ is similar to the corresponding one in (2) with exchange of time and space. So

$$\mathscr{G} = \{((nt_0,rt_1,mt_2),rx_0,m\theta_0)\},$$

with $n,m\in Z$, $r\in R$, $t_0,x_0\in R$, $t_1,t_2\in \mathbb{T}_T$, $\theta_0\in SO(2)$, $\theta_0 = 2\pi/k$, $kt_2 = 0$, $x_0\neq 0$.

$$\mathscr{G} = \{((nt_0,rt_1,r't_2),rx_0,r'\theta_0)\},$$

with $n\in Z$, $r,r'\in R$, $t_0,x_0\in R$, $t_1,t_2\in\mathbb{T}_T$, $\theta_0\in SO(2)$, $(2\pi/\theta_0)t_2 = 0$. $x_0\neq 0$. The second group gives rise to an $\alpha\rho_{x_t}\alpha_{\theta_t}$-KMS state.

(g) We have to determine a closed subgroup $\mathscr{H}$ of $T\times R\times SO(2)\sim R^2\times\mathbb{T}$ whose universal covering is $R^3$

$$T\times R\times R \xrightarrow{P} T\times R\times SO(2), \mathrm{Ker}P = Z. \qquad (75')$$

Hence $P^{-1}(\mathscr{H})$ is a closed subgroup of $T\times R\times R$, with closed projection onto $T\times R$ and intersection with $P^{-1}(T\times R)$ included into Ker$P$. We then get

$$\mathscr{G}\{((n+\alpha m+\beta p)t_0,(n+\alpha m+\delta p)x_0,n\theta_1+m\theta_2+p\theta_3)\},$$

where $n,m,p\in Z$, $t_0,x_0\in R$, $\theta_1,\theta_2,\theta_3\in SO(2)$, $\alpha$ rational, $\beta,\delta\in R$, $\mu$ and $\nu$ not both rational, $x_0$ and $t_0$ not both zero, $n\theta_1 + m\theta_2 + p\theta_3\neq 2k\pi$ for any nonzero triple of integers $n$, $m$ and $p$, if $\alpha,\beta,\delta$ not all zero, or for any nonzero couple $(n,p)$ and any integer $m$ if $\alpha = 0$, or for any nonzero couple $(m,n)$ and any integer $p$ if $\beta$ and $\alpha$ are both zero, of for any $n\neq 0$ and any $m$ and $p$ if $\alpha,\beta$ and $\delta$ are all zero. If $t_0$ (or $x_0$) are zero, we have the case of (a), and then $\theta_3 = 0$, $\delta = 0$ (or $\beta = 0$).

$$\mathscr{G}\{((n+\alpha m)t_0,(n+\beta m)x_0,n\theta_1+m\theta_2)\},$$

where $n,m\in Z$, $t_0,x_0\in R$, $\theta_1,\theta_2\in SO(2)$, $\alpha,\beta\in R$, $x_0$ and $t_0$ are not both zero, $\alpha$ rational if $x_0\neq 0$ and $\beta$ rational if $t_0 = 0$, $n\theta_1 + m\theta_2\neq 2k\pi$ for any nonzero couple of integers $n$ and $m$ if $\alpha$ and $\beta$ are not both zero or for any $n\neq 0$ and $m$ integers if $\alpha$ and $\beta$ are both zero.

$$\mathscr{G} = \{(nt_0,nx_0,n\theta_0)\},$$

where $t_0,x_0\in R$, $n\in Z$, $\theta_0\in SO(2)$, $\theta_0/2\pi$ is irrational if $x_0$ and $t_0$ are not both zero, $\theta_0 = 2\pi/k$ if $x_0$ and $t_0$ are both zero.

$$\mathscr{G} = \{(r\theta_0)\},$$

where $r\in R$ and $\theta_0\in SO(2)$. Only the last group gives rise to an $\alpha_{\theta_t}$-KMS state.

(h) The closed subgroups $\mathscr{H}$ are similar to the corresponding ones in (d). Then

$$\mathscr{G} = \{((pt_0 + (n+\alpha m)t'_0,(n+\beta m)t_1),$$
$$(px_0 + (n+\alpha m)x'_0,(n+\beta m)x_1),n\theta_1+m\theta_2)\},$$

with $t_0,t'_0,x_0,x'_0\in R$, $p,n,m\in Z$, $t_1\in\mathbb{T}_T$, $x_1\in\mathbb{T}_R$, $(t_0/x_0) = (t_1/x_1)$, $t_0t'_0 + xx'_0 = 0$, $\theta_1,\theta_2\in SO(2)$, $t'_0$ and $x'_0$ are not both zero, $x_0\neq 0$, $\lambda = (\theta_2/\theta_1)$ irrational, $\alpha$ rational, $n\theta_1 + m\theta_2\neq 2k\pi$ for any nonzero couple of integers $n$ and $m$ if $\alpha$ and $\beta$ are both zero, or else for any $n\neq 0$ and $m$ integers if $\alpha$ and $\beta$ are both zero;

$$\mathscr{G} = \{((pt_0 + nt'_0,nt_1),(px_0 + nx'_0,nx_1),n\theta_0)\},$$

with $t_0,t'_0,x_0,x'_0\in R$, $p,n,\in R$, $t_1\in\mathbb{T}_T$, $x_1\in\mathbb{T}_R$, $\theta_0\in SO(2)$, $(\theta_0/2\pi)$ irrational if $t'_0$ and $x'_0$ are not both zero, $\theta_0 = x_1 = t_1 = 2\pi/k$ if $t'_0$ and $x'_0$ are both zero, $\theta_0 = 2\pi/k$ if $t'_0$, $x'_0$, $t_1$ and $x_1$ are all zero;

$$\mathscr{G} = \{((pt_0,rt_1),(px_0,rx_1),r\theta_0)\},$$

with $t_0,x_0\in R$, $p\in Z$, $r\in R$, $t_1\in\mathbb{T}_T$, $x_1\in\mathbb{T}_R$, $x_0\neq 0$, $\theta_0\in SO(2)$. The last group gives rise to an $\alpha\rho_{x_t}\alpha_{\theta_t}$-KMS state.

(i) Here $\mathscr{H}$ is similar to the corresponding group in (a). So

$$\mathscr{G} = \{(rt_0 + (n+\alpha m)t'_0,rx_0 + (n+\alpha m)x'_0,n\theta_1 + m\theta_2)\},$$

with $t_0,x_0,t'_0,x'_0\in R$, $n,m\in Z$, $r\in R$, $\theta_1,\theta_2\in SO(2)$, $x_0\neq 0$, $t'_0$ and $x'_0$ are not both zero, $t_0t'_0 + x_0x'_0 = 0$, $\lambda = (\theta_2/\theta_1)$ irrational, $\alpha$ rational, $n\theta_1 + m\theta_2\neq 2k\pi$ for any nonzero couple of integers $n$ and $m$ if $\alpha\neq 0$ or for any $n\neq 0$ and $m$ integers if $\alpha = 0$;

$$\mathscr{G} = \{(rt_0 + nt'_0,rx_0 + nx'_0,n\theta_0)\}$$

with $t_0,x_0,t'_0,x'_0\in R$, $n\in Z$, $r\in R$, $\theta_0\in SO(2)$, $x_0\neq 0$, $t_0t'_0 + x_0x'_0 = 0$, $\theta_0/2\pi$ irrational if $t'_0$ and $x'_0$ are not both zero,

$$\mathscr{G} = (rt_0,rx_0,r'\theta_0)$$

with $t_0,x_0\in R$, $r,r'\in R$, $\theta_0\in SO(2)$, $x_0\neq 0$. These groups give rise respectively to $\alpha_t\alpha_{x_t}$-,$\alpha_t\alpha_{x_t}$-,$\alpha_t\alpha_{x_t}\alpha_{\theta_t}$-KMS states.

(j) The groups $\mathscr{H}$ are the sames as in (5). Hence,

$$\mathscr{G} = \{(((n+\alpha m)t_0,p(1+\alpha)t_1),((n+\lambda m)x_0,p(1+\lambda)x_1), p\theta_0)\},$$

with $t_0,x_0\in R$, $t_1\in\mathbb{T}_T$, $x_1\in\mathbb{T}_R$, $(t_0/x_0) = (t_1/x_1)$, $\alpha,\lambda\in R$, $\alpha\neq\lambda$, $\lambda$ irrational, $x_0\neq 0$, $n,m\in Z$, $\theta_0\in SO(2)$, $\theta_0 = (2\pi/k)$, $k(1+\alpha)t_1 = k(1+\lambda)x_1 = 0$;

$$\mathscr{G} = \{(((n+\alpha m)t_0,r(1+\alpha)t_1),((n+\lambda m)x_0,r(1+\lambda)x_1), r\theta_0)\},$$

with $t_0,x_0\in R$, $t_1\in\mathbb{T}_T$, $x_1\in\mathbb{T}_R$, $(t_0/x_0) = (t_1/x_1)$, $\alpha,\lambda\in R$, $\lambda$ irrational, $\alpha\neq\lambda$, $n,m\in Z$, $x_0\neq 0$, $\theta_0\in SO(2)$, $r\in R$, $(2\pi/\theta_0)(1+\alpha)t_1 = (2\pi/\theta_0)(1+\lambda)x_1 = 0$. The last group gives rise to an $\alpha\rho_{x_t}\alpha_{\theta_t}$-KMS state. To sum up the situation, let us list the groups giving rise to $\alpha\rho_{x_t}\alpha_{\theta_t}$-KMS states with $x_t$ and $\theta_t$ nontrivial.

From (b) we get

$$\mathscr{G} = \{(t,(nx_0,rx_1),r\theta_0)\} = \{(t,rx_0,r\theta_0)\}$$

i.e.,

$$\mathscr{G} = T\times \text{helix of axis } R. \qquad (76)$$

From (c) we get

$$\mathscr{G} = \{t,x,r\theta_0)\} = T\times R\times SO(2). \qquad (77)$$

From (e) we get

$$\mathscr{G} = \{((nt_0,mt_1),mx_0)\}\times\{(rt_2,rx_1,r\theta_0)\}. \qquad (78)$$

This group can be described as a $Z$ family of helixes with axis displayed along a $Z_0\times R_0$ array of equation in the $(T,R)$ plane,

$$t = \frac{(t_0,t_1)}{x_0} + nt_0 \quad (x_0\neq 0).$$

From (f) we get

$$\mathscr{G} = \{((nt_0,rt_1),rx_0)\}\times\{(r't_2,r'\theta_0)\}. \qquad (79)$$

This group can be described as the product of an helix along

$T$ by the real line $R_0$ of equation in the $(T\times R)$ plane, $t=((0,t_1)/x_0)x$.

From (h) we get

$$\mathscr{G}=\{(pt_0,px_0)\}\times\{(rt_1,rx_1,r\theta_0)\},$$

i.e.,

$$\mathscr{G}=\text{helix along } R_0, \quad (80)$$

where $R_0$ is the line of equation, in the $(T\times R)$ plane,

$$t=\frac{t_0}{x_0}x.$$

From (i) we get

$$\mathscr{G}=\{(rt_0,rx_0,r'\theta_0)\},$$

i.e.,

$$\mathscr{G}=R_0\times SO(2), \quad (81)$$

where $R_0$ is the line of equation, in the $(T,R)$ plane,

$$t=\frac{t_0}{x_0}x.$$

From (j) we get

$$\mathscr{G}=\{((n+\alpha m)t_0,(n+\lambda m)x_0)\}\times\{(r(1+\alpha)t_1,r(1+\lambda)x_1,r\theta_0)\}. \quad (82)$$

This group can be described as a $Z$ family of helixes with axis displayed along a $Z_0\times R_0$ array of equation, in the $(T,R)$ plane,

$$t=\frac{(1+\alpha)t_0}{(1+\lambda)x_0}x+nx_0.$$

## C. Algebraic introduction of the angular momentum

We now would like to show how, starting from $\omega|_{\mathfrak{A}_\omega}$, extremal-KMS on $\mathfrak{A}_\omega$ with respect to $\alpha_t\alpha_{\bar{x}_t}$ (and hence factorial), in the case where $\mu=0$ and $\mathscr{H}_\omega$ is nontrivial, it is possible to define the angular momentum by using objects related to $\mathfrak{A}_\omega$ only. For that purpose, let us assume the existence on $\mathfrak{A}_\omega$ of some inner automorphism $\rho$ such that

$$\rho^{-1}\circ\alpha_t\alpha_{\bar{x}_t}\circ\rho=\mathrm{Ad}v_t\circ\alpha_t\alpha_{\bar{x}_t}\ (\mathrm{Ad}v_t a=v_t a v_t^*), \quad (83)$$

where $v_t$ is a continuous one-parameter family of unitaries in $\mathfrak{A}_\omega$, fulfilling the cocycle relation,

$$v_{t+s}=v_t\circ\alpha_t\alpha_{\bar{x}_t}(v_s), \quad (84)$$

and such that $\omega\circ\rho$ and $\omega$ are quasiequivalent on $\mathfrak{A}_\omega$ (i.e., the von Neumann algebras $\pi_{\omega\circ\rho}(\mathfrak{A}_\omega)''$ and $\pi_\omega(\mathfrak{A})''$ are equal). The existence of such $\rho$ is given by the following theorem.

*Theorem VII.*1: Let $u\in\mathfrak{A}$ a unitary such that $\alpha_\theta u=e^{in\theta}u$, $\theta\in SO(2)$. Then $\rho=\mathrm{Ad}u$ is an automorphism verifying (83) and (84) with

$v_t=e^{inct}u^*\alpha_t\alpha_{\bar{x}_t}u$, where $c$ is some arbitrary constant.

*Proof*: It is just a matter of computation to show that $\rho^{-1}\circ\alpha_t\alpha_{\bar{x}_t}\circ\rho=\mathrm{Ad}(u^*\alpha_t\alpha_{\bar{x}_t}u)\circ\alpha_t\alpha_{\bar{x}_t}$ and that $u^*\alpha_t\alpha_{\bar{x}_t}u$ obey (84). Hence $v_t=e^{i\lambda t}u^*\alpha_t\alpha_{\bar{x}_t}u$ by the factoriality of $\omega$ and $\lambda$ can be written $\lambda=nc$. On the other hand, as $U_\omega(\alpha_\theta)\pi_\omega(u)\Omega_\omega=\pi_\omega(\alpha_\theta u)\Omega_\omega=e^{in\theta}\pi_\omega(u)\Omega_\omega$ and $u$ is unitary, the spectrum of $U_\omega$ is not one-sided and $G_\omega=I$. Hence $\omega$ is KMS on $\mathfrak{A}$ and $\Omega_\omega$ is separating for $\pi_\omega(\mathfrak{A})''$. Of course, $\pi_\omega|\mathfrak{A}_\omega$ on $\mathscr{H}_\omega=\{\pi_\omega(a)\Omega_\omega,a\in\mathfrak{A}_\omega\}$ is identical to $\pi_{\omega|\mathfrak{A}_\omega}$. In the same way, $\pi_{\omega\circ\rho|\mathfrak{A}_\omega}$ is identical to $\pi_\omega|\mathfrak{A}_\omega$ on $\mathscr{H}^u_\omega=\{\pi_\omega(au^*)\Omega_\omega,a\in\Omega_\omega\}$ and is a subrepresentation of $\pi_\omega|\mathfrak{A}_\omega$ on the Hilbert space $\mathscr{H}$ of $\pi_\omega$. Let $E$ the projection from $\mathscr{H}_\omega$ into $\mathscr{H}$ : $E\in\pi_\omega(\mathfrak{A}_\omega)'$ and the central support $F$ of $E$ in $\pi_\omega(\mathfrak{A}_\omega)'$ is the projection onto $[\pi_\omega(\mathfrak{A}_\omega)'\mathscr{H}_\omega]$. But

$$[\pi_\omega(\mathfrak{A}_\omega)'\mathscr{H}_\omega]\supset[\pi_\omega(\mathfrak{A}_\omega)'\Omega_\omega]\supset[\pi_\omega(\mathfrak{A})'\Omega]=\mathscr{H},$$

as $\Omega_\omega$ is separating for $\pi_\omega(\mathfrak{A})''$, and then cyclic for $\pi_\omega(\mathfrak{A})'$. So $F=I$, which means that $\pi_\omega|\mathfrak{A}_\omega$, i.e., $\pi_{\omega|\mathfrak{A}_\omega}$, is quasiequivalent to $\pi_\omega|\mathfrak{A}_\omega$, and also to $\pi_{\omega\circ\rho|\mathfrak{A}_\omega}$, which is a subrepresentation. Q.E.D.

Let $w_t\in\pi_\omega(\mathfrak{A}_\omega)''$ the Radon–Nykodym derivative[13]

$$w_t=(D(\omega\circ\rho):D\omega)_t. \quad (85)$$

The modular operators of $\omega$ and $\omega\circ\rho$ on $\pi_\omega(\mathfrak{A}_\omega)''$ being respectively the extension of $\alpha_{-\beta t}\alpha_{\bar{x}_{\beta t}}$ and $\rho^{-1}\circ\alpha_{-\beta t}\alpha_{\bar{x}_{\beta t}}\circ\rho$ (Ref. 13), we have

$$\pi_\omega(\rho^{-1}\circ\alpha_{-\beta t}\alpha_{\bar{x}_{\beta t}}\circ\rho(a))=w_t\pi_\omega(\alpha_{-\beta t}\alpha_{\bar{x}_{\beta t}}(a))w_t^*,\quad a\in\mathfrak{A}_\omega, \quad (86)$$

but on the other hand

$$\pi_\omega(\rho^{-1}\circ\alpha_{-\beta t}\alpha_{\bar{x}_{\beta t}}\circ\rho(a))=\pi_\omega(v_{-\beta t})\pi_\omega(\alpha_{-\beta t}\alpha_{\bar{x}_{\beta t}}(a))\pi_\omega(v_{-\beta t})^*. \quad (87)$$

Hence, by the factoriality of $\omega$,

$$w_t=(D(\omega\circ\rho):D_\omega)_t=e^{-i\lambda\beta t}\pi_\omega(-v_{\beta t}), \quad (88)$$

where $\lambda$ is a kind of angular momentum. More explicitly, in the case of the hypothesis of Theorem VII.1, let us consider $\omega$ and $\omega\circ\mathrm{Ad}u^*$ on $m=\pi_\omega(\mathfrak{A})''$:

$$(D(\omega\circ\mathrm{Ad}u^*):D_\omega)_t=\pi_\omega(u^*)\sigma_t^\omega\pi_\omega(u), \quad (89)$$

where $\sigma_t^\omega$ is the modular automorphism of $\omega$ on $m$, or else

$$(D(\omega\circ\mathrm{Ad}u^*):D_\omega)_t=\pi_\omega(u^*)\pi_\omega(\alpha_{-\beta t}\alpha_{\bar{x}_{\beta t}}\alpha_{-\beta vt}u),$$

$$=e^{-in\beta vt}\pi_\omega(u^*\alpha_{-\beta t}\alpha_{\bar{x}_{\beta t}}u),$$

$$=e^{-in\beta vt}e^{-in\beta ct}\pi_\omega(v_{-\beta t}), \quad (90)$$

and, by restriction to $\mathfrak{A}_\omega$,

$$w_t=(D(\omega\circ\rho):D_\omega)_t=e^{-in\beta(v+c)t}\pi_\omega(v_{-\beta t}), \quad (91)$$

where $v$ is the angular momentum and $c$ an arbitrary constant which can be chosen to equal zero independently of $\omega$.

## VIII. CONCLUSION

Besides the characterization of all $\alpha_t\alpha_{x_t}\alpha_{\theta_t}$-KMS states for the $(T\times R^3)\square SO(3)$ symmetry group and the algebraic definition of the angular momentum, one of the main conclusions of this work is the fact that the algebraic background of Ref. 1 can be generalized to a group with a semidirect product structure.

Nevertheless, this result might be improved in several ways; first of all by starting from a KMS condition with respect to time only on the $C^*$ subalgebra of fixed points under $R^3 \square K$, deduced from a stability requirement under perturbations by observables in this subalgebra.

On the other hand, we might replace SO(3) by the Lorentz group, which seems to be necessary because a space rotation of an infinite medium leads to infinite velocities.

In these both directions, we would have to replace $K$ by a noncompact group, i.e., from a mathematical point of view, to substitute the Tatsumma–Takesaki duality to the Tannaka one.

## ACKNOWLEDGMENTS

The authors are indebted to Professor D. Kastler and Professor R. Hermann for many valuable comments.

[1]H. Araki, R. Hagg, D. Kastler, and M. Takesaki, Commun. Math. Phys. **53**, 97 (1977).
[2]S. Doplicher, D. Kastler, and E. Störmer, Funct. Anal. **3**, 419 (1969).
[3]D. Kastler, G. Loupias, M. Mebkhout, and L. Michel, Commun. Math. Phys. **27**, 195 (1972).
[4]G. Loupias and M. Mebkhout, J. Math. Phys. **14**, 777 (1973).
[5]R. Hagg, D. Kastler, and E. Trych-Pohlmeyer, Commun. Math. Phys. **38**, 173 (1974).
[6]H. Araki, Publ. RIMS Kyoto University **9**, 165 (1973).
[7]D. Kastler, "Foundations of equilibrium statistical mechanics," UCLA Report 1977.
[8]O. Bratteli and D. Kastler, Commun. Math. Phys. **46**, 37 (1976).
[9]R. Hermann and D. Kastler, Commun. Math. Phys. **56**, 87 (1977).
[10]M. Takesaki, Acta Math. **119**, 273 (1967).
[11]M. Takesaki, Acta Math. **131**, 249 (1973).
[12]N. Bourbaki, *Livre* III: *Topologie Générale* (Hermann, Paris Chap. VII.
[13]A. Connes, Ann. Sci. Ecole Normale Sup. **6**, 18 (1973).

# A note on Miura's transformation

Mark J. Ablowitz[a] and Martin Kruskal

*Program in Applied Mathematics, Princeton University, Princeton, New Jersey 08540*

Harvey Segur

*Aeronautical Research Associates of Princeton, Inc., 50 Washington Road, P.O. Box 2229, Princeton, New Jersey 08540*

(Received 13 June 1977; revised manuscript received 15 November 1978)

It has been observed by Miura that the solutions of the modified Korteweg–deVries equation can be mapped into those of the Korteweg–deVries equation. In this note we show that all of the solutions of the former, decaying sufficiently rapidly as $|x|\to\infty$, map into a sparse solution set of the KdV equation. We use certain results regarding the second Painlevé transcendent to exhibit this fact.

## 1. INTRODUCTION

A key step in the development of the inverse scattering transform was the discovery[1] that every solution of the modified Korteweg–deVries equation (MKdV).

$$v_t - 6v^2 v_x + v_{xxx} = 0, \tag{1.1}$$

maps into a solution of the Korteweg–deVries equation (KdV),

$$u_t + 6uu_x + u_{xxx} = 0, \tag{1.2}$$

under the transformation

$$u = -v^2 - v_x. \tag{1.3}$$

Since (1.3) is a transformation between solutions of partial differential equations, it is usually called a Bäcklund transformation. The purpose of this note is to examine some of the consequences of this transformation.

Perhaps our most important conclusion is that the range of the transformation is quite restricted. Specifically, the set of all solutions of MKdV that evolve from smooth, rapidly decaying initial data maps into a sparse set of solutions of KdV. Hence, almost every solution of KdV that evolves from smooth, rapidly decaying initial data *cannot* be obtained from a rapidly decaying solution of MKdV by this transformation.

The restricted range of (1.3) can be anticipated from the following simple argument. Let $u(x,t_0)$ be a given smooth solution of KdV that vanishes rapidly as $|x|\to\infty$, evaluated at some $t = t_0$. Let $v(x)$ denote the corresponding solution of the ordinary differential equation (1.3), subject to the initial condition that

$$v\to 0, \quad \text{as } x\to +\infty. \tag{1.4}$$

These two conditions uniquely determine $v$, including its behavior as $x\to -\infty$. Except for very special $u$, $v$ will not vanish rapidly (i.e., faster than $1/x$) as $x\to -\infty$. Consequently, this solution of KdV cannot be the transformation of a solution of MKdV that vanishes rapidly as $|x|\to\infty$. (There are restrictions on $u$ to ensure that $v$ has no poles on the real axis; these are only inequality restrictions and do not lower the dimensionality of the function space.)

Because this argument applies at fixed $t$, the solution of (1.3) need not solve MKdV. In order to relate solutions of MKdV to solutions of KdV, we characterize each by the appropriate "scattering data." Both problems (on $-\infty < x < \infty$) can be solved by inverse scattering transforms, although the required linear eigenvalue problems are different (see, for example, Ref. 2). We denote the scattering data for KdV by $\rho(k)$; as will be shown in Sec. 2 and an Appendix, given almost any smooth initial function for KdV for which

$$\int_{-\infty}^{\infty} (1 + x^2)\, |u| dx < \infty, \tag{1.5}$$

$$\rho(0) = -1. \tag{1.6}$$

One can show that (1.6) requires that the corresponding asymptotic ($t\to\infty$) solution of KdV contains a relatively thin transition layer across which the character of the asymptotic solution changes abruptly (cf. Ref. 3). Consequently, these transition layers are a typical feature of KdV solutions; they are absent only for those very special initial conditions for which (1.6) fails.

For MKdV, the linear eigenvalue problem is different, and we denote the scattering data by $r(k)$. In this case, we assume $v$ decays rapidly as $|x|\to\infty$, so that certainly

$$\int_{-\infty}^{\infty} |v| dx < \infty. \tag{1.7}$$

By virtue of this $r(k)$ may be extended into the upper half-plane and

$$-1 < r(0) < 1. \tag{1.8}$$

It can be shown that the corresponding asymptotic solutions of MKdV contain no transition layers. All of these solutions are related by (1.3) to solutions of KdV that contain no transition layers, but the set of such KdV solutions is extremely

[a] Permanent address: Department of Mathematics, Clarkson College of Technology, Potsdam, New York 13676.

 0022-2488/79/060999-05$01.00  

sparse. Typical KdV solutions (with transition layers) *cannot* be obtained by (1.3) from rapidly decaying solutions of MKdV.

The asymptotic results for MKdV have important consequences regarding the global behavior of the second Painlevé equation,

$$w'' = zw + 2w^3. \tag{1.9}$$

Some of the recent work on (1.9) is outlined in Sec. 3.

## 2. BEHAVIOR OF THE SOLUTIONS

(1.1) may be solved in the following way[2]:
Solve

$$K(x,y;t) + F(x+y;t) - \int_x^\infty \int_x^\infty K(x,z;t)F(z+s;t) \times F(s+y;t)\,dz\,ds = 0, \tag{2.1a}$$

where

$$F(x;t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} r(k)e^{i(kx+8k^3t)}dk, \tag{2.1b}$$

and then

$$v = -2K(x,x;t). \tag{2.1c}$$

The scattering data ($r(k) = r^*(-k)$) are obtained from the initial conditions $v(x,0)$ by solving the related scattering problem at the initial instant. The scattering problem is

$$W_{1x} = -i\zeta W_1 + vW_2, \quad W_{2x} = i\zeta W_2 + vW_1. \tag{2.2}$$

If we define the (improper) eigenfunctions $\phi$, $\psi$, $\bar{\psi}$ as solutions

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix}$$

determined by

$$\phi \sim \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\zeta x} \quad \text{as } x \to -\infty, \tag{2.3a}$$

$$\psi \sim \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\zeta x} \quad \text{as } x \to +\infty, \tag{2.3b}$$

$$\bar{\psi} \sim \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-i\zeta x} \quad \text{as } x \to +\infty, \tag{2.3c}$$

and $a(\zeta)$,$b(\zeta)$ by

$$\phi = a\bar{\psi} + b\psi \to \begin{pmatrix} a & e^{-i\zeta x} \\ b & e^{i\zeta x} \end{pmatrix} \quad \text{as } x \to +\infty, \tag{2.3d}$$

then $r(\zeta) \equiv b(\zeta)/a(\zeta)$. The value $r(0)$ turns out to be crucial in the analysis. It is easily seen from (2.2) that for $\zeta = 0$,

$$\phi = \begin{pmatrix} \cosh\left(\int_{-\infty}^x v(x')dx'\right) \\ \sinh\left(\int_{-\infty}^x v(x')dx'\right) \end{pmatrix}, \tag{2.4}$$

hence

$$r_0 \equiv r(0) = \tanh\left(\int_{-\infty}^{\infty} v\,dx\right). \tag{2.5}$$

Since $\int_{-\infty}^{\infty} v\,dx < \infty$, this means $r_0$ is real and

$$-1 < r_0 < 1. \tag{2.6}$$

The condition (2.6) has important implications regarding the solutions of MKdV. The behavior of $F(x,t)$ as $t \to \infty$ given by (2.1b) can be expressed in terms of elementary functions. Following the methods of Ref. 3 we find the solution of MKdV has the following asymptotic form. [The methods in Ref. 3 require that the initial data decay faster than any exponential as $|x| \to \infty$. Consequently, the formulas (2.7), (2.8), (2.17)–(2.19) have been established only for these rapidly decaying initial functions. However, this restriction seems to be a consequence of the particular method, whereas the weaker restrictions (1.5) and (1.7) are intrinsic to the respective problems.] For $\eta \equiv x/(3t)^{1/3} \gg 1$ (bisimilarity region)

$$v \sim \left[ r\left(\frac{i}{2}\left(\frac{x}{3t}\right)^{1/2}\right) \Big/ (3t)^{1/3} \right] \mathrm{Ai}(\eta), \tag{2.7a}$$

where $\mathrm{Ai}(\eta)$ is the Airy function. For $\eta = x/(3t)^{1/3} = O(1)$ (ordinary similarity region),

$$v \sim \frac{1}{(3t)^{1/3}} g(\eta), \tag{2.7b}$$

where $g(\eta)$ is that unique solution of the Painlevé equation (1.9) with asymptotic boundary conditions,

$$g(\eta) \sim r_0 \mathrm{Ai}(\eta), \quad \text{as } \eta \to +\infty. \tag{2.7c}$$

For $\eta \ll -1$ (again a bisimilarity region),

$$v \sim \frac{(-\eta)^{1/4} d}{(3t)^{1/3}} \sin\left(\tfrac{2}{3}(-\eta)^{3/2} - \tfrac{3}{4}d^2 \ln(-\eta) + \theta_0\right) \tag{2.7d}$$

where

$$d^2 = -\frac{1}{\pi}\ln\left(1 - \left|r\left(\sqrt{-x/12t}\right)\right|^2\right). \tag{2.7e}$$

(2.7a), (2.7b), and (2.7d) are the dominant terms in the asymptotic expansions as $t \to \infty$. These expansions match in their respective overlap regions. Two points are to be noted here: (i) there are three regions to the asymptotic solution of MKdV; and (ii) (2.6) has important implications regarding the behavior of the Painlevé equation (1.9). This condition (2.6) assures that the solution of (1.9) is bounded for all $\eta$, and hence, matches (2.7d) and (2.7e). However, if $r_0 = \pm 1$ the Painlevé transcendent becomes unbounded. Although this never happens for MKdV, it virtually always happens for KdV. The behavior of the solution of (1.9) will be more fully discussed in Sec. 3, and has been noted in Refs. 4 and 3.

By virtue of Miura's transformation (1.3), we can map all the solutions of MKdV into solutions of KdV. For completeness we list the results of this computation. For $\eta \gg 1$,

$$u \sim -\left[ r\left(\frac{i}{2}\left(\frac{x}{3t}\right)^{1/2}\right) (3t)^{2/3} \right] \mathrm{Ai}'(\eta). \tag{2.8a}$$

For $\eta = O(1)$,

$$u \sim \frac{1}{(3t)^{2/3}} f(\eta), \tag{2.8b}$$

where $f(\eta)$ satisfies

$$f''' + 6ff' - (2f + \eta f') = 0, \tag{2.8c}$$

subject to the asymptotic boundary condition

$$f \sim -r_0 \mathrm{Ai}'(\eta), \quad \text{as } \eta \to +\infty. \tag{2.8d}$$

For $\eta \ll -1$,

$$u\sim\frac{d(-\eta)^{1/4}}{(3t)^{2/3}}\cos\Big(\tfrac{2}{3}(-\eta)^{3/2}-\frac{3d^2}{4}\ln(-\eta)+\theta_0\Big), \tag{2.8e}$$

where $d$ is given by (2.7e).

It is surprising that (2.8a)–(2.8e) is *virtually never an entirely correct description of the asymptotic solution of the* KdV *equation* (without solitons). To explain this we shall show that the condition (2.6) is not generic for the KdV equation. The generic case turns out to be

$$r_0=-1 \tag{2.9}$$

[whereby the solution of (2.8c) becomes unbounded and (2.7e) is singular at $x/(3t)=0$]. In order to demonstrate this, we first mention that the inverse scattering transform for (1.2), without solitons,[5] is given by

$$K(x,y;t)+B(x+y;t)+\int_x^{\infty}K(x,z;t)B(z+y;t)dz=0, \tag{2.10a}$$

$$B(x;t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\rho(k)e^{i(kx+8k^3t)}dk, \tag{2.10b}$$

$$u(x;t)=2\frac{d}{dx}K(x,x;t), \tag{2.10c}$$

where $\rho(k)$, the reflection coefficient, is obtained from the scattering problem at the initial instant,

$$\phi_{xx}+(k^2+u(x;0))\phi=0. \tag{2.11a}$$

We require $u(x;t)$ to decay sufficiently rapidly as $|x|\to\infty$, and $\phi$ to obey the boundary conditions,

$$\phi\sim e^{-ikx}+\rho(k)e^{ikx},\quad x\to+\infty, \tag{2.11b}$$

$$\phi\sim\tau(k)e^{-ikx},\quad x\to-\infty. \tag{2.11c}$$

[For appropriate initial conditions, (1.3) implies that as $t\to\infty$ (2.8a)–(2.8c) holds with $r(k)$ replaced by $\rho(k)$ and of course $r_0$ by $\rho_0\equiv\rho(0)$.]

Next we show that $\rho_0=-1$ is generic. First we consider functions $u(x,t)$ with compact support $-L\leqslant x\leqslant L$. Define the function $\mu(x;k)=\phi_x/\phi$ on $[-L,L]$. In terms of $\mu$, (2.11a) yields

$$\mu'+\mu^2+k^2+u=0. \tag{2.12a}$$

The boundary conditions (2.11b) and (2.11c) yield

$$\mu(-L;k)=-ik, \tag{2.12b}$$

$$\mu(L;k)=\frac{-ike^{-ikL}+ik\rho(k)e^{ikL}}{e^{-ikL}+\rho(k)e^{ikL}}. \tag{2.12c}$$

(2.12c) is easily inverted to express $\rho(k)$ in terms of $\mu(L;k)$,

$$\rho(k)=\frac{ik+\mu(L;k)}{ik-\mu(L;k)}e^{-2ikL}. \tag{2.12d}$$

Since (2.12a) is a first order ordinary differential equation, the boundary condition (2.12b) determines $\rho(k)$. In order to investigate the behavior of $\rho(k)$ for small $k$ we assume a perturbation expansion for $\mu$ of the form,

$$\mu=\mu^{(0)}+k\mu^{(1)}+\cdots, \tag{2.13a}$$

whereby

$$\mu^{(0)\prime}+\mu^{(0)2}+u=0, \tag{2.13b}$$

$$\mu^{(1)\prime}+2\mu^{(0)}\mu^{(1)}=0, \tag{2.13c}$$

.

.

.

Similarly, for the boundary conditions,

$$\mu^{(j)}(-L)=0(j\neq1),\quad \mu^{(1)}(-L)=-i. \tag{2.13d}$$

Thus, for $\rho(k)$ we have from (2.12d),

$$e^{2ikL}\rho(k)=\frac{ik+\mu^{(0)}(L)+k\mu^{(1)}(L)+\cdots}{ik-\mu^{(0)}(L)-k\mu^{(1)}(L)+\cdots}. \tag{2.14}$$

When no solitons are present in the initial data $\mu^{(0)}(x)$ has no poles. There are two cases: (a) if $\mu^{(0)}(L)\neq0$, then $\rho(k)=-1+O(k)$, in which case

$$\rho_0=\rho(0)=-1; \tag{2.15}$$

(b) if $\mu^{(0)}(L)=0$, then

$$\rho_0=\frac{1-i\mu^{(1)}(L)}{1+i\mu^{(1)}(L)}=\frac{1-\exp\Big(-2\int_{-L}^{L}\mu^{(0)}dx\Big)}{1+\exp\Big(-2\int_{-L}^{L}\mu^{(0)}dx\Big)}, \tag{2.16}$$

hence, $\rho_0$ is real and $-1<\rho_0<1$. $\mu^{(0)}(L)$ is a complicated functional of $u(x)$, and vanishes only for very special $u(x)$. Consequently, the usual situation is (2.15); we obtain (2.16) only in those special cases of (2.13b) for which $\mu^{(0)}(-L)=0$ and $\mu^{(0)}(L)=0$.

If the potential $u$ has noncompact support but still decays rapidly at infinity, it can be approximated arbitrarily closely by one with compact support. Hence, as shown in the Appendix, $\rho_0=-1$ is the generic case, provided only that the potential decays rapidly enough at infinity to satisfy (1.5).

It is interesting to note that if we identify $\mu^{(0)}(x)$ with a solution of MKdV, and require that $\mu^{(0)}\to0$ as $|x|\to\infty$ (the exceptional case), then (2.13b) is Miura's transformation (i.e., $u(x)=-\mu^{(0)\prime}-\mu^{(0)2}$).

When (2.15) holds the formulas (2.8) are incomplete [indeed, (2.8) is singular if $r_0=\rho_0=-1$]. In this case

$$f(\eta)\sim\frac{\eta}{2}-\frac{1}{2}(-2\eta)^{-1/2}+\cdots,\quad \text{as } \eta\to-\infty \tag{2.17}$$

(see also Sec. 3) and this indicates a breakdown of the long-time expansion for $u$,

$$u=\frac{1}{(3t)^{2/3}}f(\eta)+\frac{1}{(3t)}f_1(\eta)+\frac{1}{(3t)^{4/3}}f_2(\eta)+\cdots, \tag{2.18}$$

valid for $\eta=O(1)$. In Ref. 3 we discuss this situation in detail. A transition layer is needed to be able to pass (and therefore to have an asymptotic match) from (2.8c), (2.8d) to (2.8e), (2.7e) (since both are singular). For completeness we give the formulas governing this transition layer. The region where this occurs is $-\eta=(\ln t)^{(2+p)/3}$, $0<p<2$. There are two asymptotic formulas needed to describe this regime (i.e., two asymptotic regions). We find

$$u \sim \frac{-2\eta}{(3t)^{2/3}} g(\xi,t), \tag{2.19a}$$

where $\xi$ is related to $\eta$ by

$$(-2\eta)^{3/2} \sim 2\ln(3t)\left(1 + \frac{5}{4}\frac{\ln(2\ln 3t)}{\ln 3t} + \frac{\xi}{\ln 3t}\right). \tag{2.19b}$$

For $\xi = O(1)$,

$$g(\xi,t) \sim -\tfrac{1}{4} + \tfrac{1}{2}\text{sech}^2(\tfrac{1}{3}(\xi - \xi_0)), \tag{2.19c}$$

where $\xi_0 = -\frac{3}{2}\ln(\kappa M/2)$, $M = \frac{1}{8}(\rho_0'' + (\rho_0')^2)$, and $\kappa \doteq .80$. (2.19c) is not uniformly valid for large $\xi$. For $1 \ll \xi \ll (\ln 3t)^2$, $g(\xi,t)$ is given by

$$g \sim a(Z) + b(Z)\text{cn}^2(2K(\nu)\theta + \theta_0;\nu), \tag{2.19d}$$

where

$$a = \frac{b(1-2\nu)}{3\nu} - \frac{1}{12},$$

$$\frac{b}{\nu} = \frac{1}{(4-2\nu)},$$

$$\theta = \frac{1}{\epsilon}\int_0^Z \omega(z)e^z dz,$$

$$\frac{1}{\epsilon} = \ln 3t + \tfrac{5}{4}\ln(2\ln 3t),$$

$$Z = \ln(1 + \epsilon\xi),$$

$$\omega^2 = b/(18K^2\nu),$$

$$\frac{d\nu}{dZ} = -\frac{8}{3}\frac{(1-\nu/2)}{\nu}((1-\nu/2)E/K + \nu - 1),$$

$$\theta_0 = \text{const},$$

and where $\nu \to 1$ as $Z \to 0$, $\text{cn}(u;\nu)$ is the Jacobian elliptic cosine with modulus $\nu$, and $K,E$ are the complete elliptic integrals of the first and second kind, respectively.

## 3. THE PAINLEVÉ TRANSCENDENT OF THE SECOND KIND

A self-similar solution $w$ of MKdV satisfies (1.9), the second Painlevé equation ($P_{\text{II}}$). For each constant $r_0$ we single out a unique solution $w(z,r_0)$ by the asymptotic boundary condition

$$w(z,r_0) \sim r_0\text{Ai}(z), \quad \text{as } z \to \infty. \tag{3.1}$$

With the asymptotic results of Sec. 2, we can obtain information about the global structure of this function. This information is readily obtained from formulas (2.7c)–(2.7e).

We note that (3.1) corresponds to the rightmost portion of the simple similarity region (2.7b) of MKdV, and the leftmost portion of the exponential decay region of MKdV (2.7a). By asymptotic matching, we see that the asymptotic behavior of $w(z)$ for $z \to -\infty$, *viz*,

$$w(z,r) \sim \frac{d_0}{(-z)^{1/4}}\sin\left(\tfrac{2}{3}(-z)^{3/2} - \tfrac{3}{4}d_0^2\ln(-z) + \theta_0\right), \tag{3.2a}$$

must have its amplitude, $d_0$, obeying

$$d_0^2 = -\frac{1}{\pi}\ln(1 - r_0^2). \tag{3.2b}$$

This is the amplitude connection formula of (1.9) with asymptotic boundary condition (3.1), for $|r_0| < 1$. The connection formula for the phase $\theta_0 = \theta_0(r_0)$ is not yet known. We have verified (3.2b) numerically, as has Greene.[6]

(3.2b) suggests unusual behavior when $|r_0| \geqslant 1$. Numerical computations (Refs. 4, 6, and ourselves) support this. (As discussed earlier, $|r(k)| < 1$ in MKdV. Hence for this evolution equation, $|r_0| \geqslant 1$ does not occur.) When $r_0 = \pm 1$, $w(z)$ is on its critical branch (we refer to the case $|r_0| < 1$ as undercritical, and $|r_0| > 1$ as overcritical), and it is unbounded as $z \to -\infty$. In the critical case, the asymptotic behavior of $w(z)$ is given by

$$w(z) \sim \text{Sgn}(r_0)\left((-z/2)^{1/2} - \frac{(-z)^{-5/2}}{2^{7/2}} + O((-z)^{-11/2})\right). \tag{3.3}$$

When $|r_0| > 1$ (overcritical), $w(z)$ has a pole at a finite location, depending on $|r_0|$. The dominant terms are given by

$$w(z) \sim \text{Sgn}(r_0)\left(\frac{1}{z - z_0} - \frac{z_0}{6}(z - z_0) + O(z - z_0)^2\right). \tag{3.4}$$

In (3.4) $z_0 = z_0(r_0)$.

The fact that $P_{\text{II}}$ and MKdV are so closely related suggests that IST methods may be used to obtain solutions of $P_{\text{II}}$. This has been done.[7] For completeness, we give the linearization of (1.9) subject to (3.1). One must solve

$$K(x,y) - r_0\text{Ai}\left(\frac{x+y}{2}\right) - \left(\frac{r_0}{2}\right)^2\int_x^\infty\int_x^\infty K(x,z) \times\text{Ai}\left(\frac{z+s}{2}\right)\text{Ai}\left(\frac{s+y}{2}\right)dy\,ds = 0, \tag{3.5a}$$

for $K(x,y)$, then

$$w(x) = K(x,x) \tag{3.5b}$$

satisfies (1.9) with the asymptotic boundary condition (3.1). It should also be noted that with this observation it is easy to investigate the behavior of the equation

$$w'' = zw - 2w^3 \tag{3.6}$$

[formally obtained by taking $w \to iw$ in (1.9)]. In this case, (3.5a) with the signs of the second and third terms changed to positive gives the exact linearization, and the formula (3.2b) becomes

$$d_0^2 = \frac{1}{\pi}\ln(1 + r_0^2). \tag{3.7}$$

Moreover, by (3.5a) it can be proven that the real solution to (1.9) exists for all $z$ when $|r_0| < 1$, and a real solution to (3.6) exists for all real $z$, and for all real $r_0$ (Ref. 8).

Finally, it should be noted that to understand the properties of the similarity solution for the KdV equation (2.8b), (2.8c) it is easiest to work with the Painlevé function $w(z)$, and then use the Miura transformation as it pertains to the similarity solutions; i.e., we use $u = f(\eta)/(3t)^{2/3}$ and $v = w(\eta)/(3t)^{1/3}$ in (1.3) to obtain

$$f = -(w^2 + w'). \tag{3.8}$$

To our knowledge Whitham[4] first observed this fact regarding the similarity solutions.

## ACKNOWLEDGMENT

This work was partially supported by NSF Grant MCS 75-07508A02 and by the U.S. Army Research Office.

## APPENDIX

We have shown that if the potential in (2.11) has compact support, $\rho(0) = -1$ is generic. We show here that the same result holds even if the potentials only satisfy (1.5), by approximating them by potentials on compact support. Most of the proof requires only the weaker condition of Faddeev,[9]

$$\int_{-\infty}^{\infty} (1 + |x|)\,|u|dx < \infty, \tag{A1}$$

but the continuity of $\rho$ at $k = 0$ apparently requires (1.5) (cf. Ref. 10). First, we shall show that $\rho(0) = -1$ is generic for potentials that satisfy (1.5) and vanish for all $x < L$, for some $L < 0$ ("left-compact").

Let $u_L(x)$ vanish for all $x < L < 0$ and satisfy (1.5), and therefore also (A1). For any $\frac{1}{2} > \epsilon > 0$, there is an $X(\epsilon)$ such that

$$\int_X^{\infty} |xu_L|\,dx < \epsilon. \tag{A2}$$

Define the family of approximating potentials, each of which has compact support, by

$$\begin{aligned} u_\epsilon(x) &= u_L(x), \quad x < X(\epsilon), \\ &= 0, \qquad x > X(\epsilon). \end{aligned} \tag{A3}$$

Let $\phi(x,k)$ satisfy (2.11); equivalently,

$$\phi(x,k) = \tau(k)e^{-ikx} + \int_{-\infty}^{x} \frac{\sin k(y-x)}{k} u_L(y)\phi(y,k)dy. \tag{A4}$$

Define

$\phi_\epsilon(x,k)$ by

$$\phi_\epsilon(x,k) = \tau(k)e^{-ikx} + \int_{-\infty}^{x} \frac{\sin k(y-x)}{k} u_\epsilon(y)\phi_\epsilon(y,k)dy. \tag{A5}$$

It follows that for $x < X$,

$$\phi_\epsilon(x,k) = \phi(x,k),$$

and for $x > X$

$$\phi(x,k) - \phi_\epsilon(x,k) = \int_X^{x} \frac{\sin k(y-x)}{k} u_L(y)\phi(y,k)dy. \tag{A6}$$

According to a result of Levinson[11] (Lemma 2.0), it follows from (A1) that there exist $L_0$, $M$, dependent only on $u_L(x)$, such that for $x > L_0$,

$$|\phi(x,k)| \leqslant Mx. \tag{A7}$$

Choose $\epsilon$ small enough that $X(\epsilon) > L_0$. Then (A6) implies that for all $x > X$

$$|\phi(x,k) - \phi_\epsilon(x,k)| \leqslant \frac{M}{k}\int_X^{x} |u_L y|\,dy < \frac{\epsilon M}{k}. \tag{A8}$$

As $x \to \infty$,

$$\begin{aligned} &\phi(x,k) \to e^{-ikx} + \rho_L(k)e^{ikx}, \\ &\phi_\epsilon(x,k) \to C_1(k)e^{-ikx} + C_2(k)e^{ikx}, \end{aligned} \tag{A9}$$

where

$$\rho_\epsilon(k) = C_2(k)/C_1(k). \tag{A10}$$

Combining (A8) and (A9), one can show that for any $k > 0$,

$$|C_1(k) - 1| < \frac{\epsilon M}{k}, \quad |C_2(k) - \rho_L(k)| < \frac{\epsilon M}{k}. \tag{A11}$$

Thus, for each $k > 0$, $|\rho_\epsilon(k) - \rho_L(k)|$ can be made arbitrarily small. Choose

$$\epsilon(k) = \frac{k^2}{4M}$$

so that $X(\epsilon(k)) \to \infty$ as $k \to 0$, but $X < \infty$ for any $k > 0$. Then for all $k > 0$,

$$|\rho_\epsilon(k) - \rho_L(k)| < k.$$

But by (1.5), both $\rho_L(k)$ and $\rho_\epsilon(k)$ are continuous at $k = 0$.[10] Hence, because $\rho_\epsilon(0) = -1$ is generic for potentials on compact support, it follows that $\rho_L(0) = -1$ is generic for left-compact potentials.

Moreover, because $|\rho(k)|$ is invariant under the transformation $x \to (-x)$, and $\rho(0) = +1$ is impossible,[10] it follows that $\rho_R(0) = -1$ is generic for "right-compact" potentials [that satisfy (1.5) and vanish for $x > R$, for some $R > 0$].

Finally, let $u(x)$ satisfy (1.5), with reflection coefficient $\rho(k)$. Define a family of approximating right-compact potentials $u_{R\epsilon}(x)$ as in (A3), with reflection coefficients $\rho_{R\epsilon}(k)$. Then repeating the previous argument (A4)–(A11) shows that $|\rho(k) - \rho_{R\epsilon}(k)|$ can be made arbitrarily small for any $k \neq 0$, and in the limit $k = 0$. It follows that $\rho(0) = -1$ is generic for potentials satisfying (1.5).

[1]R.M. Miura, J. Math. Phys. **9**, 1202 (1968).
[2]M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[3]M.J. Ablowitz and H. Segur, Stud. Appl. Math. **57**, 13 (1977).
[4]R. Rosales, preprint; G. Whitham, private communication.
[5]R.M. Miura, SIAM Review **18**, 412 (1976).
[6]J.M. Greene, private communication.
[7]M.J. Ablowitz and H. Segur, Phys. Rev. Lett. **38**, 1103 (1977).
[8]M.J. Ablowitz, A. Ramani, and H. Segur, "A connection between nonlinear evolution equations and ordinary differential equations of *P*-type. II," to be published.
[9]L.D. Fadeev, Trudy Mat. Inst. Steklov, **73**, 314 (1964); AMST2, **65**, 139 (1967).
[10]P. Deift and E. Trubowitz, "Inverse scattering on the line," to appear.
[11]N. Levinson, Danske Vid. Selsk. Mat. Fys. Medd. **25**, 1 (1949).

# Some exceptional electrovac type D metrics with cosmological constant

Jerzy F. Plebánski[a)]

*Centro de Investigación y Estudios Avanzados de I.P.N., A.P. 14-740, México, D.F.*

Shahen Hacyan

*Instituto de Astronomía, Universidad Nacional Autónoma de México, A. P. 70-264, México 20, D. F.*
(Received 7 July 1978)

We investigate all type D solutions of the Einstein–Maxwell equations (with cosmological constant) such that the Debever–Penrose vectors are aligned along the two eigenvectors of the electromagnetic field, in the special case when a direct generalization of the Goldberg–Sachs theorem is not possible. A solution is found which admits no Killing vectors. We also present an extension of the Golberg–Sachs theorem valid for type D metrics.

## 1. INTRODUCTION

This work is a sequel of a previous paper (Plebański[1]) where we considered electrovac solutions with cosmological constant, $\lambda$, under the following assumptions: ($A_1$) they are of type D, ($A_2$) they have both Debever–Penrose (DP) vectors aligned along the real eigenvectors of the electromagnetic field, ($A_3$) these vectors are geodesic and shearfree, and ($A_4$) free of complex expansion. In this paper we shall use the same formalism and notation as that in Ref. 1 and other works[2–6]

If only assumptions $A_1$ and $A_2$ are made, then with the double DP vector oriented along the tetrad members $e^3$,$e^4$, the only nonzero component of the conformal curvature is $C^{(3)}$. Furthermore, as long as

$$(E^2 + \check{B}^2)^2 - (\tfrac{3}{2}C^{(3)})^2 \neq 0, \tag{1.1}$$

$A_3$ is not an independent assumption; the Bianchi identities (BI) imply that $e^3$ and $e^4$ must be geodesic and shearfree.[1] $E$ and $\check{B}$ above are the electromagnetic invariants defined by

$$\mathscr{F} := \tfrac{1}{4}f_{\mu\nu}f^{\mu\nu} + \tfrac{1}{4}f_{\mu\nu}\check{f}^{\mu\nu} =: -\tfrac{1}{2}(E + i\check{B})^2,$$

where $f_{\mu\nu}$ is the tensor of the electromagnetic field, and $\check{f}_{\mu\nu}$ its dual.

In this work we study the two exceptional cases where, still assuming $A_1$ and $A_2$, we additionally postulate that either

$$E^2 + \check{B}^2 = \tfrac{3}{2}C^{(3)} \neq 0 \tag{1.2}$$

or

$$E^2 + \check{B}^2 = -\tfrac{3}{2}C^{(3)} \neq 0. \tag{1.3}$$

($C^{(3)}$ must be real in both cases.) Then the BI *do not* imply that $e^3$ and $e^4$ are geodesic and shearfree.

This last fact makes impossible the application of a theorem by Hughston *et al.*,[7] according to which electrovac solutions which satisfy conditions $A_1$, $A_2$, and $A_3$ must admit two Killing vectors. In fact, we shall see in Sec. 2 of this paper that a solution exists, satisfying conditions $A_1$, $A_2$, and (1.2), which is totally asymmetric, i.e., it admits no Killing vectors. We shal also see, in Sec. 3, that nontrivial solutions satisfying $A_1$, $A_2$, and (1.3) *do not* exist.

## 2. THE CASE $C^{(3)} > 0$

### A. Basic equations

The general description of the consequences of $A_1$ and $A_2$ within the tetrad formalism is given in detail in Sec. 1 of Ref. 1. Imposing the additional condition (1.2), we obtain from the BI:

$$\Gamma_{422} = 0 = \Gamma_{311}, \tag{2.1}$$

$$\tfrac{1}{2}dC^{(3)} = 3C^{(3)}(\Gamma_{312}e^3 + \Gamma_{421}e^4). \tag{2.2}$$

since $C^{(3)}$ is real and not equal to 0, it immediately follows that

$$\Gamma_{312} = \Gamma_{321}, \quad \Gamma_{421} = \Gamma_{412}. \tag{2.3}$$

The Maxwell equations can be written as

$$d\ln(E + i\check{B})^{1/2} + \Gamma_{314}e^1 + \Gamma_{423}e^2 - \Gamma_{312}e^3 - \Gamma_{421}e^4 = 0.$$

Taking the real part of this equation, and remembering (2.3), we obtain

$$\tfrac{1}{2}d\ln(E^2 + \check{B}^2) + (\Gamma_{413} + \Gamma_{314})e^1 + (\Gamma_{423} + \Gamma_{324})e^2 - 2\Gamma_{312}e^3 - 2\Gamma_{421}e^4 = 0. \tag{2.4}$$

Using (1.2) and (2.3), it follows that this last equation implies

$$\bar{\Gamma}_{423} + \Gamma_{314} = 0, \tag{2.5}$$

$$\Gamma_{421} = 0 = \Gamma_{312} \tag{2.6}$$

(bars denote complex conjugation) and therefore $C^{(3)}$ and $E^2 + \check{B}^2$ must be constant. Knowing this, we can set

$$E + i\check{B} = (E^2 + \check{B}^2)^{1/2}e^{-2i\psi}, \tag{2.7}$$

reducing Maxwell equations to

$$id\psi = -\bar{\Gamma}_{423}e^1 + \Gamma_{423}e^2. \tag{2.8}$$

Thus, the connections $\Gamma_{42}$ and $\Gamma_{31}$ have the form

$$\Gamma_{42} = \Gamma_{423}e^3 + \Gamma_{424}e^4, \tag{2.9a}$$

$$\Gamma_{31} = \Gamma_{313}e^3 - \bar{\Gamma}_{423}e^4. \tag{2.9b}$$

[a)] On leave of absence from University of Warsaw, Warsaw, Poland.

  

Equations (2.8) and (2.9), together with condition $\frac{3}{2}C^{(3)} = E^2 + \check{B}^2 = \text{const}\neq 0$, and the starting point of our considerations.

Before proceeding, it is important to notice that the following gauge transformations are still possible:

phase gauge:

$$e'^1 = e^{i\phi}e^1,$$
$$e'^2 = e^{-i\phi}e^2,$$

boost gauge:

$$e'^3 = e^{\chi}e^3,$$
$$e'^4 = e^{-\chi}e^4.$$

The connections $\Gamma$ tranform as

$$\Gamma'_{423} = e^{i\phi}\Gamma_{423},$$
$$\Gamma'_{424} = e^{2\chi + i\phi}\Gamma_{424}, \qquad (2.10)$$
$$\Gamma'_{313} = e^{-2\chi - i\phi}\Gamma_{313}.$$

The optimal choice of gauge will be determined later.

We now list the structure equations relevant to our problem. The first structure equations, $de^a = e^b \wedge \Gamma^a_{\ b}$, are explicitly:

$$de^1 = \Gamma_{12}\wedge e' + 2\Gamma_{423}e^3\wedge e^4, \qquad (2.11a)$$
$$de^2 = -\Gamma_{12}\wedge e^2 + 2\overline{\Gamma}_{423}e^3\wedge e^4, \qquad (2.11b)$$
$$de^3 = \alpha\wedge e^3 + \beta\wedge e^4, \qquad (2.11c)$$
$$de^4 = \gamma\wedge e^3 - \alpha\wedge e^4, \qquad (2.11d)$$

where

$$\alpha := \overline{\Gamma}_{423}e' + \Gamma_{423}e^2 + \Gamma_{34},$$
$$\beta := \overline{\Gamma}_{424}e' + \Gamma_{424}e^2,$$
$$\gamma := \Gamma_{313}e' + \overline{\Gamma}_{313}e^2.$$

The second structure formulas are, in the particlar case that we are studying,

$$d\Gamma_{42} + \Gamma_{42}\wedge(\Gamma_{12}+\Gamma_{34}) = -2\rho e^3\wedge e^1, \qquad (2.12a)$$
$$d\Gamma_{31} + (\Gamma_{12}+\Gamma_{34})\wedge\Gamma_{31} = -2\rho e^4\wedge e^2, \qquad (2.12b)$$
$$d(\Gamma_{12}+\Gamma_{34}) + 2\Gamma_{42}\wedge\Gamma_{31} = 2\rho e^1\wedge e^2 - (10\rho + 2\lambda)e^3\wedge e^4, \qquad (2.12c)$$

where we have defined

$$\rho := -\tfrac{1}{6}(E^2 + \check{B}^2 + \lambda) = \text{const}. \qquad (2.13)$$

Using (2.9) and (2.11), we can write Eqs. (2.12a) and (2.12b) in a somewhat more explicit form:

$$[d\Gamma_{423} + \Gamma_{423}(\alpha - \Gamma_{12} - \Gamma_{34}) + \Gamma_{424}\gamma - 2\rho e^1]\wedge e^3 + [d\Gamma_{424} - \Gamma_{424}(\alpha + \Gamma_{12} + \Gamma_{34}) + \Gamma_{423}\beta]\wedge e^4 = 0 \qquad (2.14a)$$

$$[d\Gamma_{313} + \Gamma_{313}(\alpha + \Gamma_{12} + \Gamma_{34}) - \overline{\Gamma}_{423}\gamma]\wedge e^3 + [-d\overline{\Gamma}_{423} - \overline{\Gamma}_{423}(-\alpha + \Gamma_{12} + \Gamma_{34}) + \Gamma_{313}\beta - 2\rho e^2]\wedge e^4 = 0. \qquad (2.14b)$$

Now, taking the factor of $e^2\wedge e^3$ in (2.14a) and combining it with the (complex conjugated) factor of $e^1\wedge e^4$ in (2.14b) it easily follows that

$$(\Gamma_{423})^2 + \Gamma_{424}\overline{\Gamma}_{313} = 0. \qquad (2.15)$$

A similar combination of the factor of $e^1\wedge e^3$ in (2.14a) with that of $e^2\wedge e^4$ in (2.14b) gives the relation

$$2\Gamma_{423}\overline{\Gamma}_{423} + \Gamma_{424}\Gamma_{313} + \overline{\Gamma}_{424}\overline{\Gamma}_{313} = 4\rho. \qquad (2.16)$$

We will now show that the condition $\Gamma_{423}\neq 0$ leads to a contradiction, and that therefore $\Gamma_{423}$ must vanish in order to have a nontrivial solution.

## B. Proof that $\Gamma_{423} = 0$

Assume for the moment that $\Gamma_{423}\neq 0$. Then, according to (2.15), there exists a complex function $\Lambda$ such that

$$\Gamma_{424} = e^{\Lambda}\Gamma_{423}, \qquad (2.17a)$$
$$\overline{\Gamma}_{313} = -e^{-\Lambda}\Gamma_{423}. \qquad (2.17b)$$

However, under a change of phase and boost gauge, $\Lambda$ changes as

$$\Lambda' = \Lambda - 2\chi.$$

Therefore, fixing the boost gauge, we can make $\Lambda$ purely imaginary, $\Lambda = i\phi(\phi = \phi)$. The remaining freedom of phase gauge can be used to fix $\Gamma_{423}$ as purely imaginary, $\Gamma_{423} = if$ $(f = \overline{f})$. Thus, the conections take the form

$$\Gamma_{42} = if(e^3 + e^{i\phi}e^4), \qquad (2.18a)$$
$$\Gamma_{31} = if(e^{i\phi}e^3 + e^4). \qquad (2.18b)$$

Maxwell Eqs. (2.8) reduce to

$$d\psi = f(e^1 + e^2), \qquad (2.19)$$

and Condition (2.16) becomes

$$(f\sin\phi)^2 = \rho = \text{const}, \qquad (2.20)$$

which implies $\phi = \phi(f)$. It is a direct consequence of Eq. (2.20) that

$$\lambda \leqslant -(E^2 + \check{B}^2) < 0. \qquad (2.21)$$

Notice also that Eq. (2.19) implies $(e^1 + e^2)\wedge d(e^1 + e^2) = 0$, and since, according to (2.11a) and (2.11b),

$$d(e^1 + e^2) = \Gamma_{12}\wedge(e^1 - e^2), \qquad (2.22)$$

it follows that $\Gamma_{12}\wedge e^1\wedge e^2 = 0$; thus

$$\Gamma_{123} = 0 = \Gamma_{124}, \qquad (2.22a)$$
$$f_{,3} = 0 = f_{,4}. \qquad (2.22b)$$

With all the results given above, Eqs. (2.14a) and (2.14b) take the simpler form

$$[df - f^2\sin 2\phi e^1 - f\Gamma_{12}]\wedge e^3 + [d(fe^{i\phi}) - 2f^2\sin\phi e^1 - fe^{i\phi}(\Gamma_{12} + 2\Gamma_{34})]\wedge e^4 = 0, \qquad (2.23a)$$

$$[df - f^2\sin 2\phi e^2 + f\Gamma_{12}]\wedge e^4 + [d(fe^{i\phi}) - 2f^2\sin\phi e^2 + fe^{i\phi}(\Gamma_{12} + 2\Gamma_{34})]\wedge e^3 = 0. \qquad (2.23b)$$

It is now a matter of algebraic manipulation to show that these last equations, together with their complex conjugates, yield the following information:

$$\Gamma_{12} + \Gamma_{34} = \left( -\frac{f_{,1}}{f} - \frac{i}{2}\phi_{,1} \right)(e^1 - e^2), \quad (2.24)$$

$$f_{,1} = f_{,2} = \tfrac{1}{2} f^2 \sin 2\phi, \quad (2.25)$$

$$f\phi_{,1} = -\rho. \quad (2.26)$$

The next step is to use these relations in the remaining structure Eq. (2.12c). It follows that

$$2\rho = -\lambda.$$

However, from the definition (2.13) of $\rho$, this implies that

$$\lambda = \tfrac{1}{2}(E^2 + \check{B}^2) > 0,$$

which contradicts the previous result (2.21), namely that $\lambda < 0$. Thus, the conclusion is that $\Gamma_{423}$ must necessarily vanish. We study this case in the following subsection.

## C. A solution with $\Gamma_{423} = 0$ and $\Gamma_{424} \neq 0$

When $\Gamma_{423} = 0$, the condition (2.15) requires

$$\Gamma_{424}\bar{\Gamma}_{313} = 0.$$

If this condition is met when both factors vanish separately, then $\Gamma_{42} = 0 = \Gamma_{31}$, which corresponds to the case of the Bertotti–Robinson solutions.[8,9] According to the work of Ref. 1, the most general soutions of this type with both $e^3$ and $e^4$ geodesic, shearless, and free of complex expansion can be represented in the form

$$ds^2 = 2e^1 \otimes e^2 + 2e^3 \otimes e^4$$
$$= 2\phi^{-2} d\zeta d\bar{\zeta} + 2\psi^{-2} du dv, \quad (2.27)$$

where, $E + i\check{B}$ being constant,

$$\phi = 1 + \tfrac{1}{2}(\lambda + E^2 + \check{B}^2)\zeta\bar{\zeta}, \quad (2.28a)$$

$$\psi = 1 + \tfrac{1}{2}(\lambda - E^2 - \check{B}^2)uv, \quad (2.28b)$$

and the electromagnetic field is given by

$$\omega := \tfrac{1}{2}(f_{\mu\nu} + \check{f}_{\mu\nu}) dx^\mu dx^\nu$$

$$= (E + i\check{B})(e^1 \wedge e^2 + e^3 \wedge e^4) = d\aleph$$

$$\aleph := \tfrac{1}{2}(E + i\check{B})[\phi^{-1}(\zeta d\bar{\zeta} - \bar{\zeta} d\zeta) + \psi^{-1}(u dv - v du)], \quad (2.29)$$

while the nontrivial component of the conformal curvature is given by

$$C^{(3)} = -\tfrac{2}{3}\lambda. \quad (2.30)$$

A subcase of these B–R solutions with

$$E^2 + \check{B}^2 = \tfrac{3}{2} C^{(3)} = -\lambda, \quad (2.31)$$

which is possible with $\lambda < 0$ only, provides the solution of the problem studied in this paper when $\Gamma_{42} = 0 = \Gamma_{31}$. Notice that in this case we have, for the factors in the B–R solution,

$$\phi = 1, \quad \psi = 1 + \lambda uv. \quad (2.32)$$

We will now try to meet the condition (3.1) in a nontrivial manner assuming

$$\Gamma_{423} = 0, \quad \Gamma_{424} \neq 0, \quad \Gamma_{313} = 0, \quad (2.33)$$

and having, then,

$$\Gamma_{42} = \Gamma_{424} e^4, \quad \Gamma_{31} = 0. \quad (2.34)$$

According to (2.12b) this, of course, is possible only if

$$\rho = 0 \rightarrow E^2 + \check{B}^2 = -\lambda, \quad (2.35)$$

so that $\lambda$ must be negative.

Under the present assumptions, the first structure equations are

$$de^1 = \Gamma_{12} \wedge e^1, \quad (2.36a)$$

$$de^2 = -\Gamma_{12} \wedge e^2, \quad (2.36b)$$

$$de^3 = (\bar{\Gamma}_{424} e^1 + \Gamma_{424} e^2) \wedge e^4 + \Gamma_{34} \wedge e^3, \quad (2.36c)$$

$$de^4 = -\Gamma_{34} \wedge e^4, \quad (2.36d)$$

and Eqs. (2.12) reduce to:

$$d(\Gamma_{424} e^4) - (\Gamma_{12} + \Gamma_{34}) \wedge \Gamma_{424} e^4 = 0, \quad (2.37a)$$

$$d(\Gamma_{12} + \Gamma_{34}) = -2\lambda e^3 \wedge e^4. \quad (2.37b)$$

The Maxwell equations, according to (2.8), are fulfilled in this case if

$$E + i\check{B} = \text{const.} \quad (2.38)$$

We can now conveniently use the freedom of phase and boost gauges. From the imaginary part of (2.37b) we have $d\Gamma_{12} = 0 \rightarrow \Gamma_{12} = i d\phi$, $\phi = \bar{\phi}$. Consequently, by properly choosing the phase gauge we can set, without loss of generality,

$$\Gamma_{12} = 0. \quad (2.39)$$

Thus Eqs. (2.36a) and (2.36b) imply the existence of a complex coordinate $\zeta$ such that

$$e^1 = d\zeta, \quad e^2 = d\bar{\zeta}, \quad (2.40)$$

and the equation for $de^4$ implies that $e^4$ is surface orthogonal. Therefore, without loss of generality, we can fix the boost gauge in such away that

$$e^4 = dv. \quad (2.41)$$

Thus, the first structure equations reduce to

$$\Gamma_{34} \wedge dv = 0, \quad (2.42a)$$

$$de^3 = (\bar{\Gamma}_{424} d\zeta + \Gamma_{424} d\bar{\zeta}) \wedge dv + \Gamma_{34} \wedge e^3, \quad (2.42b)$$

while the second structure equations, using (2.37) amount to

$$d(\Gamma_{424} dv) = 0, \quad (2.43a)$$

$$d\Gamma_{34} = -2\lambda e^3 \wedge dv. \quad (2.43b)$$

From (2.43a) $d\Gamma_{424} \wedge dv = 0$, and it follows that

$$\Gamma_{424} = F(v), \quad (2.44)$$

where this complex function of a real variable must be not equal to 0 under the present assumptions. On the other hand, from (2.42a),

$$\Gamma_{34} = \Gamma_{344} dv, \quad (2.45)$$

and (2.43b) supplies the information that

$$(d\Gamma_{344} + 2\lambda e^3) \wedge dv = 0. \quad (2.46)$$

Therefore, denoting

$$u := -\frac{1}{2\lambda}\Gamma_{344}, \quad (2.47)$$

it follows that $e^3$ has the form of

$$e^3 = du + Xdv, \tag{2.48}$$

where $X$ is some real function. Entering with this information into (2.42b) we have

$$dX \wedge dv = [\bar{F}(v)d\zeta + F(v)d\bar{\zeta}] \wedge dv + 2\lambda u du \wedge dv, \tag{2.49}$$

which is equivalent to:

$$d[X - \zeta\bar{F}(v) - \bar{\zeta}F(v) - \lambda u^2] \wedge dv = 0. \tag{2.50}$$

Therefore, there must exist a real function $G = G(v)$ such that

$$X = \zeta\bar{F}(v) + \bar{\zeta}F(v) + \lambda u^2 + G(v), \tag{2.51}$$

and consequently the integrated tetrad of the considered case assumes the form of

$$\begin{aligned} &e^1 = d\zeta, \quad e^2 = d\bar{\zeta}, \quad e^4 = dv, \\ &e^3 = du + [\lambda u^2 + \zeta\bar{F}(v) + \bar{\zeta}F(v) + G(v)]dv, \end{aligned} \tag{2.52}$$

with the complex $F(v) \neq 0$ and $G(v)$ being arbitrary.

The tetrad is accompanied by the connections:

$$\begin{aligned} &\Gamma_{42} = F(v)dv, \quad \Gamma_{31} = 0, \\ &\Gamma_{12} = 0, \quad \Gamma_{34} = -2\lambda u dv, \end{aligned} \tag{2.53}$$

so that the double D–P vector $e^3$ has the nontrivial geodesity $\Gamma_{424} = F(v) \neq 0$.

The tetrad given above fulfills the second structure equations assuming the Einstein equations with Maxwellian sources and $\lambda$, with the constants fixed according to

$$E^2 + \check{B}^2 + \lambda = 0, \quad \tfrac{3}{2}C^{(3)} = E^2 + \check{B}^2 = -\lambda. \tag{2.54}$$

The 2 form of the electromagnetic field whose real eigenvectors coincide with the double D–P directions $e^3$ and $e^4$ is given by

$$\begin{aligned} \omega &= (E + i\check{B})(e^1 \wedge e^2 + e^3 \wedge e^4) \\ &= (E + i\check{B})(d\zeta \wedge d\bar{\zeta} + du \wedge dv), \end{aligned} \tag{2.55}$$

and is manifestly closed.

The metric of this solution,

$$\begin{aligned} ds^2 &= 2e^1 \otimes e^2 + 2e^3 \otimes e^4 \\ &= 2(d\zeta d\bar{\zeta} + du dv) + 2[\lambda u^2 + \zeta\bar{F}(v) + \bar{\zeta}F(v) + G(v)]dv^2, \end{aligned} \tag{2.56}$$

has the form of a single Kerr–Schild metric,[10,11,12] with the distinguished null vector aligned along that double D–P vector ($e^3$) which is geodesic, shearless, and free of the complex expansion.

We would now like to simplify the result described above by showing that the function $G(v)$ is irrelevant and can be absorbed by a coordinate transformation. For this purpose we first execute a transformation of the variable $u$,

$$u = A(v)u' + B(v), \tag{2.57}$$

where of course $A \neq 0$, and, otherwise, $A$ and $B$ are disposable functions of the variable $v$. We then obtain for $e^3$,

$$\begin{aligned} e^3 = {} & A du' + [\lambda A^2 u'^2 + u'(\dot{A} + 2\lambda AB) + \zeta\bar{F} + \bar{\zeta}F \\ & + (\dot{B} + \lambda B^2 + G)]dv. \end{aligned} \tag{2.58}$$

We can now choose $B(v)$ so that

$$\dot{B} + \lambda B^2 + G = 0, \tag{2.59}$$

and we select $A(u)$ as the integral of

$$\dot{A} + 2\lambda AB = 0, \tag{2.60}$$

so chosen that $A \neq 0$. Then $e^3$ reduces to

$$e^3 = A[du' + (\lambda u'^2 + \zeta\bar{F}' + \bar{\zeta}F')Adv], \tag{2.61}$$

where

$$F'(v) := A^{-2}(v)F(v). \tag{2.62}$$

We then execute the transformation of the variable $v$ defining

$$v' := \int^v A(v)dv \tag{2.63}$$

so that

$$e^4 = A^{-1}dv', \quad e^3 = A[du' + (\lambda u'^2 + \zeta\bar{F}' + \bar{\zeta}F')dv']. \tag{2.64}$$

In the result of this transformation, clearly

$$\omega = (E + i\check{B})(d\zeta \wedge d\bar{\zeta} + du' \wedge dv'), \tag{2.65a}$$

$$ds^2 = 2(d\zeta d\bar{\zeta} + du'dv') + 2(\lambda u'^2 + \zeta\bar{F}' + \bar{\zeta}F')dv'^2, \tag{2.65b}$$

and, dropping primes, it follows that without losing generality we can set in our result

$$G(v) = 0. \tag{2.66}$$

With the coordinate $v$ gauged in such a way that $G(v) = 0$, it is easy to show that if, additionally, $F(v) \to 0$, then our solution reduces precisely to the special case of the B–R solution discussed in the beginning of this section. Indeed, the transformation

$$u = (1 + \lambda\tilde{u}v)^{-1}\tilde{u} \tag{2.67}$$

gives

$$e^3 = du + \lambda u^2 dv = (1 + \lambda\tilde{u}v)^{-2}d\tilde{u}, \; e^4 = dv,$$

so that the metric and the electromagnetic field of our solution coincide here precisely with (2.27) and (2.29), with the factors $\phi$ and $\psi$ given by (2.32) after replacing $u$ in these formulas by $\tilde{u}$.

The presence in our solution of the function $F(v) \neq 0$ has an objective geometric meaning: It determines the symmetry properties of the solution. In order to see this explicitly with $F(v) \neq 0$, instead of making $G(v) = 0$ by using (2.57), we can always arrange that, by using (2.62),

$$F = \kappa e^{i\phi(v)}, \tag{2.68}$$

where, if $F = 0$, then $\kappa = 0$, and if $F \neq 0$, then $\kappa = 1$. The metric has the nontrivial components,

$$g_{\zeta\bar{\zeta}} = 1 = g_{uv}; \; g_{vv} = 2[\lambda u^2 + \kappa e^{-i\phi}\zeta + \kappa e^{i\phi}\bar{\zeta} + G(v)]. \tag{2.69}$$

When $\kappa = 1$, the real $\phi(v)$ and $G(v)$ replace as structural functions the complex $F(v)$. One can now easily work out the equation for a Killing vector of our metric,

$$\mathscr{L}_K g_{\mu\nu} = K^\rho g_{\mu\nu,\rho} + K^\rho{}_{,\mu} g_{\rho\nu} + K^\rho{}_{,\nu} g_{\rho\mu} = 0. \qquad (2.70)$$

Integrating this equation (we omit this process here for brevity) one finds that the most general Killing vector must have the form of

$$K^\mu \partial_\mu = \Big( u\dot\delta(v) - \frac{1}{2\lambda}\ddot\delta(v)\Big)\partial_u - \delta(v)\partial_v$$

$$+ (i\alpha_0\zeta + \beta_0)\partial_\zeta + (-i\alpha_0\bar\zeta + \bar\beta_0)\partial_{\bar\zeta}, \qquad (2.71)$$

where $\beta_0 = \text{const}$ is complex, and $\alpha_0 = \text{const}$ and $\delta = \delta(v)$ are real. These objects are, however, subject to the constraint conditions:

$$-\frac{1}{2\lambda}\dddot\delta - \delta\dot G - 2\dot\delta G + \kappa[\beta_0 e^{-i\phi} + \bar\beta_0 e^{i\phi}] = 0, \quad (2.72a)$$

$$\kappa\dot\delta = 0, \qquad (2.72b)$$

$$\kappa(\delta\dot\phi + \alpha_0) = 0. \qquad (2.72c)$$

If $\kappa = 0$ where, not losing generality but properly adjusting the coordinate $v$, we can set $G(v) = 0$ the only condition which remains is $\dddot\delta = 0$, so that $\delta = \frac{1}{2}pv^2 + qv + r$, and the coefficients of $p, q, r, \alpha_0, \beta_0$ and $\bar\beta_0$ in (3.48) define six Killing vectors of the special B–R metric.

If $\kappa = 1$, then $\dot\delta = 0 \to \delta = \delta_0 = \text{const}$ and the remaining constaint equations reduce to

$$\delta_0 \dot G = \beta_0 e^{-i\phi} + \bar\beta_0 e^{i\phi}, \quad \delta_0\dot\phi + \alpha_0 = 0. \qquad (2.73)$$

Consequently, if only

$$\ddot\phi \neq 0 \to \dot\phi \neq 0, \qquad (2.74)$$

then (2.72) implies $\delta_0 = 0$, $\alpha_0 = 0$, $\beta_0 = 0$, and, according to (2.71), $K^\mu = 0$, so that the metric *does not* admit any Killing vectors; it is completely asymmetric.

We can notice that, when $\phi = \rho_0 v + \phi_0$, $G = \sigma_0 e^{-i(\rho_0 v + \phi_0)} + \bar\sigma_0 e^{i(\rho_0 v + \phi_0)} + \eta_0$, with all symbols with the subscript zero being constants and $\rho_0 \neq 0$, then the studied metric has a single Killing vector,

$$K^\mu\partial_\mu = -\partial_v - i\rho_0(\zeta + \sigma_0)\partial_\zeta + i\rho_0(\bar\zeta + \bar\sigma_0)\partial_{\bar\zeta}. \quad (2.75)$$

For $\phi = \phi_0$ and $G$ being linear in $v$, there exist two Killing vectors; this case corresponds, in the proper gauge of coordinate $v$, to $G = 0$, $F =$ a complex constant, with the two obvious commuting Killing vectors,

$$\partial_v, \quad i(F\partial_\zeta - \bar F\partial_{\bar\zeta}). \qquad (2.76)$$

We shall conclude this section by observing that, when one wants to fulfill the conditions $\Gamma_{42} = 0$ and $\Gamma_{313} \neq 0$, then obviously the solution of the problem has again the general form of (3.28), but with the members of the tetrad interchanged according to the scheme:

$$e^1 \to e^2, \quad e^2 \to e^1, \quad e^3 \to e^4, \quad e^4 \to e^3,$$

i.e.,

$$e^1 = d\bar\zeta, \quad e^2 = d\zeta, \quad e^3 = dv, \qquad (2.77)$$

$$e^4 = du + [\lambda u^2 + \bar F(v)\zeta + F(v)\bar\zeta + G(v)]dv,$$

leading to $\Gamma_{313} = F(v)$.

## 3. THE CASE $C^{(3)} < 0$

In the second part of this paper, we will investigate the case (1.3) taken as the basic assumption. In this case the B.I. imply that

$$\Gamma_{424} = 0, \quad \Gamma_{313} = 0, \qquad (3.1)$$

so that both D–P vectors must be geodesic and

$$\tfrac{1}{2}dC^{(3)} = 3C^{(3)}(\Gamma_{314}e^1 + \Gamma_{423}e^2), \qquad (3.2)$$

which, with $C^{(3)} = \bar C^{(3)} \neq 0$, necessitates for consistency,

$$\Gamma_{314} = \Gamma_{413}, \quad \Gamma_{423} = \Gamma_{324}. \qquad (3.3)$$

The Maxwell equations in the form (2.4), added to their complex conjugate, yield

$$\tfrac{1}{2}d\ln(E^2 + \check B^2) + (\Gamma_{314} + \Gamma_{413})e^1 + (\Gamma_{423} + \Gamma_{324})e^2$$

$$-(\Gamma_{312} + \Gamma_{321})e^3 - (\Gamma_{421} + \Gamma_{412})e^4 = 0. \qquad (3.4)$$

By feeding here $E^2 + \check B^2 = -\frac{3}{2}C^{(3)}$ and using (3.2), one obtains, remembering (3.3),

$$-\Gamma_{314}e^1 - \Gamma_{423}e^2 - (\Gamma_{312} + \Gamma_{321})e^3 - (\Gamma_{421} + \Gamma_{412})e^4 = 0; \qquad (3.5)$$

in the case considered it necessarily follows that

$$\Gamma_{423} = 0 = \Gamma_{314}, \quad \Gamma_{421} + \Gamma_{412} = 0 = \Gamma_{312} + \Gamma_{321}. \quad (3.6)$$

Consequently, $C^{(3)} = \text{const}$ and $E^2 + \check B^2 = \text{const}$. Thus, setting (2.7) again, we obtain as the residual Maxwell equations,

$$-id\psi = \Gamma_{312}e^3 + \Gamma_{421}e^4, \qquad (3.7)$$

with both $\Gamma_{312}$ and $\Gamma_{421}$ purely imaginary.

The connections $\Gamma_{42}$ and $\Gamma_{31}$ must have the form

$$\Gamma_{42} = \Gamma_{421}e^1 + \Gamma_{422}e^2, \quad \Gamma_{31} = \Gamma_{311}e^1 + \Gamma_{312}e^2, \qquad (3.8)$$

so that in principle the geodesic vectors $e^3$ and $e^4$ can possess the nontrivial shears ($\Gamma_{422} \neq 0$, $\Gamma_{311} \neq 0$) and nontrivial twists ($\Gamma_{421} = -\bar\Gamma_{421} \neq 0$, $\Gamma_{312} = -\bar\Gamma_{312} \neq 0$).

The formulas (3.7) and (3.8), accompanied by

$$-\tfrac{3}{2}C^{(3)} = E^2 + \check B^2 = \text{const} \neq 0, \qquad (3.9)$$

form the starting point of the second part of this paper.

We notice at this point that, under the remaining tetrad gauge, the objects which enter in (3.7) and (3.8) transform according to

$$\Gamma'_{421} = e^{\kappa}\Gamma_{421}, \quad \Gamma'_{422} = e^{\kappa + 2i\phi}\Gamma_{422},$$
$$\Gamma'_{311} = e^{-\kappa - 2i\phi}\Gamma_{311}, \quad \Gamma'_{312} = e^{-\kappa}\Gamma_{312}. \qquad (3.10)$$

The optimal choice for this gauge will be determined later.

We will now list the structure equations for the case considered. One easily sees that the first structure equations are

$$de^1 = \alpha \wedge e^1 - \beta \wedge e^2, \qquad (3.11a)$$

$$de^2 = -\bar\beta \wedge e^1 - \alpha \wedge e^2, \qquad (3.11b)$$

$$de^3 = \Gamma_{34} \wedge e^3 - 2\Gamma_{421}e^1 \wedge e^2, \qquad (3.11c)$$

$$de^4 = -\Gamma_{34} \wedge e^4 + 2\Gamma_{312}e^1 \wedge e^2, \qquad (3.11d)$$

where

$$\alpha := \Gamma_{312}e^3 - \Gamma_{421}e^4 + \Gamma_{12} = -\bar{\alpha}, \qquad \beta := \bar{\Gamma}_{311}e^3 + \Gamma_{422}e^4. \tag{3.12}$$

The second Cartan structure formulas, with incorporated Einstein equations (with the Maxwellian energy–momentum tensor and the $\lambda$ term in sources), can now be stated as:

$$d\Gamma_{42} - (\Gamma_{12} + \Gamma_{34}) \wedge \Gamma_{42} = -2\mu(e^3 \wedge e^1), \tag{3.13a}$$

$$d\Gamma_{31} + (\Gamma_{12} + \Gamma_{34}) \wedge \Gamma_{31} = -2\mu(e^4 \wedge e^2), \tag{3.13b}$$

$$d(\Gamma_{12} + \Gamma_{34}) + 2\Gamma_{42} \wedge \Gamma_{31} = -(10\mu + 2\lambda)e^1 \wedge e^2 + 2\mu e^3 \wedge e^4, \tag{3.13c}$$

where we have denoted

$$\mu := -\tfrac{1}{6}(\lambda - E^2 - \check{B}^2) = \text{const}; \tag{3.14}$$

of course, deriving (3.13), we eliminated $C^{(3)}$ by using (3.9). Our problem consists now in the effective integration of the Eqs. (3.7), (3.11), and (3.13). The following relations are direct consequences of Eqs. (3.13a) and (3.13b):

$$(\Gamma_{421})^2 + \Gamma_{422}\bar{\Gamma}_{422} = 0, \tag{3.15a}$$

$$(\Gamma_{312})^2 + \Gamma_{311}\bar{\Gamma}_{311} = 0, \tag{3.15b}$$

$$2\Gamma_{421}\Gamma_{312} - \Gamma_{424}\Gamma_{311} - \bar{\Gamma}_{422}\bar{\Gamma}_{311} + 4\mu = 0 \tag{3.15c}$$

[the derivation is entirely analogous to that of Eqs. (2.15) and (2.16)]. Therefore, remembering that $\Gamma_{421}$ and $\Gamma_{312}$ are purely imaginary, there exist *real* scalars $\rho, \sigma, \tau$, and $\omega$, such that

$$\Gamma_{42} = i\rho(e^1 + e^{i\sigma}e^2), \tag{3.16a}$$

$$\Gamma_{31} = i\tau(e^{i\omega}e^1 + e^2). \tag{3.16b}$$

Under a gauge transformation, these scalars transform as

$$\rho' = e^{\kappa}\rho, \quad \tau' = e^{-\kappa}\tau, \qquad \sigma' = \sigma + 2\phi, \quad \omega' = \omega - 2\phi. \tag{3.17}$$

As this point it is convenient to consider the different branches of our problem. First there is the simplest possibility that $\rho = 0 = \tau$, namely,

$$\Gamma_{42} = 0 = \Gamma_{31}.$$

We know, however, that this corresponds to a BR solution in the special case when

$$E^2 + \check{B}^2 = -\tfrac{3}{2}C^{(3)} = \lambda.$$

The factors in the BR metric (2.27) take the form

$$\phi = 1 + \phi = 1 + \lambda\xi\bar{\xi}, \quad \psi = 1,$$

and the double DP vectors, $e^3$and $e^4$, are geodesic, shearless, and without complex expansion.

The following possibility is that at least one of the scalars $\rho$ and $\tau$ does not vanish, say $\rho \neq 0$. Then, using the freedom of gauge (3.17) we can set

$$\Gamma_{42} = i\epsilon(e^1 + e^2), \tag{3.18a}$$

$$\Gamma_{31} = i\tau(e^{i\omega}e^1 + e^2), \tag{3.18b}$$

where $\epsilon^2 = 1$, $\tau$ may or may not vanish.

It is now a matter of algebraic manipulation to show that when these relations (3.18) are substituted in the second structure Eqs. (3.13), the following equations result:

$$\lambda = -2\mu = -\tfrac{1}{2}(E^2 + \check{B}^2), \tag{3.19}$$

$$\lambda = -\epsilon\tau(1 - \cos\omega), \tag{3.20}$$

$$d\tau = \tau\sin\omega d\psi, \tag{3.21}$$

$$\tau d\omega = \lambda\epsilon d\psi, \tag{3.22}$$

$$\Gamma_{12} + \Gamma_{34} = \left(\frac{\lambda}{2}i\epsilon + \tau\sin\omega\right)e^3, \tag{3.23}$$

together with the Maxwell equations

$$-d\psi = \tau e^3 + \epsilon e^4. \tag{3.24}$$

Using this information and Eqs. (3.11) and (3.12), it follows that

$$e^1 \wedge e^2 \wedge de^1 = 0, \tag{3.25}$$

$$d(e^1 \wedge e^2) = 0. \tag{3.26}$$

The first equation implies that $e^1$ and $e^2$ must have the form

$$e^1 = \bar{A}dz + Bd\bar{z}, \qquad e^2 = \bar{B}dz + Ad\bar{z}, \tag{3.27}$$

where $A$ and $B$ are complex functions, and $z$ is a complex coordinate. Equation (3.26) implies that the determinant

$$\Delta := A\bar{A} - B\bar{B}$$

must be a function of $z$ and $\bar{z}$ only. Since we have a freedom of coordinate transformation $z \to z'(z, \bar{z})$, we can set $\Delta = 1$ without loss of generality.

In the next step we notice that, according to (3.11), we have

$$de^3 = -2i\epsilon dz \wedge d\bar{z};$$

integrating, we find

$$e^3 = du - i\epsilon(zd\bar{z} - \bar{z}dz), \tag{3.28}$$

where $u$ is some function. Finally, from (3.24) we have

$$e^4 = -\epsilon d\psi - \epsilon\tau[du - i\epsilon(zd\bar{z} - \bar{z}dz)]. \tag{3.29}$$

The condition $e^1 \wedge e^2 \wedge e^3 \wedge e^4 \neq 0$ implies that $dz \wedge d\bar{z} \wedge du \wedge d\psi \neq 0$, and therefore we can use $z, \bar{z}, u, \psi$ as independent coordinates. Furthermore, it follows that $\tau$ and $\omega$ must have the form

$$\tau = -\tfrac{1}{2}\epsilon\lambda(1 + \psi^2), \tag{3.30}$$

$$\sin\omega = \frac{2\psi}{1 + \psi^2}, \tag{3.31}$$

as a result of integrating Eqs. (3.21) and (3.22) (a constant of integration has been absorbed by a change of coordinate $\psi \to \psi$ + const). Summarizing, we have obtained the general form of the tetrad through Eqs. (3.27), (3.28), and (3.29). It remains to substitute these forms in the structure equations (3.11) in order to obtain additional constraints on the functions $A$ and $B$. It turns out that Eqs. (3.11c) and (3.11d) are automatically satisfied, whereas Eqs. (3.11a) and (3.11b) imply

$$d\bar{A} \wedge dz + dB \wedge d\bar{z}$$

$$= -\frac{i}{2}\epsilon\lambda\psi^2 e^3 \wedge e^1 + \frac{i}{2}\epsilon\lambda(-\psi^2 + 1 + 2i\psi)$$

$$\times e^3 \wedge e^2 - i\epsilon e^4 \wedge (e^1 + e^2), \tag{3.32}$$

with the additional constraint $\Delta = 1$. Some algebraic manipulations show that, according to (3.32), the functions $A$ and $B$ must have the form

$$A = (1 - i\psi) f e^{(i/2)\epsilon\lambda u} - i\psi g e^{-(i/2)\epsilon\lambda u}, \tag{3.33a}$$

$$B = i\psi f e^{(i/2)\epsilon\lambda u} + (1 + i\psi) g e^{-(i/2)\epsilon\lambda u}, \tag{3.33b}$$

where $f$ and $g$ are functions of $z,\bar{z}$ subject to the conditions

$$f_{,z} - \bar{g}_{,\bar{z}} = \mu(f\bar{z} - \bar{g}z), \tag{3.34}$$

$$f\bar{f} - g\bar{g} = 1. \tag{3.35}$$

Now, write

$$\bar{f} = F\Omega_{,z}, \tag{3.36a}$$

$$g = F\Omega_{,\bar{z}}, \tag{3.36b}$$

which is always possible for functions of two variables. Then Eqs. (3.34) and (3.35) imply:

$$F\bar{F} d\Omega \wedge d\bar{\Omega} = dz \wedge d\bar{z}, \tag{3.37}$$

$$d\Omega \wedge [\lambda F(z d\bar{z} - \bar{z} dz) + 2dF] = 0. \tag{3.38}$$

At this point, we notice that the tetrad forms (3.27), (3.28), and (3.29) are invariant under a coordinate transformation:

$$z \to z' = z'(z,\bar{z}),\ u \to u' = u + v(z,\bar{z}),$$

such that

$$z d\bar{z} - \bar{z} dz = z' d\bar{z}' - \bar{z}' dz' - dv.$$

(Since $dz \wedge d\bar{z} = dz' \wedge d\bar{z}'$, the condition $\Delta = 1$ is preserved.) Then, choosing the arbitrary function $v$ such that

$$v = 2\lambda^{-1} \ln F, \tag{3.39}$$

Eqs. (3.37) and (3.38) read, in the new system of coordinates,

$$F\bar{F} d\Omega \wedge d\bar{\Omega} = dz' \wedge d\bar{z}', \tag{3.40}$$

$$d\Omega \wedge (z' d\bar{z}' - \bar{z}' dz') = 0. \tag{3.41}$$

Equation (3.41) implies that $d\Omega$ is proportional to

$$z' d\bar{z}' - \bar{z}' dz',$$

which would imply $dz' \wedge d\bar{z}' = 0$ according to (3.40)—a contradiction!

Thus it follows that the case $C^{(3)} < 0$ cannot contain nontrivial solutions.

## 4. CONCLUSIONS

The results of this paper can be sumarized as follows: We have studied all electrovac solutions of the Einstein–Maxwell equations which are of type D, have thier DP vectors aligned along the real eigenvectors of the electromagnetic field, and are subject to the condition

$$(E^2 + \check{B}^2)^2 = (\tfrac{3}{2}C^{(3)})^2 \neq 0.$$

This condition does not allow an extension of the Goldberg–Sachs theorem.[13] We have found that the nontrivial solutions are of a very special type.

First, there are the special Bertotti–Robinson solutions, with $E^2 + \check{B}^2 = \mp\lambda$, where both $e^3$ and $e^4$ are geodesic, shearfree, and without complex expansion. There solutions exist only when $\lambda < 0$ or $\lambda > 0$, respectively.

Additionally, when $\lambda < 0$ and $E^2 + B^2 = -\lambda$, there exists an exceptional solution with $C^{(3)} = -(2/3)\lambda$ which has one DP vector geodesic, shearless and free of complex expansion, while the other DP vector has a nonzero geodesicity. This exceptional solution is

$$ds^2 = 2d\zeta d\bar{\zeta} + 2\{du + [\lambda u^2 + \zeta\bar{F}(v) + \bar{\zeta}F(v)]dv\}dv, \tag{4.1}$$

$$\tfrac{1}{2}(f_{\mu\nu} + \check{f}_{\mu\nu})dx^\mu \wedge dx^\nu = (E + i\check{B})(d\zeta \wedge d\bar{\zeta} + \mathrm{d}u \wedge dv), \tag{4.2}$$

and in general does not possess any Killing vector. It may admit one Killing vector when $F(v)$ is a particular function, two Killing vectors when $F = \text{const} \neq 0$, or six Killing vectors when $F = 0$ (reducing to a B–R solution).

The moral of these results is that a generalization of the Goldberg–Sachs[13] theorem for type D metrics is possible.

*Theorem*: If a type D electrovac solution of the Einstein–Maxwell equations has its two Debever–Penrose vectors parallel to the two eigenvectors of the electromagnetic field, then these vectors must be geodesic and shearfree if $\lambda > 0$. If $\lambda < 0$, the statement still applies, except when $E^2 + \check{B}^2 = \frac{3}{2}C^{(3)} = -\lambda$.

## ACKNOWLEDGMENTS

The author are grateful to R.P. Kerr for a discussion related to the problem studied in this paper. They also acknowledge the interest of S. Alarcón Gutiérrez in this work.

[1] J.F. Plebański "The nondiverging and nontwisting type D electrovac solutions with $\lambda$," to be published in J. Math. Phys. (1979).
[2] G. Debney, R.P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
[3] J.F. Plebański, Ann. Phys. **90**, 280 (1975).
[4] J.F. Plebański and M. Demiański, Ann. Phys. **98**, 98 (1976).
[5] G. Weir and R.P. Kerr, Proc. R. Soc. London Ser. A **335**, 31 (1977).
[6] J.F. Plebański, "Spinors, Tetrads and Forms," a monograph of Centro de Investigación y Estudios Avanzados, unpublished (1973).
[7] L.P. Hughton, R. Penrose, P. Sommers, and M. Walker, Commun. Math. Phys. **27**, 303 (1972).
[8] B. Bertotti, Phys. Rev. **116**, 1331 (1959).
[9] I. Robinson, Bull. Acad. Pol. Sci. Sec. Math. Phys. **7**, 35 (1959).
[10] R.P. Kerr and A. Schild, *Atti del Convegno sulla Relativitá Generale* (Firenze), p. 173 (1965).
[11] R.P. Kerr and A. Schild, *Proceedings of Symposia in Applied Mathematics* (Providence, R.I.), **17**, 199 (1966).
[12] J.F. Plebański and A. Schild, Nuovo Cimento B **35**, 35 (1976).
[13] J. Goldberg and R. Sachs, Acta Phys. Pol. Suppl. **22**, 13 (1962).

# Global properties of systems quantized via bundles[a]

H. D. Doebner[b)]

*International Centre for Theoretical Physics, Trieste, Italy*

J. -E. Werth

*Institut für theoretische Physik, Technische Universität Clausthal, Clausthal, Fed. Rep. Germany*
(Received 1 April 1978)

Take a smooth manifold $M$ and a Lie algebra action (G action) $\theta$ on $M$ as the geometrical arena of a physical system moving on $M$ with momenta given by $\theta$. We propose to quantize the system with a Mackey-like method via the associated vector bundle $\xi_\rho$ of a principal bundle $\xi = (P,\pi,M,H)$ with model dependent structure group $H$ and with G-action $\phi$ on $P$ lifted from $\theta$ on $M$. This (quantization) bundle $\xi_\rho$ gives the Hilbert space $\mathrm{H} = L^2(\xi_\rho,\omega)$ of the system as the linear space of sections in $\xi_\rho$ being square integrable with respect to a volume form $\omega$ on $M$; the usual position operators are obtained; $\phi$ leads to a vector field representation $D(\phi_\rho,\theta)$ of G in H and hence to momentum operators. So H carries the quantum kinematics. In this quantization the physically important connection between geometrical properties of the system, e.g., quasicompleteness of $\theta$ and $G$ maximality of $\phi_\rho$, and global properties of its quantized kinematics, e.g., skew-adjointness of the momenta and integrability of $D(\phi_\rho,\theta)$ can easily be studied. The relation to Nelson's construction of a skew-adjoint nonintegrable Lie algebra representaion and to Palais' local $G$ actions is discussed. Finally the results are applied to actions induced by coverings as examples of nonmaximal $\phi_\rho$ on $E_\rho$ lifted from maximal $\theta$ on $M$ which lead to direct consequences for the corresponding quantum kinematics.

## 1. INTRODUCTION

Consider the geometrical arena on which a classical system is described. Suppose this to be a smooth manifold $M$ or a base of a principal bundle which can be equipped with a local symmetry via a $\mathfrak{g}$-action connected with physical momenta. Then the global geometrical structure induces certain properties for the system depending on the type of theory which is constructed on the arena. Results of this type are known. We refer, e.g., to classical Maxwell fields on manifolds,[1,2] to gauge theories on principal bundles[3–5] and to spin structures on nonsimply-connected manifolds and their use in many-body physics.[6]

To quantize a (particlelike) system we use principal fibre bundles $\xi = (P,\pi,M,H)$ with projection $\pi: P\to M$, structure group $H$, and a $\mathfrak{g}$ action $(\phi,\theta)$ on $\xi$, where $\theta$ is a Lie algebra homomorphism from a finite-dimensional Lie algebra $\mathfrak{g}$ into the Lie algebra $\mathfrak{V}(M)$ of smooth vector fields on $M$ and $\phi$ denotes a lift of $\theta$ to $P$. Then any unitary finite-dimensional representation $\rho: H\to \mathrm{Aut}\,V$ with associated vector bundle $\xi_\rho = (E_\rho, \pi_\rho, M, V)$[7] and $\mathfrak{g}$ action $(\phi_\rho,\theta)$ induces a Mackey-type quantization via a vector field representation $D$ of $\mathfrak{g}$ on a Hilbert space $\mathscr{H}$ spanned by smooth sections in $\xi_\rho$. Because the quantized kinematics is completely given by $\phi_\rho$ on $\xi_\rho$, we call $\xi_\rho$ *quantization bundle*. $\phi_\rho$ leads to momentum operators and to any Borel set on $M$ corresponds a position projector.

The global group theoretical properties of the $\mathfrak{g}$ action $(\phi_\rho,\theta)$ on the quantization bundle and the representation $D$ of $\mathfrak{g}$ are directly linked: Quasicompleteness (of $\theta$) and maximality (of $\phi_\rho$) correspond to skew-adjointness and integrability of $D$. So the local $\mathfrak{g}$ actions on $E_\rho$ and $M$ give, via their physically significant singularity[8] structure, information on the quantum mechanics on the bundle, and vice versa.

The material is organized as follows: A short description of Lie algebra actions on quantization bundles and of the corresponding quantization is given in Sec. 2; Lemma 1 relates the quasicompleteness of the action to the skew-adjointness of the vector field representation. Global results are presented in Sec. 3; using a property of skew-adjoint representations (Lemma 2) we prove that a vector field representation $D$ is $G$ integrable if and only if $\phi_\rho$ is $G$ maximal (Theorem 1). Relations to Nelson's construction of a skew adjoint nonintegrable representation[9] and to Palais' local $G$ actions[10] are given. The results are applied to actions induced by coverings in Sec. 4. Two types of nonintegrability are distinguished: those which are induced from nonmaximal $\theta$ on $M$ and those which are induced from nonmaximal $\phi_\rho$ on $E_\rho$, whereas the projected symmetry $\theta$ is maximal. The proofs of Lemma 1 and Theorem 1 are given in the Appendix.

## 2. QUANTIZATION BUNDLES WITH $\mathfrak{g}$-ACTIONS

1. Let $F(X)$ be the flow of a smooth vector field $X\in\mathfrak{V}(M)$,

$$F(X):(m,t)\in D(X)\subset M\times\mathbb{R}\to M\ni F(X)(m,t)\equiv\varphi_t^X(m)$$

with (open[11]) domain $D(X,t) = \{m|(m,t)\in D(X)\}$ for $t\in\mathbb{R}$. Consider a principal fibre bundle $\xi = (P,\pi,M,H)$ with compact structure group $H$. Then a $\mathfrak{g}$ action $(\phi,\theta)$ on $\xi$ consists of $\mathfrak{g}$ actions $\phi$: $\mathfrak{g}\to\mathfrak{V}(P)$, $\theta$:$\mathfrak{g}\to\mathfrak{V}(M)$, such that for $x\in\mathfrak{g}$

(i) $D(\phi(x)) = (\pi\times \mathrm{id}_{\mathbb{R}})^{-1}D(\theta(x))$;

(ii) $\pi(F(\phi(x))(p,t)) = F(\theta(x))(\pi(p),t)$ for $(p,t)\in D(\phi(x))$;

(iii) $\phi(x)$ is $H$ invariant.

[a)]Dedicated to Günther Ludwig on the occasion of his 60th birthday.
[b)]Permanent address: Institut für Theoretische Physik, Technische Universität Clausthal, Clausthal, Federal Republic of Germany.

 0022-2488/79/061011-04$01.00  

2. Let $\rho: H \to \mathrm{Aut}\, V$ be a representation of $H$ in a finite-dimensional complex vector space $V$. Then the quantization bundle $\xi_\rho = (E_\rho, \pi_\rho, M, V)$ is the ($\rho$-) vector bundle associated with $\xi$, $E_\rho$ being the orbit space of the $H$ action on $P \times V$ given by $(p,v)h = (ph, \rho^{-1}(h)v)$; $[p,v]$ denotes the orbits and $\pi_\rho [p,v] = \pi(p)$ holds.

Any $\mathfrak{g}$ action $(\phi,\theta)$ on $\xi$ gives a $\mathfrak{g}$ action $(\phi_\rho,\theta)$ on $\xi_\rho$ via

$$F(\phi_\rho(x))([p,v]_\rho,t) = [F(\phi(x))(p,t),v]_\rho.$$

Here the $H$ invariance of $\phi(x)$ was used.

3. For a quantization consider the linear space $\mathrm{Sec}_0(\xi_\rho)$ of compactly supported smooth sections $\sigma$ in $\xi_\rho$. Equip $V$ with an inner product $\langle \cdot,\cdot \rangle_V$ and take $\rho$ to be unitary. This induces an unitary structure on $\xi_\rho$ via

$$\langle [p,v_1]_\rho, [p,v_2]_\rho \rangle_m = \langle v_1, v_2 \rangle_V,$$

$v_1, v_2 \in V$, $\pi(p) = m$. For a pre-Hilbert structure on $\mathrm{Sec}_0(\xi_\rho)$ choose a volume $\omega$ on $M$ ($M$ is assumed an orientable Riemannian or pseudo-Riemannian manifold) and define

$$\langle \sigma_1, \sigma_2 \rangle = \int_\omega \langle \sigma_1(m), \sigma_2(m) \rangle_m,$$

$\sigma_1, \sigma_2 \in \mathrm{Sec}_0(\xi_\rho)$. The corresponding Hilbert space is denoted by $L^2(\xi_\rho, \omega)$.

4. A $\mathfrak{g}$ action $(\phi_\rho,\theta)$ on $\xi_\rho$ induces a representation—called *vector field representation*—of $\mathfrak{g}$ on $\mathrm{Sec}_0(\xi_\rho)$ via a Lie algebra homomorphism $D(\phi_\rho,\theta)$: $\mathfrak{g} \to \mathrm{End}\, \mathrm{Sec}_0(\xi_\rho)$ through

$$(D(\phi_\rho,\theta)(x)\sigma)(m) = \frac{d}{dt} F(\phi_\rho(x))(\sigma(F(\theta(x))(m,t)), -t).$$

To have this representation skew-adjoint, i.e., $D(\phi_\rho,\theta)(x)$ essentially skew-adjoint for $x \in \mathfrak{g}$ on $\mathrm{Sec}_0(\xi_\rho)$, one needs some restrictions on $\theta$. $X \in \mathfrak{V}(M)$ is quasicomplete (is complete) if the set

$$E(X,t) = M - D(X,t)$$

is of measure zero (is empty) for $t \in \mathbb{R}$. So $\theta$ is quasicomplete or complete if $\theta(x)$ has this property for $x \in \mathfrak{g}$. Furthermore, $\omega$ is said to be $\theta$ invariant if $L_{\theta(x)}\omega = 0$, $x \in \mathfrak{g}$.

*Lemma* 1: Let $(\phi_\rho,\theta)$ be a $\mathfrak{g}$ action on $\xi_\rho$ such that $\theta$ is quasicomplete. Take a $\theta$-invariant $\omega$. Then the induced vector field representation $D(\phi_\rho,\theta)$ is skew-adjoint on $\mathrm{Sec}_0(\xi_\rho) \subset L^2(\xi_\rho,\omega)$.

For the proof see the Appendix.

5. $D(\phi_\rho,\theta)$ can be constructed also from $(\phi,\theta)$ on $\xi$ using the space $\mathscr{E}_0(P,\rho)$ of smooth and compactly supported equivariant functions $f: P \to V$, $f(ph) = \rho^{-1}(h) f(p)$, since the representation $D_\rho(\phi,\theta)$: $\mathfrak{g} \to \mathrm{End}\, \mathscr{E}_0(P,\rho)$ given by

$$(D_\rho(\phi,\theta)(x)f)\, p = \frac{d}{dt} f(F(\phi(x))(p,t))$$

is unitarily equivalent to $D(\phi_\rho,\theta)$ via the isomorphism $\psi_\rho: \mathscr{E}_0(P,\rho) \to \mathrm{Sec}_0(\xi_\rho)$, $(\psi_\rho(f))(m) = [p, f(p)]_\rho$ for $p \in \pi^{-1}(m)$.

6. In the special case where $(\phi,\theta)$ is the differential of a $G$ action on $\xi$ which is free and transitive on $P$, i.e., if $\xi$ is isomorphic to $(G,\pi,G/H,H)$, the quantization gives an induced representation in the sense of Mackey.[12,13]

## 3. LOOP CRITERIA

1. A skew-adjoint representation $D: \mathfrak{g} \to \mathscr{A}(\vartheta)$ is called $G$ integrable ($G$ being a connected Lie group with Lie algebra $\mathfrak{g}$), if there exists a unitary representation $U: G \to \mathfrak{U}(\mathscr{H})$ with differential $D_U = D$, i.e., if the diagram

$$\begin{array}{ccc} G & \xrightarrow{U} & \mathfrak{U}(\mathscr{H}) \\ \exp\uparrow & & \uparrow\ \mathrm{Exp} \\ \mathfrak{g} & \xrightarrow{D} & \mathscr{A}(\vartheta) \end{array}$$

is commutative. Here $\mathscr{A}(\vartheta)$ is the subset of essentially skew-adjoint operators in the Lie algebra $S(\vartheta)$ of skew-symmetric operators on a common dense invariant domain $\vartheta \subset \mathscr{H}$. As a simple geometric tool to decide on integrability we use loops in $G$ and in $\mathfrak{U}(\mathscr{H})$. Take the set $\Lambda(G)$ of all loops in $G$ starting at $e$ and consider

$$\hat{\Lambda}(G) = \{(x_1,\ldots,x_k) | k \in \mathbb{N},\ x_i \in \mathfrak{g},\ \exp x_1 \cdots \exp x_k = e\},$$

which can be regarded as a subset of $\Lambda(G)$ via the mapping $j: \hat{\Lambda}(G) \to \Lambda(G)$,

$$(j(x_1,\ldots,x_k))(t) = \exp x_1 \cdots \exp x_{n-1} \exp(kt - n + 1) x_n$$

for $t \in \Delta_n/k$, $\Delta_n = [n-1,n]$, $n = 1,\ldots,k$. Denote by $C(\mathfrak{U}(\mathscr{H}))$ and $\Lambda(\mathfrak{U}(\mathscr{H}))$ the set of curves and loops, respectively, in $\mathfrak{B}(\mathscr{H})$ starting at 1. Then there exists a natural map $\delta(D,G): \hat{\Lambda}(G) \to C(\mathfrak{U}(\mathscr{H}))$ with

$$\delta(D,G)(x_1,\ldots,x_k)(t) = \mathrm{Exp} D(x_1) \cdots \mathrm{Exp} D(x_{n-1}) \times \mathrm{Exp}(kt - n + 1) D(x_n)$$

for $t \in \Delta_n/k$, $\Delta_n = [n-1,n]$, $n = 1,\ldots,k$. This map controls the integrability of $D$. Any $g \in G$ ($G$ is connected) can be written as $g = \exp x_1 \cdots \exp x_k$ for suitable $x_i \in \mathfrak{g}$; hence the defining diagram gives

*Lemma* 2. A skew-adjoint representation $D$ of $\mathfrak{g}$ in $\mathscr{H}$ is $G$ integrable if any only if $\mathrm{Im}\,\delta(D,G) \subset \Lambda(\mathfrak{U}(\mathscr{H}))$.

2. To apply this loop criterion to vector field representations, we introduce the sets $C(P,p)$ and $\Lambda(P,p)$ of curves and loops in $P$ starting at $p$. Consider $\phi: \mathfrak{g} \to \mathfrak{V}(P)$, take the set

$$\Lambda(\phi,G,p) = \{(x_1,\ldots,x_k) \in \hat{\Lambda}(G) | \varphi_1^{\phi(x_k)} \cdots \varphi_1^{\phi(x_1)}(p) \text{ exists}\},$$

and construct, as above, $\delta(\phi,G,p): \Lambda(\phi,G,p) \to C(P,p)$ with

$$\delta(\phi,G,p)(x_1,\ldots,x_k)(t) = \varphi_{kt-n+1}^{\phi(x_n)} \varphi_1^{\phi(x_{n-1})} \cdots \varphi_1^{\phi(x_1)}(p)$$

for $t \in \Delta_n/k$, $\Delta_n = [n-1,n]$, $n = 1,\ldots,k$. Again, this $\delta$ map completely controls the integrability of skew-adjoint vector field representations.

*Theorem* 1: A skew-adjoint vector field representation $D(\phi_\rho,\theta)$ induced from a $\mathfrak{g}$ action $(\phi_\rho,\theta)$ is $G$ integrable if and only if

$$\mathrm{Im}\,\delta(\phi_\rho,G,q) \subset \Lambda(E_\rho,q) \tag{1}$$

for $q \in E_\rho$.

For the proof see the Appendix. Integrability criteria for symmetric or arbitrary skew-adjoint $D$ of $\mathfrak{g}$ can be found in Refs. 14 and 15.

3. This result relates the integrability of $D(\phi_\rho,\theta)$ to a global property of the $\mathfrak{g}$ action. Condition (1) can be identi-

fied with the infinitesimal version of Palais' maximal local $G$ action.[10] Hence we call $\phi_\rho$ on $E_\rho$ *G maximal* if condition (1) is fulfilled and the results in Ref. 10 imply that $\theta$ is $G$ maximal if and only if it is the differential of a "restriction" of a global $G$ action $\bar{\phi}$ on a manifold $\bar{M}$ which contains $M$ as an open submanifold. Furthermore, any complete $\theta$ is $\tilde{G}$ maximal; $\tilde{G}$ is the unicover of $G$.

4. The construction of a skew-adjoint non-$\mathbb{R}^2$-integrable representation by Nelson[9] takes for $\theta : \mathbb{R}^2 \to \mathfrak{B}(M)$ a non-$\mathbb{R}^2$-maximal action, which is generated (up to one point) from a covering (see Sec. 4) of the natural $\mathbb{R}^2$-action on the torus minus one point; in this case any lift of $\theta$ to an action on the total space of a bundle over $M$ is non-$\mathbb{R}^2$-maximal too. In Refs. 16 and 17 a nonmaximal $\theta$ given by a double covering of $\mathbb{R}^2 - (0,0)$ is used.

## 4. AN APPLICATION: COVERINGS

1. $G$ maximality of $\phi$ implies $G$ maximality of $\theta$; however, the converse is not true. To construct an example, take as geometrical arena a physically distinguished, connected but not simply connected manifold $M$ [e.g., $\mathbb{R}^2 - (0,0), \mathbb{R}^3$-$\mathbb{R}$] and a regular covering $p : M^N \to M$ characterized by a normal subgroup $N$ of the nontrivial fundamental group $\pi_1(M,m_0)$. Such a covering is a simple case of a nontrivial principal fibre bundle, given here as

$$\xi^{M,N} = (M^N, p, M, \pi_1, (M,m_0)/N);$$

the structure group $\pi_1(M,m_0)/N$ is discrete and the coverings are classified by normal subgroups $N \subset \pi_1(M,m_0)$.

2. Let $\theta : \mathfrak{g} \to \mathfrak{B}(M)$ be $G$ maximal. Since $p$ is a local diffeomorphism, $\theta$ induces a $\mathfrak{g}$ action

$$\theta^N : \mathfrak{g} \to \mathfrak{B}(M^N)$$

on $M^N$ such that $(\theta^N, \theta)$ is a $\mathfrak{g}$ action on $\xi^{M,N}$. Consider

$$\epsilon(\theta,G,m) = \nu(m) \circ \delta(\theta,G,m)$$

with $\nu(m) : C(M,m) \to \Omega(M,m)$ being the natural surjection onto the set of homotopy classes of curves in $C(M,m)$. Since a covering has unique path lifting, the following conditions are equivalent:

(i) $\theta^N$ is $G$ maximal;

(ii) $\mathrm{Im}\epsilon(\theta,G,p(m')) \subset p_* \pi_1(M^N,m')$ for $m' \in M^N$.

For transitive $G$ maximal $\theta$ this criterion can be simplified: The covering lift $\theta^N$ of $\theta$ is $G$ maximal iff there exists a $m_0' \in M^N$ such that

$$\mathrm{Im}\epsilon(\theta,G,p(m_0')) \subset p_* \pi_1(M^N,m_0').$$

For $m_0' \in p^{-1}(m_0)$ we have $p_* \pi_1(M^N,m_0') = N$.

3. For the construction of an associated vector bundle we use a unitary faithful finite-dimensional representation $\rho : \pi_1(M,m_0)/N \to \mathrm{Aut}V$. By Lemma 1 the corresponding vector field representation $D(\theta_\rho^N,\theta)$ of $\mathfrak{g}$ on $\mathrm{Sec}_0(\xi_\rho^{M,N})$ is skew-adjoint if $\theta$ is quasicomplete and $\omega$ is $\theta$ invariant. For the integrability Theorem 1 yields:

*Lemma* 3. A skew-adjoint vector field representation $D(\theta_\rho^N,\theta)$ induced from a transitive and $G$ maximal $\mathfrak{g}$ action $\theta$ on $M$ via a unitary and faithful $\rho$ is $G$ integrable if and only if

$$\mathrm{Im}\epsilon(\theta,G,m_0) \subset N.$$

4. As an example take $M = \mathbb{R}^2 - (0,0)$, $\omega = dx_1 \wedge dx_2$, $\mathfrak{g} = \mathbb{R}^2$ (two-dimensional Abelian Lie algebra) and define $\theta : \mathfrak{g} \to \mathfrak{B}(M)$ by

$$\theta(a,b) = a\frac{\partial}{\partial x} + b\frac{\partial}{\partial y}.$$

$\theta$ is quasicomplete and transitive, $\omega$ is $\theta$ invariant. Furthermore, $\theta$ is $G$ maximal for $G = \mathbb{R}^2$. Because

$$\pi_1(\mathbb{R}^2 - (0,0),*) = \mathrm{Im}\epsilon(\theta,\mathbb{R}^2,*) \cong \mathbb{Z},$$

Lemma 3 implies that $D(\theta_\rho^N,\theta)$ is $\mathbb{R}^2$-integrable iff $\mathbb{Z} \subset \mathrm{N}$, i.e., iff $\mathrm{N} = \mathbb{Z}$. Hence for $\mathrm{N} = 2\mathbb{Z} \subset \mathbb{Z}$ [which corresponds to a double covering of $\mathbb{R}^2 - (0,0)$], and

$$\rho : \mathbb{Z}/2\mathbb{Z} = \mathbb{Z}_2 \to \mathrm{Aut}\mathbb{C}$$

given by $\rho(0) = 1, \rho(1) = e^{i\pi}$, the skew-adjoint representation

$$D(\theta_\rho^{2\mathbb{Z}},\theta) : \mathbb{R}^2 \to \mathscr{A}(\mathrm{Sec}_0\xi^{M,2\mathbb{Z}})$$

is non-$G$-integrable for any $G$ with $\mathfrak{g} \cong \mathbb{R}^2$.

## 5. CLOSING REMARKS

1. A system with a manifold $M$ as configuration space and a $\mathfrak{g}$ action $\theta$ on $M$ connected with physical momenta can be quantized through a quantization bundle $\xi_\rho = (E_\rho, \pi_\rho, M, V)$. This is an associated vector bundle of a principle fibre bundle $\xi = (P,\pi,M,H)$ with structure group $H$ and $\mathfrak{g}$ action $(\phi,\theta)$, where $\phi$ is a lift of $\theta$. The physical interpretation of $H$ and its representation $\rho$ is model dependent. In special cases (as for quantum mechanics on homogeneous spaces[18]) the Lie algebra of $H$ may be chosen as a subalgebra of $\mathfrak{g}$.

The quantization method leads to compactly supported sections in $\xi_\rho$ as states and, with a volume form $\omega$ on $M$, to the Hilbert space $L^2(\xi_\rho,\omega)$; the Borel sets on $M$ correspond to projection operators and the $\mathfrak{g}$ action leads to a vector field representation $D(\phi_\rho,\theta)$ of $\mathfrak{g}$. The generators of $D(\phi_\rho,\theta)$ can be interpreted as observables (momentum operators) if they are essentially skew-adjoint. This is the case if $\theta$ on $M$ is quasicomplete and if the form $\omega$ is $\theta$ invariant.

As the $\mathfrak{g}$ action on $E_\rho$ does not need to be $G$ maximal, also $D(\phi_\rho,\theta)$ on $\mathrm{Sec}_0(\xi_\rho) \subset L^2(\xi_\rho,\omega)$ is not necessarily the differential of a unitary representation of a Lie group $G$ with Lie algebra $\mathfrak{g}$, i.e., $D(\phi_\rho,\theta)$ may be non-$G$-integrable. However, its integrability depends on the geometry of $(\phi_\rho,\theta)$ on $\xi_\rho$ only. The key is the loop criterion (see Lemmas 2, and 3) or, as an equivalent condition the $G$ maximality of $(\phi_\rho,\theta)$. So the integrability for vector field representations is reduced to a pure geometrical problem. Skew-adjoint representations of $\mathfrak{g}$ not being vector field representations will not appear as a reuslt of this geometrical quantization procedure.

2. For the $M$ given, some physical information is needed to choose a $\mathfrak{g}$ action $\theta$. If $\theta$ is taken to be $G$ maximal, it can be considered as a "restriction" of a complete $\mathfrak{g}$ action on some manifold $\bar{M} \supset M$ (see Sec. 3.3) and it is justified to call points in $\bar{M} - M = M'$ "singularities" of $\theta$; they cannot be reached by the system moving on $M$. The skew-adjointness of

$D(\phi_\rho,\theta)$ is directly connected with $M'$ and is assured if the integral curves ending in a singularity build a set of measure zero; under some restrictions (c.f. the examples given above) this is equivalent to

$$\dim\bar{M} - \dim M' \geqslant 2.$$

This condition is related to the construction of nonintegrable representations via coverings (Sec. 4), because only for

$$\dim\bar{M} - \dim M' = 2.$$

the homomorphism

$$i* : \pi_1(\bar{M} - M', m) \to \pi_1(\bar{M}, m)$$

is not injective in general,[19] which gives the possibility to use the non-simply-connectedness of $\bar{M} - M'$ (Lemma 3).

## ACKNOWLEDGMENTS

We acknowledge fruitful discussions with W. Greub, J. Hennig, F. Pasemann, and J. Tolar. One of us (H.D.D.) would like to thank Professor Abdus Salam, the Internaional Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste.

## APPENDIX

1. *Proof of Lemma* 1: Put ($x \in g$)

$$\Gamma(\theta(x),t) = \{\sigma \in \mathrm{Sec}_0(\xi_\rho) | \mathrm{supp}\sigma \subset D(\theta(x),t)\}.$$

Define $U_t(\phi_\rho,\theta)(x)$: $\Gamma(\theta(x),-t) \to \Gamma(\theta(x),t)$ by $[\sigma \in \Gamma(\theta(x),-t)]$,

$$(U_t(\phi_\rho,\theta)(x)\sigma)(m) = \begin{cases} F(\phi_\rho(x))(\sigma(F(\theta(x))(m,t)), -t) & \text{if } m \in D(\theta(x),t), \\ 0_m, & \text{otherwise,} \end{cases}$$

which is bijective and, moreover, isometric since

(i) $M - D(\theta(x),t)$ is of measure zero;

(ii) $L_{\theta(x)}\omega = 0$;

(iii) for $m \in D(\theta(x),t)$ the association

$$q \in \pi_\rho^{-1}(m) \to F(\phi_\rho(x))(q,t) \in \pi_\rho^{-1}(F(\theta(x))(m,t))$$ is unitary.

$\Gamma(\theta(x),t)$ is dense in $\mathrm{Sec}_0(\xi_\rho)$, so $U_t(\phi_\rho,\theta)(x)$ extends to a unitary operator on $L^2(\xi_\rho,\omega)$ and

$$t \in \mathbb{R} \to U_t(\phi_\rho,\theta)(x) \in \mathfrak{U}(\mathscr{H})$$

is a strongly continuous one-parameter unitary group. By Stone's theorem, there is a (unique) skew-adjoint operator $A(x)$ on $\vartheta(A(x)) \subset L^2(\xi_\rho,\omega)$, $\vartheta(A(x)) \supset \mathrm{Sec}_0(\xi_\rho)$, such that

$$A(x)|\mathrm{Sec}_0(\xi_\rho) = D(\phi_\rho,\theta)(x).$$

Thus $D(\phi_\rho,\theta)(x)$ is skew-symmetric on $\mathrm{Sec}_0(\xi_\rho)$. It is even essentially skew-adjoint on $\mathrm{Sec}_0(\xi_\rho) \subset L^2(\xi_\rho,\omega)$, which can be shown directly, using the method indicated in Ref. 9. By definition, $\mathrm{Sec}_0(\xi_\rho)$ is a common invariant domain for all $D(\phi_\rho,\theta)(x)$, $x \in g$.

2. *Proof of Theorem* 1: Consider the sets ($x_i \in g$)

$$D(\theta;x_1,\ldots,x_k) = \{m \in M | \varphi_1^{\theta(x_k)} \cdots \varphi_1^{\theta(x_1)}(m) \text{ exists}\}$$

and

$$\Gamma(\theta;x_1,\ldots,x_k) = \{\sigma \in \mathrm{Sec}_0(\xi_\rho) | \mathrm{supp}\sigma \subset D(\theta;x_1,\ldots,x_k)\}.$$

Since $D(\theta(x),t)$ is open in $M$, the same holds for $D(\theta;x_1,\ldots,x_k)$. By the quasicompleteness of $\theta$.

$$E(\theta;x_1,\ldots,x_k) := M - D(\theta;x_1,\ldots,x_k)$$

is a set of measure zero. So $\Gamma(\theta;x_1,\ldots,x_k)$ is dense in $\mathrm{Sec}_0(\xi_\rho)$. Because of Lemma 2 we have to prove that

(i) $\mathrm{Im}\delta(D(\phi_\rho,\theta),G) \subset A(\mathfrak{U}(L^2(\xi_\rho,\omega)))$

and

(ii) $\mathrm{Im}\delta(\phi_\rho,G,q) \subset A(E_\rho,q)$ for $q \in E_\rho$

are equivalent. Now condition (i) is equivalent to

$$(*)\ \mathrm{Exp}D(\phi_\rho,\theta)(x_1)\cdots\mathrm{Exp}D(\phi_\rho,\theta)(x_k)\sigma = \sigma$$

for $(x_1,\ldots,x_k) \in \hat{A}(G)$, $\sigma \in \Gamma(\theta;x_1,\ldots,x_k)$. Stone's theorem gives (compare the discussion in the proof of Lemma 1)

$$(\mathrm{Exp}D(\phi_\rho,\theta)(x)\sigma)(m) = (U_1(\phi_\rho,\theta)(x)\sigma)(m)$$
$$= F(\phi_\rho(x))(\sigma(F(\theta(x))(m,1)), -1)$$

for $x \in g$, $\sigma \in \Gamma(\theta;x)$, $m \in D(\theta;x)$. Hence (*) is equivalent to

$$\delta(\phi_\rho,G,\sigma(\delta(\theta,G,\pi_\rho(q))(x_1,\ldots,x_k)(1)))(-x_k,\ldots,-x_1)(1) = \sigma(\pi_\rho(q))$$

for

$$q \in D(\phi_\rho;x_1,\ldots,x_k),\ \sigma \in \Gamma(\theta;x_1,\ldots,x_k),\ (x_1,\ldots,x_k) \in \hat{A}(G).$$

Because $D(\phi_\rho;x_1,\ldots,x_k)$ is open in $E_\rho$, (*) is equivalent to

$$\delta(\phi_\rho,G,q)(x_1,\ldots,x_k)(1) = q$$

for $(x_1,\ldots,x_k) \in \hat{A}(G)$, $q \in D(\phi_\rho;x_1,\ldots,x_k)$. But this is equivalent to condition (ii).

[1] C.M. Misner and J.A. Wheeler, Ann. Phys. (N.Y.) **2**, 525–603 (1957).
[2] L.L. Henry, J. Math. Phys. **18**, 662 (1977).
[3] M.E. Mayer, Lecture Notes Math. **570**, 307–49 (1977).
[4] M.F. Atiyah and R.S. Ward, Commun. Math. Phys. **55**, 117–24 (1977); R.S. Ward, Phys. Lett. **61**, 81 (1977); M.F. Atiyah, N.J. Hitchin, and I.M. Singer, Proc. Natl. Acad. Sci. USA **74**, 2662–63 (1977).
[5] L. O'Raifeartaigh (preprint).
[6] H.-R. Petry, in Proceedings of the Informal Meeting on Differential Geometric Methods in Physics, Clausthal, 1977, pp. 161–70.
[7] W. Greub *et al.*, *Connections, Curvature and Cohomology* (Academic, New York, 1973).
[8] C.J.S. Clarke and B.G. Schmidt, Gen. Rel. Grav. **8**, 129–37 (1977).
[9] E. Nelson, Ann. Math. **70**, 572 (1959).
[10] R.S. Palais, Mem. Am. Math. Soc. **22** (1957).
[11] S. Lang, *Introduction to Differentiable Manifolds* (Interscience, New York, 1962).
[12] G.W. Mackey, Bull. Am. Math. Soc. **69**, 628 (1963).
[13] D.J. Simms, Lecture Notes Math. **52**, 1–90 (1968).
[14] M. Flato *et al.*, Ann. Sci. ENS **5** (3) 423–34 (1972).
[15] J. Simon, Commun. Math. Phys. **28**, 39–46 (1972).
[16] H.D. Doebner and J.-E. Werth, ICTP, Trieste, preprint IC/76/25 (1976).
[17] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I* (Academic, New York, 1972).
[18] H.D. Doebner and J. Tolar, J. Math. Phys. **16**, 975–84 (1975).
[19] C. Godbillon, *Topologie Algébrique* (Hermann, Paris, 1971).

# SU(4) Clebsch–Gordan coefficients for the formation of baryonium and exotic baryons[a)]

Ronald Anderson[b)] and G. C. Joshi

*School of Physics, University of Melbourne, Parkville, Victoria, Australia 3052*
(Received 5 June 1978)

We give the SU(4) Clebsch–Gordan coefficients required for the formation of exotic hadronic states of the form $(QQ\bar{Q}\bar{Q})$ and $(QQ\bar{Q}QQ)$.

## I. INTRODUCTION

With the recent experimental evidence for narrow resonances in baryon ($B$)–antibaryon ($\bar{B}$) scattering, interest has been renewed in the "exotic" multiquark hadrons of four and five quarks.[1,2] The most studied of these states is the four quark state popularly known as Baryonium.[3-9] The structures suggested for such hadrons are states composed of combinations of diquarks.[10,11] Two quarks with relatively small angular momentum form a diquark ($D$) which combines with other diquarks or quarks. For Baryonium the structure is that of a diquark–antidiquark pair, while for the five quark state ($B_5$), diquark–antiquark–diquark. The diquarks can be of two types depending on whether they lie in the $\bar{3}$ or 6 representations of the SU(3) color group.[12] The series of Baryonium states constructed from the first type of diquark have small decay widths in mesonic channels but normal hadronic decays into $B\bar{B}$ channels. Similarly the corresponding $B_5$ series of states show a reluctance for decaying into mesons but couple strongly to $B\bar{B}B$ (or $B$-Baryonium) channels. Figure 1 illustrates a string picture realization of a quark–gluon theory for such states constructed from the diquarks in the $\bar{3}$ representation.[4,5,7] The second series of states constructed from the diquarks in the color 6 representation are reluctant to decay into either mesonic or $B\bar{B}$ channels. Whenever possible however, these states will cascade via pion emission into a resonance of the same type (color 6).

In this paper we present the Clebsch–Gordan coefficients necessary for the decompositon of the products $D\times\bar{D}$ (Baryonium) and $(D\times\bar{q})\times D$, ($B_5$). A knowledge of such decompositions is required for a detailed treatment of mass formulas and the question of identifying particle states with particular representations and "mixtures" between various representations in the above products.[13] The form in which the coefficients have been given is with respect to a decomposition in SU(3). Extensive tables already exist for the decomposition of the products of the 15 and the various 20 representations of SU(4),[14,15] and for a number of the SU(3) coefficients required in our calculations we refer the reader to those contained in the paper of Haacke, Moffat, and Savaria.[14] Following other work in such calculations we use the techniques developed by Biedenharn and Baird,[16] and Louck.[17] In the next section we present the formalism and notation, and in Sec. III our phase conventions.

[a)]Supported in part by the Australian Research Grants Committee.
[b)]Work supported by an Australian Postgraduate Research Award.

## II. FORMALISM AND NOTATION

Taking the quark flavors as forming the first fundamental representation of SU(4), then the diquarks occur in the 6 and 10 representations. The antidiquarks occur in the conjugate representations. The 6 in SU(4) is self-conjugate. The following products give the representations in which the Baryonium states lie:

$$6\times 6 = 1 + 15 + 20'', \qquad (2.1)$$

$$6\times \overline{10} = 15 + \overline{45}, \qquad (2.2)$$

$$10\times 6 = 15 + 45, \qquad (2.3)$$

$$10\times \overline{10} = 1 + 15 + 84. \qquad (2.4)$$

The diquark–antiquark products necessary for the formation of the $B_5$ states are given by:

$$6\times\bar{4} = 4 + \overline{20}', \qquad (2.5)$$

$$10\times\bar{4} = 4 + 36. \qquad (2.6)$$

Hence for $B_5$ we have:

$$4\times 6 = \bar{4} + 20', \qquad (2.7)$$

$$4\times 10 = 20' + 20, \qquad (2.8)$$

$$\overline{20}'\times 6 = \bar{4} + 20' + \overline{36} + \overline{60}, \qquad (2.9)$$

$$\overline{20}'\times 10 = \bar{4} + 20' + \overline{36} + 140'', \qquad (2.10)$$

$$36\times 6 = 20' + 20 + \overline{36} + 140'', \qquad (2.11)$$

$$36\times 10 = 20' + 20 + 60 + 120 + 140''. \qquad (2.12)$$

Table I gives the SU(3) decompositions for these SU(4) representations. The basis scheme of Gelfand and Zetlin[16-18] allows an economical way of expressing the states within a representation and in such a scheme the computation involved in obtaining the coefficients is greatly simplified. The SU(4) states in this basis are represented by a triangular array known as the "Gelfand pattern,"

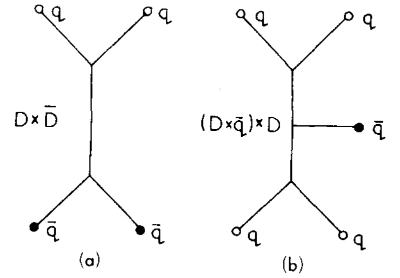


FIG. 1. (a) The string pictures of Baryonium, and (b) $B_5$.

 0022-2488/79/061015-15$01.00  

$$\begin{array}{ccccccc} m_{14} & & m_{24} & & m_{34} & & m_{44} \\ & m_{13} & & m_{23} & & m_{33} & \\ & & m_{12} & & m_{22} & & \\ & & & m_{11} & & & \end{array}$$

The set of integers $(m_{14},m_{24},m_{34})$ describe the Young's tableau associated with the representation. The integer $m_{i4}$ is the number of boxes in the $i$th row in the tableau. For SU(4), $m_{44}$ is zero. The remaining integers enumerate the states in a given representation. They are subject to the constraint

$$m_{i,j+1} \geqslant m_{i,j} \geqslant m_{i+1,j+1}. \tag{2.13}$$

In terms of the quantum numbers, the states within the representations will be specified by

$$|(R,Z),r,Y,I,I_z\rangle, \tag{2.14}$$

where $R$ labels the SU(4) representation, $r$ the SU(3) subgroup of SU(4), and $Y$, $I$ and $I_z$ the hypercharge, isospin, and the third component of isospin, respectively. $Z$ is related to the charm quantum number via

$$C = aZ + bN. \tag{2.15}$$

$N$ is the baryon number. The constants $a$ and $b$ depend on the quark model used. For the fractionally charged GIM model,[19] $a = -1$ and $b = \frac{3}{4}$. The values of $R$, $r$, and $2I+1$ are related to the integers $m_{ij}$ by the Weyl dimensionality formula for $n = 4$, 3, and 2, respectively,

$$D(n) = \prod_{i<j}^{n} \frac{(m_{i,n} - m_{j,n} + j - i)}{1!2!\cdots(n-1)!}. \tag{2.16}$$

The three conserved quantum numbers $Z$, $Y$, and $I_z$ are given by $Q(n)$ for $n = 4,3$, and 2, respectively, where

$$Q(n) = \sum_{i=1}^{n-1} m_{i,n-1} - \frac{(n-1)}{n} \sum_{i=1}^{n} m_{i,n}. \tag{2.17}$$

The Clebsch–Gordan coefficients are defined by the transformation law,

$$\begin{aligned} |(R,Z),r,Y,I,I_z\rangle &= \sum_{r_1\cdots I_{2z}} \left[\begin{array}{cc|c} R_1 & R_2 & R \\ r_1,Z_1; & r_2,Z_2 & r,Z \end{array}\right] \left[\begin{array}{cc|c} r_1 & r_2 & r \\ I_1,Y_1; & I_2,Y_2 & I,Y \end{array}\right] \\ &\quad \times C^{I,I_1,I_2}_{I_z,I_{1z},I_{2z}} |(R_1,Z_1),r_1,Y_1,I_1,I_{1z}\rangle |(R_2,Z_2),r_2,Y_2,I_2,I_{2z}\rangle, \end{aligned} \tag{2.18}$$

with the following constraints;

$$\begin{aligned} &Z = Z_1 + Z_2, \quad Y = Y_1 + Y_2, \quad I_z = I_{1z} + I_{2z}, \\ &I = |I_1 - I_2| \cdots I_1 + I_2, \end{aligned} \tag{2.19}$$

where

$\left[\begin{array}{cc|c} R_1 & R_2 & R \\ r_1,Z_1; & r_2,Z_2 & r,Z \end{array}\right]$ is the SU(3) singlet factor,

$\left[\begin{array}{cc|c} r_1 & r_2 & r \\ I_1,Y_1; & I_2,Y_2 & I,Y \end{array}\right]$ is the SU(2) singlet factor,

$C^{I,I_1,I_2}_{I_z,I_{1z},I_{2z}}$ is the SU(2) Clebsch–Gordan coefficient.

The singlet factors are computed by considering the matrix elements of the infinitestimal generator $E_{ij}$ of SU(4). These satisfy the communtation relations

$$[E_{i,j},E_{k,l}] = \delta_{j,k}E_{i,l} - \delta_{i,k}E_{k,j}. \tag{2.20}$$

The generators, $E_{ij}$ $(i,j = 1,\ldots,4)$ define the raising operators for the six subgroups:

$$\begin{aligned} &E_{12} = I_+, \quad E_{14} = K_+, \\ &E_{13} = V_+, \quad E_{24} = L_+, \\ &E_{23} = U_+, \quad E_{34} = M_+, \end{aligned} \tag{2.21}$$

$K$, $L$, $M$ being the operators for three SU(2) subgroups of SU(4), while $I$, $V$, and $U$ are the familiar $I$-spin, $U$-spin, and $V$-spin subgroups of SU(3).

For the matrix elements of the operators in the Gelfand–Zetlin basis we have used the expression given by Louck in Ref. 17 [Eq. (2.62)]. Tables II together with the phase factors in Tables III (defined below) contain the SU(3) singlet factors required for the formation of the diquark states and the decompositions in Eqs. (2.1)–(2.12). Tables IV contain the SU(2) singlet factors associated with these decompositions. The remaining SU(2) singlet factors which are required may be found in Ref. 14.

## III. PHASE CONVENTIONS

Within a given multiplet in SU(4) or SU(3) the relative phases are determined by the signs of the operators $E_{i,i+1}$. In the convention we have followed, these operators have positive matrix elements for $i = 1,2$, and 3. A further condition is required to determine the relative phase between different representations in a decomposition of a product. For this the highest Clebsch–Gordan singlet factor has been chosen to be positive. For a given $R$ the highest SU(3) singlet factor is the one with the highest $Z$, then $r$ in which $r_1$ is a maximum in $r_1 \times r_2$. If this is insufficient, then the ordering is on $Z_1$, then $r_2$. Following Haacke *et al.*[14] the highest of two SU(3) multiplets is the one with the highest $Y$ value. For two multiplets with the same highest $Y$ value, then the highest is the one with the highest $I$ value at this $Y$. Similarly the highest SU(2) singlet factor for a given $r$ is the factor with the highest $Y$ in which $I_1$ is a maximum in $I_1 \times I_2$. If this is insufficient, then the ordering is on $Y_1$, and $I_2$.

The phase factors, $\epsilon_1$, $\epsilon_3$ given in Table V contain the symmetry properties of the SU(2) singlet factors, where

$$\left[\begin{array}{cc|c} r_1 & r_2 & r \\ I_1,Y_1; & I_2,Y_2 & I,Y \end{array}\right] = \epsilon_1(-1)^{I-I_1-I_2} \left[\begin{array}{cc|c} r_2 & r_1 & r \\ I_2,Y_2; & I_1,Y_2 & I,Y \end{array}\right], \tag{3.1}$$

$$\left[\begin{array}{cc|c} r_1 & r_2 & r \\ I_1,Y_1; & I_2,Y_2 & I,Y \end{array}\right] = \epsilon_3(-1)^{I-I_1-I_2} \left[\begin{array}{cc|c} \bar{r}_1 & \bar{r}_2 & \bar{r} \\ I_1,-Y_1; & I_2-Y_2 & I,-Y \end{array}\right]. \tag{3.2}$$

The factors $\eta_1$, $\eta_3$ associated with the SU(3) singlet factors are given in Table III, where

$$\left[\begin{array}{cc|c} R_1 & R_2 & R \\ r_1,Z_1; & r_2,Z_2 & r,Z \end{array}\right] = \eta_1\epsilon_1 \left[\begin{array}{cc|c} R_2 & R_1 & R \\ r_2,Z_2; & r_1,Z_1 & r,Z \end{array}\right], \tag{3.3}$$

$$\left[\begin{array}{cc|c} R_1 & R_2 & R \\ r_1,Z_1; & r_2,Z_2 & r,Z \end{array}\right] = \eta_3\epsilon_3 \left[\begin{array}{cc|c} \bar{R} & \bar{R}_2 & \bar{R} \\ \bar{r}_1,-Z_1; & \bar{r}_2,-Z_2 & \bar{r},-Z \end{array}\right]. \tag{3.4}$$

TABLE I. SU(3) decompositions ($r$) of SU(4). The $m_{i,4}$, $i = 1,\ldots,3$ associated with the SU(4) representations are also given.

| SU(4) Rep. | $r$ | $Z$ |
|---|---|---|
| 4[1,0,0] | $3$ | $\frac{1}{4}$ |
| | $1$ | $-\frac{3}{4}$ |
| 6[1,1,0] | $\bar{3}$ | $\frac{1}{2}$ |
| | $3$ | $-\frac{1}{2}$ |
| 10[2,0,0] | $6$ | $\frac{1}{2}$ |
| | $3$ | $-\frac{1}{2}$ |
| | $1$ | $-\frac{3}{2}$ |
| 15[2,1,1] | $3$ | $1$ |
| | $\underline{1} + 8$ | $0$ |
| | $\bar{3}$ | $-1$ |
| 20[3,0,0] | $10$ | $\frac{3}{4}$ |
| | $6$ | $-\frac{1}{4}$ |
| | $3$ | $-\frac{3}{4}$ |
| | $1$ | $-\frac{9}{4}$ |
| 20′[2,1,0] | $\underline{8}$ | $\frac{3}{4}$ |
| | $\bar{3} + 6$ | $-\frac{1}{4}$ |
| | $3$ | $-\frac{5}{4}$ |
| 20″[2,2,0] | $\bar{6}$ | $1$ |
| | $8$ | $0$ |
| | $6$ | $-1$ |
| 36[3,1,1] | $6$ | $\frac{5}{4}$ |
| | $3 + 15$ | $\frac{1}{4}$ |
| | $\underline{1} + 8$ | $-\frac{3}{4}$ |
| | $\bar{3}$ | $-\frac{7}{4}$ |
| 45[3,1,0] | $15$ | $1$ |
| | $\underline{8} + 10$ | $0$ |
| | $\bar{3} + 6$ | $-1$ |
| | $3$ | $-2$ |
| 60[3,2,0] | $\overline{15}$ | $\frac{5}{4}$ |
| | $\bar{6} + 15$ | $\frac{1}{4}$ |
| | $8 + 10$ | $-\frac{3}{4}$ |
| | $6$ | $-\frac{7}{4}$ |
| 84[4,2,2] | $6$ | $2$ |
| | $3 + 15$ | $1$ |
| | $\underline{1} + \underline{8} + 27$ | $0$ |
| | $\bar{3} + \overline{15}$ | $-1$ |
| | $\bar{6}$ | $-2$ |
| 120[5,1,1] | $15'$ | $\frac{7}{4}$ |
| | $10 + 35$ | $\frac{3}{4}$ |
| | $6 + 24$ | $-\frac{1}{4}$ |
| | $3 + 15$ | $-\frac{5}{4}$ |
| | $1 + 8$ | $-\frac{9}{4}$ |
| | $\bar{3}$ | $-\frac{13}{4}$ |
| 140″[4,2,1] | $15$ | $\frac{7}{4}$ |
| | $\underline{8} + 10 + \underline{27}$ | $\frac{3}{4}$ |
| | $\bar{3} + 6 + \overline{15}$ $+ 24$ | $-\frac{1}{4}$ |
| | $3 + \bar{6} + 15$ | $-\frac{5}{4}$ |
| | $8$ | $-\frac{9}{4}$ |

TABLE II A. SU(3) singlet factors: $4 \times 4 = 6 + 10$.

$r = 6, \quad Z = \frac{1}{2}$

| $r_1,Z_1;r_2,Z_2$ | 10 |
|---|---|
| $3,\frac{1}{4};3,\frac{1}{4}$ | 1 |

$r = \bar{3}, \quad Z = \frac{1}{2}$

| $r_1,Z_1;r_2,Z_2$ | 6 |
|---|---|
| $3,\frac{1}{4};3,\frac{1}{4}$ | 1 |

$r = 3, \quad Z = -\frac{1}{2}$

| $r_1,Z_1;r_2,Z_2$ | 6 | 10 |
|---|---|---|
| $1,-\frac{3}{4};3,\frac{1}{4}$ | $-1/\sqrt{2}$ | $1/\sqrt{2}$ |
| $3,\frac{1}{4};1,-\frac{3}{4}$ | $1/\sqrt{2}$ | $1/\sqrt{2}$ |

$r = 1, \quad Z = -\frac{3}{2}$

| $r_1,Z_1;r_2,Z_2$ | 10 |
|---|---|
| $1,-\frac{3}{4};1,-\frac{3}{4}$ | 1 |

TABLE II B. SU(3) singlet factors: $4\times 10 = 20' + 20$.

$r=10,\quad Z=\frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20 |
|---|---|
| $3,\frac{1}{4};6,\frac{1}{2}$ | 1 |

$r=8,\quad Z=\frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20′ |
|---|---|
| $3,\frac{1}{4};6,\frac{1}{2}$ | 1 |

$r=6,\quad Z=-\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20′ | 20 |
|---|---|---|
| $1,-\frac{3}{4};6,\frac{1}{2}$ | $-\sqrt{2/3}$ | $1/\sqrt{3}$ |
| $3,\frac{1}{4};3,-\frac{1}{2}$ | $1/\sqrt{3}$ | $\sqrt{2/3}$ |

$r=\bar{3},\quad Z=-\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20′ |
|---|---|
| $3,\frac{1}{4};3,-\frac{1}{2}$ | 1 |

$r=3,\quad Z=-\frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20′ | 20 |
|---|---|---|
| $1,-\frac{3}{4};3,-\frac{1}{2}$ | $-1/\sqrt{3}$ | $\sqrt{2/3}$ |
| $3,\frac{1}{4};1,-\frac{3}{2}$ | $\sqrt{2/3}$ | $1/\sqrt{3}$ |

$r=1,\quad Z=-\frac{9}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20 |
|---|---|
| $1,-\frac{3}{4};1,-\frac{3}{2}$ | 1 |

TABLE II C. SU(3) singlet factors: $6\times 4 = \bar{4} + 20'$.

$r=8,\quad Z=\frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20′ |
|---|---|
| $\bar{3},\frac{1}{2};3,\frac{1}{4}$ | 1 |

$r=1,\quad Z=\frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\bar{4}$ |
|---|---|
| $\bar{3},\frac{1}{2};3,\frac{1}{4}$ | 1 |

$r=6,\quad Z=-\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20′ |
|---|---|
| $3,-\frac{1}{2};3,\frac{1}{4}$ | 1 |

$r=\bar{3},\ Z=-\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\bar{4}$ | 20′ |
|---|---|---|
| $3,-\frac{1}{2};3,\frac{1}{4}$ | $\sqrt{2/3}$ | $1/\sqrt{3}$ |
| $\bar{3},\frac{1}{2};1,-\frac{3}{4}$ | $-1/\sqrt{3}$ | $\sqrt{2/3}$ |

$r=3, Z=-\frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20′ |
|---|---|
| $3,-\frac{1}{2};1,-\frac{3}{4}$ | 1 |

TABLE II D. SU(3) singlet factors: $6\times 6 = 1 + 15 + 20''$.

$r=\bar{6},\quad Z=1$

| $r_1,Z_1;r_2,Z_2$ | 20″ |
|---|---|
| $\bar{3},\frac{1}{2};\bar{3},\frac{1}{2}$ | 1 |

$r=3,\quad Z=1$

| $r_1,Z_1;r_2,Z_2$ | 15 |
|---|---|
| $\bar{3},\frac{1}{2};\bar{3},\frac{1}{2}$ | 1 |

$r=8,\quad Z=0$

| $r_1,Z_1;r_2,Z_2$ | 15 | 20″ |
|---|---|---|
| $3,-\frac{1}{2};\bar{3},\frac{1}{2}$ | $1/\sqrt{2}$ | $1/\sqrt{2}$ |
| $\bar{3},\frac{1}{2};3,-\frac{1}{2}$ | $-1/\sqrt{2}$ | $1/\sqrt{2}$ |

$r=1,\quad Z=0$

| $r_1,Z_1;r_2Z_2$ | 1 | 15 |
|---|---|---|
| $3,-\frac{1}{2};\bar{3},\frac{1}{2}$ | $-1/\sqrt{2}$ | $-1/\sqrt{2}$ |
| $\bar{3},\frac{1}{2};3,-\frac{1}{2}$ | $+1/\sqrt{2}$ | $-1/\sqrt{2}$ |

$r=6,\quad Z=-1$

| $r_1,Z_1;r_2,Z_2$ | 20″ |
|---|---|
| $3,-\frac{1}{2};3,-\frac{1}{2}$ | 1 |

$r=\bar{3},\quad Z=-1$

| $r_1,Z_1;r_2,Z_2$ | 15 |
|---|---|
| $3,-\frac{1}{2};3,-\frac{1}{2}$ | $-1$ |

TABLE II E. SU(3) singlet factors: $6\times 10 = 15 + 45$.

$r=15, \quad Z=1$

| $r_1,Z_1;r_2,Z_2$ | 45 |
|---|---|
| $\bar{3},\frac{1}{2};6,\frac{1}{2}$ | 1 |

$r=3, \quad Z=1$

| $r_1,Z_1;r_2,Z_2$ | 15 |
|---|---|
| $\bar{3},\frac{1}{2};6,\frac{1}{2}$ | 1 |

$r=10, \quad Z=0$

| $r_1,Z_1;r_2,Z_2$ | 45 |
|---|---|
| $3,-\frac{1}{2};6,\frac{1}{2}$ | 1 |

$r=8, \quad Z=0$

| $r_1,Z_1;r_2,Z_2$ | 15 | 45 |
|---|---|---|
| $3,-\frac{1}{2};6,\frac{1}{2}$ | $\sqrt{3}/2$ | $\frac{1}{2}$ |
| $\bar{3},\frac{1}{2};3,-\frac{1}{2}$ | $-1/2$ | $\sqrt{3}/2$ |

$r=1, \quad Z=0$

| $r_1,Z_1;r_2,Z_2$ | 15 |
|---|---|
| $\bar{3},\frac{1}{2};3,-\frac{1}{2}$ | 1 |

$r=6, \quad Z=-1$

| $r_1,Z_1;r_2,Z_2$ | 45 |
|---|---|
| $3,-\frac{1}{2};3,-\frac{1}{2}$ | 1 |

$r=\bar{3}, \quad Z=-1$

| $r_1,Z_1;r_2,Z_2$ | 15 | 45 |
|---|---|---|
| $3,-\frac{1}{2};3,-\frac{1}{2}$ | $1/\sqrt{2}$ | $1/\sqrt{2}$ |
| $\bar{3},\frac{1}{2};1,-\frac{3}{2}$ | $-1/\sqrt{2}$ | $1/\sqrt{2}$ |

$r=3, \quad Z=-2$

| $r_1,Z_1;r_2,Z_2$ | 45 |
|---|---|
| $3,-\frac{1}{2};1,-\frac{3}{2}$ | 1 |

TABLE II F. SU(3) singlet factors: $10\times\bar{4} = 4 + 36$.

$r=6, \quad Z=\frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | 36 |
|---|---|
| $6,\frac{1}{2};1,\frac{3}{4}$ | 1 |

$r=15, \quad Z=\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | 36 |
|---|---|
| $6,\frac{1}{2};\bar{3},-\frac{1}{4}$ | 1 |

$r=3, \quad Z=\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | 4 | 36 |
|---|---|---|
| $3,-\frac{1}{2};1,\frac{3}{4}$ | $-1/\sqrt{5}$ | $2/\sqrt{5}$ |
| $6,\frac{1}{2};\bar{3},-\frac{1}{4}$ | $2/\sqrt{5}$ | $1/\sqrt{5}$ |

$r=8, \quad Z=-\frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | 36 |
|---|---|
| $3,-\frac{1}{2};\bar{3},-\frac{1}{4}$ | 1 |

$r=1, \quad Z=-\frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | 4 | 36 |
|---|---|---|
| $1,-\frac{3}{2};1,\frac{3}{4}$ | $-\sqrt{2/5}$ | $\sqrt{3/5}$ |
| $3,-\frac{1}{2};\bar{3},-\frac{1}{4}$ | $\sqrt{3/5}$ | $\sqrt{2/5}$ |

$r=\bar{3}, \quad Z=-\frac{7}{4}$

| $r_1,Z_1;r_2,Z_2$ | 36 |
|---|---|
| $1,-\frac{3}{2};\bar{3},-\frac{1}{4}$ | 1 |

TABLE II G. SU(3) singlet factors: $10\times\overline{10} = 1 + 15 + 84$.

$r=6,\quad Z=2$

| $r_1,Z_1;r_2,Z_2$ | 84 |
|---|---|
| $6,\frac{1}{2};1,\frac{3}{2}$ | 1 |

$r=15,\quad Z=1$

| $r_1,Z_1;r_2,Z_2$ | 84 |
|---|---|
| $6,\frac{1}{2};\bar{3},\frac{1}{2}$ | 1 |

$r=3,\quad Z=1$

| $r_1,Z_1;r_2,Z_2$ | 15 | 84 |
|---|---|---|
| $3,-\frac{1}{2};1,\frac{3}{2}$ | $-1/\sqrt{3}$ | $\sqrt{2/3}$ |
| $6,\frac{1}{2};\bar{3},\frac{1}{2}$ | $\sqrt{2/3}$ | $1/\sqrt{3}$ |

$r=27,\quad Z=0$

| $r_1,Z_1;r_2,Z_2$ | 84 |
|---|---|
| $6,\frac{1}{2};\bar{6},-\frac{1}{2}$ | 1 |

$r=8,\quad Z=0$

| $r_1,Z_1;r_2,Z_2$ | 15 | 84 |
|---|---|---|
| $3,-\frac{1}{2};\bar{3},\frac{1}{2}$ | $-1/\sqrt{6}$ | $\sqrt{5/6}$ |
| $6,\frac{1}{2};\bar{6},-\frac{1}{2}$ | $\sqrt{5/6}$ | $1/\sqrt{6}$ |

$r=1,\quad Z=0$

| $r_1,Z_1;r_2,Z_2$ | 1 | 15 | 84 |
|---|---|---|---|
| $1,-\frac{3}{2};1,\frac{3}{2}$ | $1/\sqrt{10}$ | $-1/\sqrt{2}$ | $2/\sqrt{10}$ |
| $3,-\frac{1}{2};\bar{3},\frac{1}{2}$ | $-\sqrt{3/10}$ | $1/\sqrt{6}$ | $2\sqrt{2/15}$ |
| $6,\frac{1}{2};\bar{6},-\frac{1}{2}$ | $\sqrt{3/5}$ | $1/\sqrt{3}$ | $1/\sqrt{15}$ |

$r=\overline{15},\quad Z=-1$

| $r_1,Z_1;r_2,Z_2$ | 84 |
|---|---|
| $3,-\frac{1}{2};\bar{6},-\frac{1}{2}$ | 1 |

$r=\bar{3},\quad Z=-1$

| $r_1,Z_1;r_2,Z_2$ | 15 | 84 |
|---|---|---|
| $1,-\frac{3}{2};\bar{3},\frac{1}{2}$ | $-1/\sqrt{3}$ | $\sqrt{2/3}$ |
| $3,-\frac{1}{2};\bar{6},-\frac{1}{2}$ | $\sqrt{2/3}$ | $1/\sqrt{3}$ |

$r=\bar{6},\quad Z=-2$

| $r_1,Z_1;r_2,Z_2$ | 84 |
|---|---|
| $1,-\frac{3}{2};\bar{6},-\frac{1}{2}$ | 1 |

TABLE II H. SU(3) singlet factors: $20' \times 6 = 4 + \overline{20}' + 36 + 60$.

$r = \overline{15}, \quad Z = \frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | 60 |
|---|---|
| $8,\frac{3}{4};\overline{3},\frac{1}{2}$ | 1 |

$r = 6, \quad Z = \frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | 36 |
|---|---|
| $8,\frac{3}{4};\overline{3},\frac{1}{2}$ | 1 |

$r = \overline{3}, \quad Z = \frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{20}'$ |
|---|---|
| $8,\frac{3}{4};\overline{3},\frac{1}{2}$ | 1 |

$r = 15, \quad Z = \frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | 36 | 60 |
|---|---|---|
| $6,-\frac{1}{4};\overline{3},\frac{1}{2}$ | $1/\sqrt{2}$ | $1/\sqrt{2}$ |
| $8,\frac{3}{4};3,-\frac{1}{2}$ | $-1/\sqrt{2}$ | $1/\sqrt{2}$ |

$r = \overline{6}, \quad Z = \frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{20}'$ | 60 |
|---|---|---|
| $\overline{3},-\frac{1}{4};\overline{3},\frac{1}{2}$ | 1/2 | $\sqrt{3}/2$ |
| $8,\frac{3}{4};3,-\frac{1}{2}$ | $\sqrt{3}/2$ | $-\frac{1}{2}$ |

$r = 3, \quad Z = \frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | 4 | $\overline{20}'$ | 36 |
|---|---|---|---|
| $\overline{3},-\frac{1}{4};\overline{3},\frac{1}{2}$ | $-1/\sqrt{15}$ | $-1/\sqrt{3}$ | $-\sqrt{3/5}$ |
| $6,-\frac{1}{4};\overline{3},\frac{1}{2}$ | $-\sqrt{2/5}$ | $1/\sqrt{2}$ | $-1/\sqrt{10}$ |
| $8,\frac{3}{4};3,-\frac{1}{2}$ | $2\sqrt{2/15}$ | $1/\sqrt{6}$ | $-\sqrt{3/10}$ |

$r = 10, \quad Z = -\frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | 60 |
|---|---|
| $6,-\frac{1}{4};3,-\frac{1}{2}$ | 1 |

$r = 8, \quad Z = -\frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{20}'$ | 36 | 60 |
|---|---|---|---|
| $3,-\frac{5}{4};\overline{3},\frac{1}{2}$ | $\sqrt{3}/2\sqrt{2}$ | $\frac{1}{2}$ | $\sqrt{3}/2\sqrt{2}$ |
| $\overline{3},-\frac{1}{4};3,-\frac{1}{2}$ | $-\frac{1}{4}$ | $-\sqrt{3}/2\sqrt{2}$ | $\frac{3}{4}$ |
| $6,-\frac{1}{4};3,-\frac{1}{2}$ | $\frac{3}{4}$ | $-\sqrt{3}/2\sqrt{2}$ | $-\frac{1}{4}$ |

$r = 1, \quad Z = -\frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | 4 | 36 |
|---|---|---|
| $3,-\frac{5}{4};\overline{3},\frac{1}{2}$ | $-\sqrt{3/5}$ | $-\sqrt{2/5}$ |
| $\overline{3},-\frac{1}{4};3,-\frac{1}{2}$ | $\sqrt{2/5}$ | $-\sqrt{3/5}$ |

$r = 6, \quad Z = -\frac{7}{4}$

| $r_1,Z_1;r_2,Z_2$ | 60 |
|---|---|
| $3,-\frac{5}{4};3,-\frac{1}{2}$ | 1 |

$r = \overline{3}, \quad Z = -\frac{7}{4}$

| $r_1,Z_1;r_2,Z_2$ | 36 |
|---|---|
| $3,-\frac{5}{4};3,-\frac{1}{2}$ | $-1$ |

TABLE III. SU(3) singlet factors: $20' \times \overline{10} = 4 + \overline{20}' + 36 + \overline{140}''$.

$r = 8, \quad Z = \frac{9}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{140}''$ |
|---|---|
| $8,\frac{3}{4};1,\frac{3}{2}$ | 1 |

$r = \overline{15}, \quad Z = \frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{140}''$ |
|---|---|
| $8,\frac{3}{4};\bar{3},\frac{1}{2}$ | 1 |

$r = 6, \quad Z = \frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | 36 | $\overline{140}''$ |
|---|---|---|
| $6,-\frac{1}{4};1,\frac{3}{2}$ | $-1/\sqrt{2}$ | $1/\sqrt{2}$ |
| $8,\frac{3}{4};\bar{3},\frac{1}{2}$ | $1/\sqrt{2}$ | $1/\sqrt{2}$ |

$r = \bar{3}, Z = \frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{20}'$ | $\overline{140}''$ |
|---|---|---|
| $\bar{3},-\frac{1}{4};1,\frac{3}{2}$ | $-1/\sqrt{3}$ | $\sqrt{2/3}$ |
| $8,\frac{3}{4};\bar{3},\frac{1}{2}$ | $\sqrt{2/3}$ | $1/\sqrt{3}$ |

$r = 24, \quad Z = \frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{140}''$ |
|---|---|
| $8,\frac{3}{4};\bar{6},-\frac{1}{2}$ | 1 |

$r = 15, \quad Z = \frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | 36 | $\overline{140}''$ |
|---|---|---|
| $6,-\frac{1}{4};\bar{3},\frac{1}{2}$ | $-1/2$ | $\sqrt{3}/2$ |
| $8,\frac{3}{4};\bar{6},-\frac{1}{2}$ | $\sqrt{3}/2$ | $\frac{1}{2}$ |

$r = \bar{6} \quad Z = \frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{20}'$ | $\overline{140}''$ |
|---|---|---|
| $\bar{3},-\frac{1}{4};\bar{3},\frac{1}{2}$ | $-1\sqrt{6}$ | $\sqrt{5/6}$ |
| $8,\frac{3}{4};\bar{6},-\frac{1}{2}$ | $\sqrt{5/6}$ | $1/\sqrt{6}$ |

$r = 3, \quad Z = \frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | 4 | $\overline{20}'$ | 36 | $\overline{140}''$ |
|---|---|---|---|---|
| $3,-\frac{5}{4};1,\frac{3}{2}$ | $\sqrt{2/15}$ | $-1/\sqrt{3}$ | $-1/\sqrt{5}$ | $1/\sqrt{3}$ |
| $\bar{3},-\frac{1}{4};\bar{3},\frac{1}{2}$ | $1/\sqrt{5}$ | 0 | $-\sqrt{3/10}$ | $-1/\sqrt{2}$ |
| $6,-\frac{1}{4};\bar{3},\frac{1}{2}$ | $-\sqrt{2/15}$ | $1/\sqrt{3}$ | $-3/2\sqrt{5}$ | $1/2\sqrt{3}$ |
| $8,\frac{3}{4};\bar{6},-\frac{1}{2}$ | $2\sqrt{2/15}$ | $1/\sqrt{3}$ | $1/2\sqrt{5}$ | $1/2\sqrt{3}$ |

$r = 27, \quad Z = -\frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{140}''$ |
|---|---|
| $6,-\frac{1}{4};\bar{6},-\frac{1}{2}$ | 1 |

$r = \overline{10}, \quad Z = -\frac{3}{4}$

| $r_{12},Z_1;r_2,Z_2$ | $\overline{140}''$ |
|---|---|
| $\bar{3},-\frac{1}{4};\bar{6},-\frac{1}{2}$ | 1 |

$r = 8, \quad Z = -\frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{20}'$ | 36 | $\overline{140}''$ |
|---|---|---|---|
| $3,-\frac{5}{4};\bar{3},\frac{1}{2}$ | $-\frac{1}{2}$ | $-1/2\sqrt{2}$ | $\sqrt{5}/2\sqrt{2}$ |
| $\bar{3},-\frac{1}{4};\bar{6},-\frac{1}{2}$ | $+1/2\sqrt{2}$ | $\frac{3}{4}$ | $\sqrt{5/4}$ |
| $6,-\frac{1}{4};\bar{6},-\frac{1}{2}$ | $\sqrt{5}/2\sqrt{2}$ | $-\sqrt{5/4}$ | $\frac{1}{4}$ |

$r = 1, \quad Z = -\frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | 4 | 36 |
|---|---|---|
| $3,-\frac{5}{4};\bar{3},\frac{1}{2}$ | $1/\sqrt{5}$ | $-2/\sqrt{5}$ |
| $6,-\frac{1}{4};\bar{6},-\frac{1}{2}$ | $-2/\sqrt{5}$ | $-1/\sqrt{5}$ |

$r = \overline{15}, \quad Z = -\frac{7}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{140}''$ |
|---|---|
| $3,-\frac{5}{4};\bar{6},-\frac{1}{2}$ | 1 |

$r = \bar{3}, \quad Z = -\frac{7}{4}$

| $r_1,Z_1;r_2,Z_2$ | 36 |
|---|---|
| $3,-\frac{5}{4};\bar{6},-\frac{1}{2}$ | $-1$ |

TABLE II J. SU(3) singlet factors: $36\times 6 = 20 + 20' + \overline{36} + 140''$.

$r = 15,\quad Z = \frac{7}{4}$

| $r_1,Z_1;r_2,Z_2$ | $140''$ |
|---|---|
| $6,\frac{5}{4};\bar{3},\frac{1}{2}$ | 1 |

$r = 3,\quad Z = \frac{7}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{36}$ |
|---|---|
| $6,\frac{5}{4};\bar{3},\frac{1}{2}$ | 1 |

$r = 27,\quad Z = \frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | $140''$ |
|---|---|
| $15,\frac{1}{4};\bar{3},\frac{1}{2}$ | 1 |

$r = 10,\quad Z = \frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20 | $140''$ |
|---|---|---|
| $6,\frac{5}{4};3,-\frac{1}{2}$ | $1/\sqrt{3}$ | $\sqrt{2/3}$ |
| $15,\frac{1}{4};\bar{3},\frac{1}{2}$ | $\sqrt{2/3}$ | $-1/\sqrt{3}$ |

$r = 8,\quad Z = \frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | $20'$ | $\overline{36}$ | $140''$ |
|---|---|---|---|
| $3,\frac{1}{4};\bar{3},\frac{1}{2}$ | $-\frac{1}{4}$ | $\sqrt{5}/4\sqrt{2}$ | $5/4\sqrt{2}$ |
| $6,\frac{5}{4};3,-\frac{1}{2}$ | $-\sqrt{5}/2\sqrt{3}$ | $\sqrt{3}/2\sqrt{2}$ | $-\sqrt{5}/2\sqrt{6}$ |
| $15,\frac{1}{4};\bar{3},\frac{1}{2}$ | $5/4\sqrt{3}$ | $3\sqrt{5}/4\sqrt{6}$ | $-1/4\sqrt{6}$ |

$r = 1,\quad Z = \frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{36}$ |
|---|---|
| $3,\frac{1}{4},\bar{3},\frac{1}{2}$ | 1 |

$r = 24,\quad Z = -\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | $140''$ |
|---|---|
| $15,\frac{1}{4};3,-\frac{1}{2}$ | 1 |

$r = 6,\quad Z = -\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | $20'$ | 20 | $140''$ |
|---|---|---|---|
| $3,\frac{1}{4};3,-\frac{1}{2}$ | $\frac{1}{6}$ | $\sqrt{5}/3\sqrt{2}$ | $\frac{5}{6}$ |
| $8,-\frac{3}{4};\bar{3},\frac{1}{2}$ | $\sqrt{5}/3\sqrt{2}$ | $\frac{2}{3}$ | $-\sqrt{5}/3\sqrt{2}$ |
| $15,\frac{1}{4};3,-\frac{1}{2}$ | $\frac{5}{6}$ | $-\sqrt{5}/3\sqrt{2}$ | $\frac{1}{6}$ |

$r = \overline{15},\quad Z = -\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{36}$ | $140''$ |
|---|---|---|
| $8,-\frac{3}{4};\bar{3},\frac{1}{2}$ | $\frac{1}{2}$ | $\sqrt{3}/2$ |
| $15,\frac{1}{4};3,-\frac{1}{2}$ | $\sqrt{3}/2$ | $-\frac{1}{2}$ |

$r = \bar{3},\quad Z = -\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | $20'$ | $\overline{36}$ | $140''$ |
|---|---|---|---|
| $1,-\frac{3}{4};\bar{3},\frac{1}{2}$ | $-\frac{1}{3}$ | $1/\sqrt{3}$ | $\sqrt{5}/3$ |
| $3,\frac{1}{4};3,-\frac{1}{2}$ | $-1/\sqrt{3}$ | $\frac{1}{2}$ | $-\sqrt{5}/2\sqrt{3}$ |
| $8,-\frac{3}{4};\bar{3},\frac{1}{2}$ | $\sqrt{5}/3$ | $\sqrt{5}/2\sqrt{3}$ | $-\frac{1}{6}$ |

$r = 15,\quad Z = -\frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | $140''$ |
|---|---|
| $8,-\frac{3}{4};3,-\frac{1}{2}$ | 1 |

$r = \bar{6},\quad Z = -\frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{36}$ | $140''$ |
|---|---|---|
| $\bar{3},-\frac{7}{4};\bar{3},\frac{1}{2}$ | $1/\sqrt{2}$ | $1/\sqrt{2}$ |
| $8,-\frac{3}{4};3,-\frac{1}{2}$ | $1/\sqrt{2}$ | $-1/\sqrt{2}$ |

$r = 3,\quad Z = -\frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | $20'$ | 20 | $140''$ |
|---|---|---|---|
| $1,-\frac{3}{4};3,-\frac{1}{2}$ | $\sqrt{2}/3\sqrt{3}$ | $\sqrt{5}/3\sqrt{3}$ | $2\sqrt{5}/3\sqrt{3}$ |
| $\bar{3},-\frac{7}{4};\bar{3},\frac{1}{2}$ | $-\sqrt{5}/3$ | $-\sqrt{2}/3$ | $\sqrt{2}/3$ |
| $8,-\frac{3}{4};3,-\frac{1}{2}$ | $\sqrt{10}/3\sqrt{3}$ | $-4/3\sqrt{3}$ | $1/3\sqrt{3}$ |

$r = 8,\quad Z = -\frac{9}{4}$

| $r_1,Z_1;r_2,Z_2$ | $140''$ |
|---|---|
| $\bar{3},-\frac{7}{4};3,-\frac{1}{2}$ | 1 |

$r = 1,\quad Z = -\frac{9}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20 |
|---|---|
| $\bar{3},-\frac{7}{4};3,-\frac{1}{2}$ | $-1$ |

TABLE II K. $36\times 10 = 20 + 20' + \overline{60} + 120 + 140''$.

$r = 15', \quad Z = \frac{7}{4}$

| $r_1,Z_1;r_2,Z_2$ | 120 |
|---|---|
| $6,\frac{5}{4};6,\frac{1}{2}$ | 1 |

$r = 15, \quad Z = \frac{7}{4}$

| $r_1,Z_1;r_2,Z_2$ | 140″ |
|---|---|
| $6,\frac{5}{4};6,\frac{1}{2}$ | 1 |

$r = \overline{6}, \quad Z = \frac{7}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{60}$ |
|---|---|
| $6,\frac{5}{4};6,\frac{1}{2}$ | 1 |

$r = 35, \quad Z = \frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | 120 |
|---|---|
| $15,\frac{1}{4};6,\frac{1}{2}$ | 1 |

$r = 27, \quad Z = \frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | 140″ |
|---|---|
| $15,\frac{1}{4};6,\frac{1}{2}$ | 1 |

$r = \overline{10}, \quad Z = \frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{60}$ |
|---|---|
| $15,\frac{1}{4};6,\frac{1}{2}$ | 1 |

$r = 10, \quad Z = \frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20 | 120 | 140″ |
|---|---|---|---|
| $3,\frac{1}{4};6,\frac{1}{2}$ | $-1/\sqrt{21}$ | $\sqrt{15}/2\sqrt{7}$ | $-\sqrt{5}/2\sqrt{3}$ |
| $6,\frac{5}{4};3,-\frac{1}{2}$ | $\sqrt{5/21}$ | $\sqrt{3/7}$ | $1/\sqrt{3}$ |
| $15,\frac{1}{4};6,\frac{1}{2}$ | $\sqrt{5/7}$ | $-1/2\sqrt{7}$ | $-\frac{1}{2}$ |

$r = 8, \quad Z = \frac{3}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20′ | $\overline{60}$ | 140″ |
|---|---|---|---|
| $3,\frac{1}{4};6,\frac{1}{2}$ | $-1/4\sqrt{6}$ | $-\sqrt{15}/4\sqrt{2}$ | $5/4\sqrt{3}$ |
| $6,\frac{5}{4};3,-\frac{1}{2}$ | $\sqrt{5}/2\sqrt{6}$ | $\sqrt{3}/2\sqrt{2}$ | $\sqrt{5}/2\sqrt{3}$ |
| $15,\frac{1}{4};6,\frac{1}{2}$ | $5/4\sqrt{2}$ | $-\sqrt{5}/4\sqrt{2}$ | $-\frac{1}{4}$ |

$r = 24, \quad Z = -\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | 120 | 140″ |
|---|---|---|
| $8,-\frac{3}{4};6,\frac{1}{2}$ | $1/\sqrt{2}$ | $-1/\sqrt{2}$ |
| $15,\frac{1}{4};3,-\frac{1}{2}$ | $1/\sqrt{2}$ | $1/\sqrt{2}$ |

$r = \overline{15}, \quad Z = -\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{60}$ | 140″ |
|---|---|---|
| $8,-\frac{3}{4};6,\frac{1}{2}$ | $-1\sqrt{2}$ | $1/\sqrt{2}$ |
| $15,\frac{1}{4};3,-\frac{1}{2}$ | $1/\sqrt{2}$ | $1/\sqrt{2}$ |

$r = 6, \quad Z = -\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20′ | 20 | $\overline{60}$ | 120 | 140″ |
|---|---|---|---|---|---|
| $1,-\frac{3}{4};6,\frac{1}{2}$ | $1/3\sqrt{6}$ | $-2\sqrt{2}/3\sqrt{21}$ | $\sqrt{5}/3\sqrt{2}$ | $5/3\sqrt{14}$ | $-5/3\sqrt{6}$ |
| $3,\frac{1}{4};3,-\frac{1}{2}$ | $-1/2\sqrt{2}$ | $1/\sqrt{14}$ | $-\sqrt{5}/2\sqrt{6}$ | $5/\sqrt{42}$ | 0 |
| $6,\frac{5}{4};1,-\frac{3}{2}$ | $\sqrt{5}/3\sqrt{2}$ | $2\sqrt{5}/3\sqrt{14}$ | $1/\sqrt{6}$ | $\sqrt{5}/\sqrt{42}$ | $\sqrt{5}/3\sqrt{2}$ |
| $8,-\frac{3}{4};6,\frac{1}{2}$ | $-5/6\sqrt{3}$ | $10/3\sqrt{21}$ | $\sqrt{5/6}$ | $-2/3\sqrt{7}$ | $-1/3\sqrt{3}$ |
| $15,\frac{1}{4};3,-\frac{1}{2}$ | $5/6\sqrt{2}$ | $5/3\sqrt{14}$ | $-\sqrt{5}/2\sqrt{6}$ | $-1/\sqrt{42}$ | $-\sqrt{2/3}$ |

TABLE II K.

$r=\bar{3},\quad Z=-\frac{1}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20′ | 140″ |
|---|---|---|
| $3,\frac{1}{4};3,-\frac{1}{2}$ | $1/\sqrt{6}$ | $\sqrt{5/6}$ |
| $8,-\frac{3}{4};6,\frac{1}{2}$ | $\sqrt{5/6}$ | $-1/\sqrt{6}$ |

$r=15,\quad Z=-\frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | $\overline{60}$ | 120 | 140″ |
|---|---|---|---|
| $\bar{3},-\frac{7}{4};6,\frac{1}{2}$ | $1/\sqrt{3}$ | $1/\sqrt{6}$ | $-1/\sqrt{2}$ |
| $8,-\frac{3}{4};3,-\frac{1}{2}$ | $-1/\sqrt{3}$ | $2/\sqrt{6}$ | 0 |
| $15,\frac{1}{4};1,-\frac{3}{2}$ | $1/\sqrt{3}$ | $1/\sqrt{6}$ | $1/\sqrt{2}$ |

$r=3,\quad Z=-\frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20′ | 20 | 120 | 140″ |
|---|---|---|---|---|
| $1,-\frac{3}{4};3,-\frac{1}{2}$ | $-2/3\sqrt{3}$ | $1/3\sqrt{21}$ | $\sqrt{10/21}$ | $-\sqrt{10}/3\sqrt{3}$ |
| $3,\frac{1}{4};1,-\frac{3}{2}$ | $\frac{1}{3}$ | $4/3\sqrt{7}$ | $\sqrt{5/14}$ | $\sqrt{5}/3\sqrt{2}$ |
| $\bar{3},-\frac{7}{4};6,\frac{1}{2}$ | $-\sqrt{5/3}$ | $2\sqrt{5}/3\sqrt{7}$ | $-1/\sqrt{14}$ | $1/3\sqrt{2}$ |
| $8,-\frac{3}{4};3,-\frac{1}{2}$ | $\sqrt{5}/3\sqrt{3}$ | $4\sqrt{5}/3\sqrt{21}$ | $-\sqrt{2/21}$ | $-2\sqrt{2}/3\sqrt{3}$ |

$r=\bar{6},\quad Z=-\frac{5}{4}$

| $r_1,Z_1;r_2,Z_2$ | 140″ |
|---|---|
| $8,-\frac{3}{4};3,-\frac{1}{2}$ | 1 |

$r=8,\quad Z=-\frac{9}{4}$

| $r_1,Z_1;r_2,Z_2$ | 120 | 140″ |
|---|---|---|
| $\bar{3},-\frac{7}{4};3,-\frac{1}{2}$ | $1/\sqrt{2}$ | $-1/\sqrt{2}$ |
| $8,-\frac{3}{4};1,-\frac{3}{2}$ | $1/\sqrt{2}$ | $1/\sqrt{2}$ |

$r=1,\quad Z=-\frac{9}{4}$

| $r_1,Z_1;r_2,Z_2$ | 20 | 120 |
|---|---|---|
| $1,-\frac{3}{4};1,-\frac{3}{2}$ | $\sqrt{2/7}$ | $\sqrt{5/7}$ |
| $\bar{3},-\frac{7}{4};3,-\frac{1}{2}$ | $\sqrt{5/7}$ | $-\sqrt{2/7}$ |

$r=\bar{3},\quad Z=-\frac{13}{4}$

| $r_1,Z_1;r_2,Z_2$ | 120 |
|---|---|
| $\bar{3},-\frac{7}{4};1,-\frac{3}{2}$ | 1 |

TABLE III. Phase factors associated with the SU(3) singlet factors.

| $R_1$ | $R_2$ | $R$ | $\eta_1$ | $\eta_3$ |
|---|---|---|---|---|
| 4 | 4 | 6 | −1 | 1 |
| | | 10 | 1 | 1 |
| 4 | 10 | 20′ | −1 | 1 |
| | | 20 | 1 | 1 |
| 6 | 6 | 15 | −1 | −1 |
| | | 20″ | 1 | 1 |
| 6 | 4 | $\bar{4}$ | −1 | 1 |
| | | 20′ | 1 | 1 |
| 6 | 10 | 15 | −1 | 1 |
| | | 45 | 1 | 1 |
| 10 | $\bar{4}$ | 4 | −1 | −1 |
| | | 36 | 1 | 1 |
| 10 | $\overline{10}$ | 1 | 1 | 1 |
| | | 15 | −1 | −1 |
| | | 84 | 1 | 1 |
| 20′ | 6 | 4 | −1 | 1 |
| | | $\overline{20}'$ | −1 | 1 |
| | | 36 | −1 | −1 |
| | | 60 | 1 | 1 |
| 20′ | $\overline{10}$ | 4 | 1 | −1 |
| | | $\overline{20}'$ | −1 | −1 |
| | | 36 | −1 | 1 |
| | | $\overline{140}''$ | 1 | 1 |
| 36 | 6 | 20′ | −1 | −1 |
| | | 20 | −1 | 1 |
| | | $\overline{36}$ | −1 | 1 |
| | | 140″ | 1 | 1 |
| 36 | 10 | 20′ | 1 | −1 |
| | | 20 | −1 | −1 |
| | | $\overline{60}$ | 1 | 1 |
| | | 120 | 1 | 1 |
| | | 140″ | −1 | 1 |

TABLE IV A. SU(2) singlet factors: $15 \times 3 = 6 + \overline{15} + 24$.

$I = \frac{3}{2}$, $Y = \frac{5}{2}$

| $I_1, Y_1; I_2, Y_2$ | 24 |
|---|---|
| $1, \frac{4}{3}; \frac{1}{2}, \frac{1}{3}$ | 1 |

$I = \frac{1}{2}$, $Y = \frac{5}{3}$

| $I_1, Y_1; I_2, Y_2$ | $\overline{15}$ |
|---|---|
| $1, \frac{4}{3}; \frac{1}{2}, \frac{1}{3}$ | 1 |

$I = 2$, $Y = \frac{2}{3}$

| $I_1, Y_1; I_2, Y_2$ | 24 |
|---|---|
| $\frac{3}{2}, \frac{1}{3}; \frac{1}{2}, \frac{1}{3}$ | 1 |

$I = 1$, $Y = \frac{2}{3}$

| $I_1, Y_1; I_2, Y_2$ | 6 | $\overline{15}$ | 24 |
|---|---|---|---|
| $\frac{1}{2}, \frac{1}{3}; \frac{1}{2}, \frac{1}{3}$ | $1/\sqrt{15}$ | $-2/3\sqrt{2}$ | $4\sqrt{2}/3\sqrt{5}$ |
| $1, \frac{4}{3}; 0, -\frac{2}{3}$ | $-\sqrt{2/5}$ | $1/\sqrt{3}$ | $2/\sqrt{15}$ |
| $\frac{3}{2}, \frac{1}{3}; \frac{1}{2}, \frac{1}{3}$ | $2\sqrt{2/15}$ | $\frac{2}{3}$ | $1/3\sqrt{5}$ |

$I = 0$, $Y = \frac{2}{3}$

| $I_1, Y_1; I_2, Y_2$ | $\overline{15}$ |
|---|---|
| $\frac{1}{2}, \frac{1}{3}; \frac{1}{2}, \frac{1}{3}$ | 1 |

$I = \frac{3}{2}$, $Y = -\frac{1}{3}$

| $I_1, Y_1; I_2, Y_2$ | $\overline{15}$ | 24 |
|---|---|---|
| $1, \frac{2}{3}; \frac{1}{2} \frac{1}{3}$ | $-1/\sqrt{3}$ | $\sqrt{2/3}$ |
| $\frac{3}{2}, \frac{1}{3}; 0 - \frac{2}{3}$ | $\sqrt{2/3}$ | $1/\sqrt{3}$ |

$I = \frac{1}{2}$, $Y = -\frac{1}{3}$

| $I_1, Y_1; I_2, Y_2$ | 6 | $\overline{15}$ | 24 |
|---|---|---|---|
| $0, -\frac{2}{3}; \frac{1}{2}, \frac{1}{3}$ | $1/\sqrt{10}$ | $-1/\sqrt{2}$ | $\sqrt{2/5}$ |
| $\frac{1}{2}, \frac{1}{3}; 0, -\frac{2}{3}$ | $-\sqrt{3/10}$ | $1/\sqrt{6}$ | $2\sqrt{2/15}$ |
| $1, -\frac{2}{3}; \frac{1}{2}, \frac{1}{3}$ | $\sqrt{3/5}$ | $1/\sqrt{3}$ | $1/\sqrt{15}$ |

$I = 1$, $Y = -\frac{4}{3}$

| $I_1, Y_1; I_2, Y_2$ | $\overline{15}$ | 24 |
|---|---|---|
| $\frac{1}{2}, -\frac{5}{3}; \frac{1}{2}, \frac{1}{3}$ | $-\sqrt{2/3}$ | $\sqrt{1/3}$ |
| $1, -\frac{2}{3}; 0, -\frac{2}{3}$ | $1/\sqrt{3}$ | $\sqrt{2/3}$ |

$I = 0$, $Y = -\frac{4}{3}$

| $I_1, Y_1; I_2, Y_2$ | 6 | 24 |
|---|---|---|
| $0, -\frac{2}{3}; 0, -\frac{2}{3}$ | $-1/\sqrt{5}$ | $2/\sqrt{5}$ |
| $\frac{1}{2}, -\frac{5}{3}; \frac{1}{2}, \frac{1}{3}$ | $2/\sqrt{5}$ | $1/\sqrt{5}$ |

$I = \frac{1}{2}$, $Y = -\frac{7}{3}$

| $I_1, Y_1; I_2, Y_2$ | 24 |
|---|---|
| $\frac{1}{2}, -\frac{5}{3}; 0, -\frac{2}{3}$ | 1 |

TABLE IV B. SU(2) singlet factors: $15\times\bar{3} = 8 + 10 + 27$.

$I=1,\quad Y=2$

| $I_1,Y_1;I_2,Y_2$ | 27 |
|---|---|
| $1,\frac{4}{3};0,\frac{2}{3}$ | 1 |

$I=\frac{3}{2},\quad Y=1$

| $I_1,Y_1;I_2,Y_2$ | 10 | 27 |
|---|---|---|
| $1,\frac{4}{3};\frac{1}{2},-\frac{1}{3}$ | $-1/\sqrt{2}$ | $1/\sqrt{2}$ |
| $\frac{3}{2},\frac{1}{3};0,\frac{2}{3}$ | $1/\sqrt{2}$ | $1/\sqrt{2}$ |

$I=\frac{1}{2},\quad Y=1$

| $I_1,Y_1;I_2,Y_2$ | 8 | 27 |
|---|---|---|
| $\frac{1}{2},\frac{1}{3};0,\frac{2}{3}$ | $1/\sqrt{5}$ | $2/\sqrt{5}$ |
| $1,\frac{4}{3};\frac{1}{2},-\frac{1}{3}$ | $2/\sqrt{5}$ | $-1/\sqrt{5}$ |

$I=2,\quad Y=0$

| $I_1,Y_1;I_2,Y_2$ | 27 |
|---|---|
| $\frac{3}{2},\frac{1}{3};\frac{1}{2}-\frac{1}{3}$ | 1 |

$I=1,\quad Y=0$

| $I_1,Y_1;I_2,Y_2$ | 8 | 10 | 27 |
|---|---|---|---|
| $\frac{1}{2},\frac{1}{3};\frac{1}{2}-\frac{1}{3}$ | $-1/3\sqrt{5}$ | $-\frac{2}{3}$ | $2\sqrt{2/15}$ |
| $1,-\frac{2}{3};0,\frac{2}{3}$ | $2/\sqrt{15}$ | $1/\sqrt{3}$ | $\sqrt{2/5}$ |
| $\frac{3}{2},\frac{1}{3};\frac{1}{2},-\frac{1}{3}$ | $4\sqrt{2}/3\sqrt{5}$ | $-\sqrt{2}/3$ | $-1/\sqrt{15}$ |

$I=0,\quad Y=0$

| $I_1,Y_1;I_2,Y_2$ | 8 | 27 |
|---|---|---|
| $0,-\frac{2}{3};0,\frac{2}{3}$ | $\sqrt{2/5}$ | $\sqrt{3/5}$ |
| $\frac{1}{2},\frac{1}{3};\frac{1}{2},-\frac{1}{3}$ | $\sqrt{3/5}$ | $-\sqrt{2/5}$ |

$I=\frac{3}{2},\quad Y=-1$

| $I_1,Y_1;I_2,Y_2$ | 27 |
|---|---|
| $1,-\frac{2}{3};\frac{1}{2},-\frac{1}{3}$ | 1 |

$I=\frac{1}{2},\quad Y=-1$

| $I_1,Y_1;I_2,Y_2$ | 8 | 10 | 27 |
|---|---|---|---|
| $0,-\frac{2}{3};\frac{1}{2}-\frac{1}{3}$ | $-1/\sqrt{15}$ | $-1/\sqrt{3}$ | $\sqrt{3/5}$ |
| $\frac{1}{2},-\frac{5}{3};0,\frac{2}{3}$ | $2\sqrt{2/15}$ | $1/\sqrt{6}$ | $\sqrt{3/10}$ |
| $1,-\frac{2}{3};\frac{1}{2},-\frac{1}{3}$ | $\sqrt{2/5}$ | $-1/\sqrt{2}$ | $-1/\sqrt{10}$ |

$I=1,\quad Y=-2$

| $I_1,Y_1;I_2,Y_2$ | 27 |
|---|---|
| $\frac{1}{2}-\frac{5}{3};\frac{1}{2},-\frac{1}{3}$ | 1 |

$I=0\quad Y=-2$

| $I_1,Y_1;I_2,Y_2$ | 10 |
|---|---|
| $\frac{1}{2},-\frac{5}{3};\frac{1}{2},-\frac{1}{3}$ | $-1$ |

TABLE IV C. SU(2) singlet factors: $15\times 6 = 8 + 10 + \overline{10} + 27 + 35$.

$I=2,\quad Y=2$

| $I_1,Y_1;I_2,Y_2$ | 35 |
|---|---|
| $1,\frac{4}{3};1,\frac{2}{3}$ | 1 |

$I=1,\quad Y=2$

| $I_1,Y_1;I_2,Y_2$ | 27 |
|---|---|
| $1,\frac{4}{3};1,\frac{2}{3}$ | 1 |

$I=0,\quad Y=2$

| $I_1,Y_1;I_2,Y_2$ | $\overline{10}$ |
|---|---|
| $1,\frac{4}{3};1,\frac{2}{3}$ | 1 |

$I=\frac{5}{2},\quad Y=1$

| $I_1,Y_1;I_2,Y_2$ | 35 |
|---|---|
| $\frac{3}{2},\frac{1}{3};1,\frac{2}{3}$ | 1 |

$I=\frac{3}{2},\quad Y=1$

| $I_1,Y_1;I_2,Y_2$ | 10 | 27 | 35 |
|---|---|---|---|
| $\frac{1}{2},\frac{1}{3};1,\frac{2}{3}$ | $\frac{1}{3}$ | $-1/\sqrt{3}$ | $\sqrt{5}/3$ |
| $1,\frac{4}{3};\frac{1}{2},-\frac{1}{3}$ | $-1/\sqrt{3}$ | $\frac{1}{2}$ | $\sqrt{5}/2\sqrt{3}$ |
| $\frac{3}{2},\frac{1}{3};1,\frac{2}{3}$ | $\sqrt{5}/3$ | $\sqrt{5}/2\sqrt{3}$ | $\frac{1}{6}$ |

$I=\frac{1}{2},\quad Y=1$

| $I_1,Y_1;I_2,Y_2$ | 8 | $\overline{10}$ | 27 |
|---|---|---|---|
| $\frac{1}{2},\frac{1}{3};1,\frac{2}{3}$ | $1/3\sqrt{5}$ | $-\frac{2}{3}$ | $4/\sqrt{30}$ |
| $1,\frac{4}{3};\frac{1}{2},-\frac{1}{3}$ | $-2/\sqrt{15}$ | $1/\sqrt{3}$ | $\sqrt{2/5}$ |
| $\frac{3}{2},\frac{1}{3};1,\frac{2}{3}$ | $4\sqrt{2}/3\sqrt{5}$ | $\sqrt{2}/3$ | $1/\sqrt{15}$ |

$I=2,\quad Y=0$

| $I_1,Y_1;I_2,Y_2$ | 27 | 35 |
|---|---|---|
| $1,-\frac{2}{3};1,\frac{2}{3}$ | $-1/\sqrt{2}$ | $1/\sqrt{2}$ |
| $\frac{3}{2},\frac{1}{3};\frac{1}{2},-\frac{1}{3}$ | $1/\sqrt{2}$ | $1/\sqrt{2}$ |

$I=1,\quad Y=0$

| $I_1,Y_1;I_2,Y_2$ | 8 | 10 | $\overline{10}$ | 27 | 35 |
|---|---|---|---|---|---|
| $0,-\frac{2}{3};1,\frac{2}{3}$ | $-\sqrt{2}/3\sqrt{5}$ | $\frac{1}{3}$ | $\sqrt{2}/3$ | $-\sqrt{2/5}$ | $\sqrt{2}/3$ |
| $\frac{1}{2},\frac{1}{3};\frac{1}{2},-\frac{1}{3}$ | $\sqrt{5}/3\sqrt{3}$ | $-\sqrt{2}/3\sqrt{3}$ | $-2/3\sqrt{3}$ | 0 | $4/3\sqrt{3}$ |
| $1,-\frac{2}{3};1,\frac{2}{3}$ | $-2\sqrt{2}/3\sqrt{5}$ | $\frac{2}{3}$ | $-\sqrt{2}/3$ | $1/\sqrt{10}$ | $1/3\sqrt{2}$ |
| $1,\frac{4}{3};0,-\frac{4}{3}$ | $-4/3\sqrt{5}$ | $-\sqrt{2}/3$ | $\frac{1}{3}$ | $1/\sqrt{5}$ | $\frac{1}{3}$ |
| $\frac{3}{2},\frac{1}{3};\frac{1}{2},-\frac{1}{3}$ | $4\sqrt{2}/3\sqrt{15}$ | $2/3\sqrt{3}$ | $2\sqrt{2}/3\sqrt{3}$ | $\sqrt{3/10}$ | $1/3\sqrt{6}$ |

$I=0,\quad Y=0$

| $I_1,Y_1;I_2,Y_2$ | 8 | 27 |
|---|---|---|
| $\frac{1}{2},\frac{1}{3};\frac{1}{2},-\frac{1}{3}$ | $-1/\sqrt{5}$ | $2/\sqrt{5}$ |
| $1,-\frac{2}{3};1,\frac{2}{3}$ | $2/\sqrt{5}$ | $1/\sqrt{5}$ |

$I=\frac{3}{2},\quad Y=-1$

| $I_1,Y_1;I_2,Y_2$ | $\overline{10}$ | 27 | 35 |
|---|---|---|---|
| $\frac{1}{2},-\frac{5}{3};1,\frac{2}{3}$ | $1/\sqrt{3}$ | $-1/\sqrt{2}$ | $1/\sqrt{6}$ |
| $1,-\frac{2}{3};\frac{1}{2},-\frac{1}{3}$ | $-1/\sqrt{3}$ | 0 | $2/\sqrt{6}$ |
| $\frac{3}{2},\frac{1}{3};0,-\frac{4}{3}$ | $1/\sqrt{3}$ | $1/\sqrt{2}$ | $1/\sqrt{6}$ |

TABLE IVC.

$I=\frac{1}{2},\quad Y=-1$

| $I_1, Y_1; I_2, Y_2$ | 8 | 10 | 27 | 35 |
|---|---|---|---|---|
| $0, -\frac{2}{3}; \frac{1}{2}, -\frac{1}{3}$ | $1/\sqrt{5}$ | 0 | $-\sqrt{3/10}$ | $1/\sqrt{2}$ |
| $\frac{1}{2}, -\frac{5}{3}; 1, \frac{2}{3}$ | $-2\sqrt{2/15}$ | $1/\sqrt{3}$ | $-1/\sqrt{20}$ | $1/2\sqrt{3}$ |
| $\frac{1}{2}, \frac{1}{3}; 0, -\frac{4}{3}$ | $-\sqrt{2/15}$ | $-1/\sqrt{3}$ | $1/\sqrt{5}$ | $1/\sqrt{3}$ |
| $1, -\frac{2}{3}; \frac{1}{2}, -\frac{1}{3}$ | $\sqrt{2/15}$ | $1/\sqrt{3}$ | $3/\sqrt{20}$ | $1/2\sqrt{3}$ |

$I=1,\quad Y=-2$

| $I_1, Y_1; I_2, Y_2$ | 27 | 35 |
|---|---|---|
| $\frac{1}{2}, -\frac{5}{3}; \frac{1}{2} -\frac{1}{3}$ | $-1/\sqrt{2}$ | $1/\sqrt{2}$ |
| $1, -\frac{2}{3}; 0, -\frac{4}{3}$ | $1/\sqrt{2}$ | $1/\sqrt{2}$ |

$I=0,\quad Y=-2$

| $I_1, Y_1; I_2, Y_2$ | 10 | 35 |
|---|---|---|
| $0, -\frac{2}{3}; 0, -\frac{4}{3}$ | $-1/\sqrt{3}$ | $\sqrt{2/3}$ |
| $\frac{1}{2}, -\frac{5}{3}; \frac{1}{2}, -\frac{1}{3}$ | $\sqrt{2/3}$ | $1/\sqrt{3}$ |

$I=\frac{1}{2},\quad Y=-3$

| $I_1, Y_1; I_2, Y_2$ | 35 |
|---|---|
| $\frac{1}{2}, -\frac{5}{3}; 0, -\frac{4}{3}$ | 1 |

TABLE V. Phase factors associated with the SU(2) singlet factors.

| $r_1$ | $r_2$ | $r$ | $\epsilon_1$ | $\epsilon_2$ |
|---|---|---|---|---|
| 15 | 3 | 6 | −1 | −1 |
| | | $\overline{15}$ | −1 | 1 |
| | | 24 | 1 | 1 |
| 15 | $\bar{3}$ | 8 | −1 | −1 |
| | | 10 | −1 | 1 |
| | | 27 | 1 | 1 |
| 15 | 6 | 8 | 1 | −1 |
| | | 10 | −1 | −1 |
| | | $\overline{10}$ | 1 | 1 |
| | | 27 | −1 | 1 |
| | | 35 | 1 | 1 |


[1]Proceedings of the 3rd European Symposium on Antinucleon–Nucleon Interactions, Stockholm, July 1976.
[2]L. Montanet, Cern preprint, CERN/EP/Phys 77-22.
[3]F. Myhrer, Cern preprint. Ref. Th 2348-CERN (1977).
[4]Chan Hong-Mo and H. Hogaasen, Phys. Lett. B **72**, 121, 400 (1978); Nucl. Phys. B **136**, 401 (1978).
[5]H.G. Dosch and M.G. Schmidt, Phys. Lett. B **68**, 89 (1977).
[6]G.C. Rossi and G. Veneziano, Nucl. Phys. B **123**, 507 (1977).
[7]G. Veneziano, Cern preprint, Rf. Th. 2425-CERN (1977).
[8]M. Imadin, S. Otsuki, and F. Toyoda, Prog. Theor. Phys. **55**, 551 (1976); **57**, 517 (1977); M. Uehara, Saga University preprint, Saga-77-3.
[9]G.F. Chew, LBL Report 5391 (1976).
[10]C. Rosenzweig, Phys. Rev. Lett. **36**, 697 (1976).
[11]The diquark model for nonexotic baryons has been explored before, see, for example, D.B. Lichtenberg and L.J. Tassie, Phys. Rev. **155**, 1601 (1967); D.B. Lichtenberg, Phys. Rev. **178**, 2197 (1969); M.I. Pavkovic, Phys. Rev. D **13**, 2128 (1976); and R.S. Kaushal and D.S. Kulshreshtha, Ann. Phys. (N.Y.) **108**, 198 (1977).
[12]Y. Nambu and M.-Y Han, Phys. Rev. D **14**, 1459 (1976).
[13]R.L. Jaffe, Phys. Rev. D **15**, 267 (1977); 281 (1977).
[14]E.M. Haacke, J.W. Moffat, and P. Savaria, J. Math. Phys. **17**, 2041 (1976).
[15]V. Rabl, G. Campbell Jr., and K.C. Wali, J. Math. Phys. **16**, 2494 (1975).
[16]L.C. Biedenharn, J. Math. Phys. **4**, 436 (1963); G.E. Baird and L.C. Biedenharn, **4**, 1449 (1963).
[17]J.D. Louck, Am. J. Phys. **38**, 3 (1970).
[18]I.M. Gel'fand, R.A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (Pergamon, Oxford, England, and Macmillan, New York, 1963).
[19]S.L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. D **2**, 1285 (1970).

# Invariant imbedding equations in general geometries: Numerical solution in spherical geometry

S. K. Shenoy[a] and A. J. Mockel

*University of Florida, Department of Nuclear Engineering, Gainesville, Florida 32611*
(Received 28 March 1977)

A method for deriving the invariant imbedding equations in general geometries by the transfer matrix method is developed. Time-independent monoenergetic neutron transport in a nonmultiplying medium is considered. Specifically, invariant imbedding equations are derived for the plane, spherical, and cylindrical geometries. The reflection and transmission functions are evaluated numerically using the discrete ordinates method, for the case of a spherical shell with a perfectly absorbing core. A source of error in the results obtained is pointed out.

## I. INTRODUCTION

The method of invariant imbedding, a method based on the invariance principles introduced by Ambarzumian[1] and Chandrasekhar,[2] is today a general tool of mathematical physics. The method has been applied to monoenergetic neutron transport and a wide range of other processes by Bellman *et al.*[3–5] Polyenergetic neutron transport calculations in a slab have been performed by Mockel.[6,7] Reactor physics calculations using the method have been performed by Shimizu and Aoki.[8]

The nonlinear, initial value equations obtained by the invariant imbedding method have many advantages for digital computer applications in comparison with the conventional Boltzmann equations, which are linear (and hence simpler for analytical purposes) boundary value equations.

Relatively little work has been done concerning the rigorous derivation and numerical solution of invariant imbedding equation in curved geometries. Baily[9] derived the equations for the reflection functions in a rigorous fashion in plane and spherical geometries, from the corresponding Boltzmann formulations. In this work we shall use the transfer-matrix method outlined by Aronson and Yarmush[10] to derive the equations for both the reflection and transmission functions in an arbitrary general geometry. Specifically, plane, spherical, and cylindrical cases will be considered. Steady state, monoenergetic neutron transport in a homogeneous medium is considered, but nonhomogeneous cases can be considered without much difficulty. It is assumed that only neutron absorption and isotropic scattering reactions take place in the medium.

## II. DEFINITIONS

Consider bodies [Fig. 1(a), (b), and (c)] in which the surface is divided into two nonoverlapping complimentary faces, 1 and 2.

A particle is said to be reflected from the body if it emerges from the same face at which it enters. If the particle emerges from the other face, it is said to be transmitted.

[a]Present address: Building 130, Department of Nuclear Energy, Brookhaven National Laboratory, Upton, N.Y. 11973.

Two general types of integral operators which operate on incident neutron currents to provide the emergent currents can be defined; $R_*$, the reflection operator and $T_*$, the transmission operator. Operators which provide currents emergent at face 2 are denoted by positive superscripts and those which provide currents emergent at face 1 are denoted by negative superscripts.

Let $I_1$ and $I_2$ denote the incident neutron currents at face 1 and 2 respectively. Let $I_1'$ and $I_2'$ denote the emergent currents. Assuming a linear system, we can write in operator notation,

$$I_2'(\mu_1) = T_*^+ I_1 + R_*^+ I_2, \tag{1}$$

$$I_1'(\mu_1) = T_*^- I_2 + R_*^- I_1, \tag{2}$$

where

$$T_*^+ I_1 = 2\pi \int_0^1 T_*^+(x,\mu_1;\mu_0) I_1(\mu_0)\, d\mu_0, \tag{3}$$

$$R_*^+ I_2 = 2\pi \int_0^1 R_*^+(x,\mu_1;\mu_0) I_2(\mu_0)\, d\mu_0, \text{ etc.} \tag{4}$$

The transmission function $T_*^+(x,\mu_1;\mu_0)$, which is the kernel of the operator $T_*^+$, is defined as the probability den-

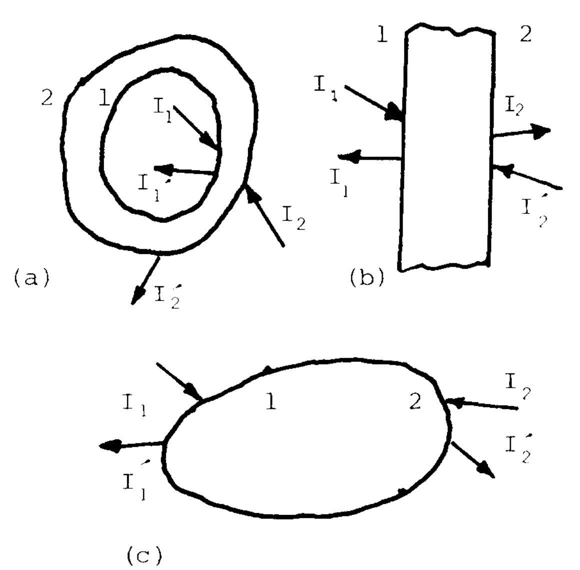


FIG. 1. (a) hollow shell, (b) infinite slab, (c) single surfaced body.

 0022-2488/79/061030-11$01.00  

sity for a neutron impinging on face 1 in the direction $\mu_0$ to be transmitted (i.e., emerge at face 2) in the direction $\mu_1$. Here $x$ is a measure of the size of the medium, e.g., the thickness in the case of a slab, or the outer radius in the case of a spherical or cylindrical shell. $\mu_0$ and $\mu_1$ are the absolute values of the cosines of the angles between the neutron directions and the normals to the surfaces of the medium. For the sake of convenience $T_*^-(x,\mu_1;\mu_0)$ may sometimes be written as $T_*^-(x)$ or $T_*^-(\mu_1;\mu_0)$. Similarly, $R_*^-(x,\mu_1;\mu_0)$ can be defined as the probability density for a neutron impinging on face 1 to be reflected (i.e., reemerge at face 1) in the direction $\mu_1$. In a similar fashion $R_*^+(x,\mu_1;\mu_0)$ and $T_*^- x,\mu_1;\mu_0)$ can be defined. Collectively, the above functions are known as the response functions. Note that for a homogeneous slab $R_*^+ = R_*^-$ and $T_*^+ = T_*^-$. For a medium of zero thickness (in any geometry), $R_*^+ = R_*^- = 0$ and $T_*^+ = T_*^- = I$, where $I$ is the unit matrix.

## III. THE TRANSFER MATRIX

Let us consider a family of concentric similar shells [e.g., Fig. 1(a)] whose inner and outer faces are given respectively by

$$\rho_{\text{in}} = r_1 f(\theta,\phi), \qquad \rho_{\text{out}} = r_2 f(\theta,\phi),$$

where $r$ can be regarded as the radius of the surface and denotes which member of the family of concentric surfaces is under consideration. $\rho$, $\theta$, and $\phi$ are spherical coordinates with some point in the center as origin.

Equations 1 and 2 can be solved to give

$$\begin{bmatrix} I_2' \\ I_2 \end{bmatrix} = H(r_2,r_1)\begin{bmatrix} I_1 \\ I_1' \end{bmatrix}, \tag{5}$$

where

$$H(r_2,r_1) = \begin{bmatrix} T_*^+ - R_*^+ U R_*^- & R_*^+ U \\ & \\ -UR_*^- & U \end{bmatrix}. \tag{6}$$

Here $U = (T_*^-)^{-1}$. The $H$ matrix or the transfer matrix is a $2\times 2$ matrix of operators which operates on the column matrix of currents on one face of the body to give the column matrix of currents on the other face.

Consider a composite shell whose surfaces have radii $r_1$, $r_2$, and $r_3$. The composition law of $H$ matrices can be written as

$$H(r_3,r_1) = H(r_3,r_2)H(r_2,r_1). \tag{7}$$

The above equation holds good whether or not $r_2$ is in between $r_3$ and $r_1$. Generalizing for a shell with $n$ concentric surfaces

$$H(r_n,r_1) = H(r_n,r_{n-1})H(r_{n-1},r_{n-2})\cdots H(r_2,r_1), \tag{8}$$

where $r_n,r_{n-1}\cdots r_2,r_1$ are the radii of concentric similar surfaces. Equation (8) enables us to express the transfer matrix (and hence the reflection and transmission operators) of a given shell in terms of those of concentric similar shells imbedded in it.

From Eq. (6) it can be seen that for a shell of zero thickness $H(r,r) = I$, where $I$ is the unit matrix. Now, consider a shell of thickness $\epsilon$, where $\epsilon$ is infinitesmally small. We can write

$$H(r+\epsilon,r) - I = \epsilon(Gr + Cr)I, \tag{9}$$

where $Gr$ and $Cr$ are independent of $\epsilon$. $\epsilon Cr$ is the change in the $H$ matrix from the unit matrix $I$ due to the collision effects alone, and $\epsilon Gr$ is that due to the geometric effects alone. To first order in $\epsilon$, there is no interaction between the collision and geometric effects. Hence, to first order in $\epsilon$, an expression for $Cr$ can be obtained by studying neutron transport in a slab. An expression for $Gr$ in a given geometry can be obtained by considering neutron transport in a transparent medium (i.e., no collisions) of that geometry.

Making use of the composition law of $H$ matrices, Eq. (9) can be written as

$$H(r+\epsilon,a)H(a,r) - H(r,a)H(a,r) = \epsilon MH(r,a)H(a,r), \tag{10}$$

where $M = (Gr + Cr)$ and $a$ is any standard radius. In the limit of $\epsilon \to 0$, we get the differential equation,

$$\frac{dH(r,a)}{dr} = MH(r,a). \tag{11}$$

This is the matrix form of the invariant imbedding equations. Next we shall find explicit expressions for $Cr$ and Gr.

## IV. THE FORM OF *Cr*

An expression for $Cr$ can be obtained by considering neutron transport in a homogeneous slab. For a slab we can write $H(r,a) = H(x)$, where $x$ is the thickness of the slab. Since $Gr = 0$ for the slab case, from Eq. (11) we get

$$\frac{dH(x)}{dx} = CrH(x). \tag{12}$$

Since $H(0) = I$, Eq. (12) has the solution

$$\begin{aligned} H(x) &= \exp(Crx) \\ &= I + Crx, \end{aligned} \tag{13}$$

to first order in $x$.

Another expression for $H(x)$ can be obtained in the following way. For a homogeneous slab $T_*^+(x) = T_*^-(x) = T_*(x)$ and $R_*^+(x) = R_*^-(x) = R_*x$. Also, $T_*(0) = I$ and $R_*(0) = 0$. To first order in $x$ we can write

$$T_*(x) = I - \alpha x, \tag{14}$$

$$R_*(x) = \beta x, \tag{15}$$

where $\alpha$ and $\beta$ are operators related to the differential cross sections. Substituting (14) and (15) in (6), $H$ can be written as

$$H(x) = I - \begin{bmatrix} \alpha & -\beta \\ \beta & -\alpha \end{bmatrix} x. \tag{16}$$

Comparing (13) and (16),

$$Cr = -\begin{bmatrix} \alpha & -\beta \\ \beta & -\alpha \end{bmatrix}. \tag{17}$$

Expressions for $\alpha$ and $\beta$ in terms of the neutron collision cross sections can be obtained from the transport equation.

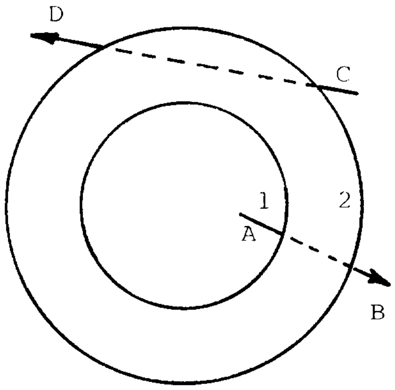


FIG. 2. Transmission and reflection by a transparent medium.

The steady state transport equation for monoenergetic neutron currents in a homogeneous slab, in which only neutron scattering and absorption reactions take place, is

$$\frac{\partial I(S,\mu)}{\partial S} = -\frac{\sigma I(S,\mu)}{\mu} + \frac{\gamma\sigma}{2}\int_{-1}^{+1}\frac{I(S,\mu')}{\mu'}\,d\mu', \tag{18}$$

where

$S$ = depth in slab,

$\sigma$ = total macroscopic cross section,

$\mu$ = cosine of the angle between direction of neutron motion and the normal to the slab,

$\gamma$ = average albedo for single neutron collisions.

Differentiating Eq. (1) and making use of Eqs. (14) and (15) we get

$$\frac{\partial I_2'}{\partial x} = -\alpha I_1 + \beta I_2. \tag{19}$$

More explicitly, since $I$ is to be identified with $I_1$ for $\mu > 0$ and with $I_2$ for $\mu < 0$, Eq. (19) can be written as

$$\frac{\partial I(S,\mu)}{\partial S} = -2\pi\int_0^1 K_\alpha(\mu_1,\mu')I(S,\mu')\,d\mu' + 2\pi\int_{-1}^{0} K_\beta(\mu_1,\mu')I(S,\mu')\,d\mu', \tag{20}$$

for $\mu > 0$. $K_\alpha$ and $K_\beta$ are respectively the kernels of $\alpha$ and $\beta$.

Comparing Eq. (20) with Eq. (18), we get

$$K_\alpha(\mu,\mu') = \frac{\sigma}{2\pi\mu'}\delta(\mu'-\mu) - \frac{\gamma\sigma}{4\pi\mu'} \tag{21}$$

and

$$K_\beta(\mu,\mu') = \frac{\gamma\sigma}{4\pi\mu'}. \tag{22}$$

## V. THE FORM OF *Gr*

We will follow the method indicated in Ref. 10 to find the form of $Gr$. In the absence of collision, i.e., for a transparent medium, $H(r+\epsilon,r)$ is a function of $\epsilon/r$ only for any $\epsilon$. We can write

$$Gr = G/r, \tag{23}$$

where $G$ is independent of $r$.

Consider a neutron path $AB$ (Fig. 2) in a transparent medium. The operator for this transmission process is $T_*^+$. Now consider a neutron entering at $B$ and leaving at $A$. The operator for this process is $T_*^-$. But for a transparent medium the neutron has just retraced the path of the first neutron. Hence we must have $T_*^- T_*^+ = I$ or $T_*^+ = (T_*^-)^{-1} = U$. Hence we can write

$$UR_*^- = T_*^+R_*^- \quad \text{and} \quad R_*^+U = R_*^+T_*^+.$$

It is apparent that $R_*^- = 0$ for a transparent medium of the geometry shown in Fig. 2. Now, let us examine $R_*^+T_*^+$. $R_*^+$ is the operator for a neutron impinging on face 2 to reemerge at face 2, like path $CD$. Since, if we reverse path $AB$, the neutron emerges from face 1 and not face 2, we can conclude that for a transparent medium $R_*^+T_*^+ = 0$. Applying the above relations in Eq. (6), we get

$$H_{tr}(r_2,r_1) = \begin{bmatrix} T_{*_{tr}}^+ & 0 \\ 0 & T_{*_{tr}}^+ \end{bmatrix}, \tag{24}$$

$Cr = 0$ for a transparent medium. Using (23) in (11) we get

$$\frac{dH(r,a)}{dr} = \frac{G}{r}H(r,a). \tag{25}$$

Since $H(a,a) = I$, we get using (24) and (25),

$$G = a\begin{bmatrix} \eta & 0 \\ 0 & \eta \end{bmatrix}, \tag{26}$$

where

$$\eta = \lim_{\epsilon\to 0}\left(\frac{d}{d\epsilon}\right)T_{*_{tr}}^+(a+\epsilon). \tag{27}$$

Hence,

$$Gr = \frac{G}{r} = \frac{a}{r}\begin{bmatrix} \eta & 0 \\ 0 & \eta \end{bmatrix}. \tag{28}$$

$Gr = 0$, for the infinite slab case since $\eta = 0$ in this case.

## VI. OPERATOR FORM OF INVARIANT IMBEDDING EQUATIONS

Substituting for $Cr$ and $Gr$ in Eq. (11) we get

$$\frac{dH(r,a)}{dr} = \begin{bmatrix} -\alpha + \frac{a\eta}{r} & \beta \\ -\beta & \alpha + \frac{a\eta}{r} \end{bmatrix} H(r,a)$$

$$
= \begin{bmatrix} \left(-\alpha + \frac{a}{r}\eta\right)(T_*^+ - R_*^+ UR_*^-) - \beta UR_*^- & \left(-\alpha + \frac{a}{r}\eta\right)R_*^+ U + \beta U \\ -\beta(T_*^+ - R_*^+ UR_*^-) - \left(\alpha + \frac{a\eta}{r}\right)UR_*^- & -\beta R_*^+ U + \left(\alpha + \frac{a\eta}{r}\right)U \end{bmatrix}. \tag{29}
$$

Differentiating Eq. (6)

$$
\frac{\partial H(r,a)}{\partial r} = \begin{bmatrix} \frac{\partial T_*^+}{\partial r} - \frac{\partial R_*^+}{\partial r}UR_*^- + R_*^+ U\frac{\partial T_*^-}{\partial r}UR_*^- - R_*^+ U\frac{\partial R_*^-}{\partial r} & \frac{\partial R_*^+}{\partial r}U - R_*^+ U\frac{\partial T_*^-}{\partial r}U \\ U\frac{\partial T_*^-}{\partial r}UR_*^- - U\frac{\partial R_*^-}{\partial r} & -U\frac{\partial T_*^-}{\partial r}U \end{bmatrix}. \tag{30}
$$

Above we have used the relation $\partial U/\partial r = -U(\partial T_*^-/\partial r)U$.

Equating the matrices in Eqs. (29) and (30) and after some rearrangement we get the four equations.

$$
\frac{\partial R_*^+}{\partial r} = R_*^+\beta R_*^+ - R_*^+\alpha - \alpha R_*^+ - R_*^+\frac{a}{r}\eta + \frac{a}{r}\eta R_*^+ + \beta, \tag{31}
$$

$$
\frac{\partial T_*^-}{\partial r} = T_*^-\beta R_*^+ - T_*^-\alpha - T_*^- a\frac{\eta}{r}, \tag{32}
$$

$$
\frac{\partial T_*^+}{\partial r} = R_*^+\beta T_*^+ - \alpha T_*^+ + \frac{a\eta}{r}T_*^+, \tag{33}
$$

$$
\frac{\partial R_*^-}{\partial r} = T_*^+\beta T_*^-. \tag{34}
$$

The above four equations are the operator form of the invariant imbedding equations in a general geometry. The expressions for the operators $\alpha$ and $\beta$ have already been derived. An integral representation for $\eta$ has to be derived for the particular geometry under consideration and is independent of the neutron scattering properties of the medium.

The response functions are a type of Green's functions of the Boltzmann equation. Hence, the reciprocity relations which are applicable to Green's functions are also applicable to the response functions. A discussion of the reciprocity relations can be found Ref. 11.

## VII. INVARIANT IMBEDDING EQUATIONS IN SPECIFIC GEOMETRIES

### A. Infinite slab

Consider an infintie slab of thickness $x$ [Fig. 1(b)]. The transmitted neutron currents consist partly of neutrons which stream through the medium without any collisions and partly of those which have suffered collisions in the medium. Thus, we write

$$
T_*^-(x,\mu_1;\mu_0) = T^-(x,\mu_1;\mu_0) + \frac{1}{2\pi}e^{-(\sigma x/\mu_1)}\delta(\mu_0 - \mu_1), \tag{35}
$$

$$
T_*^+(x,\mu_1;\mu_0) = T^+(x,\mu_1;\mu_0) + \frac{1}{2\pi}e^{-(\sigma x/\mu_1)}\delta(\mu_0 - \mu_1). \tag{36}
$$

The first and the second term in Eqs. (35) and (36) are the collided and uncollided parts respectively. Since the reflected currents in the slab case do not have uncollided parts, we can write $R_*^+(x,\mu_1;\mu_0) = R^+(x,\mu_1;\mu_0)$ and $R_*^-(x,\mu_1;\mu_0) = R^-(x,\mu_1;\mu_0)$.

Since $\eta = 0$, for the slab case, we have from Eq. (31)

$$
\begin{aligned}
\frac{\partial R_*^+(x,\mu_1;\mu_0)}{\partial x} &= R_*^+\beta R_*^+ - R_*^+\alpha - \alpha R_*^+ + \beta \\
&= 2\pi\int_0^1 R_*^+(\mu_1,S)\left(2\pi\int_0^1 K_\beta(\mu_1,\mu')R_*^+(\mu',\mu_0)\,d\mu'\right)dS - 2\pi\int_0^1 R_*^+(\mu_1,S)K_\alpha(S,\mu_0)\,dS \\
&\quad - 2\pi\int_0^1 K_\alpha(\mu_1,S)R_*^+(S',\mu_0)\,dS + \beta.
\end{aligned} \tag{37}
$$

Substituting for $K_\alpha$ and $K_\beta$ from Eqs. (21) and (22), and with some rearrangement of terms

$$\frac{\partial R^{+}}{\partial x}(x,\mu_1;\mu_0) = \frac{\gamma\sigma}{4\pi\mu_0} - \left(\frac{\sigma}{\mu_1} + \frac{\sigma}{\mu_0}\right)R^{+}(x,\mu_1;\mu_0) + \frac{\gamma\sigma}{2}\int_0^1 \frac{R^{+}}{\lambda}(\lambda,\mu_0)\,d\lambda$$

$$+ \frac{\gamma\sigma}{2}\left[\frac{1}{\mu_0} + 2\pi\int_0^1 \frac{R^{+}}{\lambda}(\lambda;\mu_0)\,d\lambda\right]\int_0^1 R^{+}(\mu_1;S)\,dS. \tag{38}$$

Similarly, substituting for $T_{*}^{-}(x,\mu_1;\mu_0)$, $\alpha$ and $\beta$ in Eq. (32) we get

$$\frac{\partial T^{-}(x,\mu_1;\mu_0)}{\partial x} = \frac{\gamma\sigma}{2}\left[\int_0^1 T^{-}(\mu_1,S)\,dS + \frac{1}{2\pi}e^{-(\sigma x/\mu_1)}\right]\left[2\pi\int_0^1 \frac{R^{+}}{\lambda}(\lambda;\mu_0)\,d\lambda + \frac{1}{\mu_0}\right] - \frac{\sigma}{\mu_0}T^{-}(\mu_1;\mu_0). \tag{39}$$

In a similar fashion, from (33) and (34) respectively, we obtain

$$\frac{\partial T^{+}}{\partial x}(x,\mu_1;\mu_0) = \frac{\gamma\sigma}{2}\left[\int_0^1 \frac{T^{+}(S;\mu_0)}{S}\,dS + \frac{1}{2\pi}\frac{e^{-\sigma x/\mu_0}}{\mu_0}\right]\left[1 + 2\pi\int_0^1 R^{-}(\mu_1;S)\,dS\right] - \frac{\sigma}{\mu_1}T^{+}(\mu_1,\mu_0) \tag{40}$$

and

$$\frac{\partial R^{-}}{\partial x}(x,\mu_1;\mu_0) = \frac{\gamma\sigma}{2}\left[\int_0^1 \frac{T^{+}}{S}(S;\mu_0)\,dS + \frac{1}{2\pi}\frac{e^{-\sigma x/\mu_0}}{\mu_0}\right]\left[2\pi\int_0^1 T^{-}(\mu_1,S)\,dS + e^{-\sigma x/\mu_1}\right]. \tag{41}$$

Since there can be no collisions in a slab of zero thickness, the initial conditions are

$$R^{+}(0) = T^{-}(0) = T^{+}(0) = R^{-}(0) = 0. \tag{42}$$

The reciprocity relations are

$$\mu_0 R^{+}(\mu_1;\mu_0) = \mu_1 R^{+}(\mu_0;\mu_1), \tag{43}$$

$$\mu_0 T^{+}(\mu_1;\mu_0) = \mu_1 T^{-}(\mu_0;\mu_1), \tag{44}$$

$$\mu_0 R^{-}(\mu_1;\mu_0) = \mu_1 R^{-}(\mu_0;\mu_1). \tag{45}$$

It can be easily shown that the above reciprocity relations are satisfied in Eqs. (38), (39), (40), and (41). From (44) it is seen that $T^{+}(\mu_1;\mu_0)$ can be evaluated directly from $T^{-}(\mu_1;\mu_0)$ instead of having to solve Eq. (40).

### B. Spherical shell with a perfectly absorbing core

Separating the response functions into a collided and an uncollided part as before,

$$R_{*}^{+}(x,\mu_1;\mu_0) = R^{+}(x,\mu_1;\mu_0) + \begin{cases} \dfrac{1}{2\pi}e^{-\sigma(2x\mu_1)}\delta(\mu_0 - \mu_1) & \text{if } x^2(1-\mu_0^2) \geqslant a^2 \\ 0 & \text{if } x^2(1-\mu_0^2) < a^2 \end{cases}, \tag{46}$$

$$T_{*}^{-}(x,\mu_1;\mu_0) = T^{-}(x,\mu_1;\mu_0) + \frac{1}{2\pi}e^{-\sigma\{[x^2 - a^2(1-\mu_1^2)]^{1/2} - a\mu_1\}}\delta\left[\mu_0\left(1 - \frac{a^2}{x^2}(1-\mu_1^2)\right)^{1/2}\right], \tag{47}$$

$$T_{*}^{+}(x,\mu_1;\mu_0) = T^{+}(x,\mu_1;\mu_0) + \frac{1}{2\pi}e^{-\sigma\{[x^2 - a^2(1-\mu_0^2)]^{1/2} - a\mu_0\}}\delta\left(\mu_0 - \left(1 - \frac{a^2}{x^2}(1-\mu_1^2)\right)^{1/2}\right), \tag{48}$$

and

$$R_{*}^{-}(x,\mu_1;\mu_0) = R^{-}(x,\mu_1;\mu_0), \tag{49}$$

where $x$ is the outer radius of the shell and $a$ is the inner radius. The quantities within brackets in the exponentials of (46), (47), and (48) are the distances inside the shell which the uncollided neutrons stream through.

From Eq. (27), the integral representation of $\eta$ can be written as

$$\eta = \lim_{\epsilon\to 0}\left(\frac{d}{d\epsilon}\right)2\pi\int_0^1 T_{tr}^{+}(a+\epsilon,\mu_1;\lambda)\,d\lambda. \tag{50}$$

For a transparent medium, the number of neutrons incident on the inner surface of the shell must equal the number emerging at the outer surface.

$$a^2 I(\mu_0) = (a+\epsilon)^2 I(\mu_1), \quad \text{where} \quad \mu_1 = \left(1 - \frac{a^2}{(a+\epsilon)^2}(1-\mu_0^2)\right)^{1/2}. \tag{51}$$

Hence, we can write

$$T_{tr}^{+}(\mu_1;\mu_0)=\frac{a^2}{(a+\epsilon)^2}\cdot\frac{1}{2\pi}\cdot\delta\Big[\mu_1-\Big(1-\frac{a^2}{(a+\epsilon)^2}(1-\mu_0^2)\Big)^{1/2}\Big]. \quad (52)$$

Then

$$\eta=\lim_{\epsilon\to 0}\Big(\frac{d}{d\epsilon}\Big)\cdot\Big[\frac{a^2}{(a+\epsilon)^2}\int_0^1\delta\Big(\mu_1-\Big(1-\frac{a^2}{(a+\epsilon)^2}(1-\lambda_0^2)\Big)^{1/2}d\lambda\Big)\Big]. \quad (53)$$

Using this representation, the terms containing $\eta$ in Eq. (31) can be derived as

$$\frac{a}{x}\eta R_*^{+}=-\Big(\frac{1-\mu_1^2}{x\mu_1}\Big)\frac{\delta}{\delta\mu_1}R^{+}(\mu_1;\mu_0)+\Big(\frac{1-\mu_1^2}{x\mu_1^2}\Big)R^{+}(\mu_1;\mu_0), \quad (54)$$

$$\frac{a}{x}R_*^{+}\eta=-\frac{2}{x}R^{+}(\mu_1;\mu_0)+\Big(\frac{1-\mu_0^2}{x\mu_0}\Big)\frac{\partial R^{+}}{\partial\mu_0}(\mu_1;\mu_0). \quad (55)$$

Deriving the other remaining terms and rearranging, we get from Eq. (31),

$$\frac{\partial R^{+}(\mu_1;\mu_0)}{\partial x}+\frac{1-\mu_0^2}{x\mu_0}\frac{\partial R^{+}}{\partial\mu_0}(\mu_1;\mu_0)+\frac{1-\mu_1^2}{x\mu_1}\frac{\partial R^{+}}{\partial\mu_1}(\mu_1;\mu_0)-\frac{1+\mu_1^2}{x\mu_1^2}R^{+}(\mu_1;\mu_0)+\Big(\frac{\sigma}{\mu_1}+\frac{\sigma}{\mu_0}\Big)R^{+}(\mu_1;\mu_0)$$

$$=\frac{\gamma\sigma}{2}\Big[\int_0^1\frac{1}{\lambda}R^{+}(\lambda;\mu_0)\,d\lambda+\frac{1}{2\pi\mu_0}\begin{Bmatrix}1 & \text{if } x^2(1-\mu_0^2)<a^2\\ 1+e^{-2\sigma x\mu_0} & \text{if } x^2(1-\mu_0^2)\geqslant a^2\end{Bmatrix}\Big]$$

$$\Big[2\pi\int_0^1 R^{+}(\mu_1;S)\,dS+1+\begin{Bmatrix}0 & \text{if } x^2(1-\mu_1^2)\leqslant a^2\\ e^{-2\sigma x\mu_1} & \text{if } x^2(1-\mu_1^2)>a^2\end{Bmatrix}\Big]. \quad (56)$$

Similarly, from Eqs. (32), (33), and (34), we get respectively,

$$\frac{\partial T^{-}(\mu_1;\mu_0)}{\partial x}+\frac{(1-\mu_0^2)}{x\mu_0}\frac{\partial T^{-}(\mu_1;\mu_0)}{\partial\mu_0}-\Big(\frac{2}{x}-\frac{\sigma}{\mu_0}\Big)T^{-}(\mu_1;\mu_0)$$

$$=\frac{\gamma\sigma}{2}\Big[2\pi\int_0^1 T^{-}(\mu_1;\lambda)\,d\lambda+\mu_1 x\frac{e^{-\sigma([x^2-a^2(1-\mu_1^2)]^{1/2}-a\mu_1)}}{[x^2-a^2(1-\mu_1^2)]^{1/2}}\Big]$$

$$\times\Big[\int_0^1\frac{R^{+}(s;\mu_0)}{s}ds+\frac{1}{2\pi\mu_0}\begin{Bmatrix}1 & \text{if } x^2(1-\mu_0^2)<a^2\\ 1+e^{-2\sigma x\mu_0} & \text{if } x^2(1-\mu_0^2)\geqslant a^2\end{Bmatrix}\Big], \quad (57)$$

$$\frac{\partial T^{+}(\mu_1;\mu_0)}{\partial x}+\frac{(1-\mu_1^2)}{x\mu_1}\frac{\partial T^{+}(\mu_1;\mu_0)}{\partial\mu_1}-\Big\{\frac{(1-\mu_1^2)}{x\mu_1^2}-\frac{\sigma}{\mu_1}\Big\}T^{+}(\mu_1;\mu_0)$$

$$=\frac{\gamma\sigma}{2}\Big[\int_0^1\frac{T^{+}}{\lambda}(\lambda;\mu_0)\,d\lambda+\frac{a^2}{2\pi x}\cdot\frac{e^{-\sigma([x^2-a^2(1-\mu_0^2)]^{1/2}-a\mu_0)}}{[x^2-a^2(1-\mu_0^2)]^{1/2}}\Big]$$

$$\times\Big[2\pi\int_0^1 R^{+}(\mu_1;s)\,ds+\begin{Bmatrix}1 & \text{if } x^2(1-\mu_1^2)\leqslant a^2\\ 1+e^{-2\sigma x\mu_1} & \text{if } x^2(1-\mu_1^2)>a^2\end{Bmatrix}\Big], \quad (58)$$

and

$$\frac{\partial R^{-}(\mu_1;\mu_0)}{\partial x}=\frac{\gamma\sigma}{2}\Big[\int_0^1\frac{T^{+}(\lambda;\mu_0)\,d\lambda}{\lambda}+\frac{1}{2\pi}\frac{a^2}{x}\frac{e^{-\sigma([x^2-a^2(1-\mu_0^2)]^{1/2}-a\mu_0)}}{[x^2-a^2(1-\mu_0^2)]^{1/2}}\Big]$$

$$\times\Big[2\pi\int_0^1 T^{-}(\mu_1;s)\,ds+\frac{x\mu_1 e^{-\sigma([x^2-a^2(1-\mu_1^2)]^{1/2}-a\mu_1)}}{[x^2-a^2(1-\mu_1^2)]^{1/2}}\Big]. \quad (59)$$

When $x=a$, i.e., for a shell of zero thickness,

$$R^{+}(\mu_1;\mu_0)=T^{-}(\mu_1;\mu_0)=T^{+}(\mu_1;\mu_0)=R^{-}(\mu_1;\mu_0)=0,$$

the reciprocity relations are

$$\mu_0 R^{+}(\mu_1;\mu_0) = \mu_1 R^{+}(\mu_0;\mu_1), \tag{60}$$

$$x^2\mu_0 T^{+}(\mu_1;\mu_0) = a^2\mu_1 T^{-}(\mu_0;\mu_1), \tag{61}$$

and

$$\mu_0 R^{-}(\mu_1;\mu_0) = \mu_1 R^{-}(\mu_0;\mu_1). \tag{62}$$

The above reciprocity relations are found to be satisfied in Eqs. (56)–(59), providing us with a further check for the correctness of our equations.

For the remaining cases we shall confine ourselves to the $R^{+}(\mu_1;\mu_0)$ function.

## C. Hollow spherical shells and solid spheres

This is very similar to the last case. We write

$$R_{\cdot}^{+}(\mu_1;\mu_0) = R^{+}(\mu_1;\mu_0) + \frac{1}{2\pi}\begin{Bmatrix} e^{-2\sigma x\mu_1}\delta(\mu_0-\mu_1) & \text{if } x^2(1-\mu_1^2) > a^2 \\ e^{-2\sigma\{x\mu_1 - [a^2 - x^2(1-\mu_1^2)]^{1/2}\}}\delta(\mu_0-\mu_1) & \text{if } x^2(1-\mu_1^2) \leqslant a^2 \end{Bmatrix}. \tag{63}$$

Proceeding as before,

$$\frac{\partial R^{+}}{\partial x}(\mu_1;\mu_0) + \frac{(1-\mu_0^2)}{x\mu_0}\frac{\partial R^{+}}{\partial\mu_0}(\mu_1;\mu_0) + \frac{(1-\mu_1^2)}{x\mu_1}\frac{\partial R^{+}}{\partial\mu_1}(\mu_1;\mu_0) - \frac{(1+\mu_1^2)}{x\mu_1^2}R^{+}(\mu_1;\mu_0) + \left(\frac{\sigma}{\mu_0} + \frac{\sigma}{\mu_1}\right)R^{+}(\mu_1;\mu_0)$$

$$= \frac{\gamma\sigma}{2}\left(\int_0^1 \frac{R^{+}}{\lambda}(\lambda;\mu_0)\,d\lambda + \frac{1}{2\pi\mu_0}\begin{Bmatrix} 1 + e^{-2\sigma\{x\mu_0 - [a^2 - x^2(1-\mu_0^2)]^{1/2}\}} & \text{if } x^2(1-\mu_0^2) < a^2 \\ 1 + e^{-2\sigma x\mu_0} & \text{if } x^2(1-\mu_0^2) \geqslant a^2 \end{Bmatrix}\right)$$

$$\times\left(2\pi\int_0^1 R^{+}(\mu_1;s)\,ds + 1 + \begin{bmatrix} e^{-2\sigma\{x\mu_1 - [a^2 - x^2(1-\mu_1^2)]^{1/2}\}} & \text{if } x^2(1-\mu_1^2) \leqslant a^2 \\ e^{-2\sigma x\mu_1} & \text{if } x^2(1-\mu_1^2) > a^2 \end{bmatrix}\right). \tag{64}$$

The initial condition is $R^{+}(a,\mu_1;\mu_0) = 0$.

If $a = 0$ in the above equation, we have the equation for solid spheres.

## VIII. INFINITE CYLINDRICAL SHELL WITH A PERFECTLY ABSORBING CORE

In this case the direction of the neutron current is defined by two angular variables, $\theta$ and $\phi$, as shown in Fig. 3. $\phi$ is the angle between the direction of the axis $Z$ of the cylinder and the neutron current direction $\Omega$. $\theta$ is the angle between the planes formed by $\Omega$ and $Z$ vectors and $Z$ and $r$ vectors. The $r$ vector also denotes the normal to the surface. $\mu$ is the cosine of the angle between $\Omega$ and $r$. It is easily seen that

$$\mu = \psi\sin\phi, \tag{65}$$

where $\psi = \cos\theta$. We have to express $\alpha$, $\beta$, and $\eta$ in terms of $\psi$ and $\phi$ instead of in terms of $\mu$.

Proceeding as before, the transport equation for a homogeneous slab is

$$\frac{\partial I(x,\psi,\phi)}{\partial x} + \frac{\sigma I(x,\psi,\phi)}{\psi\sin\phi} = \frac{\gamma\sigma}{4\pi}\int_{-1}^{+1} \mathrm{d}\psi'\int_0^{2\pi}\frac{I(x,\psi',\phi')}{\psi'\sin\phi'}\,\mathrm{d}\phi'. \tag{66}$$

Comparing this with

$$\frac{\partial I(x,\psi,\phi)}{\partial x} = -\int_0^1 d\psi'\int_0^{\pi} K_\alpha(\psi_1,\phi_1,\psi',\phi')I(\psi',\phi')\,d\phi' + \int_{-1}^{0} d\psi'\int_{\pi}^{2\pi} K_\beta(\psi_1,\phi_1,\psi',\phi')I(\psi',\phi')\,d\phi', \tag{67}$$

we get

$$K_\alpha(\psi_1,\phi_1,\psi_0,\phi_0) = \frac{\sigma}{\psi_0\sin\phi_0}\delta(\psi_1-\psi_0)\delta(\phi_1-\phi_0) - \frac{\gamma\sigma}{4\pi\psi_0\sin\phi_0}, \tag{68}$$

and

$$K_\beta(\psi_1,\phi_1,\psi_0,\phi_0) = \frac{\gamma\sigma}{4\pi\psi_0\sin\phi_0}. \tag{69}$$

Separating the collided and uncollided parts of the reflection function,

$$R_*^{+}(x_1,\psi_1,\phi_1;\psi_0,\phi_0) = R^{+}(x,\psi_1,\phi_1;\psi_0,\phi_0) + \begin{cases} e^{-(2\sigma x\psi_1 \csc\phi_1)}\delta(\psi_0-\psi_1)\delta(\phi_0-\phi_1) & \text{if } x^2(1-\psi_0^2)>a^2 \\ 0 & \text{if } x^2(1-\psi_0^2)\leqslant a^2 \end{cases}, \tag{70}$$

where $2x\psi_1 \csc\phi_1$ is the distance traveled by a streaming neutron which impinges and reemerges at the outer surface of the cylinder,

$$\eta = \lim_{\epsilon\to 0}\frac{d}{d\epsilon}\left[\frac{a^2}{(a+\epsilon)^2}\int_0^\pi d\phi'\int_0^1 \delta\left(\mu_1-\left(1-\left(\frac{a}{a+\epsilon}\right)^2(1-\lambda^2)\right)^{1/2}\right)\right]\delta(\phi_1-\phi')\,d\lambda. \tag{71}$$

Substituting for $K_\alpha$, $K_\beta$, and $\eta$ in (31), we get

$$\begin{aligned}
&\frac{\partial R^{+}}{\partial x}(\psi_1,\phi_1;\psi_0,\phi_0) + \frac{(1-\psi_0^2)}{\psi_0 x}\frac{\partial R^{+}}{\partial \psi_0}(\psi_1,\phi_1;\psi_0,\phi_0) + \frac{(1-\psi_1^2)}{\psi_1 x}\frac{\partial R^{+}}{\partial \psi_1}(\psi_1,\phi_1;\psi_0,\phi_0) \\
&\quad - \left(\frac{1+\psi_1^2}{x\psi_1^2}\right)R^{+}(\psi_1,\phi_1;\psi_0,\phi_0) + \sigma\left(\frac{\csc\phi_0}{\psi_0}+\frac{\csc\phi_1}{\psi_1}\right)R^{+}(\psi_1,\phi_1;\psi_0,\phi_0) \\
&\quad = \frac{\gamma\sigma}{4\pi}\left[\int_0^\pi d\phi'\int_0^1 \frac{\csc\phi'}{\Psi'}R^{+}(\Psi',\phi';\Psi_0,\phi_0)d\Psi' + \frac{1}{\Psi_0\sin\phi_0}\begin{cases} 1 & \text{if } x^2(1-\psi_1^2)\leqslant a^2 \\ 1+e^{-2\sigma x\psi_0\csc\phi_0} & \text{if } x^2(1-\Psi_0^2)>a^2 \end{cases}\right] \\
&\quad \times\left[\int_0^\pi d\phi'\int_0^1 R^{+}(\psi_1,\phi_1;\psi',\phi')\,d\psi' + 1 + \begin{cases} 0 & \text{if } x^2(1-\psi_1^2)\leqslant a^2 \\ e^{-2\sigma x\psi_1\csc\phi_1} & \text{if } x^2(1-\psi_1^2)>a^2 \end{cases}\right].
\end{aligned} \tag{72}$$

The initial condition is $R^{+}(a,\psi_1,\phi_1;\psi_0,\phi_0)=0$, where $a$ is the inner radius of the shell. The reciprocity relation is

$$\psi_0\sin\phi_0 R^{+}(\psi_1,\phi_1;\psi_0,\phi_0) = \psi_1\sin\phi_1 R^{+}(\psi_0,\phi_0;\psi_1,\phi_1). \tag{73}$$

*Comparison of results*: The invariant imbedding equations obtained here by the transfer matrix method must be compared with results obtained by others. Equations (38), (56), and (64) agree with Eqs. (4.20), (5.15) and (6.10) derived by Bailey.[9] Equation (72) differs from that given in Ref. 4 by the presence of the term $-[(1+\psi_1^2)/x\psi_1^2]R^{+}(\psi_1,\phi_1;\psi_0,\phi_0)$, which was overlooked in the original derivation.

## IX. NUMERICAL SOLUTION FOR THE CASE OF A SPHERICAL SHELL WITH A PERFECTLY ABSORBING CORE

In this section a procedure for solving Eqs. (56) and (57) and a method for checking the accuracy of the results will be presented. $\mu_0$ and $\mu_1$ are divided into $NF$ number of intervals over the range 0–1 and the response functions are evaluated only in these discrete directions. We denote these directions by ${}^1\mu, {}^2\mu, \cdots, {}^{NF}\mu$, such that $1 > {}^1\mu > {}^2\mu > \cdots > {}^{NF}\mu > 0$. For the sake of simplicity we denote $F({}^N\mu_1, {}^M\mu_0) = F_{NM}$ where $F(\mu_1,\mu_0)$ is any one of the four response functions. Integrals in (56) and (57) are evaluated using the Gaussian quadrature,

$$\int_0^1 F({}^N\mu_1,\mu')\,d\mu' = \sum_{i=1}^{NF} WiF_{Ni}. \tag{74}$$

The partial derivatives with respect to the direction cosines are evaluated using differential quadratures,[12]

$$\left[\frac{\partial F({}^N\mu_1,\mu_0)}{\partial\mu_0}\right]_{\mu_0={}^M\mu} = \sum_{j=1}^{NF} K_j^{M}F_{Nj}, \tag{75}$$

and

$$\left[\frac{\partial F(\mu_1,{}^M\mu_0)}{\partial\mu_1}\right]_{\mu_1={}^N\mu} = \sum_{j=1}^{NF} K_j^{N}F_{jM}. \tag{76}$$

The discrete direction cosines and corresponding Gaussian weight factors $Wi$ can be found in any standard book of mathematical functions. A method of evaluating the coefficients $K_j^i$ for $j=1,2,\cdots,NF$ and $i=1,2,\cdots,NF$ can be found in Ref. 12.

After the above substitutions and some rearrangement

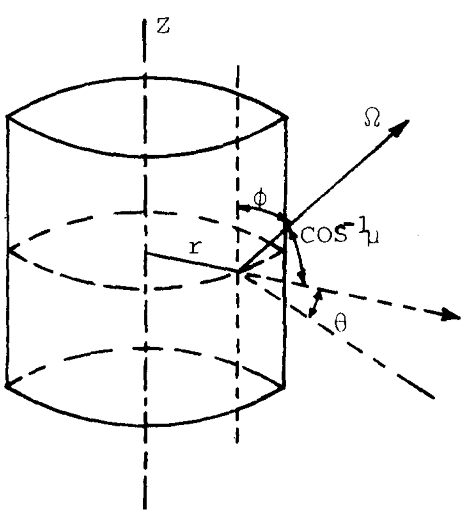


FIG. 3. Cylindrical coordinates.

of terms, Eq. (56) may be written as

$$\frac{\partial R^{+}_{NM}}{\partial x} = -\left(\frac{1-{}^{M}\mu^{2}}{x\,{}^{M}\mu}\right)\sum_{j=1}^{NF} K_{j}^{N} R^{+}_{Nj} - \frac{1}{x}\left[\sum_{j=1}^{NF} K_{j}^{M} \times \frac{R^{+}_{jM}}{{}^{j}\mu} - \sum_{j=1}^{NF} K_{j}^{Mj}\mu R^{+}_{jM}\right] - \left[\frac{\sigma}{M_{\mu}} + \frac{\sigma}{N_{\mu}}\right] R^{+}_{NM}$$
$$+ \frac{\gamma\sigma\pi}{{}^{M}\mu} \times \left[\sum_{j=1}^{NF} W_{K} R^{+}_{MK} + \frac{1}{2\pi} EP\phi W(M)\right] \times \left[\sum_{j=1}^{NF} W_{K} R^{+}_{NK} + \frac{1}{2\pi} EP\phi W(N)\right], \quad (77)$$

for $M = 1,2,\ldots,NF$ and $N = 1,2,\cdots,NF$.

Here,

$$EP\phi W(J) = \begin{Bmatrix} 1 & \text{if } x^{2}(1-{}^{J}\mu^{2}) \leqslant a^{2} \\ 1+\exp(-2\sigma x\,{}^{J}\mu) & \text{if } x^{2}(1-{}^{J}\mu^{2}) > a^{2} \end{Bmatrix}.$$

Equation (57) may be discretized in a similar fashion. As a result of discretization we have on our hands first order ordinary differential equations, which we solve employing Gill's formulation of the fourth order Runge–Kutta method. The reciprocity relation ${}^{M}\mu R^{+}_{NM} = {}^{N}\mu R^{+}_{MN}$ may be used to reduce the computing time. We start with a shell of zero thickness, i.e., $x = a$. The initial condition is $R^{+}_{NM} = T^{-}_{NM} = 0$ for $N = 1,2,\cdots NF$ and $M = 1,2,\cdots,NF$. We then proceed outwards by increasing $x$ in small steps. For better accuracy these steps are kept very small at the beginning stages of computation.

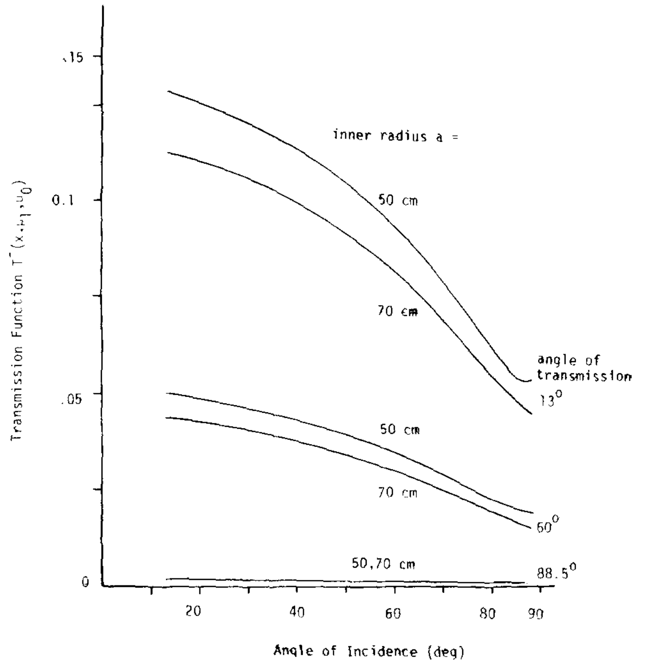


FIG. 5. The transmission function $T^{-}(x,\mu_1;\mu_0)$ vs angle of incidence for various angles of reflection for spherical shells. Thickness of shells = 5 cm $\sigma = 1$ cm$^{-1}$.

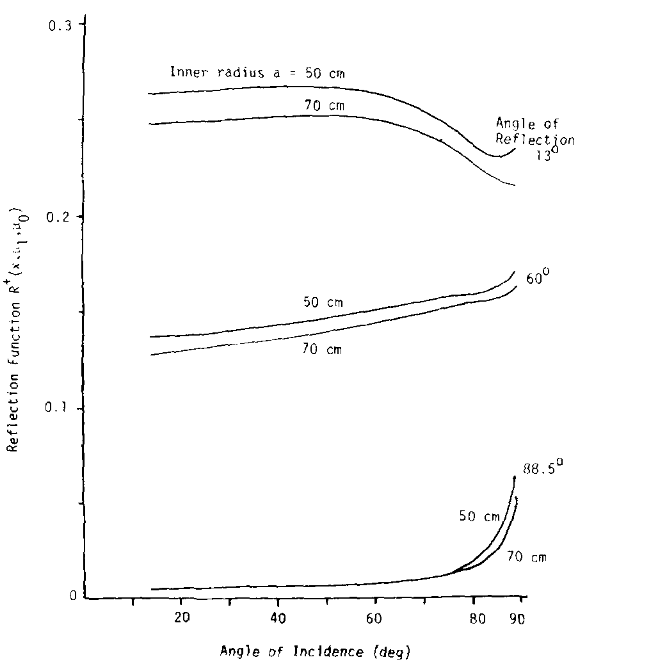


FIG. 4. The reflection function $R^{+}(x,\mu_1;\mu_0)$ vs angle of incidence for various angles of reflection for spherical shells. Thickness of shells = 5 cm, $\sigma = 1$ cm$^{-1}$.

If we take $\gamma = 1$, i.e., there are only neutron scattering reactions in the medium, then in the steady state the number of neutrons emerging from the medium must be equal to the number of neutrons entering the medium. Assuming uniform unit current incident on the outer surface of the shell, i.e., $2\pi\int_0^1 I(\mu_0)\,d\mu_0 = 1$, the above relation can be expressed as

$$\int_0^1\int_0^1 R_{*}^{+}(\mu_1,\mu_0)\,d\mu_0\,d\mu_1 + \frac{a^2}{x^2}\int_0^1\int_0^1 T_{*}^{-}(\mu_1,\mu_0)\,d\mu_0\,d\mu_1 = \frac{1}{2\pi}. \quad (78)$$

We denote the calculated value of the left side of Eq. (78) for a given $X$ by INT($\mathfrak{X}$). An estimate of the percentage errors in the numerical values obtained is given by

$$\mathrm{ERINT}(\mathfrak{X}) = \frac{100(\mathrm{INT}(\mathfrak{X}) - (1/2\pi))}{(1/2\pi)}. \quad (79)$$

Equations (78) and (79) are useful to check the accuracy of the numerical results obtained and keep track of the propagation of error with increasing thickness of the shell.

## X. DISCUSSION OF NUMERICAL RESULTS

The reflection and transmission functions were calculated using a IBM-370 computer system. The values of $\sigma$ and $\gamma$ were assumed to be unity. Figures 4 and 5 show the reflection and transmission functions respectively. Two cases are shown; one with $a = 50$ cm and the other with $a = 70$ cm. The thickness of the shell was 5 cm in each case. The number of discrete directions $NF$ was seven in both the cases.

The distribution of $R^{+}(\mu_1;\mu_0)$ function shown in Fig. 4 can be compared with the $r(\mu_1;\mu_0)$ function published by

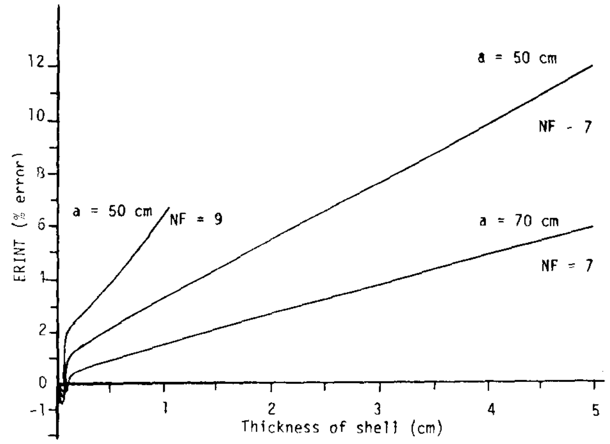


FIG. 6. Propagation of integrated error ERINT with the thickness of the shell.

Bellman *et al.* (Ref. 5, Fig. 2). The two functions are related as follows:

$$r(\mu_1;\mu_0) = \pi(\mu_0/\mu_1)R^+(\mu_1;\mu_0),$$
$$= \pi R^+(\mu_0;\mu_1).$$

After interchanging the angles of incidence and reflection in Fig. 4 and multiplication by $\pi$, the $R^+(\mu_1;\mu_0)$ function is seen to have a similar distribution and magnitude as the $r(\mu_1;\mu_0)$ function. A direct comparison of the numerical values will not be made since the thickness of the shells in the two cases are different.

In Ref. 5, the reflected current was not separated into the collided and uncollided parts. Hence, $r(\mu_1;\mu_0)$ is in fact the total reflection function, where as $R^+(\mu_1;\mu_0)$ consists of only the collided part. In the present cases, the uncollided part of the reflected currents form a small portion of the total reflected currents. Hence, the overall shapes of $r(\mu_1;\mu_0)$ and $R^+(\mu_1;\mu_0)$ are similar.

The collided and uncollided parts of the emergent currents, in general, do not have the same angular distributions. Hence, it is felt that in order to obtain the correct numerical values of their distributions, the response functions should be separated into the collided and the uncollided parts. The total value of the response functions may be obtained by adding the collided and the uncollided parts.

Fig. 6 shows the propagation of the percentage error ERINT($\mathfrak{X}$) with increasing thickness of the shell. A positive value of ERINT($\mathfrak{X}$) indicates that there are more neutrons emerging from the body than the number incident on the body. A negative value would indicate the opposite. The curves in Fig. 6 typify some peculiar characteristics of error propagation.

It is seen from Fig. 6 that in all the three cases shown there is a sharp initial dip in the ERINT($\mathfrak{X}$) function followed by a continuous rise with increasing thickness of the shell. It was also seen that for a given inner radius $a$, (e.g., $a = 50$ cm) and for a given thickness of the shell, the error was greater for the case with $NF = 9$ than for $NF = 7$. This result is surprising, since usually in integrating with Gaussian quadratures we expect a higher accuracy with finer angular mesh, i.e., larger value of $NF$. The error was found to be still larger with $NF = 13$ (not shown in Fig. 6), and the result became rapidly unstable with $NF = 25$.

For a given order of Gaussian quadrature, i.e., for a given value of $NF$, the error increased with decreasing inner radius, $a$ of the shell. It is seen from Fig. 6 that for $NF = 7$, the error is larger for this case with $a = 50$ cm than with $a = 70$ cm. At very small inner radii (e.g., $a = 10$ cm the solution was found to become unstable.

A careful step by step investigation of the results leads to the following explanation of the above patterns of error propagation. The explanation is more qualitative and heuristic than rigorous.

After the start of the computation process till the thickness of the shell reaches such a value as to satisfy the condition,

$$X^2(1 - {}^{NF}\mu^2) \geqslant a^2, \tag{80}$$

the exponential terms $\exp(-2\sigma x/{}^j\mu)$ in Eq. (77) are ignored. This is equivalent to ignoring the uncollided part of the $R_*^+(X,\mu_1;\mu_0)$ function. But, the uncollided current may be a significant part of the total reflected current for shells of small thicknesses. This explains the sharp dip in the ERINT($\mathfrak{X}$) function at the beginning. The thickness at which the curves turn upwards (in Fig. 6) corresponds to the value of $X$ which satisfies Eq. (80). After this point it appear that the uncollided reflected current is overestimated in our formulation. This would introduce a positive error in the computed values of the reflection and transmission functions, which error accumulates steadily with increasing thickness of the shell.

This propagating error seems to stem from our effort to handle by discrete ordinate method the uncollided portion of the reflected current, which is proportional to $\exp(-2\sigma X\mu)$ and is a rapidly varying function of both $X$ and $\mu$. That the uncollided reflected current is indeed responsible for the propagating positive error is supported by the observation that in the cases where the solution becomes unstable, the instability first appears with the value of $R^+({}^{NF}\mu, {}^{NF}\mu_0)$. This is because the uncollided current is first taken into account in the direction ${}^{NF}\mu$. This reasoning can also explain why a larger error (value of ERINT) was observed with $NF = 9$ than with $NF = 7$ in Fig. 6, since for Gaussian directions $[{}^{NF}\mu]_{NF=9} < [{}^{NF}\mu]_{NF=7}$ the exponential $\exp(-2\sigma x\, {}^{NF}\mu)$ is much greater for the case with $NF = 9$ than with $NF = 7$. Hence, for two cases with the same inner radius and thickness of the shell, much higher importance is given to the uncollided portion of the reflected current in the case with $NF = 9$ than with $NF = 7$. As the value of $NF$ is increased, the error too increases and the solution begins to become unstable. In a similar fashion we can explain why for a given $NF$ and thickness of the shell, the error is smaller for a larger inner radius. It is easy to see that as the inner radius of the shell becomes larger, the uncollided reflected current becomes less and less important. For the slab case ($a = \infty$), the uncollided portion of the reflected current is zero.

In conclusion it can be said that in curved geometries the uncollided part of the reflection function presents a difficulty in solving the invariant imbedding equations by the discrete ordinate method, since it introduces an error which progressively increases as the thickness of the shell is increased.


[1]V.A. Ambarzumian, *Theoretical Astrophysics* (Pergamon, New York, 1958).

[2]S. Chandrasekhar, *Radiative Transfer* (Dover, New York, 1960).

[3]R. Bellman, R. Kalaba, and G.M. Wing, "Invariant Imbedding and Mathematical Physics. I. Particle Processes," J. Math. Phys. **1**, 280 (1960).

[4]R. Bellman, R. Kalaba, and G.M. Wing, "Invariant Imbedding and Neutron Transport Theory," J. Math. and Mech. **8**, 575 (1959).

[5]R. Bellman H. Kagiwada, and R. Kalaba, "Invariant Imbedding and Radiative Transfer in Spherical Shells," J. Comp. Phys. **2**, 245 (1967).

[6]A.J. Mockel, "Invariant Imbedding and Polyenergetic Neutron Transport Theory-Part I: Theory," Nucl. Sci. Eng. **29**, 43 (1967).

[7]A.J. Mockel, "Invariant Imbedding and Polyenergetic Neutron Transport Theory-Part II: Numerical Results," Nucl. Sci. Eng. **29**, 51 (1967).

[8]A. Shimizu and K. Aoki, *Application of Invariant Imbedding to Reactor Physics* (Academic, New York, 1972).

[9]P.B. Bailey, "A Rigorous Derivation of Some Invariant Imbedding Equations of Transport Theory," J. Math. Anal. Appl. **8**, 144 (1964).

[10]R. Aronson and D.L. Yarmush, "Transfer Matrix Method for Gamma-Ray and Neutron Penetration," J. Math. Phys. **7**, 221 (1966).

[11]G.I. Bell and S. Glasstone, *Nuclear Reactor Theory* (Van Nostrand, New York, 1970).

[12]R. Bellman, B.G. Kashef, and J. Casti, "Differential Quadrature: A Technique for the Rapid Sclution of Non-Linear Partial Differential Equations," J. Comp. Phys. **10**, 40 (1972).

# SU(2,1) generation of electrovacs from Minkowski space[a)]

I. Hauser and F. J. Ernst

*Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616*
(Received 28 June 1978)

For every nonnull Killing vector **K** of any given electrovac, there exists a group of transformations $\mathscr{H}_{\mathbf{K}}$ of the gravitational and electromagnetic potentials of Ernst. This is the group which is a nonlinear representation of SU(2,1) and was developed by Kinnersley on the basis of work by Ehlers, Harrison, and Geroch. For every **K** of Minkowski space (MS), we compute the set $\mathscr{H}_{\mathbf{K}}$(MS) of all electrovacs derived from MS by noniterative application of $\mathscr{H}_{\mathbf{K}}$; the results include appropriate null tetrads, the connection forms, the conform tensors, and (in the discussion) the group of all motions of very member of every $\mathscr{H}_{\mathbf{K}}$(MS). Each conform tensor is type $N_{pp}$ (plane gravitational wave) or type D, and the principal null vector(s) are also eigenvectors of the Maxwell field. Except for those **K** which represent infinitesimal rotations about a timelike 2-surface of MS followed by null translations in that 2-surface, each **K** has a corresponding MS Killing vector **L** such that the $G_2$ generated by **K** and **L** has nonnull surfaces of transitivity and is invertible. The discussion covers properties of the principal null rays and the Maxwell fields, Killing tensors of the results (one of the $N_{pp}$ families admits an irreducible Killing tensor of Segre characteristic [(11)(11)]), and the precise conditions under which a Killing vector of an electrovac is also an MS Killing vector. Also, some deductions are made concerning the Petrov class and principal null ray properties of the second generation electrovacs which would result from further applications of SU(2,1). Those points of MS which are possible singularities of electrovacs in $\mathscr{H}_{\mathbf{K}}$(MS) are classified. The conditions under which an electrovac in $\mathscr{H}_{\mathbf{K}}$(MS) has all of $R^4$ (except for curvature singularities) as its domain are found; in particular, such an extension to $R^4$ exists whenever the one-parameter group generated by **K** has no fixed points or whenever one restricts $\mathscr{H}_{\mathbf{K}}$ to the Ehlers or Harrison transformations.

## 1. INTRODUCTION

This paper is about a specific application of known transformation groups which can be used to construct a family of electrovacs[1] from any given electrovac $V_4$ with at least one nonnull Killing vector. There exists exactly one of these groups $\mathscr{H}_{\mathbf{K}}$ for each choice of the nonnull Killing vector **K**; we let

$$\mathscr{H}_{\mathbf{K}}(V_4)$$

denote the family[2] of electrovacs constructed from $V_4$ by the application of $\mathscr{H}_{\mathbf{K}}$.

Some details about the group $\mathscr{H}_{\mathbf{K}}$ are given in later sections, but a brief summary of its historical origins will be sufficient identification for the present. The key subsets of $\mathscr{H}_{\mathbf{K}}$ were originally discovered by Ehlers[3] and by Harrison,[4] and the union of these sets with various gauge transformations, duality rotations, and uniform conformal mappings ultimately provided a full set of generators for $\mathscr{H}_{\mathbf{K}}$. The theory of that restriction of $\mathscr{H}_{\mathbf{K}}$ which is a nonlinear representation of SU(1,1) and which transforms vacuums into vacuums was developed by Geroch.[5] The extension to electrovacs is a nonlinear representation of SU(2,1), and its theory was developed by Kinnersley.[6] Pertinent calculational techniques including the use of differential forms to facilitate applications of $\mathscr{H}_{\mathbf{K}}$ and including corresponding transformations of connection forms, the Maxwell field, and the conform tensor have been described by the authors in a preceding paper,[7] which is designated as (I) in the sequel.

The objective of this paper is the application of $\mathscr{H}_{\mathbf{K}}$ to Minkowski space (MS) for every **K** in MS. To be certain that our objective is clear, we caution the reader that only *noniterated* applications of the group are intended; e.g., we are *not* going to compute $\mathscr{H}_{\mathbf{A}}(V_4')$ where $V_4'$ is any member of $\mathscr{H}_{\mathbf{K}}$ (MS), and **A** is any nonnull KV (Killing vector) of $V_4'$. Nor do we believe that there is a point to any finite succession of detailed calculations of this kind as long as there is no evidence that a resulting electrovac will be asymptotically flat[8] or have some redeeming physical or formal feature. Our purpose has nothing to do with providing a first step to a blind program of such iterations.

Apart from their didactic value as illustrations of the techniques developed in (I), our results proved to exhibit those interesting features which we anticipated and were seeking and which made the calculations worthwhile.

For example, even before we made any detailed calculations, we had used the transformation formulas[7] of (I) to study the principal null ray properties of the electrovacs resulting from applications of the groups $\mathscr{H}_{\mathbf{K}}$. *Every electrovac in $\mathscr{H}_{\mathbf{K}}$(MS) was proven to have a type $N_{pp}$ gravitational field* (type $N$ conform tensor with parallel propagating null rays) *if the* MS *Killing bivector*

$$\omega_\alpha{}^\beta := \nabla_\alpha K^\beta$$

*has only zero eigenvalues, and to have a type D gravitational field otherwise.* Now that we are armed with a detailed knowledge of $\mathscr{H}_{\mathbf{K}}$(MS), we can similarly *predict* what would happen to principal null ray properties if we were to apply a second transformation $V_4' \to \mathscr{H}_{\mathbf{A}}(V_4')$ where **A** is any nonnull KV of any given electrovac $V_4'$ in $\mathscr{H}_{\mathbf{K}}$(MS). Some interesting results, based on a simple partial analysis, are given in Sec. 7.

[a)]Work supported in part by the National Science Foundation under Grant No. PHY 75-08750.

  

Other points of interest concern the type $D$ electrovacs in the sets $\mathscr{H}_K$(MS). With the exception of one $\mathscr{H}_K$(MS), all of them turn out to be in the Hamilton–Jacobi (H–J) separable class of Carter[9] and *to have either nonzero expansion for both principal null congruences or zero expansion for both principal null congruences. A notable example of those which have no diverging principal null rays is Melvin's magnetic universe*, which was discovered by Bonner[10] and analyzed by Melvin[11] and Thorne.[12] It was, in fact, shown in a previous paper by Ernst[13] that this electrovac is a member of $\mathscr{H}_K$(MS) where **K** is any generator of a rotation about a timelike 2-surface in MS; that Melvin's magnetic universe is type $D$ was proven by Wild.[14] The family which contains Melvin's magnetic universe is discussed in Sec. 7I.

The exceptions which are not amongst Carter's type $D$ electrovacs each have *exactly one nondiverging principal null congruence* and are in a broad class of nondiverging type $D$ electrovacs which has been obtained by Kinnersley[15] and Plebański.[16] They are like the type $D$ electrovacs of Carter[9] in that each admits a Killing tensor whose Segre characteristic is [(11),(11)], each admits two (no more than that in our case, it so happens) independent commuting Killing vectors which commute with the Killing tensor, and each admits separability of the H–J equation after multiplication by an integrating factor. *They differ from the Carter class in that their two-parameter Abelian isometry groups have null surfaces of transitivity and are not invertible.* Further details about this family are given in Sec. 7H.

We have also found every KV of every member of $\mathscr{H}_K$ (MS), and we have thus been able to determine exactly which Killing vectors of MS survive as Killing vectors after each transformation. Moreover, amongst two of the type $N$ families and two of the type $D$ families, we found *electrovacs which have (regardless of the choice of gauge)*[2] *two or three Killing vectors which are not Killing vectors for* MS. Each type $N$ electrovac in the aforesaid families[17] has a $G_5$ (five-parameter group of motions), and one of them[17] has a $G_6$; they are discussed in Sec. 7C. The members of the two type $D$ families each have a $G_4$ and are discussed in Secs. 7E and 7I.

There is another type $N$ family (Sec. 7B) whose group of motions is an Abelian $G_2$ with a timelike 2-surface of transitivity. *Each member of this family was found to have an irreducible Killing tensor* $\kappa^{\alpha\beta}$ *whose Segre characteristic is* [(1)(1)(2)]. As a consequence, the H–J equation for geodesic orbits is completely separable after multiplication by an integrating factor. Complete H–J separability is, in fact, common to all of the electrovacs in all of the $\mathscr{H}_K$(MS).

In spite of the obviousness of our objective, previous calculations of $\mathscr{H}_K$(MS) have been reported as far as we know only for those **K** which generate either a rotation about a timelike 2-surface[13] or a rotation about a null 2-surface.[7] It is our hope that the more complete coverage of this paper will be helpful in the ultimate objective of developing nontrivial extensions of $\mathscr{H}_K$ which can be used to construct new physically interesting solutions of the Einstein–Maxwell equations. The feasibility of such objectives has been conclusively demonstrated by the recent striking work of Kinnersley and Chitre on axially symmetric stationary spacetimes.[18]

In Sec. 2, we shall classify the pertinent MS Killing vectors from the viewpoint of the conjugacy class structure of the extended Poincaré group (which includes the inversions).[19] The conjugacy class structure is important in this paper, because

$$\mathscr{H}_K(\mathrm{MS}) = \mathscr{H}_A(\mathrm{MS})$$

if the one parameter groups generated by **K** and **A** are conjugate and if isometric electrovacs with corresponding charged particle orbits are regarded as identical. This fact greatly simplifies our task, since we need use no more than one representative KV from each conjugacy class.

We continue Sec. 2 by selecting an appropriate curvilinear coordinate system in $R^4$ for each representative **K**. This will actually consist of one or more similar charts with respective domains $M_1,\ldots,M_n$ chosen so that solutions for certain crucial **K**-related 1-form potentials[7] $M^{rs}$ will exist over these domains. *The group* $\mathscr{H}_K$ *will first be applied only to the restrictions* $V_{4i}$ *of Minkowski space to* $M_i$, *and not generally to* MS *as a whole*. However, we shall see in Sec. 7K that, with appropriate choices of the gauges for the potentials $M^{rs}$, *the domain of applicability can be extended to* MS *whenever* **K** *has no zeros or wherever one of the transformation parameters* ($b_O$) *has unit modulus.*

Corresponding to each given **K** and to each $R^4$ submanifold $M_i$, an appropriate null tetrad **k**, **m**, **t**, **t*** which commutes with **K** is constructed in Sec. 3. Instructive unified expressions (i.e., ones whose forms do not depend on the particular **K**) for the potentials $M^{rs}$ in Minkowski space are derived in Sec. 4. The equations which use these potentials to transform the MS null tetrads $k, m, t, t^*$ into corresponding null tetrads $k', m', t', t'^*$ of the electrovacs in $\mathscr{H}_K$(MS) are also given.

In Sec. 5, the *final results* are listed for the various **K**, and formulas for translating to the notations of Carter[9] are given in some cases where they apply. Some general properties of the results are summarized. The transformed connections forms, Killing bivectors, conform tensors, and Maxwell fields are listed in Sec. 6.

The results are analyzed and discussed in Sec. 7A–7K. Section 7A reports on some common features of the resulting metrics. A key property of the Poincaré group accounts for most of these features. This property appears to have escaped notice until now and is elaborated upon in a theorem in Sec. 7A.

Properties of the results which are specific to each representative **K** are reported in Secs. 7B–7I. Most of the topics which will be discussed in these sections have already been mentioned.

In Sec. 7J, those point sets in MS which are candidates for singularities of the $V_4'$ in $\mathscr{H}_K$(MS) are classified according to their dimension and signature in the original spacetime MS. The dependence of singularity existence on parameter values is considered without getting into a detailed analysis of specific cases.

In the final Sec. 7K, *we express the results in terms of charts which are Cartesian relative to the original space time*

TABLE I. Canonical expressions for $\mathbf{a}$, $\boldsymbol{\omega}$ in those Minkowski space Killing vectors $\mathbf{K} = \mathbf{a} + q^a\omega_a{}^b\mathbf{e}_b$ for which $\boldsymbol{\omega} = \frac{1}{2}\omega^{ab}\mathbf{e}_a \wedge \mathbf{e}_b$ is not zero. The canonical expressions are obtained by appropriate choices of the origin of Cartesian coordinates $q^a$, of the corresponding orthonormal tetrad of vectors $\mathbf{e}_a$, and of a real constant factor of $\mathbf{K}$. The $x^\alpha$ ($\alpha = 1,2,3,4$) are suitable curvilinear coordinates defined in the text. $\mathscr{E}_3$, $\mathscr{E}_4$ occur in $W^E := -2(\omega - i{}^*\omega) = d\mathscr{E}_3 dx^3 + d\mathscr{E}_4 dx^4$ where $\omega := \frac{1}{2}dq^a \wedge dq^b\omega_{ab}$, and $\mathscr{E}_4$ is that complex potential $\mathscr{E}$ of Ernst such that $\mathrm{R1}\mathscr{E} = -\mathbf{K}\cdot\mathbf{K}$. $\epsilon_0 = \pm 1$ is the sign of $(q^3 - q^4)(q^3 + q^4)$.

| Geometric Character of **K** | Parameter values | **a** | $\boldsymbol{\omega}$ | $x^\alpha$ | $\mathscr{E}_3$ | $\mathscr{E}_4 = \mathscr{E}$ |
|---|---|---|---|---|---|---|
| $(N,N)$ | | $\mathbf{e}_m$ | $\mathbf{e}_k \wedge \mathbf{e}_2$ | $x,y,\rho,\sigma$ | $0$ | $-2i(x+iy)$ |
| $(S,N)$ | $a=1$ | $a\mathbf{e}_1$ | $\mathbf{e}_k \wedge \mathbf{e}_2$ | $\rho,\sigma,\xi,\eta$ | $-i\mathscr{E}_4^*$ | $-(\sigma - ia)^2 - 2a^2$ |
| $(0,N)$ | $a=0$ | | | | | |
| $(0,P)$ | $0<\alpha<\pi/2$ | $0$ | $\mathbf{e}_3 \wedge \mathbf{e}_4\cos\alpha + \mathbf{e}_1 \wedge \mathbf{e}_2\sin\alpha$ | $r,s,\Phi,X$ | $-e^{-i\alpha}\cdot(\epsilon_0 s^2\sin\alpha + ir^2\cos\alpha)$ | $e^{-i\alpha}\cdot(\epsilon_0 s^2\cos\alpha - ir^2\sin\alpha)$ |
| $(S,S)$ | $a>0$ | $a\mathbf{e}_1$ | $\mathbf{e}_3 \wedge \mathbf{e}_4$ | $y,s,x,\chi$ | $2iy$ | $\epsilon_0 s^2 - a^2 + 2iay$ |
| $(0,S)$ | $a=0$ | | | | | |
| $(T,T)$ | $a>0$ | $a\mathbf{e}_4$ | $\mathbf{e}_1 \wedge \mathbf{e}_2$ | $z,r,T,\phi$ | $-2iz$ | $a^2 - r^2 - 2iaz$ |
| $(S,T)$ | $a>0$ | $a\mathbf{e}_3$ | $\mathbf{e}_1 \wedge \mathbf{e}_2$ | $T,r,z,\phi$ | $2iT$ | $-a^2 - r^2 + 2iaT$ |
| $(N,T)$ | $a=1$ | $a\mathbf{e}_k$ | $\mathbf{e}_1 \wedge \mathbf{e}_2$ | $\sigma,r,\rho,\phi$ | $-2i\sigma$ | $-r^2 - 2ia\sigma$ |
| $(0,T)$ | $a=0$ | | | | | |

MS. These charts are used to discuss when and how it is possible to "join" the several electrovacs $V'_{4i}$ $(i = 1,\dots,n)$ which are defined over different submanifolds of $R^4$ and which are generated by applying a single element of $\mathscr{H}_{\mathbf{K}}$ to $V_{41},\dots,V_{4n}$. As has already been noted *the "joining" is possible whenever the one parameter group generated by* **K** *has no fixed points or (for any* **K***) whenever one of the transformation parameters* ($b_O$) *has unit modulus.* The joining is accomplished by using an atlas with Cartesian charts and by selecting appropriate gauges.

## 2. THE CHARTS

Let $q^a$ ($a = 1,2,3,4$) denote any Cartesian coordinate system for Minkowski space, $\mathbf{e}_a$ denote the corresponding orthonormal tetrad $(+\ +\ +\ -)$ of vectors, and $\mathbf{e}_k,\mathbf{e}_m,\mathbf{e}_t,\mathbf{e}_t^*$ denote the null tetrad

$$\mathbf{e}_k := \frac{1}{\sqrt{2}}(\mathbf{e}_3 + \mathbf{e}_4), \quad \mathbf{e}_m := \frac{1}{\sqrt{2}}(\mathbf{e}_3 - \mathbf{e}_4),$$
$$\mathbf{e}_t := \frac{1}{\sqrt{2}}(\mathbf{e}_1 + i\mathbf{e}_2), \quad \mathbf{e}_t^* := \frac{1}{\sqrt{2}}(\mathbf{e}_1 - i\mathbf{e}_2). \tag{1}$$

The set of all Killing vectors in MS is the set of all vector fields

$$\mathbf{K} = \mathbf{a} + q^a\omega_{ab}\mathbf{e}^b \tag{2}$$

such that $\mathbf{a}$ is any uniform vector field, and

$$\boldsymbol{\omega} = \tfrac{1}{2}\omega_{ab}\mathbf{e}^a \wedge \mathbf{e}^b \tag{3}$$

is any uniform bivector field. We shall exclude the pure translations $\mathbf{K} = \mathbf{a}$ from further consideration, because they are either null or yield the known trivial result[20]

$$\mathscr{H}_{\mathbf{a}}(\mathrm{MS}) = \{\mathrm{MS}\}.$$

For any given $\mathbf{K}$ such that $\boldsymbol{\omega} \neq \mathbf{0}$, we can select the Cartesian coordinate system and the normalization and sign of $\mathbf{K}$ so that it has a *canonical form* defined by the expressions for $\mathbf{a}$, $\boldsymbol{\omega}$ in one of the rows of Table I. It is clear that any Killing vectors $\mathbf{K}$ and $\mathbf{A}$ have the same canonical form (meaning the same values for the parameter in the 2nd column of Table I, as well as the same canonical expressions for $\mathbf{a}$ and $\boldsymbol{\omega}$) if and only if there exists an isometry $\psi$:MS→MS such that $\psi_* \mathbf{K} = \beta\,\mathbf{A}$ where $\beta$ is a nonzero real constant. So, $\mathbf{K}$ and $\mathbf{A}$ have the same canonical form if and only if the one-parameter groups which they generate are conjugate subsets of the extended Poincaré group.[19]

The first column of Table I contains labels which identify the type of *infinitesimal* rigid motion induced by **K**. The labels $(N,N)$, $(S,N)$, $(0,N)$ designate rotations about a null 2-plane $\Sigma$ followed respectively by null, spacelike, and zero translations which map $\Sigma \to \Sigma$. The label $(0,P)$ designates a rotation about a point (a screw transformation, as it is called). $(S,S)$ and $(0,S)$ are rotations about a spacelike 2-plane $\Sigma$ followed respectively by spacelike and zero translations which map $\Sigma \to \Sigma$. Finally, $(T,T)$,$(S,T)$,$(N,T)$,$(0,T)$ are rotations about a timelike 2-plane $\Sigma$ followed respectively by timelike, spacelike, null, and zero translations which map $\Sigma \to \Sigma$. The $(S,N)$ and $(0,N)$ items are placed in the same row of Table I, because we find it convenient to use the same curvilinear coordinate system $x^\alpha$ ($\alpha = 1,2,3,4$) for both cases. Like remarks hold for $(S,S)$,$(0,S)$ and for $(N,T)$,$(0,T)$, *but we could equally well regard* $(0,S)$ *as the limit of* $(0,P)$ *as* $\alpha \to 0$, *and we can regard* $(0,T)$ *as the limit of* $(0,P)$ *as* $\alpha \to \frac{1}{2}\pi$, *or of* $(T,T)$ *or* $(S,T)$ *as* $a \to 0$.

To help us define the various curvilinear coordinate systems[21] which are listed in column 5 of Table I, let

$$q_k := \frac{1}{\sqrt{2}}(q^3 - q^4), \quad q_m := \frac{1}{\sqrt{2}}(q^3 + q^4),$$
$$q_t := \frac{1}{\sqrt{2}}(q^1 + iq^2). \tag{4}$$

The coordinates for $(N,N)$ are defined by

$$q_k = \sigma, \quad q_m = \rho - y\sigma - \tfrac{1}{6}\sigma^3,$$
$$q^1 = x, \quad q^2 = y + \tfrac{1}{2}\sigma^2. \tag{5a}$$

For $(S,N)$ and $(0,N)$,

$$q_k = \sigma, \quad q_m = \rho - a\xi\eta - \tfrac{1}{2}\sigma(\xi^2 + \eta^2),$$

$$q^1 = \sigma\xi + a\eta, \quad q^2 = \sigma\eta + a\xi. \tag{5b}$$

Observe that the $\rho$ as defined by Eq. (5b) is not the same as the $\rho$ defined by Eq. (5a). For (0,$P$), we have

$$X: = \chi\cos\alpha + \phi\sin\alpha, \quad \Phi: = -\chi\sin\alpha + \phi\cos\alpha, \tag{5c}$$

where $\chi$, $\phi$ as well as $r$, $s$ are defined by

$$q_k = \frac{1}{\sqrt{2}}\epsilon_k s e^{-\chi}, \quad q_m = \frac{1}{\sqrt{2}}\epsilon_m s e^{\chi},$$

$$\epsilon_k: = \mathrm{sgn} q_k, \quad \epsilon_m: = \mathrm{sgn} q_m,$$

$$q_t = \frac{1}{\sqrt{2}} r e^{i\phi}, \quad \phi_1 < \phi < \phi_2. \tag{5d}$$

Above, $r > 0$ and $s > 0$. Equations (5c) and (5d) actually define four distinct charts over disjoint open subsets of MS corresponding to the different $(\epsilon_k,\epsilon_m)$. The parameter $\epsilon_0 = \pm 1$ which appears in parts of the last two columns of Table I is defined by

$$\epsilon_0: = \epsilon_k\epsilon_m. \tag{5e}$$

For ($S$,$S$) and (0,$S$) $s$, $\chi$ are defined by Eqs. (5d), and $x$,$y$ are defined by

$$q^1 = x + a\chi, \quad q^2 = y. \tag{5f}$$

Finally, for ($T$,$T$), ($S$,$T$), ($N$,$T$), (0,$T$), $r$ and $\phi$ are defined by Eqs. (5d), and $z$, $T$, $\sigma$, and $\rho$ are defined by

$$\begin{aligned} &q^3 = z + a^3\phi, \quad q^4 = T + a^4\phi, \\ &a^3 = 0, \quad a^4 = a > 0, \quad \text{for } (T,T), \\ &a^3 = a > 0, \quad a^4 = 0 \quad \text{for } (S,T), \\ &a^3 = a^4 = 2^{-1/2} \quad \text{for } (N,T), \\ &a^3 = a^4 = 0 \quad \text{for } (0,T), \\ &\sigma = \frac{1}{\sqrt{2}}(z - T), \quad \rho = \frac{1}{\sqrt{2}}(z + T). \end{aligned} \tag{5g}$$

$\sigma$ and $\rho$ are defined as above only for ($N$,$T$) and (0,$T$).

The above equations do not constitute complete definitions of our charts since the domains are not yet specified. The fact is that the equations corresponding to each choice of **K** are to be regarded as defining a family of charts (which may have only one member) with respective domains $M_1,\ldots,M_n$ which are simply connected open submanifolds of $R^4$ (the manifold of MS). $M_1 + \cdots + M_n$ is equal to $R^4$ except for that set of measure zero on which the Jacobian of the equations

$$q^a = q^a(x^1,x^2,x^3,x^4)$$

vanishes.

For example, when **K** is type ($N$,$N$) there is only one chart $M_1$ whose domain is $R_4$. For ($S$,$N$), there are three charts $M_1,M_2,M_3$ separated by the null hyperplanes $q_k = \pm 1$. When **K** is type (0,$N$), however, there are two charts $M_1,M_2$ separated by the null hyperplane $q_k = 0$. *In the case of* (0,$P$), *at least eight charts* $M_1,\ldots,M_8$ *are needed.* The first four correspond to given $(\phi_1,\phi_2)$ and to the four distinct pairs of signs $(\epsilon_k,\epsilon_m)$; $M_1 + M_2 + M_3 + M_4$ is a wedge subspace of $R^4$ from which the null hyperplane $q_k = 0$, the null hyperplane $q_m = 0$, and the timelike 2-surface $q^1 = q^2 = 0$ are excluded. Exactly the same statements hold for $M_5,\ldots,M_8$ except that a different range of $\phi$ is chosen so as to provide full angular coverage of $R^4$. These examples should suffice.

For a given **K**, consider the electrovacs

$$V_{4i}: = (M_i,g_i,0)$$

whose metrics $g_i$ are the restrictions of the Minkowski metric to $M_i$, and whose Maxwell 2-form equals 0. By a construction which is spelled out in Secs. 4 and 5, these MS subspaces $V_{4i}$ will be transformed by any given element of $\mathscr{H}_{\mathbf{K}}$ into $n$ electrovacs

$$V'_{4i} = (M'_i,q'_i,F'_i).$$

$M'_i$ equals $M_i$ minus those points at which the transformed metric $g'_i$ and the transformed Maxwell 2-form $F'_i$ have their singularities.

For given **K**, $\mathscr{H}_{\mathbf{K}}(V_{4i})$ will denote the set of all $V'_{4i}$ generated by the elements of **K**, and $\mathscr{H}_{\mathbf{K}}$(MS) will denote the union over $i$ of these sets $\mathscr{H}_{\mathbf{K}}(V_{4i})$. *Until Sec. 7K, we shall use* $V'_4 = (M',g',F')$ *as a generic symbol to denote any of the electrovacs* $V'_{4i},\ldots,V'_{4n}$, *and we shall work simultaneously with all of the n charts.*

We have still not explained how we arrived at our paticular choices of curvilinear coordinates. There were systematic guidelines. First, we chose a coordinate tetrad $\mathbf{X}_\alpha$ which included as many MS Killing vectors as possible. In all cases, it will be noted that $\mathbf{X}_3$ and $\mathbf{X}_4$ are MS Killing vectors, and

$$\mathbf{X}_4: = \mathbf{K}. \tag{6}$$

In every case except (0,$P$), $\mathbf{X}_1$ is also an MS Killing vector. $\mathbf{X}_2$ is not a KV for any of the cases.

The above criterion does not, of course, lead to a unique choice of the coordinate tetrad. Another criterion is best explained after we construct, for each **K**, a new null tetrad **k**, **m**, **t**, **t*** which has the desirable property

$$\mathscr{L}_{\mathbf{K}}\mathbf{k} = \mathscr{L}_{\mathbf{K}}\mathbf{m} = \mathscr{L}_{\mathbf{K}}\mathbf{t} = 0. \tag{7}$$

## 3. THE NULL TETRADS[22]

Let $\mathbf{e}_k,\mathbf{e}_m,\mathbf{e}_t,\mathbf{e}_t^*$ be a uniform null tetrad such that **a** and **ω** are given by Table I. Then, for ($N$,$N$), ($S$,$N$),(0,$N$), it was found that the *simplest* transformation to a new null tetrad which satisfies Eq. (7) is given by the null rotation

$$\begin{aligned} &\mathbf{k} = \mathbf{e}_k, \quad \mathbf{t} = \mathbf{e}_t - \lambda\mathbf{e}_k, \\ &\mathbf{m} = \mathbf{e}_m + \lambda^*\mathbf{e}_t + \lambda\mathbf{e}_t^* - \lambda\lambda^*\mathbf{e}_k, \\ &\lambda: = (i/\sqrt{2})\sigma \quad \text{for } (N,N), \\ &\lambda: = (\xi + i\eta)/\sqrt{2} \quad \text{for } (S,N) \text{ and } (0,N). \end{aligned} \tag{8a}$$

In all other cases,

$$\begin{aligned} &\mathbf{k} = \mathbf{e}_k\exp(\epsilon_1\chi), \quad \mathbf{m} = \mathbf{e}_m\exp(-\epsilon_1\chi), \\ &\mathbf{t} = \mathbf{e}_t\exp(-i\epsilon_2\phi) \end{aligned} \tag{8b}$$

where $\epsilon_1 = \epsilon_2 = 1$ for (0,$P$), $\epsilon_1 = 1$ and $\epsilon_2 = 0$ for ($S$,$S$) and (0,$S$), and $\epsilon_1 = 0$ and $\epsilon_2 = 1$ for the four (··,$T$) cases.

The components of **K** relative to these new null tetrads are required for frequent later use. The following expansions of $\mathbf{K}$ in terms of $\mathbf{k}$, $\mathbf{m}$, $\mathbf{t}$, $\mathbf{t}^*$ are derived from Eqs. (2), (5), (8), and the expressions for $\mathbf{a}$ and $\boldsymbol{\omega}$ in Table I.

For $(N,N)$,

$$\mathbf{K} = -y\mathbf{k} + \mathbf{m}. \tag{9a}$$

For $(S,N)$ and $(0,N)$,

$$\sqrt{2}\mathbf{K} = (-i\sigma + a)\mathbf{t} + (i\sigma + a)\mathbf{t}^*. \tag{9b}$$

For $(O,P)$,

$$\sqrt{2}\mathbf{K} = s(\epsilon_m\mathbf{k} - \epsilon_k\mathbf{m})\cos\alpha - ir(\mathbf{t} - \mathbf{t}^*)\sin\alpha. \tag{9c}$$

For $(S,S)$ and $(0,S)$,

$$\sqrt{2}\mathbf{K} = s(\epsilon_m\mathbf{k} - \epsilon_k\mathbf{m}) + a(\mathbf{t} + \mathbf{t}^*). \tag{9d}$$

For $(T,T)$, $(S,T)$, $(N,T)$, $(0,T)$,

$$\sqrt{2}\mathbf{K} = (a^3 + a^4)\mathbf{k} + (a^3 - a^4)\mathbf{m} - ir(\mathbf{t} - \mathbf{t}^*). \tag{9e}$$

$a^3$,$a^4$ are defined by Eqs. (5g).

Next, we obtained the expressions for the covectors (images in the dual space) of $\mathbf{k}$, $\mathbf{m}$, $\mathbf{t}$ in terms of $dq^a$, and various natural groupings of these terms helped us to choose almost all of the $x^\alpha$. We follow our usual convention[22] of generally employing boldface letters to denote tangent vectors and the corresponding light face letters to denote their covectors. Thus, from Eqs. (1) and (4),

$$e_k = dq_k, \quad e_m = dq_m, \quad e_t = dq_t. \tag{10}$$

From Eqs. (8a), (8b), (10) and some of Eqs. (5a)–(5g), we obtained neat expressions for $k$, $m$, $t$ after grouping terms. For $(N,N)$,

$$k = d\sigma, \quad m = d\rho - yd\sigma, \sqrt{2}t = dx + idy. \tag{11a}$$

For $(S,N)$ and $(0,N)$,

$$k = d\sigma, \quad m = d\rho,$$

$$\sqrt{2}t = \sigma d\xi + ad\eta + i(\sigma d\eta + ad\xi). \tag{11b}$$

For $(0,P)$,

$$\sqrt{2}k = \epsilon_k(ds - sd\chi), \quad \sqrt{2}m = \epsilon_m(ds + sd\chi),$$

$$\sqrt{2}t = dr + ird\phi. \tag{11c}$$

For $(S,S)$ and $(0,S)$, $k$ and $m$ are as in Eqs. (11c), and

$$\sqrt{2}t = dx + idy + ad\chi. \tag{11d}$$

For $(T,T)$, $(S,T)$, $(N,T)$, and $(0,T)$, $t$ is as in Eqs. (11c), and

$$\begin{aligned} \sqrt{2}k &= dz - dT + (a^3 - a^4)d\phi, \\ \sqrt{2}m &= dz + dT + (a^3 + a^4)d\phi, \end{aligned} \tag{11e}$$

where $a^3$, $a^4$ are defined for each case by Eqs. (5g). The above Eqs. (11a)–(11e) helped us to decide on most choices of $x^\alpha$ in Table I. For example, consider $(0,N)$; it is easy to see why we settled on $y$, $x$, $\rho$, $\sigma$ from Eqs. (11a). *Equations* (11) *will be used to compute our final results in Sec.* 5.

The most forceful guide for our final selection of $x^\alpha$ is their role in the complex scalar fields $\mathscr{E}_3$ and $\mathscr{E}_4$ which are given in Table I. This role is acted out in the first three phases of the calculation of $\mathscr{H}_\mathbf{K}$(MS). Each of these phases is an application to a special case of the techniques developed in (I).[7]

## 4. THE POTENTIALS

The first phase is the calculation of the 2-forms

$$W^E := -4\mathbb{P}\omega, \quad \omega := \tfrac{1}{2}d\wedge K. \tag{12}$$

$\mathbb{P}$ is a projection operator which extracts a self-dual part of any 2-form on which it operates. To help us be more specific, we introduce the following basis[22] for self-dual 2-forms:

$$B_1 := kt, \quad B_0 := km + tt^*, B_{-1} := mt^*. \tag{13}$$

(We follow our usual convention of omitting the wedge symbol in exterior products of forms.) Then, for *spin weight* $A = 1,0,-1$,

$$^*B_A = iB_A, \quad \mathbb{P}B_A = B_A, \quad \mathbb{P}B^*_A = 0. \tag{14}$$

From the expressions for $\boldsymbol{\omega}$ in Table I and from Eqs. (1), (8a), (8b), and (12)–(14), we compute $W^E$ in terms of the basis $B_A$. Then, we use Eqs. (11a) – (11e) to compute $W^E$ in terms of the basis $dx^\alpha dx^\beta$. *For all cases*, we obtain[23]

$$W^E = d\mathscr{E}_3 dx^3 + d\mathscr{E}_4 dx^4, \tag{15}$$

$\mathscr{E}_3$, $\mathscr{E}_4$ depend at most on $x^1$, $x^2$. Equation (15) defines $\mathscr{E}_3$, $\mathscr{E}_4$ up to arbitrary constants which are discussed below.

The second phase is the calculation of the complex scalar potential $\mathscr{E}^E = \mathscr{E}$ which is defined by the equation

$$d\mathscr{E} = K\ulcorner W^E, \tag{16a}$$

$$\mathrm{R1}\mathscr{E} = f := -\mathbf{K}\cdot\mathbf{K}. \tag{16b}$$

The *step product* ($\ulcorner$) operation is bilinear with respect to both factors and is defined for any 1-forms $u$, $v$, $w$ by

$$w\ulcorner(uv) = u(w\cdot v) - v(w\cdot u). \tag{17}$$

Since

$$K\cdot d = \frac{\partial}{\partial x^4},$$

Eqs. (15)–(17) yield

$$\mathscr{E} = \mathscr{E}_4 \tag{18}$$

after selecting the real part of the arbitrary constant in $\mathscr{E}_4$ so as to satisfy Eq. (16b). As regards the arbitrary constant in $\mathscr{E}_3$ and the imaginary part of the arbitrary constant in $\mathscr{E}_4$, we select them to achieve a certain simplicity of form.[23]

The third phase is the calculation of the one-form potentials $M^{rs}$ which are defined by the equations[7]:

$$\begin{aligned} &M^{rs} = (M^{sr})^*, \quad r,s \text{ have letter values } O,E, \\ &dM^{OO} = 0, \quad K\cdot M^{OO} = 1, \\ &dM^{OE} = W^E, \quad K\cdot M^{OE} = \mathscr{E}, \\ &dM^{EE} = \mathscr{E}^* W^E + \mathscr{E} W^{E*}, \ K\cdot M^{EE} = \mathscr{E}^*\mathscr{E}. \end{aligned} \tag{19}$$

From the expressions for $\mathscr{E}_3$, $\mathscr{E}_4$ in Table I and from Eqs. (15), (16), (18), it is readily verified that solutions of Eqs. (19) are given by[24]

$$\begin{aligned} &M^{OO} = dx^4, \quad M^{OE} = \mathscr{E}_3 dx^3 + \mathscr{E}_4 dx^4, \\ &M^{EE} = \tfrac{1}{2}(\mathscr{E}^*_3\mathscr{E}_4 + \mathscr{E}_3\mathscr{E}^*_4)dx^3 + |\mathscr{E}_4|^2 dx^4, \end{aligned} \tag{20}$$

*for every* **K**. These solutions may be subject to gauge transformations[24]

$$M^{rs} \to M^{rs} + dF^{rs}, \quad K \cdot dF^{rs} = 0,$$

which leads to electrovacs locally isometric to those which we shall obtain.

We are now ready to consider how the potentials $\mathscr{E}^r$, $M^{rs}$ are used to construct the various transformations induced by any element of the group $\mathscr{H}_{\mathbf{K}}$.

Let $\theta_a$ be any given null tetrad and $K$ be the covector of **K** in the MS subspace $V_4$ [a generic symbol for any $V_{4i} = (M_i, g_i, 0)$].. Let $V'_4 = (M', g', F')$ denote any $\mathscr{H}_{\mathbf{K}}(V_4)$ member with manifold $M'$, metric $g'$, and Maxwell 2-form $F'$. Basic to all of our final calculations are the following equations[7] which supply us with a null tetrad $\theta'_a$ and the Killing covector $K'$ in the transformed electrovac $V'_4$:

$$\theta'_a = |\Lambda|\theta_a - |\Lambda|^{-1}K_a\Delta, \quad K' - K = f|\Lambda|^{-2}\Delta, \tag{21}$$

where

$$\Lambda = b_O + b_E\mathscr{E}, \tag{22a}$$

$$b^*_O b_E + b_O b^*_E = -\tfrac{1}{2}|b_M|^2, \tag{22b}$$

$$\Delta := b^*_r b_s M^{rs} - M^{OO}. \tag{23}$$

Equations (22a) and (23) are specializations of equations[7] in (I) to the case where the "old" electrovac is a vacuum. $b_O, b_E$ are complex parameters of the transformation and are independent except for the statement that not both can be zero and except for the inequality implied by Eq. (22b); the parameter $|b_M|$ is a factor in $F'$, which we consider in Sec. 6. Components $K_a$ and $K'_a$ of the respective 1-forms $K$ and $K'$ are defined by the equations

$$K = K^a\theta_a, \quad K' = K^{a'}\theta'_a. \tag{24}$$

$K_a$ and $K'_a$ are related by

$$K'_a = |\Lambda|^{-1}K_a, \tag{25}$$

*which holds for any initial electrovac if we use the general expression for $\Lambda$ as given in* (I).

For the present problem, $\Lambda$ is given by Eq. (22a), and we are primarily interested in components relative to the tetrads

$$\theta_k := k, \quad \theta_m := m, \quad \theta_t := t, \quad \theta_n := t^*, \tag{26a}$$

$$\theta'_k = : k', \quad \theta'_m = : m', \quad \theta'_t = : t', \quad \theta'_n = : t'^*.$$

Note our use of scripts $a = k, m, t, n$ for null tetrad components. Since $k \cdot m = t \cdot t^* = 1$,

$$K^k = K_m, \quad K^m = K_k, \quad K^t = K_n, \quad K^n = K_t.$$

Thus

$$K = K_m k + K_k m + K_n t + K_t t^*. \tag{26b}$$

Since $K$ is real, $K_n = (K_t)^*$.

## 5. RESULTS

We now apply Eqs. (21)–(23) to the calculation of $k'$, $m'$, $t'$ in terms of $dx^\alpha$. In this calculation, $K_a$ $(a = k,m,t,n)$ can be obtained by inspection of Eqs. (9) and (26). Then, Eqs. (11), (20), and the expressions for $\mathscr{E}_3$, $\mathscr{E}_4$ in Table I are used. The results are given below for $(N,N)$, $(S,N)$, and $(0,N)$ Killing vectors in Minkowski space. As we shall see, $V'_4$ has a type $N$ gravitational field in these cases.

$(N,N)$ results:

$$k' = |\Lambda|^{-1}d\sigma, \quad m' = |\Lambda|d\rho - |\Lambda|^{-1}y\,d\sigma, \tag{27a}$$

$$\sqrt{2}t' = |\Lambda|(dx + i\,dy),$$

where

$$\Lambda = b_O - 2ib_E(x + iy). \tag{27b}$$

$(S,N),(0,N)$ results:

$$k' = |\Lambda|d\sigma, \quad m' = |\Lambda|d\rho, \tag{28a}$$

$$\sqrt{2}t' = |\Lambda|(\sigma + ia)d\xi + i|\Lambda|^{-1}(\sigma - ia)(d\eta - J\,d\xi),$$

where

$$\Lambda = b_O + b_E(-\sigma^2 - a^2 + 2ia\sigma), \tag{28b}$$

$$J = i[b^*_O b_E(\sigma^2 + a^2 + 2ia\sigma) - b_O b^*_E(\sigma^2 + a^2 - 2ia\sigma)]$$

$$+ |b_E|^2 4a\sigma(\sigma^2 + a^2).$$

In all remaining cases, we use the notation $\lambda^{(j)}$ for certain real fields which depend at most on the single coordinate $x^j$ ($j = 1$ or 2). $k', m', t'$ are generally numerical multiples of the 1-forms

$$\{j.\delta\} := 2^{-1/2}\{|\Lambda|dx^j + \delta|\Lambda|^{-1}y^{(j)}[f^{(3-j)}dx^3 + a^{(j)}dx^4]\}, \tag{29a}$$

where

$$j = 1 \text{ or } 2, \quad \delta = \pm 1 \text{ or } \pm i, \quad |\Lambda|^2 = \lambda^{(2)} - \lambda^{(1)}, \tag{29b}$$

and $y^{(j)}, f^{(j)}, a^{(j)}$ are defined as follows.

$(0,P)$:

$$y^{(j)} = x^j, \quad a^{(1)} = \sin\alpha, \quad a^{(2)} = \cos\alpha,$$

$$f^{(1)} = [\lambda^{(1)} - |b_O \sin\alpha|^2]\csc\alpha, \tag{30a}$$

$$f^{(2)} = [\lambda^{(2)} - |b_O \sin\alpha|^2]\sec\alpha.$$

All cases except $(0,P)$:

$$y^{(1)} = 1, \quad y^{(2)} = x^2, \quad a^{(1)} = a, \quad a^{(2)} = 1, \tag{30b}$$

$$f^{(1)} = a^{-1}\lambda^{(1)}, \quad f^{(2)} = \lambda^{(2)}.$$

$\lambda^{(j)}$ depends on $\alpha$ in Eqs. (30a) and on $a$ in Eqs. (30b) in such manner that $f^{(j)}$ is defined as $\alpha \to 0$, $\alpha \to \frac{1}{2}\pi$, or $a \to 0$, respectively. $k', m', t', x^\alpha, \lambda^{(1)}, \lambda^{(2)}$ are given below. As we shall see, $V'_4$ has a type $D$ gravitational field in all of these cases.

$(0,P)$ results:

$$k' = \epsilon_k\{2.-1\}, \quad m' = \epsilon_m\{2.1\}, \quad t' = \{1.i\}, \tag{31a}$$

where

$$x^1 = r, \quad x^2 = s, \quad x^3 = \Phi, \quad x^4 = X,$$

$$\lambda^{(1)} = |b_O|^2 - |b_O - i\widetilde{b}_E r^2 \sin\alpha|^2,$$

$$\lambda^{(2)} = |b_O + \widetilde{b_E}\epsilon s^2\cos\alpha|^2, \quad \widetilde{b_E} := b_E \exp(-i\alpha).$$

$(S,S)$,$(0,S)$ results:

$$k' = \epsilon_k\{2.-1\}, \quad m' = \epsilon_m\{2.1\}, \quad t' = i\{1.-i\}, \tag{31b}$$

where

$$x^1 = y, \quad x^2 = s, \quad x^3 = x, \quad x^4 = \chi,$$
$$\lambda^{(1)} = |b_O|^2 - |b_O + 2ib_E ay|^2,$$
$$\lambda^{(2)} = |b_O + b_E(\epsilon_0 s^2 - a^2)|^2.$$

$(T,T)$ results:

$$k' = \{1.-1\}, \quad m' = \{1.1\}, \quad t' = \{2.i\}, \tag{31c}$$

where

$$x^1 = z, \quad x^2 = r, \quad x^3 = T, \quad x^4 = \phi,$$
$$\lambda^{(1)} = |b_O|^2 - |b_O - 2ib_E az|^2,$$
$$\lambda^{(2)} = |b_O + b_E(a^2 - r^2)|^2.$$

$(S,T)$ results:

$$k' = -\{1.-1\}, \quad m' = \{1.1\}, \quad t' = \{2.i\}, \tag{31d}$$

where

$$x^1 = T, \quad x^2 = r, \quad x^3 = z, \quad x^4 = \phi,$$
$$\lambda^{(1)} = |b_O|^2 - |b_O + 2ib_E aT|^2,$$
$$\lambda^{(2)} = |b_O - b_E(a^2 + r^2)|^2.$$

$(N,T)$,$(0,T)$ results:

$$k' = |\Lambda|dx^1, \quad t' = \{2.i\},$$
$$m' = |\Lambda|^{-1}(\lambda^{(2)}dx^3 + adx^4), \tag{31e}$$

where

$$x^1 = \sigma, \quad x^2 = r, \quad x^3 = \rho, \quad x^4 = \phi,$$
$$\lambda^{(1)} = |b_O|^2 - |b_O - 2ib_E a\sigma|^2,$$
$$\lambda^{(2)} = |b_O - b_E r^2|^2.$$

Equations (27) – (31), above, collectively constitute self-contained results for all $(k',m',t')$, except that the values of the parameters, $\alpha$, $a$ are given in Table I. We can see from Table I that $(0,S)$ results can be obtained by letting $\alpha \to 0$ in Eqs. (31a), as well as $a \to 0$ in Eqs. (31b). Likewise, $\alpha \to \frac{1}{2}\pi$ in Eqs. (31a), and $a \to 0$ in Eqs. (31c), (31d), and (31e) yield $(0,T)$ results. The electrovacs which are constructed from the two $(0,S)$ limits are isometric, and the same holds for the different $(0,T)$ limits.

The line elements which are constructed from our type $D$ results are easily compared with the line element derived by Hauser and Malhiot[25] for spacetimes which have Killing tensors of Segre characteristic [(11)(11)]. *$\lambda^{(1)}$ and $\lambda^{(2)}$ are their notations for the eigenvalues of the Killing tensors* [except that in the $(\cdot\cdot,T)$ cases, we must interchange our scripts $1 \leftrightarrow 2$ and $3 \leftrightarrow 4$ to get agreement with their usages]; their $\rho^2$ is our $|\Lambda|^2$.

It may be useful to see exactly how some of our type $D$ electrovac results fit into the extensive family of solutions derived by Carter.[9] We restrict our comparison to the cases for which $d\lambda^{(1)}$ and $d\lambda^{(2)}$ are not zero, and for which the surfaces of transitivity of $\mathbf{X}_3$, $\mathbf{X}_4$ are timelike. *For these cases, Carter's coordinates $\mu$, $\lambda$, $\tau$, $\sigma$ and his metrical fields $\Delta_\mu$, $\Delta_\lambda$ are given in terms of our coordinates and parameters as follows.*

$(0,P)$ with $\epsilon_0 = 1$, comparison with Carter:

$$\Delta_\mu = [\gamma + (2|b_E|^{-1})^{1/2}\mu]\sin\alpha = (r\sin\alpha)^2,$$
$$\Delta_\lambda = [-\beta + (2|b_E|^{-1})^{1/2}\lambda]\cos\alpha = (s\cos\alpha)^2,$$
$$\tau = (2|b_E|)^{-1/2}[X - |b_E|^2(\beta^2\tan\alpha - \gamma^2\cot\alpha)\Phi],$$
$$\sigma = -(2|b_E|)^{1/2}(\sin\alpha\cos\alpha)^{-1}\Phi,$$

where

$$\beta + i\gamma := b_O b_E^{-1} e^{i\alpha}, \quad \beta,\gamma = \text{real}.$$

In the following, $\beta,\gamma$ are defined as above with $\alpha = 0$.

$(S,S)$ with $\epsilon_0 = 1$, comparison with Carter:

$$\Delta_\mu = 1, -\gamma + (2a|b_E|^{-1})^{1/2}\mu = 2ay,$$
$$\Delta_\lambda = 1 - a^{-2}[\beta + (2a|b_E|^{-1})^{1/2}]\lambda = (s/a)^2,$$
$$\tau = (2a|b_E|)^{-1/2}(a\chi + |b_E|^2\gamma^2 x),$$
$$\sigma = -(2a|b_E|)^{1/2}x.$$

$(T,T)$, comparison with Carter:

$$\Delta_\lambda = 1, \quad \gamma + (2a|b_E|^{-1})^{1/2}\lambda = 2az,$$
$$\Delta_\mu = 1 + a^{-2}[\beta + (2a|b_E|^{-1})^{1/2}]\mu,$$
$$\tau = (2a|b_E|)^{-1/2}(a\phi + |b_E|^2\gamma^2 T),$$
$$\sigma = (2a|b_E|)^{1/2}T.$$

## 6. CONNECTION FORMS, KILLING BIVECTORS, CONFORM TENSORS, AND MAXWELL FIELDS

We next compute the connection 1-forms $v'_A$, the Killing bivector components $\omega'_A$, the Maxwell 2-form components $F'_A$, and the Weyl conform tensor components[2]

$$C'_{-A-B} := C^{AB\prime}$$

in $V'_4$. The scripts $A$ and $-A-B$ $(A,B = 1,0,-1)$ are spin weights; as regards their raising and lowering, $\omega_1 = \omega^{-1}$, $\omega_{-1} = \omega^1$, $\omega_0 = -2\omega^0$. The components are relative to the self-dual basis[22]

$$B'_A := \tfrac{1}{2}B_A^{ab}\theta'_a\theta'_b \quad (a,b = k,m,t,n), \tag{32}$$

where $\theta'_a$ is defined by Eq. (26a), and $B_A^{ab}$ are the same numerical coefficients as those used in Eq. (13). Our connection forms are defined by[22]

$$v'_A := \tfrac{1}{2}B_A^{ab}v'_{ab}, \quad d\theta'_a = v'_{ab}\theta^{b\prime}. \tag{33}$$

$\mathrm{Rl}\, v'_{1n}, \mathrm{Im}\, v'_{1n}, v'_{1t}, v'_{1k}$ are the expansion, twist, shear, and geodesy of the null congruence of $\mathbf{k}'$; for $\mathbf{m}'$, the same roles are played by $\mathrm{Rl}\, v'_{-1t}, \mathrm{Im}\, v'_{-1t}, v'_{-1n}, v'_{-1m}$. The following results are obtained by using the transformation formulas of (I) together with Eqs. (8), (9), and the expressions for $\boldsymbol{\omega}$ in Table I.

$(N,N)$ results:

$$v'_1 = 0, \quad v'_0 = -d\ln\Lambda,$$
$$v'_{-1} = -i2^{-1/2}|\Lambda|^{-1}(1 - 4b_E\Lambda^{-1}y)k'. \tag{34a}$$

$$\omega_1' = \omega_0' = 0, \tag{34b}$$

$$\omega'_{-1} = -i2^{-1/2}|A|^{-2}(1 - 2b_E f A^{-1}).$$

$$F_1' = F_0' = 0, \quad F'_{-1} = -i2^{-1/2} b_M^* A^{-2}. \tag{34c}$$

$C'_{-A-B} = 0$ except

$$C'_{-2} = -6b_E A^{-1}|A|^{-2}(1 - 2b_E f A^{-1}). \tag{34d}$$

$(S,N)$, $(0,N)$ results:

Same as $(N,N)$ Eqs. (34a)–(34d) except

$$v'_{-1} = |A|^{-1}[(\sigma^2 - a^2)^{-1}(iat' + \sigma t'^*) - 2b_E A^{-1}(\sigma + ia)t']. \tag{35}$$

$(0,P)$, $(\cdot\cdot,S)$, $(\cdots,T)$ results

$$v_1' = 2\widetilde{b}_E|A|^{-1}A^{-1}(K_t k' - K_k t'),$$

$$v_0' = |A|^{-1}\{-2^{-1/2}[\epsilon_1 s^{-1}(\epsilon_k k' - \epsilon_m m') + \epsilon_2 r^{-1}(t' - t'^*)] + 2\widetilde{b}_E A^{-1}(K_m k' + K_k m' + K_n t' + K_t t'^*)\},$$

$$v'_{-1} = -2\widetilde{b}_E|A|^{-1}A^{-1}(K_n m' - K_m t^*). \tag{36a}$$

$$\omega_1' = \omega'_{-1} = 0,$$

$$\omega_0' = |A|^{-2}(1 - 2b_E f A^{-1})\exp(-i\alpha). \tag{36b}$$

$$F_1' = F'_{-1} = 0, \quad F_0' = b_M^* A^{-2}\exp(-i\alpha). \tag{36c}$$

$C'_{-A-B} = 0$ except

$$C_0' = 2b_E|A|^{-2}A^{-1}(1 - 2b_E f A^{-1})\exp(-2i\alpha). \tag{36d}$$

In Eqs. (36a)–(36d),

$$\widetilde{b}_E := b_E\exp(-i\alpha),$$

$$0 < \alpha < \tfrac{1}{2}\pi, \quad \epsilon_1 = \epsilon_2 = 1, \quad \text{for } (0,P),$$

$$\alpha = 0, \quad \epsilon_1 = 1, \quad \epsilon_2 = 0 \quad \text{for } (\cdot\cdot,S), \tag{36e}$$

$$\alpha = \tfrac{1}{2}\pi, \quad \epsilon_1 = 0, \quad \epsilon_2 = 1 \quad \text{for } (\cdot\cdot,T).$$

Concerning various symbols in the above Eqs. (34)–(36), we recall that $f = \mathrm{R1}\mathscr{E}$ and that $\mathscr{E}$ is given by Table I. Also, $A = b_O + b_E\mathscr{E}$ and is given in Eqs. (30) and (31). $K_a$ can be obtained by inspection of Eqs. (9) [see Eq. (26b)].

As regards the parameter $b_M$, which occurs in Eqs. (34c) and (36c), it modulus is uniquely determined by $b_O$ and $b_E$ according to Eq. (22b). Its phase is independent of $b_O$ and $b_E$ and is responsible for the duality orientation of $F'$.

It is always true that the substitution

$$b_r \to b_r\exp(ic), \quad (r = O,E,M),$$

where $c$ is any real number, does not alter $g'$. This can be verified for our particular problem by inspecting Eqs. (21)–(23). It follows that we can choose $b_O$ or $b_E$ (but not both) to be real and nonnegative without loss of generality; i.e., the set $\mathscr{H}_{\mathbf{K}}(\mathrm{MS})$ is not diminished by that constraint.

We now briefly comment on the algebraic classification of the various gravitational and electromagnetic fields. These comments are based on Eqs. (34a), (35), (36c), and (36d). We see that all of the $(\cdot\cdot,N)$ results have $N_{pp}$ gravitational and Maxwell fields with a common set of null rays. As regards the rest, the $(0,P)$, $(\cdot\cdot,S)$, and $(\cdot\cdot,T)$ results have type $D$ gravitational fields, whose principal null vectors are also eigenvectors of the Maxwell field. Further details will be given in the discussion of individual cases.

## 7. ANALYSIS AND DISCUSSION

### A. Some common features of the metrics

The metric $g'$ can be computed from Eqs. (27)–(31), and

$$g' = k'\otimes m' + m'\otimes k' + t'\otimes t'^* + t'^*\otimes t'.$$

The components relative to our charts satisfy the following conditions for all type $N$ cases and for the type $D$ cases $(0,P)$, $(S,S)$, $(0,S)$, $(T,T)$, $(S,T)$:

$$g'_{\alpha\beta} = 0, \quad \text{if} \quad \alpha = 1,2 \text{ and } \beta = 3,4. \tag{37}$$

For $(N,T)$, however,

$$g'_{11} = g'_{12} = g'_{32} = g'_{42} = 0,$$
$$g'_{33}g'_{44} - (g'_{34})^2 = 0, \quad g_{41} = a. \tag{38}$$

As regards $(0,T)$, it's a hybrid. The form of its metric depends on the choice of the coordinate pair $(x^1,x^3)$. On the one hand, $(0,T)$ may be obtained as the limits of $(0,P)$ as $\alpha \to \frac{1}{2}\pi$, or of $(T,T)$ or $(S,T)$ as $a \to 0$. On the one hand, we may let $a \to 0$ in the $(N,T)$ expressions (as we have preferred to do in our own work).

The results expressed by Eq. (37) could have been predicted from the first two conclusions of the following theorem, which can be verified by inspecting Table I and Eqs. (11).

*Theorem*: For any given nonnull Killing vector $\mathbf{K}$ in MS, there exists another Killing vector $\mathbf{L}$ (which may be null) such that the following statements hold:

(1) $\mathbf{K}\wedge\mathbf{L}\neq 0$, $[\mathbf{K},\mathbf{L}] = 0$.

(2) If $\mathbf{K}\wedge\mathbf{L}$ is not null, there exists a coordinate basis $\mathbf{X}_\alpha$ for which $\mathbf{X}_3 = \mathbf{L}$, $\mathbf{X}_4 = \mathbf{K}$, and the components of the MS metric $g$ satisfy equations of the form (37).

(3) If $\mathbf{K}\wedge\mathbf{L}$ is null, there exists a coordinate basis $\mathbf{X}_\alpha$ for which $\mathbf{X}_1 = \mathbf{L}, \mathbf{X}_4 = \mathbf{K}$, and the components for the MS metric $g$ satisfy equations of the form (38).

There is another theorem which is well known from the study[5] of axially symmetric stationary vacuums and which also holds for those axially symmetric stationary electrovacs which have been considered, for example, by Kinnersley and Chitre.[18] The theorem claims, in effect, that if the above statements (1) and (2) hold for the Killing vectors $\mathbf{K}$ and $\mathbf{L}$ and for the metric $g$ of the axially symmetric stationary electrovac $V_4$, then there exists a gauge for $M^{rs}$ such that

(A) $\mathbf{L}$ is a KV of every $V_4'$ in $\mathscr{H}_{\mathbf{K}}(V_4)$,

(B) Equations (37) hold (relative to the coordinate basis $\mathbf{X}_\alpha$) for every $V_4'$ in $\mathscr{H}_{\mathbf{K}}(V_4)$.

Of course, both of the above statements are true for those of our results which correspond to nonnull $\mathbf{K}\wedge\mathbf{L}$ (i.e., nonnull surfaces of transitivity).

*The point is that, if Eqs. (37) are replaced by Eqs. (38), they are also true for those of our results which correspond to null* $\mathbf{K}\wedge\mathbf{L}$. Statement (A) can be proven directly from Eq. (15) or can be verified by inspecting Eqs. (31e), and (B) has already been discussed. *The truth of these statements when* $V_4 = \mathrm{MS}$ *raises the question of whether they hold more generally* and whether methods similar to those employed for axially symmetric stationary electrovacs can be applied to an appropriately restricted (but nontrivial) class of electrovacs which admit two-parameter Abelian isometry groups with null surfaces of transitivity.

Other interesting properties of our metrics derive from a basic theorem[26] concerning any electrovac $V_4 = (M,g,F)$ with a nonnull Killing vector $\mathbf{K}$, and any member $V_4' = (M',g',F')$ of $\mathscr{H}_{\mathbf{K}}(V_4)$.

*Theorem*: If $\mathbf{A}$ and $\mathbf{B}$ are any tangent vector fields in $M$, then at all points in the intersection of $M'$ with the domains of $\mathbf{A}$ and of $\mathbf{B}$,

$$(\mathbf{K}\wedge\mathbf{A})\urcorner(\mathbf{K}\wedge\mathbf{B}) = (\mathbf{K}\wedge\mathbf{A})\urcorner'(\mathbf{K}\wedge\mathbf{B}). \tag{39}$$

The step product employed above is defined for any tangent vectors $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, by

$$(\mathbf{A}\wedge\mathbf{B})\urcorner(\mathbf{C}\wedge\mathbf{D}) = (\mathbf{A}\urcorner\mathbf{C})(\mathbf{B}\urcorner\mathbf{D}) - (\mathbf{A}\urcorner\mathbf{D})(\mathbf{B}\urcorner\mathbf{C}),$$

$$\mathbf{A}\urcorner\mathbf{C} := g(\mathbf{A},\mathbf{C}).$$

If $g$ is replaced by $g'$, then $\urcorner'$ replaces $\urcorner$.

We are presently interested in the application of the above theorem to the components of $g$ and $g'$ relative to a coordinate basis $\mathbf{X}_\alpha$ such that $\mathbf{X}_4 = \mathbf{K}$. We have then that all components of

$$\Delta_{\alpha\beta} := g_{\alpha\beta}g_{44} - g_{\alpha 4}g_{\beta 4} = g'_{\alpha\beta}g'_{44} - g'_{\alpha 4}g'_{\beta 4} \tag{40}$$

are invariant under $\mathscr{H}_{\mathbf{K}}$. In particular, consider those of our metrics which satisfy $\Delta_{33}\neq 0$, whereupon Eqs. (37) and (25) yield

$$g'_{44} = |\Lambda|^{-2}g_{44}, \tag{41a}$$

$$g'_{ij} = |\Lambda|^2 g_{ij}, \quad i=1,2, \quad j=1,2. \tag{41b}$$

The above equations give us a better feeling for the relations between our metrics $g'_{\alpha\beta}$ and the original MS metric $g_{\alpha\beta}$. Clearly, the same result must hold for any axially symmetric stationary spacetime and is, in fact, already well known.

Next, consider those of our metrics which satisfy $\Delta_{33} = 0$. Equation (41a) is still true. From Eqs. (38), we obtain the sample relations

$$(g'_{14})^2 = (g_{14})^2, \quad g'_{22} = |\Lambda|^2 g_{22}.$$

Note that the first of the above equations is consistent with the value $g'_{14} = a$ given in Eqs. (38).

The $\mathscr{H}_{\mathbf{K}}$-invariant $\Delta_{33}$ is positive, negative, or zero according as the 2-surface of transitivity generated by $\mathbf{K}$ and $\mathbf{L}$ is spacelike, timelike, or null. In the sequel, we use the notation[27]

$$\epsilon := -\operatorname{sgn}\Delta_{33} = 1, -1, \text{ or } 0. \tag{42}$$

The value of $\epsilon$ can be different in different chart domains $M_i$ corresponding to the same $\mathbf{K}$.

As is clear from Eqs. (34d) and (36d), if $b_E = 0$, then $V_4'$ is a Minkowski space whch can differ at most from the original one by a uniform conformal mapping combined with a gauge transformation. *In the following* Secs. B–I *which are concerned with properties of the specific results corresponding to different Killing vector types, we shall assume* $b_E\neq 0$ *unless we explicitly say otherwise.*

## B. When K is *(N,N)*

Equations (34) directly imply that $V_4'$ *has a type* $N_{pp}$ *gravitational field*. Moreover, *the Maxwell field is zero or is also type N, and its null rays are the same as the principal null rays of the conform tensor.*

Equations (27) supply $k', m', t'$ in terms of the differentials of the coordinates $x,y,\rho,\sigma$. Now, though this null tetrad and chart are suited to our particular problem, they are not the ones used in any conventional formulation of the *general* type $N_{pp}$ electrovac solution. A fairly conventional form of the general solution[28] is given by a null tetrad $k^{(2)}, m^{(2)}, t^{(2)}, t^{(2)*}$ and coordinates $z, z^*, \rho, \sigma$ (complex $z$) such that

$$k^{(2)} = d\sigma, \quad t^{(2)} = dz,$$

$$m^{(2)} = d\rho + (H + H^* - \tfrac{1}{2}\Psi\Psi^*)d\sigma,$$

$$H = H(z,\sigma), \quad \Psi = \Psi(z,\sigma).$$

The corresponding connection forms, Weyl conform tensor components, and Maxwell field components are

$$v_1^{(2)} = v_0^{(2)} = 0, \quad v_{-1}^{(2)} = -H_z + \tfrac{1}{2}\Psi_z\Psi^*,$$

$$C_i^{(2)} = 0 \text{ if } i\neq -2, \quad C_{-2}^{(2)} = H_{zz} - \tfrac{1}{2}\Psi_{zz}\Psi^*,$$

$$F_1^{(2)} = F_0^{(2)} = 0, \quad F_{-1}^{(2)} = -\tfrac{1}{2}\Psi_z,$$

where, e.g., $H_z := \partial H/\partial z$. *For the special case of the* $(N,N)$ *result* given by Eqs. (27), $\rho$ and $\sigma$ are the same coordinates as in Eqs. (37), and

$$z = 2^{-1/2}[b_O(x+iy) - ib_E(x+iy)^2],$$

$$H = -(4b_E^*\Lambda)^{-1}, \quad \Psi = -2b_M H.$$

Thus, $H$ and $\Psi$ are independent of $\sigma$ in our $(N,N)$ case.

We next consider the Killing vectors of $V_4'$. We give no derivations and simply state that the *group of all motions of* $V_4'$ *is a* $G_2$ whose generators are

$$\mathbf{X}_\rho = \mathbf{e}_k = |\Lambda|\mathbf{k}', \quad \mathbf{X}_\sigma = \mathbf{K}, \tag{43}$$

which are also Killing vectors in MS. Here, $\epsilon = 1$.

Consider the set

$$\mathscr{H}_{\mathbf{A}}\mathscr{H}_{\mathbf{K}}(\mathrm{MS}) := \text{the union of the sets } \mathscr{H}_{\mathbf{A}}(V_4')$$
$$\text{for all } V_4' \text{ in } \mathscr{H}_{\mathbf{K}}(\mathrm{MS})$$
$$(\text{allowing } b_E = 0),$$

where

$$\mathbf{A} = c_1\mathbf{X}_\rho + c_2\mathbf{X}_\sigma, \quad c_2\neq 0,$$

$$c_1, c_2 = \text{real constants}.$$

It can easily be shown from the $(N,N)$ expressions for $\mathbf{a}$, $\boldsymbol{\omega}$ in Table I and from Eqs. (1) and (43) that $\mathbf{A}$ is type $(N,N)$ for all $c_1$ and all $c_2\neq 0$. (To prove this, use a translation of the origin along the 2-axis.) Therefore,

$$\mathscr{H}_{\mathbf{A}}(\mathrm{MS}) = \mathscr{H}_{\mathbf{K}}(\mathrm{MS}).$$

Moreover, from the group concept,

$$\mathscr{H}_{\mathbf{A}}\mathscr{H}_{\mathbf{A}} = \mathscr{H}_{\mathbf{A}}(\mathrm{MS}).$$

We conclude that

$$\mathscr{H}_{\mathbf{A}}\mathscr{H}_{\mathbf{K}}(\mathrm{MS}) = \mathscr{H}_{\mathbf{K}}(\mathrm{MS}).$$

So, the $(N,N)$ results represents a "dead end" in the sense that no new electrovacs are generated by a second application of any of the groups $\mathscr{H}_{\mathbf{A}}$; one must first return to a Minkowski space.

$V'_4$ turns out to have an irreducible Killing tensor[29] whose Segre characteristic is [(1)(1)(2)]. The Hamiltonian for geogesic orbits is given by

$$H := g^{\alpha\beta}p_\alpha p_\beta = (\lambda^{(2)} - \lambda^{(1)})^{-1}(H_1 + H_2) + 2p_3p_4,$$

$$H_1 := p_1p_1, \quad H_2 := p_2p_2 + 2yp_3p_3,$$

$$\lambda^{(1)} := |b_O|^2 - |b_O - 2ib_E x|^2, \quad \lambda^{(2)} := |b_O + 2ib_E y|^2,$$

$$x^1 := x, \quad x^2 := y, \quad x^3 := \rho, \quad x^4 := \sigma.$$

The Killing tensor $\kappa^{\alpha\beta}$ is given by

$$\kappa := \kappa^{\alpha\beta}p_\alpha p_\beta = (\lambda^{(2)} - \lambda^{(1)})^{-1}(\lambda^{(2)}H_1 + \lambda^{(1)}H_2).$$

Thus, we have four independent constants of geodesic motion, viz., $p_3, p_4, H, \kappa$. The fields $\lambda^{(1)}$ and $\lambda^{(2)}$ are nondegenerate eigenvalues of $\kappa^\beta_\alpha$ corresponding to eigenvectors $\mathbf{t}' - \mathbf{t}'^*$ and $\mathbf{t}' + \mathbf{t}'^*$, respectively; observe that $\lambda^{(2)} - \lambda^{(1)} = |A|^2$. The third eigenvalue is 0; it has index 2 and corresponds to the eigenvector $\mathbf{k}$ and the generalized eigenvector $\mathbf{m}$.

## C. When K is *(S,N)* or *(0,N)*

*All statements made in the first paragraph of* Sec. 7B *concerning the type* $N_{pp}$ *characters of the gravitational and Maxwell fields also apply here.*

As regards the problem of transforming the forms of Eqs. (28) into the conventional forms of Eqs. (37), the transformation leads to unwieldy expressions, and there are no ignorable coordinates in the resulting line element. We see no advantage in using the conventional forms for the $(S,N)$, $(0,N)$ results.

The Killing vector structure turns out to be rich. $V'_4$ has five independent Killing vectors:

$$\mathbf{X}^{(1)} := \mathbf{X}_\rho = \mathbf{e}_k, \quad \mathbf{X}^{(2)} := \mathbf{X}_\xi, \quad \mathbf{X}^{(3)} := \mathbf{X}_\eta = \mathbf{K},$$

$$\mathbf{X}^{(4)} := \xi\mathbf{X}_\rho - f_{11}(\sigma)\mathbf{X}_\xi - f_{12}(\sigma)\mathbf{X}_\eta,$$

$$\mathbf{X}^{(5)} := \eta\mathbf{X}_\rho - f_{12}(\sigma)\mathbf{X}_\xi - f_{22}(\sigma)\mathbf{X}_\eta,$$

where $f_{ij}(\sigma)$ are defined by (dots denote derivatives with respect to $\sigma$)

$$\dot f_{11} = (\sigma^2 - a^2)^{-2}(\sigma^2 + a^2),$$

$$\dot f_{12} = (\sigma^2 - a^2)^{-2}[(\sigma^2 + a^2)J - 2a\sigma|A|^2],$$

$$\dot f_{22} = (\sigma^2 - a^2)^{-2}[(\sigma^2 + a^2)(J^2 + |A|^4) - 4a\sigma J|A|^2].$$

The corresponding Killing bivector components, relative to the same basis $B'_4$ which was used in Eqs. (35) and defined by Eq. (32), all satisfy

$$\omega_1^{(i)\prime} = \omega_0^{(i)\prime} = 0, \quad (1 \leqslant i \leqslant 5). \tag{44}$$

$\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \mathbf{X}^{(3)}$ are our old friends $\mathbf{X}_1, \mathbf{X}_3, \mathbf{X}_4$, *but* $\mathbf{X}^{(4)}$ *and* $\mathbf{X}^{(5)}$ *are not generally Killing vectors in* MS. Specifically (recall that we are assuming $b_E \neq 0$), $\mathbf{X}^{(4)}$ is an MSKV if and only if $a = 0$ and $b_O b_E^{-1}$ is real, and $\mathbf{X}^{(5)}$ is not an MSKV for any parameter values (with $b_E \neq 0$).

The Killing vectors are not invariant under a gauge transformation $M^{rs} \to M^{rs} + dF^{rs}$. They undergo a corresponding transformation $\mathbf{X}^{(i)} \to \mathbf{Y}^{(i)}$. We have investigated this transformation in detail and find that there is a gauge for which $\mathbf{Y}^{(4)}$ is an MSKV, but $\mathbf{Y}^{(2)}$ is not an MSKV in that gauge. As regards $\mathbf{Y}^{(5)}$, it is not an MSKV for any gauge.

With only one exception, the above $G_5$ group of motions is all $V'_4$ has. *The exception occurs when*

$$b_O = 0, a = 0,$$

*whereupon* $V'_4$ *is a vacuum metric which admits a* $G_6$. The sixth KV is

$$\mathbf{X}^{(6)} = \tfrac{1}{5}(-\sigma\mathbf{X}_\sigma + 5\rho\mathbf{X}_\rho + 3\xi\mathbf{X}_\xi - \eta\mathbf{X}_\eta),$$

with a corresponding bivector whose components satisfy

$$\omega_1^{(6)\prime} = 0, \quad \omega_0^{(6)\prime} = 1, \quad \omega_{-1}^{(6)\prime} \neq 0. \tag{45}$$

There is one parameter $b_E$ which is left when we set $b_O = a = 0$. However, it can be absorbed by a coordinate transformation, and then there are none.

The vacuum subcase of the $(S,N)$, $(0,N)$ results in a particular example of a known class of $N_{pp}$ gravitational fields which admit a $G_5$. The special case $b_O = a = 0$ may be in one of the families of type $N_{pp}$ vacuums which are given explicitly and are known to have a $G_6$. Derivations and some details concerning these $G_5$ and $G_6$ admitting solutions can be found in Petrov's book.[17]

Next, we make some observations about the set

$$\mathscr{H}_{\mathbf{A}}\mathscr{H}_{\mathbf{K}}(\mathrm{MS}),$$

where

$$\mathbf{A} = \sum_i \{c_i\mathbf{X}^{(i)}\}, \quad c_i = \text{real constant},$$

$$\text{at least one } c_i \neq 0 \text{ for } i > 1.$$

Let $V''_4$ denote any member of the above set. The following statements hold:

(1) Suppose $c_i = 0$ for all $i > 3$, and $|c_1| \neq |c_2|$ if $a = 1$. Then, $\mathbf{A}$ is $(0,N)$ if $\mathbf{K}$ is $(0,N)$, and $\mathbf{A}$ is $(S,N)$ if $\mathbf{K}$ is $(S,N)$. Moreover,

$$\mathscr{H}_{\mathbf{A}}\mathscr{H}_{\mathbf{K}}(\mathrm{MS}) = \mathscr{H}_{\mathbf{K}}(\mathrm{MS}).$$

(2) Suppose $c_6 = 0$, and $V'_4$ or $V''_4$ is a vacuum. Then $V''_4$ is either a Minkowski space or has a type $N_{pp}$ gravitational field.

(3) Suppose $c_6 = 0$, and $V'_4$ is a vacuum. Then the Maxwell field of $V''_4$ is zero or is type $N_{pp}$ with null rays which coincide with those of the gravitational field.

(4) Recall that $\mathbf{X}^{(6)}$ is a KV of $V'_4$ if and only if $b_O = a = 0$. If $b_O = a = 0$ and $c_6 \neq 0$, then $V''_4$ is type II and has a principal null vector $\mathbf{k}''$ which is generally (i.e., except perhaps for special parameter values) diverging and twisting. $F''$ is not $N$ and has $\mathbf{k}''$ as an eigenvector.

The proof of statement (1) is similar to the proof of the analogous statement in Sec. 7B. The other statements are proven with the aid of the transformation formulas in (I)[7] and Eqs. (35), (44), and (45).

## D. When K is *(0,P)*

From Eqs. (36) and (9c), $V_4'$ has a type $D$ gravitational field, both principal null congruences have twist as well as expansion, and $\mathbf{k}'$ and $\mathbf{m}'$ are eigenvectors of the Maxwell field as well as principal null vectors of the conform tensor. The Maxwell field is nonsingular almost everywhere, i.e., its eigenvalues are not zero except on a set of measure zero.

The group of all motions is $G_2$ generated by $\mathbf{X}_3$ and $\mathbf{X}_4$. As inspection of Eqs. (11c) will corroborate, the trichotomic variable $\epsilon$ which was defined by Eq. (42) and which identifies the signature of the 2-surfaces of transitivity is equal, for $(0,P)$, to

$$\epsilon = \epsilon_0 := \epsilon_k \epsilon_m.$$

Let $V_4(\epsilon_k,\epsilon_m,\phi_1,\phi_2)$ denote the restriction of MS to that simply connected open subset of $R^4$ in which $q_k$ and $q_m$ [see Eqs. (4) and (5d)] have respective signs $\epsilon_k$ and $\epsilon_m$, and in which $\phi_1<\phi<\phi_2$. Then, $\mathscr{H}_\mathbf{K}(\text{MS})$ is the union of at least eight electrovac families

$$\mathscr{H}_\mathbf{K}[V_4(\epsilon_k,\epsilon_m,\phi_1,\phi_2)].$$

The two electrovacs corresponding to the same value of $\alpha$, $b_O$, $b_E$, $b_M$, $\phi_1$, $\phi_2$, and $\epsilon$ are clearly isometric. The problem of "joining" the electrovacs corresponding to the same values of $\alpha$, $b_O$, $b_E$, and $b_M$ is covered in Sec. 7K.

We have made only a limited analysis of what happens if we use the KV

$$\mathbf{A} = c_1\mathbf{X}_3 + c_2\mathbf{X}_4, \quad c_1\neq 0,$$

to generate new electrovacs $V_4''$ by applying $\mathscr{H}_\mathbf{A}$ to $V_4'$. A key role in that analysis is played by the fact that the Killing bivector $\mathbf{v}$ corresponding to $\mathbf{X}_4$ has components

$$v_1' = \epsilon v_{-1}'\neq 0, \quad v_0'\neq 0,$$

relative to the same basis $B_A'$ which was used for Eqs. (36). Upon restricting ourselves to the case where $V_4'$ is a vacuum, we have been able to prove that $V_4''$ is never type $N$ or type III. Also, $V_4''$ is type I whenever $|b_E|$ is chosen "sufficiently small" but positive, and $V_4'$ is a vacuum; this suggests that $V_4''$ will generally be type I except perhaps for special parameter values.

## E. When K is *(S,S)* or *(0,S)*

The pertinent equations for this case are (30b), (31b), and (36).

With some exceptions, the statements made in the first paragraph of the preceding Sec. 7D are also valid here. The exceptions occur when $a=0$ and $b_O b_E^{-1}$ is real, in which case *both principal null congruences have zero twist*, but do have nonzero expansion; if, in addition, $b_M^* b_E^{-2}$ is real or imaginary, then the Maxwell field is singular, but is not null.

For any given values of the parameters $a$, $b_O$, $b_E$, $b_M$, there are exactly four electrovacs defined over disjoint submanifolds of $R^4$ in which

$$\epsilon = \epsilon_0 := \epsilon_k\epsilon_m = \pm 1.$$

The situation is similar to that described in the second paragraph of the preceding Sec. 7D, but the complications arising from the use of the coordinate $\phi$ in Sec. 7D can be deleted when discussing $(S,S)$ and when discussing $(0,S)$ regarded as a limiting case of $(S,S)$.

The group of all motions of any $(S,S)$ $V_4'$ is the $G_2$ generated by $\mathbf{X}_3$ and $\mathbf{X}_4$, and exactly the same remarks apply here as those made in the final paragraph of the preceding Sec. 7D concerning the generation of further electrovacs $V_4''$.

In contrast, the group of all motions of any $(0,S)$ $V_4'$ is a $G_4$ generated by

$$\begin{aligned}
&\mathbf{X}^{(1)} := \mathbf{X}_\chi, \quad \mathbf{X}^{(2)} := \mathbf{X}_x, \quad \mathbf{X}^{(3)} = \mathbf{X}_y + 2bx\mathbf{X}_\chi,\\
&\mathbf{X}^{(4)} := x\mathbf{X}_y - y\mathbf{X}_x + b\,(x^2-y^2)\mathbf{X}_\chi, \qquad (46)\\
&b := i(b_O^* b_E - b_O b_E^*).
\end{aligned}$$

$\mathbf{X}^{(3)}$ and $\mathbf{X}^{(4)}$ are MS Killing vectors if and only if $b=0$. If we select a different gauge, then $\mathbf{X}^{(2)}$, $\mathbf{X}^{(3)}$, $\mathbf{X}^{(4)}$ are transformed and generally become different vector fields, but the number of them which are also MS Killing vectors is not thereby increased (though it can decrease). The discussion on this point is similar to that given in Sec. 7C.

The bivectors corresponding to the Killing vector $\mathbf{X}^{(i)}$ satisfy the equalities

$$\begin{aligned}
&\omega_1^{(1)\prime} = \omega_{-1}^{(1)\prime} = 0,\\
&\omega_1^{(2)\prime} = \epsilon\omega_{-1}^{(2)\prime} = b_E\,\epsilon_k s A^{-1},\\
&\omega_1^{(3)\prime} = -\epsilon\omega_{-1}^{(3)\prime} = i\omega_1^{(2)\prime},\\
&\omega_1^{(4)\prime} = -\epsilon\omega_{-1}^{(4)\prime} = ir\omega_1^{(2)\prime}.\\
&\omega_0^{(i)\prime}\neq 0 \text{ for all } i.
\end{aligned}$$

In view of the above, we expect type I (except perhaps for special parameter values) if one computes $\mathscr{H}_\mathbf{A}(V_4')$, where

$$\mathbf{A} = \sum_i c_i\mathbf{X}^{(i)}, \quad c_i\neq 0 \text{ for at least one } i>1.$$

## F. When K is *(T,T)*

The pertinent equations are (30b), (31c), and (36). The first paragraph of Sec. 7D on the principal null rays and on the Maxwell field of $(0,P)$ are valid here. The group of all motions is a $G_2$, and the statements concerning $\mathscr{H}_\mathbf{A}(V_4')$ in the last paragraph of Sec. 7D also apply.

For any given values of the parameters $a$, $b_O$, $b_E$, $b_M$, there are at least two electrovacs in $\mathscr{H}_\mathbf{K}(\text{MS})$ defined on overlapping submanifolds of $R^4$ in which

$$\phi_1<\phi<\phi_2, \quad r>0.$$

The situation is similar to that described in the second paragraph of Sec. 7D except that, here,

$$\epsilon = 1$$

throughout each domain.

## G. When K is *(S, T)*

The $(T,T)$ text can be adopted almost verbatim. The only significant difference is that $\epsilon = -1$ for $(S,T)$.

## H. When K is *(N, T)*

We have already discussed some unusual features of these type $D$ electrovacs in Sec. 7A. There is only a $G_2$ with null surfaces ($\epsilon = 0$) of transitivity. The chart domains are as described for $(T,T)$.

As regards the principal null rays, $\mathbf{k}'$ has diverging rays which are also twisting, but $\mathbf{m}'$ is nondiverging. The Maxwell field has the same null rays and is nonsingular.

The bivectors $\boldsymbol{\omega}'$ and $\mathbf{v}'$ corresponding to the Killing vectors $\mathbf{X}_3$ and $\mathbf{X}_4$ have components which satisfy:

$$\omega'_1 = \omega'_{-1} = 0, \quad \omega'_0 \neq 0,$$

$$v'_1 = 0, \quad v'_{-1} = \sqrt{2} b_E \Lambda^{-1} r, \quad v'_0 \neq 0.$$

If we let

$$\mathbf{A} = c_1 \mathbf{X}_3 + c_2 \mathbf{X}_4, \quad c_1 \neq 0,$$

then it follows that the members of $\mathscr{H}_{\mathbf{A}}(V'_4)$ are *algebraically special. In fact, they are type* II except perhaps for special parameter values.

## I. When K is *(0, T)*

This $\mathscr{H}_{\mathbf{K}}(\text{MS})$ is the family which contains Melvin's magnetic universe[10–14] ($b_O = 1$ and $b_E = \frac{1}{4}B^2$, where $B$ is the magnetic field magnitude). Both principal null congruences are nondiverging. $\mathbf{k}'$ and $\mathbf{m}'$ are both eigenvectors of the Maxwell field, which is singular if and only if $b_O b_E^{-1}$ and $ib_M^* b_E^{-2}$ are real.

The $(0,T)$ case is exceptional in the variety of ways $V'_4$ can be labeled $\epsilon = 1, -1$, or 0 depending on which MS Killing vector is paired with $\mathbf{K}$ to give us $\mathbf{X}_3$ and $\mathbf{X}_4$. [Recall the definition of $\epsilon$ in Eq. (42).] $\epsilon = 1$ for that $(0,T)V'_4$ which is obtained by taking the limit of a $(T,T)V'_4$ as $a \to 0$, and $\epsilon = -1$ for that $(0,T)\,V'_4$ which is the limit of an $(S,T)\,V'_4$ as $a \to 0$. In the case of the limit of a $(0,P)\ V'_4$ as $\alpha \to \frac{1}{2}\pi$, we obtain a family of four electrovacs defined over different domains of $R^4$, and $\epsilon = \epsilon_k \epsilon_m$ in the respective domains. Finally, $\epsilon = 0$ if we obtain our $(0,T)\ V'_4$ as the limit of an $(N,T)\ V'_4$ as $a \to 0$. The various $(0,T)$ electrovacs which are thus derived are mutually isometric in a local sense. The different $\epsilon$ and the different forms obtained for the metrics are merely due to the different choices of $\mathbf{X}_3$, We will stick with

$$\mathbf{X}_3 = \mathbf{X}_\rho = \mathbf{e}_k$$

the the immediate sequel.

The group of all motions is a $G_4$ with generators

$$\mathbf{X}^{(1)} := \mathbf{X}_\phi, \quad \mathbf{X}^{(2)} := \mathbf{X}_\rho,$$

$$\mathbf{X}^{(3)} := \mathbf{X}_\sigma - 2b\rho\mathbf{X}_\phi, \quad \mathbf{X}^{(4)} := \rho\mathbf{X}_\rho - \sigma\mathbf{X}_\sigma,$$

where $b$ is defined in Eqs. (46). $\mathbf{X}^{(3)}$ and $\mathbf{X}^{(4)}$ are MS Killing vectors if and only if $b = 0$, whereupon $\mathbf{X}^{(4)}$ reduces to the familiar boost generator

$$z\partial_T - T\partial_z.$$

If we had selected a different gauge, then $\mathbf{X}^{(2)}, \mathbf{X}^{(3)}, \mathbf{X}^{(4)}$ would generally be different than the present triad, but the number of them which are MS Killing vectors would be (if $b \neq 0$) greater than or equal to 2 in any case.

The Killing bivector components satisfy the equalities:

$$\omega_1^{(1)\prime} = \omega_{-1}^{(1)\prime} = 0,$$

$$\omega_1^{(2)\prime} = 0, \quad \omega_{-1}^{(2)\prime} = \sqrt{2} b_E r \Lambda^{-1},$$

$$\omega_1^{(3)\prime} = \omega_{-1}^{(2)\prime}, \quad \omega_{-1}^{(3)\prime} = 0,$$

$$\omega_1^{(4)\prime} = -\sigma\omega_{-1}^{(2)\prime}, \quad \omega_{-1}^{(4)\prime} = \rho\omega_{-1}^{(2)\prime}.$$

Suppose we use a linear combination

$$\mathbf{A} = \sum_i c_i \mathbf{X}^{(i)}, \quad c_i \neq 0 \text{ for at least one } i > 1,$$

to construct a new electrovac $V''_4$ by applying $\mathscr{H}_{\mathbf{A}}$ to $V'_4$. If we restrict ourselves to the case where $V'_4$ is a vacuum and let $c_3 = c_4 = 0$ or $c_2 = c_4 = 0$, then $V''_4$ is type II. We have not done anything further in the way of detailed analysis for other values of $c_i$, but it is clear that $V''_4$ will be type I except perhaps for special parameters values.

## J. Singularities

We next consider the problem of singularities in $V'_4$. This is an immense topic which can property be treated only by a detailed analysis of each separate case. However, we have not carried out any thorough analysis. Nor are we sure that our particular type $N_{pp}$ and $D$ metrics merit a thorough analysis any more than any other type $N_{pp}$ and $D$ metrics. All we shall do here is go over some key points which stress the dependence of the singularities, *especially with regard to their existence and their distribution in* $R^4$, on the type of KV and on the transformation parameters $b_O$ and $b_E$. When we refer to singularities here, we mean, of course, the singularities of $V'_4$ and not of any extensions of $V'_4$.

Let the set of all zeros of $\Lambda$ be denoted by $Z(\Lambda)$. Consider the example of the $(N,N)$ results in Eqs. (27a), (27b), and (34a)–(34d). $Z(\Lambda)$ is the timelike 2-surface (in MS) of transitivity

$$y = -\tfrac{1}{2}\beta, \quad x = \tfrac{1}{2}\gamma,$$

$$\beta := \text{Rl}(b_O b_E^{-1}), \quad \gamma := \text{Im}(b_O b_E^{-1}).$$

Are any points on this 2-surface "genuine" singularities of $V'_4$? Since the curvature tensor is type $N$, it cannot by itself help us answer this question. However, $V'_4$ has two Killing vectors and a Killing tensor which make the H–J equation for geodesic orbits separable. So, the question can be answered by analysis. We have made only a few exploratory calculations which show that at least some zeros of $\Lambda$ may be regarded as singularities when $\beta > 0$. Specifically, if $\beta > 0$, there are at least some timelike geodesic line segments of *finite proper length* which have zeros of $\Lambda$ as limit points. Also, when $\beta > 0$, the mutually orthogonal vectors $K^{\alpha'}$ and $\omega^{\alpha'}{}_\beta K^{\beta'}$ are spacelike, and the corresponding Riemannian curvature blows up as we approach any zero of $\Lambda$.

For $(S,N)$ and $(0,N)$, $Z(\Lambda)$ is the set of points such that

$$\sigma^2 + (a)^2 = \beta, \quad 2a\sigma = -\gamma \quad (a = 1,0).$$

$Z(\Lambda)$ is empty except when

$$\beta = 1 + \tfrac{1}{4}\gamma^2 \geqslant 1 \quad \text{if } a = 1,$$

$$\beta \geqslant 0, \quad \gamma = 0 \quad \text{if } a = 0.$$

When $Z(\Lambda)$ is not empty, it consists of the null hypersurfaces of transitivity (in MS)

$$\sigma = -\tfrac{1}{2}\gamma \quad \text{if } a = 1,$$

$$\sigma = \pm\beta \quad \text{if } a = 0.$$

We have not further looked into the singularity question in these cases, though the analysis would not be difficult in view of the profusion of Killing vectors.

The zeros of $\Lambda$ are assured singularities of each of the type $D$ electrovacs, since $C_0$ as given by Eqs. (36d) is a curvature invariant which blows up as $\Lambda \to 0$. For the $(0,P)$ results, $Z(\Lambda)$ is the set of points in $R^4$ such that

$$\epsilon s^2 = -\beta + \gamma \tan\alpha, \quad r^2 = \beta + \gamma \cot\alpha.$$

For $(S,S)$ and $(0,S)$,

$$\epsilon s^2 = -\beta + (a)^2, \quad 2ay = -\gamma, \quad a \geqslant 0.$$

For the $(\cdot\cdot,T)$ cases,

$$r^2 = (a^4)^2 - (a^3)^2 + \beta, \quad 2(a^3q^4 - a^4q^3) = -\gamma.$$

As before, $\beta$ and $\gamma$ are the real and imaginary parts of $b_O \times b_E^{-1}$. The parameters $a^3$, $a^4$, and the coordinates $r, s, y$ are defined by Eqs. (5).

There is a host of possibilities for these type $D$ cases, and the reader can easily work out a classification scheme for the zeros of $\Lambda$. As an example, consider $(T,T)$ for which $a^3 = 0$ and $a^4 = a > 0$. If $(a)^2 + \beta$ is negative, $Z(\Lambda)$ is clearly empty. If $(a)^2 + \beta$ is zero, $Z(\Lambda)$ is a timelike line in MS. If $(a)^2 + \beta$ is positive, $Z(\Lambda)$ is a timelike 2-surface in MS. As regards the geometric character of $Z(\Lambda)$ relative to $V_4'$ itself, that is a problem which we have not investigated.

## K. Extension to $R^4 - Z(\Lambda)$

The charts which were introduced in Sec. 2 and which we have been using until now were chosen:

(1) to include **K** as one of the coordinate basis vectors,

(2) to make the isometries as manifest as possible,

(3) to facilitate the calculation of $M^{rs}$,

(4) to obtain the results in forms which are similar to and can easily be compared with standard expressions for types $N_{pp}$ and $D$ electrovacs (especially the latter).

Though no explicit criteria were formulated concerning the gauges for $M^{rs}$, we did select them so as to satisfy the above conditions (2)–(4).

However, in spite of their advantages, some of those charts and gauges fail to give us what may be called a "global view" of the transformations $\mathscr{H}_{\mathbf{K}}$. In this final section, we shall attempt to supply such a view, and we shall select new charts and gauges which serve *that end* though they do not satisfy all of our previous criteria.

For given **K** and given $b_O, b_E, b_M$, and $\alpha$ or $a$, the electrovacs which we obtained constitute a family

$$V_{41}', \ldots, V_{4n}'$$

defined over respective domains $M_1', \ldots, M_n'$ of $R^4$. Here,

$$M_i' = M_i - Z(\Lambda),$$

where $M_1, \ldots, M_n$ are the domains of the charts which we defined in Sec. 2. We now want to investigate the following problems:

(1) When is it possible to *join* $V_{41}', \ldots, V_{4n}'$, i.e., to form a single electrovac $V_4'$ whose domain is $M' = R^4 - Z(\Lambda)$ and whose restriction to $M_i'$ is $V_{4i}'$?

(2) What is a suitable atlas and a suitable orthonormal or null tetrad for $V_4'$?

Consider the first question. The new metric is related to the old metric by the equation

$$g' = |\Lambda|^2 g - K \otimes \Delta - \Delta \otimes K - |\Lambda|^{-2} f \Delta \otimes \Delta$$

as can be seen from Eqs. (21) and (16b). The MS metric $g$ and the Killing covector $K$ each has $R^4$ as its domain. Therefore, *the various electrovacs $V_{4i}'$ can be joined if $\Delta$ can be extended to $R^4$*. Therefore, the answer to the first question hinges on the domains of the potentials $M^{rs}$.

As regards the second question, an obvious atlas is the set of Cartesian coordinate systems $q^a$ or the corresponding null coordinate systems $q_k, q_m, q_l, q_n$ for which **K** has the canonical form prescribed by Table I. When the transformation (21) is applied to the MS orthonormal tetrad $dq^a$, we obtain

$$e^{a'} = |\Lambda| dq^a - |\Lambda|^{-1} K^a \Delta. \qquad (47)$$

If we grant that $\Delta$ has $R^4$ as its domain, then the above equation supplies us with an orthonormal tetrad for $V_4'$ such that $M'$ is the domain of this tetrad. The Killing vector components $K^a$ in Eq. (47) are relative to the MS orthonormal basis $\mathbf{e}_a$ and can be obtained by inspecting the third and fourth columns of Table I.

Now, we get down to specifics.

The chart chosen for $(N,N)$ in Eqs. (5a) already has $R^4$ as its domain. Therefore, the same is true for the potentials $M^{rs}$ given by Eqs. (20) and the first line of Table I. Case $(N,N)$ needs no doctoring in this section.

As regards the other cases, we shall spare the reader all details concerning the use of Eqs. (20) and (5) and of Table I to compute $M^{rs}$ in terms of Cartesian or null coordinates. The important point is how the gauges of Eqs. (20) are altered. For some cases, we shall impose gauge transformations such that

$$\Delta \to \Delta + dF, \quad F = F(x^1, x^2, x^3).$$

For $(S,N)$,

$$F = (|b_O|^2 - 1)\sigma\xi.$$

For $(S,S)$,

$$F = (|b_O - a^2 b_E|^2 - 1)a^{-1}x.$$

For $(T,T)$, $(S,T)$, $(N,T)$,

$$F=(|b_0-\mathbf{a}\cdot\mathbf{a}b_E|^2-1)a^{-2}(a^3q^3+a^4q^4),$$

where

$$\mathbf{a}\cdot\mathbf{a}=(a^3)^2-(a^4)^2,\quad a=[(a^3)^2+(a^4)^2]^{1/2}. \tag{48}$$

$F=0$ for all other cases. $a^3,a^4$ in Eqs. (48) are defined by Eqs. (5g). The parameter $a$ is the same one which appears in Table I and which we have used throughout the paper. Thus, $a=1$ for $(N,T)$.

With the above alterations of gauge and with the use of the Cartesian coordinate $q^\alpha$, we obtain *new* expressions for $\Delta$.

For $(S,N)$,

$$\Delta=(|b_O|^2-1)dq^1+b_O^*b_E[-i(q^1+iq^2)d\sigma-(dq^1-idq^2)]$$
$$+b_Ob_E^*[i(q^1-iq^2)d\sigma-(dq^1+idq^2)]$$
$$+|b_E|^2\{[-2\sigma q^1-(\sigma^2+1)q^2]d\sigma$$
$$+(3\sigma^2+1)dq^1+\sigma(\sigma^2+3)dq^2\} \tag{49a}$$

For $(S,S)$,

$$\Delta=[(|b_O-a^2b_E|^2-1)a^{-1}+2bq^2+4|b_E|^2a(q^2)^2]dq^1$$
$$+[-\tfrac{1}{2}|b_M|^2+|b_E|^2(\epsilon_0s^2-2a^2)](q^3dq^4-q^4dq^3). \tag{49b}$$

For $(T,T)$, $(S,T)$, $(N,T)$

$$\Delta=[|b_O-\mathbf{a}\cdot\mathbf{a}b_E|^2-1+2b(a^3q^4-a^4q^3)$$
$$+4|b_E|^2(a^3q^4-a^4q^3)^2]a^{-2}(a^3dq^3+a^4dq^4)$$
$$+[\tfrac{1}{2}|b_M|^2+|b_E|^2(r^2+2\mathbf{a}\cdot\mathbf{a})](q^1dq^2-q^2dq^1). \tag{49c}$$

For $(0,N)$,

$$\Delta=(|b_O|^2-1)d(\sigma^{-1}q^2)-i[b_O^*b_E(q^1+iq^2)$$
$$-b_Ob_E^*(q^1-iq^2)]d\sigma$$
$$+|b_E|^2(-q^2\sigma^2d\sigma+\sigma^3dq^2). \tag{49d}$$

For $(0,P)$,

$$\Delta=\{(|b_O|^2-1)(\epsilon_0s^2)^{-1}\cos\alpha+b_O^*\widetilde{b}_E+b_O\widetilde{b}_E^*$$
$$+|b_E|^2\epsilon_0s^2\cos\alpha\}(q^3dq^4-q^4dq^3)$$
$$+\{(|b_O|^2-1)r^{-2}\sin\alpha-ib_O^*\widetilde{b}_E+ib_O\widetilde{b}_E^* \tag{49e}$$
$$+|b_E|^2r^2\sin\alpha\}(q^1dq^2-q^2dq^1).$$

To obtain the $(0,S)\Delta$, set $\alpha=0$ in the above Eq. (49e). To obtain the $(0,T)\Delta$, set $a=\tfrac{1}{2}\pi$. We recall that[30]

$$\sigma:=2^{-1/2}(q^3-q^4),\quad \epsilon_0s^2:=(q^3)^2-(q^4)^2,$$
$$r^2:=(q^1)^2+(q^2)^2,\quad q^3dq^4-q^4dq^3=\epsilon_0s^2d\chi, \tag{50}$$
$$q^1dq^2-q^2dq^1=r^2d\phi,\quad \epsilon_0=\pm1.$$

The parameters $\mathbf{a}\cdot\mathbf{a}$, $a^\alpha$, $\widetilde{b}_E$, $b$, $|b_M|^2$ are defined by Eqs. (48), (5g), (36e), (46), and (22b), respectively.

Inspection of Eqs. (49) enables us to draw the following conclusion.

*If* **K** *has no zeros, as is true for* $(N,N)$, $(S,N)$, $(S,S)$, $(T,T)$, $(S,T)$, $(N,T)$, *then there exists a choice of gauges for* $M^{rs}$ *such that each member of* $\mathscr{H}_{\mathbf{K}}(\mathrm{MS})$ *has*

$$R^4-Z(A)$$

*as its domain. If we restrict* $\mathscr{H}_{\mathbf{K}}$ *by the condition* $|b_O|=1$, *then the same conclusion applies to* $(0,N)$, $(0,P)$, $(0,S)$, *and* $(0,T)$. If **K** has at least one zero and if $|b_O|\neq1$, the points at which $s=0$ or $r=0$ (as the case may be) are excluded[31] from the domains of the electrovacs $V'_{4i}$, and the situation is not significantly different from that described in preceding sections where we used the curvilinear coordinates $x^\alpha$.

[1]The class of all electrovacs is understood to include all vacuums.

[2]This family is uniquely determined up to a gauge transformation, which is discussed later.

[3]J. Ehlers, *Les théories relativistes de la gravitation* (CNRS, Paris, 1959).

[4]B.K. Harrison, J. Math. Phys. **9**, 1744 (1968).

[5]R. Geroch, J. Math. Phys. **12**, 918 (1971); **13**, 394 (1972).

[6]W. Kinnersley, J. Math. Phys. **14**, 651 (1973).

[7]I. Hauser and F.J. Ernst, J. Math. Phys. **19**, 1316 (1978).

[8]No asymptotically flat electrovac has ever been constructed from MS by a *finite* number of such iterations, and many believe that it cannot be done. At the same time, no one has proven to date it is impossible.

[9]B. Carter, Commun. Math. Phys. **10**, 280 (1968); Phys. Lett. A **26**, 399 (1968).

[10]W.B. Bonnor, Proc. Phys. Soc. London Ser. A **67**, 225 (1953).

[11]M.A. Melvin, Phys. Rev. **139**, B225 (1965).

[12]K.S. Thorne, Phys. Rev. **139**, B244 (1965).

[13]F.J. Ernst, J. Math. Phys. **17**, 54 (1976).

[14]W. Wild, Bull. Am. Phys. Soc. **22**, 105 (1977).

[15]W. Kinnersley, J. Math. Phys. **10**, 1195 (1969), and Ph.D. Thesis, California Institute of Technology (1968).

[16]J.F. Plebański, Ann. Phys. **90**, 190 (1975) and a recent preprint.

[17]The vacuum subcases of these electrovacs are particular examples of a type $N_{pp}$ family of gravitational fields which admit a $G_5$ and which are given by A.Z. Petrov, *Einstein Spaces* (Pergamon, London, 1969), p. 186. That one of our type $N_{pp}$ spacetimes which admits a $G_6$ is a vacuum and is a member of the general class of type II or $\mathrm{II}_d$ gravitational fields of maximum mobility; this class is discussed by Petrov on pp. 149–54 of the same reference. As a note of caution, Petrov uses the designation $T_2$ to include both type II and type $\mathrm{II}_d$ (i.e., $N$) spacetimes; nor does he point out the *pp*-character of the solutions on p. 186.

[18]W. Kinnersley, J. Math. Phys. **18**, 1529 (1977); W. Kinnersley and D.M. Chitre, J. Math. Phys. **18**, 1538 (1977), and J. Math. Phys. **19**, 1926 (1978); **19**, 2037 (1978).

[19]We have no reference on the conjugacy classes of the extended Poincaré group, but we are certain that they must have been studied and worked out in the past.

[20]Strictly, $\mathscr{H}_{\mathbf{K}}(\mathrm{MS})$ contains, because of the gauge arbitrariness, submanifolds of MS as well as the original MS.

[21]The terms of the form $ax^\alpha$ in Eqs. (5b), (5f), and (5g) arise from our insistence that **K** be one of the coordinate basis vectors. Alternative coordinate systems, which the reader may prefer, are given in Sec. 7K of this paper.

[22]For exact correspondences of our null tetrad formalism with those of others, see F. Ernst, J. Math. Phys. **19**, 489 (1978).

[23]Equation (15) is also the correct expression for $W^E$ for any axially symmetric stationary electrovac such that $x^3,x^4$ are ignorable coordinates, $g_{\alpha\beta}=0$ if $\alpha=1,2$ and $\beta=3,4$, and the Maxwell 2-form $F$ is given by $dA_3dx^3+dA_4dx^4$ where $A_3,A_4$ depend only on $x^1,x^2$. For an axially symmetric stationary vacuum, $\mathrm{R1}\mathscr{E}_4=-g_{44}$ and $\mathrm{R1}\mathscr{E}_3=-g_{43}$ if we make appro-

priate selections of the arbitrary constants in $\mathscr{E}_3$, $\mathscr{E}_4$. In our current work in this paper, we let R1$\mathscr{E}_3$ differ from $-g_{43}$ by a constant ($\pm a$) in some cases.

[24]These solutions for $M^{OO}$ and $M^{OE}$, as well as for the $dx^4$ term of $M^{EE}$, are also valid for any axially symmetric stationary electrovac; (see Ref. 23). However, our $dx^3$ in $M^{EE}$ is peculiar to the present problem. Our current choices of gauge for the Minkowski space potentials will be altered in Sec. 7K to enable us to extend the domains of some of our results.

[25]I. Hauser and R.J. Malhiot, J. Math. Phys. **19**, 187 (1978).

[26]This theorem can be proven directly from the invariance of $K_\alpha K_\beta - (\mathbf{K}\cdot\mathbf{K})g_{\alpha\beta}$ under the group $\mathscr{H}_\mathbf{K}$.

[27]This is the same $\epsilon$ which was used by Ref. 25.

[28]H.W. Brinkmann, Proc. Nat. Acad. Sci. (U.S.) **9**, 1 (1923); W. Kundt, Z. Physik **163**, 77 (1961). For additional references see *Gravitation: an Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), specifically pp. 85–101 in the article by J. Ehlers and W. Kundt. Also see W. Kinnersley, in *General Relativity and Gravitation*, edited by G. Shaviv and J. Rosen (Wiley, New York, 1975).

[29]For a discussion of Killing tensors and their relation to separability, see, e.g., N.M.J. Woodhouse, Commun. Math. Phys. **44**, 9 (1975); W. Dietz, J. Phys. A **9**, 519 (1976); C.D. Collinson and J. Fugère, J. Phys. A **10**, 745 (1977).

[30]The reader may prefer to use the expressions given by Eqs. (47) and (49) in place of some or all of our results in Sec. 6. In that case, examples of suitable coordinate choices are $q^1,q^2,s,\chi$ for $(S,S)$, $q^3,q^4,r,\phi$ for $(S,T)$ and $(T,T)$, and $s,\chi,r,\phi$ for $(0,P)$.

[31]The excluded points at $r=0$ generally constitute a conical singularity. The excluded points at $s=0$ are, in some cases, at infinity with respect to the geodesics of the electrovacs $V'_{4r}$.

# Note on the stability of the Schwarzschild metric[a)]

Robert M. Wald

*Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637*

(Received 28 September 1978; revised manuscript received 27 November 1978)

It is shown that the standard arguments for the stability of the Schwarzschild metric can be made into a rigorous proof that the numerical values of linear perturbations of Schwarzschild must remain uniformly bounded for all time.

In many instances one has a solution describing an equilibrium configuration and one wishes to determine its stability, that is, determine whether an initially small disturbance of the solution will remain small or grow large with time. An important example of this problem in general relativity is the Schwarzschild spacetime. As is well known, the exactly spherical collapse must produce a Schwarzschild black hole, since the Schwarzschild metric is the only vacuum, spherically symmetric solution of Einstein's equation. The stability of this process is a very important issue in the theory of gravitational collapse. Can an initially small deviation from spherical collapse or an initially small, nonspherical perturbation of an already formed Schwarzschild black hole result in a large change at late times, perhaps converting the black hole into a naked singularity? Within the context of linear perturbation theory, extensive analyses by numerous authors (see Vishveshwara,[1] Price[2] and references cited therein) have demonstrated beyond a reasonable doubt that the answer to this question is no; the Schwarzschild solution is stable. The purpose of this note is to remove even the unreasonable doubt on this conclusion (still, however, in the context of linear perturbation theory) by filling in some mathematical gaps in previous arguments. The methods of this paper are generally applicable to cleaning up stability analyses where naive stability arguments similar to the type described below exist (for example, for a proof of stability of the Reissner–Nordström metric); however, they are not likely to contribute new ideas to stability proofs where the naive arguments are not applicable (for example, for a proof of stability of the Kerr metric).

The equations governing scalar, electromagnetic,[3] or gravitational perturbations[4,5] of Schwarzschild can be expressed (after separation of the angular variables) in the form

$$\frac{\partial^2 f}{\partial t^2} = \frac{\partial^2 f}{\partial r_*^2} - V(r_*) f, \tag{1}$$

where the function $f$ characterizes the perturbation, $r_*$ is the Regge–Wheeler[4] coordinate, and $V$ is a smooth, positive potential which goes to zero at $r_* \to \infty$ (infinity) and $r_* \to -\infty$ (horizon). Hence, the operator $A$ defined by

$$A = -\frac{d^2}{dr_*^2} + V \tag{2}$$

is a positive, self-adjoint operator on the Hilbert space $L^2(r_*)$ of square integrable functions of $r_*$. To show stability of the Schwarzschild metric, we should prove the following: Given well-behaved initial data for $f$ at $t = 0$—say, for simplicity, $C^\infty$ initial data of compact support in $r_*$—then $f$ remains bounded for all time. We shall give a complete proof of this statement below, but first we review a naive stability argument and an energy integral argument. Then, we obtain a bound on the square integral of $f$ using spectral theory methods (which may be viewed as a mathematically precise version of the naive stability argument) and, finally, using this result, give a complete proof of stability.

## NAIVE STABILITY ARGUMENT

In this approach, one first postulates that every solution of Eq. (1) can be expressed as a superposition of solutions with time dependence $e^{\alpha t}$ ($\alpha$ complex). Suppose $\mathrm{Re}\,\alpha > 0$. Solution of Eq. (1) asymptotically as $r_* \to \pm\infty$ shows that the solution must be exponentially growing or decreasing with $r_*$ in these limits. One asserts that an initially well-behaved solution cannot be constructed by including the exponentially growing (in $r_*$) solutions in the superposition, so only the exponentially decreasing behavior need be considered. [Alternatively, one appeals to ingoing wave (at the horizon) and outgoing wave (at infinity) boundary conditions to eliminate the exponentially growing solution in $r_*$]. But, for the exponentially decreasing in $r_*$ solutions $h$, we have, by Eq. (1),

$$\alpha^2 \int |h|^2 dr_* = -\int \bar{h} A h \, dr_*. \tag{3}$$

Since $A$ is a positive, self-adjoint operator, the right-hand side is real and negative, which contradicts the assumption $\mathrm{Re}\,\alpha > 0$. Thus, there are no physically relevant exponentially growing modes. One then asserts that superposition of the oscillating modes ($\alpha$ imaginary) cannot yield an initially well-behaved perturbation which becomes large at late time. (Note, however, that the individual oscillating modes are themselves badly behaved on the horizon.)

## ENERGY INTEGRAL METHOD

Many of the unproven assertions of the naive stability argument can be circumvented by the following energy integral approach. Multiplying Eq. (1) by $\dot{f} = \partial \bar{f}/\partial t$ and integrating over $r_*$, we obtain

[a)]Supported in part by the National Science Foundation under grant PHY 76-81102 A01 with the University of Chicago and by the Sloan Foundation.

  

$$\int \dot{\bar{f}} \ddot{f} \, dr_* = -\int \dot{\bar{f}} A f \, dr_*. \quad (4)$$

Adding this equation to its complex conjugate and using the self-adjointness of $A$, we find

$$\frac{\partial}{\partial t}\left(\int |\dot{f}|^2 \, dr_* + \int \bar{f} A f dr_*\right) = 0 \quad (5)$$

or

$$\int |\dot{f}|^2 \, dr_* + \int \bar{f} A f = C, \quad (6)$$

where $C$ is a constant. Since $\int \bar{f} A f \geqslant 0$ since $A$ is positive, $C$ is a bound for $\int |\dot{f}|^2 \, dr_*$. Thus, in particular, uniform exponential growth in time of initially well-behaved perturbations are ruled out. However, this approach still has two important deficiencies: (1) Growth of perturbations linearly or slower with $t$ is not easily ruled out. (2) The bounds in any case are on integrals of the perturbation; the possibility that the perturbation is blowing up in ever narrowing regions is not ruled out.

The first deficiency is rectified (with, however, a substantial increase in the required mathematical machinery) as follows.

## SPECTRAL THEORY METHOD[6]

The spectral theorem states that any self-adjoint operator, $A$, can be approximated by a sum of projection operators; more precisely, there is a one-parameter family of projection operators $E_\lambda$ such that $A$ can be written as

$$A = \int_{-\infty}^{\infty} \lambda dE_\lambda, \quad (7)$$

where the precise meaning of the integral (as well as other properties of $E_\lambda$) are given, for example, in Refs. 7 and 8. For a positive, self-adjoint operator, as is the case here, only positive values of $\lambda$ contribute (i.e., $E_\lambda = 0$ for $\lambda \leqslant 0$). Functions of a self-adjoint operator $A$ may be defined using the spectral representation, Eq. (7),

$$F(A) \equiv \int F(\lambda) \, dE_\lambda \quad (8)$$

Let $f_0(r_*)$ and $\dot{f}_0(r_*)$ denote the ($C^\infty$, compact support) initial data for $f$ and let

$$g_t = [\cos A^{1/2} t] f_0 + [A^{-1/2} \sin A^{1/2} t] \dot{f}_0, \quad (9)$$

where the operators in Eq. (9) are defined by their spectral representation, e.g.,

$$A^{-1/2} \sin A^{1/2} t = \int_0^\infty \frac{\sin \lambda^{1/2} t}{\lambda^{1/2}} dE_\lambda. \quad (10)$$

Now, $f_0$ and $\dot{f}_0$ (being $C^\infty$ and of compact support) lie in $\operatorname{dom} A \supset \operatorname{dom} A^{1/2}$; it is easy to show that $[\cos A^{1/2} t] f_0$ and $[A^{-1/2} \sin A^{1/2} t] \dot{f}_0$ also lie in $\operatorname{dom} A^{1/2}$. The same argument as is used in the proof of Stone's theorem[7,8] then proves that $dg_t/dt$ exists and

$$\frac{dg_t}{dt} = -(A^{1/2} \sin A^{1/2} t) f_0 + (\cos A^{1/2} t) \dot{f}_0. \quad (11)$$

Similarly, we find

$$\frac{d^2 g_t}{dt^2} = -(A \cos A^{1/2} t) f_0 - (A^{1/2} \sin A^{1/2} t) \dot{f}_0$$

$$= -A g_t. \quad (12)$$

Thus, $g_t$ is a solution of Eq. (1). Furthermore, by Eq. (9) $g_t$ at $t = 0$ reduces to $f_0$ and by Eq. (10) $dg_t/dt$ at $t = 0$ is $\dot{f}_0$. Since solutions of Eq. (1) are uniquely determined by their initial data, we must have $f = g$. Thus, we have proven that

$$f(t, r_*) = (\cos A^{1/2} t) f_0(r_*) + (A^{-1/2} \sin A^{1/2} t) \dot{f}_0(r_*). \quad (13)$$

It is easy to show that $\cos A^{1/2} t$ and $A^{-1/2} \sin A^{1/2} t$ are bounded operators with norm less than one. Thus,

$$\int |f(t, r_*)|^2 \, dr_* \leqslant 2\|f_0\|^2 + 2\|\dot{f}_0\|^2, \quad (14)$$

which gives the desired bound of the square integral of $f$ (as opposed to $\dot{f}$) at time $t$ in terms of its initial data.

Note that this analysis may be viewed as a mathematically precise version of the naive stability argument. In essence, Eq. (13) states that every initially well-behaved perturbation is a superposition of oscillating modes, and Eq. (14) shows that the square integral of such a superposition remains bounded for all time.

## COMPLETE STABILITY PROOF

We now complete the stability proof by showing that the numerical values of the perburbation are bounded by integrals of it and its derivatives, which, in turn, can be bounded by the above methods. The main result needed for this completion is the following lemma (see, e.g., Ref. 9).

*Lemma*: Let $\psi: \mathbb{R} \to \mathbb{C}$ be $C^\infty$ and of compact support. Then for all $x$, $|\psi(x)|^2 \leqslant \frac{1}{2}[\int |\psi|^2 dx + \int |\psi'|^2 dx]$, where $\psi' = d\psi/dx$.

*Proof*:

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} \hat{\psi}(k) \, dk$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} (1 + k^2)^{-1/2} (1 + k^2)^{1/2} \hat{\psi}(k) \, dk. \quad (15)$$

Hence, by the Schwartz inequality,

$$|\psi(x)|^2 \leqslant \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{dk}{1 + k^2} \int_{-\infty}^{\infty} (1 + k^2) |\hat{\psi}(k)|^2 \, dk$$

$$= \frac{1}{2}\left[\int_{-\infty}^{\infty} |\hat{\psi}(k)|^2 + \int_{-\infty}^{\infty} |k\hat{\psi}(k)|^2 \, dk\right]$$

$$= \frac{1}{2}\left[\int_{-\infty}^{\infty} |\psi(x)|^2 \, dx + \int_{-\infty}^{\infty} |\psi'(x)|^2 \, dx\right]. \quad (16)$$

[In $n$ dimensions, square integrals of the first $s$ derivatives of $\psi$ bound $\psi(x)$ if $s > n/2$.]

Applying this lemma to the perturbation $f(t, r_*)$, we have

$$|f(t,r_*)|^2 \leq \tfrac{1}{2}\int |f(t,r_*)|^2\, dr_* + \tfrac{1}{2}\int |f'(t,r_*)|^2\, dr_*. \quad (17)$$

But, from the definition of $A$ and the fact that $V$ is positive, we have

$$\int |f'(t,r_*)|^2\, dr_* \leq \int \bar{f}(t,r_*)\, Af(t,r_*)\, dr_*. \quad (18)$$

Furthermore, by Eq. (6) we have,

$$\int \bar{f}(t,r_*)\, Af(t,r_*)\, dr_* \leq \int \bar{f}_0 A f_0 + \|\dot{f}_0\|^2. \quad (19)$$

[Equation (19) also follows from Eq. (13).] Equations (17)–(19), together with Eq. (14), show that, for all $t$, $r_*$,

$$|f(t,r_*)|^2 \leq \int |f_0|^2\, dr_* + \tfrac{1}{2}\int \bar{f}_0 A f_0\, dr_* + \frac{3}{2}\int |\dot{f}_0|^2\, dr_*, \quad (20)$$

which uniformly bounds $f$ outside the horizon for all time in terms of the initial data $f_0, \dot{f}_0$. By continuity, this bound also applies on the horizon. (Derivatives of $f$ can also be bounded in a similar manner.) This completes the stability proof.


[1]C.V. Vishveshwara, Phys. Rev. D **1**, 2870 (1970). See also V. Moncrief, Ann. Phys. **88**, 323 (1973).
[2]R. Price, Phys. Rev. D **5**, 2419 (1972).
[3]R. Ruffini, J. Tiomno, and C.V. Vishveshwara, Lett. Nuovo Cimento **3**, 211 (1972).
[4]T. Regge and J.A. Wheeler, Phys. Rev. **108**, 1063 (1957); L.A. Edelstein and C.V. Vishveshwara, Phys. Rev. D **1**, 3514 (1970).
[5]F. Zerilli, Phys. Rev. Lett. **24**, 737 (1968).
[6]R.M. Wald, Ph.D. thesis, Princeton University, 1972 (unpublished).
[7]M. Reed and B. Simon, *Functional Analysis* (Academic, New York, 1972).
[8]F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Ungar, New York, 1955).
[9]P. Gilkey, *The Index Theory and the Heat Equation* (Publish or Perish, Boston, 1975).

# Integral bounds for $N$-body total cross sections[a)]

T. A. Osborn

*Cyclotron Laboratory, Department of Physics, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2*

D. Bollé[b)]

*Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium*

(Received 19 September 1978;revised manuscript received 9 November 1978)

We study the behavior of the total cross sections in the three- and $N$-body scattering problem. Working within the framework of the time-dependent two-Hilbert space scattering theory, we give a simple derivation of integral bounds for the total cross section for all processes initiated by the collision of two clusters. By combining the optical theorem with a trace identity derived by Jauch, Sinha, and Misra, we find, roughly speaking, that if the local pairwise interaction falls off faster than $r^{-3}$, then $\sigma_{\rm tot}(E)$ must decrease faster than $E^{-1/2}$ at high energy. This conclusion is unchanged if one introduces a class of well-behaved three-body interactions.

## I. INTRODUCTION

We investigate the behavior of total cross sections in the $N$-body scattering problem. For all collisions initiated by the scattering of two clusters we will obtain integral bounds on these cross sections. Our result is based on combining two structures common in scattering theory. The first is a trace theorem developed by Jauch *et al.*[1] to study time delay in the two-body problem. The second is the time dependent two-Hilbert space formalism for nonrelativistic $N$-body scattering.[2] More precisely, within the two-Hilbert space theory one can obtain an integral expression in time for the $S$ matrix. This integral has a kernel that is related to a trace class operator. The trace theorem of Jauch *et al.* allows one to exhibit a bound for the energy integral of the imaginary part of the forward $S$ matrix element. By employing the optical theorem this bound is transformed into a bound on the total cross section.

In the two-body problem the calculation outline above has been carried out by Martin and Misra.[3] Let us briefly recall their results. Take $R(z)$ and $R_0(z)$ to be the exact and free resolvents corresponding to the Hamiltonians $H = H_0 + V$ and $H_0$. For $\text{Im}\, z \neq 0$, one can prove under suitable, but mild conditions on the potential $V$ that $R_0^{\nu}(z) VR^{\nu}(z)$ is trace class when $\nu > \frac{3}{4}$. Using this fact and the program outlined above one arrives at a bound on $\sigma_{\rm tot}(E)$ that requires the cross section to decrease faster than $E^{-1/2}$ for high energy. We find the same conclusion holds for the $N$-particle case. Furthermore, if we introduce a class of well-behaved three-body potentials, these integral bounds on the cross section are unchanged.

Section II describes the elements of the $N$-body two-Hilbert space theory we use and gives the statement of the trace theorem. Section III obtains the integral bounds for the cross sections, under the assumption that a certain operator is trace class. In Sec. IV we prove the trace-class character of this operator in the $N$-body problem. Finally, we consider the effect of adding three-body forces to the system.

a)Work supported in part by a grant from the National Research Council of Canada and by a NATO Research Grant.

b)Bevoegdverklaard Navorser NFWO, Belgium.

## II. $N$-BODY SCATTERING THEORY

In this section we first define the two-Hilbert space multichannel theory that describes the scattering solutions of the $N$-body problem. The basis theory outlined here is the same as that found in the work of Chandler and Gibson.[2] However, we use a notation that explicitly exhibits the individual asymptotic channels.

Let $\mathscr{H}$ denote the $N$-body Hilbert space after the total center of mass motion has been removed. $\mathscr{H}$ consists of square integrable functions of the $3(N-1)$ coordinates describing the particle configuration in coordinate or momentum space. Take $V_{ij}$ to be the local pairwise interaction between particles $i$ and $j$, and let $H_0$ denote the self-adjoint operator representing the free $N$-particle kinetic energy. The full Hamiltonian is

$$H = H_0 + \sum_{i>j} V_{ij}. \tag{2.1}$$

The second space is formed by the direct product of all the asymptotic channel spaces $\mathscr{H}_\alpha$. Each distinct channel is labeled by an index $\alpha$. The symbol $\alpha$ denotes both a partition, $A$, of the $N$ particles into $N_\alpha$ clusters and the specification of the eigenfunction of each cluster. Set $A = \{al{:}l = 1, N_\alpha\}$. Here $al$ is the set containing the labels of the $nl$ particles in the $l^{\rm th}$ cluster of partition $A$. The internal cluster wavefunctions are taken to be $\psi^\alpha_{al}$. Consider cluster $al$. Let $h_0(al)$ denote the Hamiltonian for the internal kinetic energy and $v(al)$ the internal cluster potential. Then $\psi^\alpha_{al}$ is an eigenfunction of $h(al) = h_0(al) + v(al)$ with binding energy $-\epsilon^\alpha_{al}$. The total binding energy of all clusters in channel $\alpha$ is the sum

 0022-2488/79/061059-05$01.00  

$$\epsilon^{\alpha} = \sum \epsilon^{\alpha}_{al}, \tag{2.2}$$

where the index $al$ runs over all clusters in $A$.

In terms of these cluster properties one may construct the chanel spaces $\mathscr{H}_\alpha$ and the associated channel Hamiltonians $\widetilde{H}_\alpha$. Consider each cluster to be a point particle with the mass of all of its constituent particles and let $H_0[A]$ denote the relative motion kinetic energy operator for these $N_\alpha$ bodies. Then $\widetilde{H}_\alpha$ is just

$$\widetilde{H}_\alpha = H_0[A] - \epsilon^{\alpha}. \tag{2.3}$$

This Hamiltonian gives the energy available to the $N_\alpha$ clusters when they are outside of each others' force fields and freely moving. The space $\mathscr{H}_\alpha$ is the space of square-integrable functions in the relative coordinates that determine the positions of the centers of mass of the $N_\alpha$ clusters. The operator $\widetilde{H}_\alpha$ is the Laplacian in $3(N_\alpha - 1)$ variables displaced by a fixed energy $\epsilon^{\alpha}$, so it will only have an absolutely continuous spectrum.

The mapping between $\mathscr{H}_\alpha$ and $\mathscr{H}$ is given by the Kato identification operator $J_\alpha$. Let $f_\alpha$ be any function in $\mathscr{H}_\alpha$. Set $F_\alpha$ to be

$$F_\alpha = \Big(\prod_{al} \psi^{\alpha}_{al}\Big) f_\alpha, \tag{2.4}$$

then $J_\alpha$ is defined as

$$F_\alpha = J_\alpha f_\alpha. \tag{2.5}$$

The wave operator is the basic object in multichannel scattering. This operator maps $\mathscr{H}_\alpha$ into $\mathscr{H}$. Define

$$U_\alpha(t) = e^{+itH} J_\alpha e^{-it\widetilde{H}_\alpha}, \tag{2.6}$$

then

$$\Omega^{(\pm)}_\alpha = \operatorname*{s-lim}_{t\to-\infty} U_\alpha(\pm t). \tag{2.7}$$

The wave operators satisfy the indentities

$$H\Omega^{(\pm)}_\alpha = \Omega^{(\pm)}_\alpha \widetilde{H}_\alpha, \tag{2.8}$$

$$\Omega^{(\pm)\dagger}_\alpha \Omega^{(\pm)}_\alpha = \delta_{\alpha\beta} I_\alpha, \tag{2.9}$$

where $I_\alpha$ is the identity operator on $\mathscr{H}_\alpha$. The first relation here, the intertwining property, is equivalent to the requirement of energy conservation in the scattering process. The second property is channel orthogonality. From the wave operator one defines the $S$ matrix by

$$S_{\alpha\beta} = \Omega^{(-)\dagger}_\alpha \Omega^{(+)}_\beta. \tag{2.10}$$

Clearly, $S_{\alpha\beta}$ maps any initial wavepacket in $\mathscr{H}_\beta$ into the resultant wavepacket in $\mathscr{H}_\alpha$. The intertwining property implies

$$S_{\alpha\beta}\widetilde{H}_\beta = \widetilde{H}_\alpha S_{\alpha\beta}. \tag{2.11}$$

This is the abstract form of energy conservation for the $S$ matrix. Finally, let us define a channel Hamiltonian in $\mathscr{H}$. Define the sum of all pairwise potentials as $V$. Let

$$V_\alpha = \sum_{i,j\in al} V_{ij}. \tag{2.12}$$

Here the sum runs over all $al$ in $A$. The $\alpha$ channel Hamiltonian is

$$H_\alpha = H_0 + V_\alpha. \tag{2.13}$$

The definition of $J_\alpha$ implies

$$H_\alpha J_\alpha = J_\alpha \widetilde{H}_\alpha. \tag{2.14}$$

For short range potentials, $V_{ij}$, that are multiplication-operators in coordinate space by $L^2$ functions, proofs of the existence, intertwining, and channel orthogonality of the waveoperators may be found in Hack[4] and Hunziker.[5] The only technical feature of two-Hilbert space scattering for short ranged interactions that has not been given adequate proof is asymptotic completeness. Faddeev[6] and more recently Ginibre and Moulin[7] have provided a proof when $N = 3$.

Before turning to the trace theorem we recall the direct integral representation of a Hilbert space. Consider a self-adjoint operator $h$ on a Hilbert space $\mathscr{H}'$. Consider the situation where $h$ has only an absolutely continuous spectrum $\Lambda$. The direct integral representation of $\mathscr{H}'$ is denoted by

$$\mathscr{H}' = \int_\Lambda^{\oplus} \mathscr{H}_\lambda \, d\lambda. \tag{2.15}$$

The direct integral is a generalization of the direct sum to a continuum of Hilbert spaces $\mathscr{H}_\lambda$, $\lambda$ being a continuous parameter in the measure space $\Lambda$. Each $\psi$ in $\mathscr{H}'$ is associated with an equivalence class of functions on $\mathscr{H}_\lambda$, $\{\psi_\lambda\}$, where

$$\psi = \int_\Lambda^{\oplus} \psi_\lambda d\lambda \tag{2.16}$$

The direct integral bears the same relation to the direct sum as a function of a real variable does to a sequence. For example, the scalar product for $\psi,\phi$ in $\mathscr{H}'$ is related to $\{\psi_\lambda\}$, $\{\phi_\lambda\}$ by

$$(\psi,\phi) = \int_\Lambda (\psi_\lambda,\phi_\lambda)_\lambda \, d\lambda, \tag{2.17}$$

where $(\ ,\ )_\lambda$ is the scalar product in $\mathscr{H}_\lambda$. This correspondence between $\psi$ and $\{\psi_\lambda\}$ is an isomorphism. For more details we refer to Ref. 8. Note that $h\psi$ corresponds to $\{\lambda\psi_\lambda\}$ and $e^{-iht}\psi$ to $\{e^{-i\lambda t}\psi_\lambda\}$. In our applications $\lambda$ is the energy of the system and $\mathscr{H}_\lambda$ may be intuitively thought of as the space describing the degrees of freedom remaining in the system once the energy has been fixed. In the terminology of nuclear physics, the utility of the direct product representation is that it allows one to rigorously define what is meant by on-energy-shell quantities.

The theorem proved by Jauch *et al.* is then:

Let $V_t = e^{-iht}$ be a unitary group with self-adjoint generator $h$ and absolutely continuous spectrum $\Lambda$, and $\Gamma$ an arbitrary trace class operator on $\mathscr{H}'$. Then there exists a dense set $D \subseteq \mathscr{H}'$ such that for all $f,g \in D$,

$$G[f,g] = \int_{-\infty}^{\infty} (f, V_t^\dagger \Gamma V_t g) dt \tag{2.18}$$

defines a sesquilinear function on $D \times D$. If $f_\lambda$ and $g_\lambda$ are the components of $f$ and $g$ in the direct integral representation with respect to the spectral family of $h$, then

$$G[f,g] = \int_\Lambda (f_\lambda, G_\lambda g_\lambda) d\lambda, \tag{2.19}$$

where $G_\lambda$ is an essentially unique family of trace-class operators in $\mathscr{H}_\lambda$ for almost all $\lambda$ and

$$\int_\Lambda \mathrm{tr} G_\lambda d\lambda = 2\pi \mathrm{Tr}\Gamma. \tag{2.20}$$

Finally,

$$\int_\Lambda \|G_\lambda\|_1 d\lambda \leqslant 2\pi \|\Gamma\|_1, \tag{2.21}$$

where $\|\Gamma\|_1$ denotes the trace norm of $\Gamma$ and $\|G_\lambda\|_1$ is the trace norm in $\mathscr{H}_\lambda$ of $G_\lambda$.

The proof of the theorem proceeds by direct verification for rank one operators, by immediate extension to finite rank operators, and finally to arbitrarty trace-class operators represented as limits of finite rank operators.

## III. TOTAL CROSS-SECTION BOUNDS

Determining the cross section bounds requires two steps. First, we give the time integral formula for the elastic $S$ matrix. Then we combine the trace theorem with the $N$-body optical theorem to obtain the bound. Consider the $S$ matrix first. We extend Ikebe's representation[9] from the two-body problem to the $N$-body problem. Let $f_\alpha$ be in the domain of $\widetilde{H}_\alpha$ and $Jf_\alpha$ in the domain of $H_\alpha$. Let $F$ be any function in $\mathscr{H}$, then

$$\frac{d}{dt}(F, U_\alpha(t) f_\alpha) = i(F, e^{itH}[HJ_\alpha - J_\alpha \widetilde{H}_\alpha] e^{-it\widetilde{H}_\alpha} f_\alpha). \tag{3.1}$$

Because of Eq. (2.14) the square bracket operator can be expressed as

$$HJ_\alpha - H_\alpha J_\alpha = (V - V_\alpha) J_\alpha = V^\alpha J_\alpha, \tag{3.2}$$

where $V^\alpha$ is defined as the sum of all interactions acting between the clusters in partition $A$. Now integrate Eq. (3.1) with respect to $t$ and use Eq. (2.7) to obtain

$$(F, \Omega_\alpha^{(-)} f_\alpha) - (F, \Omega_\alpha^{(+)} f_\alpha) = \int_{-\infty}^{\infty} \frac{d}{dt}(f, U_\alpha(t) f_\alpha) dt. \tag{3.3}$$

Choose $F = \Omega_\alpha^{(-)} f'_\alpha$, where $f'_\alpha \in \mathscr{H}_\alpha$. Define the elastic $\alpha$ channel $t$ matrix to be

$$S_{\alpha\alpha} = I_\alpha - iT_{\alpha\alpha}. \tag{3.4}$$

This $t$ matrix is a bounded operator on $\mathscr{H}_\alpha$. Recalling properties (2.8–2.10) allows us to write Eq. (3.3) as

$$(f'_\alpha, T_{\alpha\alpha} f_\alpha) = \int_{-\infty}^{\infty} (f'_\alpha, e^{it\widetilde{H}_\alpha} (\Omega_\alpha^{(-)\dagger} V^\alpha J_\alpha) e^{-it\widetilde{H}_\alpha} f_\alpha) dt. \tag{3.5}$$

We are now in a position to apply the trace theorem. However the operator $\Omega_\alpha^{(-)\dagger} V^\alpha J_\alpha$ is not in general trace class on $\mathscr{H}_\alpha$. It will become trace class if multiplied on the left and the right by resolvent operators. Define the function $R^\nu(x;z)$ to be

$$R^\nu(x;z) = \frac{1}{(x-z)^\nu} \quad x \in \mathbb{R}, z \in \mathbb{C}, \ \nu > 0. \tag{3.6}$$

Under suitable constraints on the pair interaction $V_{ij}$ we establish in Sec. IV that for $z$ not in the spectrum of $H$,

$$\|R^\nu(\widetilde{H}_\alpha; z^*) \Omega_\alpha^{(-)\dagger} V^\alpha J_\alpha R^\nu(\widetilde{H}_\alpha; z)\|_1 < \infty, \tag{3.7}$$

when $\nu > \frac{3}{4}$ and $\alpha$ is a two-cluster channel.

Given inequality (3.7) let us see what bound results for $T_{\alpha\alpha} - T^\dagger_{\alpha\alpha}$. In Eq. (3.5) replace $f'_\alpha$ by $R^\nu(\widetilde{H}_\alpha; z) f'_\alpha$ and $f_\alpha$ by $R^\nu(\widetilde{H}_\alpha; z) f_\alpha$. Then Eq. (3.5) becomes

$$(f'_\alpha, R^\nu(\widetilde{H}_\alpha; z^*) T_{\alpha\alpha} R^\nu(\widetilde{H}_\alpha; z) f_\alpha) = \int_{-\infty}^{\infty} (f'_\alpha, e^{it\widetilde{H}_\alpha} \Gamma e^{-it\widetilde{H}_\alpha} f_\alpha), \tag{3.8}$$

where $\Gamma$ is the operator in the trace norm of expression (3.7). The operator $T_{\alpha\alpha}$ appearing in the left-hand side of Eq. (3.8) is bounded and commutes with $\widetilde{H}_\alpha$. Therefore, the direct integral representation for the left hand side of Eq. (3.8) is

$$(f'_\alpha, R^\nu(\widetilde{H}_\alpha; z^*) T_{\alpha\alpha} R^\nu(\widetilde{H}_\alpha; z) f_\alpha) = \int_{-\epsilon^\alpha}^{\infty} \left( f'_{\alpha E}, \frac{\tau_{\alpha\alpha}(E)}{|E-z|^{2\nu}} f_{\alpha E} \right) dE, \tag{3.9}$$

where $\{f'_{\alpha E}\}$ and $\{f_{\alpha E}\}$ are the direct integral components of $f'_\alpha$ and $f_\alpha$ and $\tau_{\alpha\alpha}(E)$ is the component representation of $T_{\alpha\alpha}$.

Statement (2.19) of the trace theorem allows us to represent the right-hand side of (3.8) as

$$\int_{-\epsilon^\alpha}^{\infty} (f'_{\alpha E}, G(E) f_{\alpha E}) dE,$$

where $G(E)$ is defined by $\Gamma$. Let $\mathscr{H}_{\alpha E}$ be the direct integral component of $\mathscr{H}_\alpha$. Then on $\mathscr{H}_{\alpha E}$, $G(E)$ is trace class. The theorem of Jauch *et al.* tells us that $G(E)$ is essentially unique, so that for almost all $E$,

$$\tau_{\alpha\alpha}(E) = |E-z|^{2\nu} G(E). \tag{3.10}$$

This relation implies $\tau_{\alpha\alpha}(E)$ is trace class in $\mathscr{H}_{\alpha E}$. The inequality (2.21) reads then

$$\int_{-\epsilon^\alpha}^{\infty} \frac{|\mathrm{tr}\tau_{\alpha\alpha}(E)|}{|E-z|^{2\nu}} dE \leqslant \int_{-\epsilon^\alpha}^{\infty} \frac{\|\tau_{\alpha\alpha}(E)\|_1}{|E-z|^{2\nu}} dE \leqslant 2\pi \|\Gamma\|_1, \tag{3.11}$$

where we used the fact that $|\mathrm{tr}\tau_{\alpha\alpha}(E)| \leqslant \|\tau_{\alpha\alpha}(E)\|_1$. Next, the optical theorem states that

$$\mathrm{Im}\ \mathrm{tr}\tau_{\alpha\alpha}(E) = \mathrm{const} E \sigma^\alpha_{\mathrm{tot}}(E), \tag{3.12}$$

where $\sigma^\alpha_{\mathrm{tot}}(E)$ is the sum of all cross sections energetically open at $E$. In the three-body problem a proof of Eq. (3.12) may be found in Ref. 10. For $N \geqslant 4$ we assume the optical theorem remains valid. From a technical point of view Eq. (3.12) is almost equivalent to the statement that the multichannel $S$ matrix, $S_{\alpha\beta}$, is unitary. This in turn is closely related to asymptotic completeness. The constant in Eq. (3.12) is a function of the masses of the $N$ particles. Thus inequality (3.11) becomes

$$\int_{-\epsilon^\alpha}^{\infty} \frac{E\sigma^\alpha_{\mathrm{tot}}(E)}{|E-z|^{2\nu}} dE < \infty, \quad \nu > \tfrac{3}{4}. \tag{3.13}$$

At high energy the integral bound (3.13) requires the $\sigma^\alpha_{\mathrm{tot}}(E)$ fall off faster than $E^{-1/2}$.

## IV. TRACE-CLASS OPERATORS

In this section we prove that the operator $\Gamma$ is trace class in the $N$-body scattering problem with two- and three-body potentials. It is assumed that our local two-body potentials, $v_{ij}(y)$, satisfy the condition

$$(1 + 3y^2)^{\mu} v_{ij}(y) \in L^2(y), \quad 1 \geqslant \mu > \tfrac{3}{4}. \tag{A1}$$

The variable $y$ is the vector separation of particles $i$ and $j$. Of course (A1) implies that $v_{ij} \in L^2(y)$, which was the condition needed for our statement of the waveoperator properties in Sec. III. When $\mu$ is near $\frac{3}{4}$, (A1) is satisfied if the power behavior of $v_{ij}(y)$ for large $y$ is like $|y|^{-3-\delta}$, where $\delta$ is an arbitrarily small positive number. For $L^2$ potentials it is known[11] that $N$ body boundstate wavefunctions, $\psi^{\alpha}_{al}$ will satisfy

$$\int |(1 + 3y_i^2)^n \psi^{\alpha}_{al}(y_i, y_i^c)|^2 dy_i dy_i^c < \infty. \tag{4.1}$$

Here $y_i$ is the position of the $i$th particle relative to the center of mass of the boundstate cluster $al$. The $y_i^c$ denote all the remaining coordinates and $n$ is any positive number.

We summarize our first result:

*Lemma* 1: Let $v_{ij}$ satisfy (A1) for all $i,j$. Then the operator

$$\Gamma = R^{\nu}(\widetilde{H}_{\alpha}; z^*) \Omega_{\alpha}^{(-)\dagger} V^{\alpha} J_{\alpha} R^{\nu}(\widetilde{H}_{\alpha}; z), \tag{4.2}$$

is trace class on $\mathscr{H}_{\alpha}$ when $\nu > \frac{3}{4}$ and $z$ is to the left of the spectrum of $H$, for each two-cluster channel $\alpha$.

*Proof*: We show that $\Gamma$ can be written as a product of two Schmidt operators. We first note that the intertwining property allows us to write $\Gamma$ as

$$\Gamma = \Omega_{\alpha}^{(-)\dagger} R^{\nu}(H; z^*) V^{\alpha} P_{\alpha} R^{\nu}(H_{\alpha}; z) J_{\alpha}. \tag{4.3}$$

The operator $P_{\alpha}$ is the projection operator defined by the range of $J_{\alpha}$. So $P_{\alpha} J_{\alpha} = J_{\alpha}$. Both $J_{\alpha}$ and $\Omega_{\alpha}^{(-)}$ are isometries mapping $\mathscr{H}_{\alpha}$ into $\mathscr{H}$. Thus it suffices to shown that

$$\Gamma' = R^{\nu}(H; z^*) V^{\alpha} P_{\alpha} R^{\nu}(H_{\alpha}; z) \tag{4.4}$$

is trace class on $\mathscr{H}$. We introduce an operator $M$ on $\mathscr{H}$ that is defined as the multiplication operator with the function $m(x) = (1 + x^2)^{-\mu}$, $\mu > \frac{3}{4}$. The variable $x$ is the vector distance between the centers of mass of the two clusters in channel $\alpha$. Referring to Eq. (2.4), we note that $x$ is independent of the internal cluster coordinates in $\psi^{\alpha}_{al}$. Thus, $M$ commutes with $P_{\alpha}$ and we may write

$$\Gamma' = [R^{\nu}(H; z^*) V^{\alpha} M^{-1} P_{\alpha}] [P_{\alpha} M R^{\nu}(H_{\alpha}; z)]. \tag{4.5}$$

To compute the fractional resolvent in the left bracket it is convenient to use

$$R^{\nu}(H; -l) = \frac{\sin\nu\pi}{\pi} \int_0^{\infty} R(H; -l-\lambda) \lambda^{-\nu} d\lambda. \tag{4.6}$$

In Eq. (3.6), $z^* = -l$, where $l$ is positive and chosen to be greater than the lower bound of $H$. Now define $A = R^{\nu}(H; z^*) V^{\alpha} M^{-1} P_{\alpha}$ and $B = P_{\alpha} M R^{\nu}(H_0; z)$. Clearly it is sufficient to show $A$ and $B$ are Schmidt class.

Let us consider $B$ first. Label the two clusters in channel $\alpha$ as 1 and 2. The boundstate eigenfuctions are denoted by $\psi_1(y_1)$ and $\psi_2(y_2)$. Here $y_1$ and $y_2$ specify all the relative particle coordinates in the two clusters. The variable $x$ is the vector giving the separation of the center of mass of the two clusters. The variables $x, y_1, y_2$ have momentum conjugates $p, q_1, q_2$. In momentum space the kernel of $P_{\alpha} M$ is found to be

$$\langle p q_1 q_2 | P_{\alpha} M | p' q_1' q_2' \rangle = \widetilde{\psi}_1(q_1) \widetilde{\psi}_1^*(q_1') \widetilde{\psi}_2(q_2) \widetilde{\psi}_2^*(q_2') \widetilde{m}(p - p'), \tag{4.7}$$

where $\widetilde{\psi}_{1,2}$ are the Fourier transforms of $\psi_{1,2}$, and $\widetilde{m}$ is the Fourier transform of $m(x)$. Thus the Schmidt norm of $B$ is given by

$$\|B\|_2^2 = \int \frac{|\widetilde{\psi}_1(q_1) \widetilde{\psi}_1(q_1') \widetilde{\psi}_2(q_2) \widetilde{\psi}_2(q_2') \widetilde{m}(p - p')|^2}{|p^2 + q_1^2 + q_2^2 - z|^{2\nu}} \times dp\, dp'\, dq_1\, dq_2 dq_1'\, dq_2', \tag{4.8}$$

where $p^2$, $q_1^2$, and $q_2^2$ denote the kinetic energy for relative cluster motion, the internal motion of cluster 1 and 2, respectively. Define $I(\nu)$ to be the integral $\int |p^2 - z|^{-2\nu} dp$. The integral $I(\nu)$ is finite if $\nu > \frac{3}{4}$. The total kinetic energy satisfies the obvious inequality $p^2 + q_1^2 + q_2^2 \geqslant p^2$. This implies Eq. (4.8) gives the bound

$$\|B\|_2 \leqslant \|m\| I(\nu)^{1/2}, \tag{4.9}$$

where we have used $\|m\| = \|\widetilde{m}\|$ and $\|\widetilde{\psi}_{1,2}\| = \|\psi_{1,2}\| = 1$.

It remains to show that $A$ is Schmidt class. The interaction $V^{\alpha}$ is the finite sum of terms $V_{ij}$ where $i$ is in cluster 1 and $j$ is in cluster 2. Consider one of these terms $V_{ij}$, determined by the potential function $v_{ij}(\eta_{ij})$. Here $\eta_{ij}$ is the relative separation of particles $i$ and $j$. Let $y_i$ and $y_j$ be the positions of particle $i$ and $j$ measured from the center of mass of clusters 1 and 2. Denote By $y_i^c$ and $y_j^c$ the remaining independent cluster coordinates. The coordinate space kernel for the operator $V_{ij} M^{-1} P_{\alpha}$ is

$$\langle x, y_1, y_2 | V_{ij} M^{-1} P_{\alpha} | x', y_1', y_2' \rangle = \delta(x - x') D(x, y_1, y_2) \psi_1^*(y_1') \psi_2^*(y_2'), \tag{4.10}$$

where

$$D(x, y_1, y_2) = v_{ij}(\eta_{ij})(1 + x^2)^{\mu} \psi_1(y_1) \psi_2(y_2). \tag{4.11}$$

We observe that condition (A1) for $v_{ij}$ implies $D$ is $L^2$. To see this note that $x = y_i + y_j + \eta_{ij}$. Thus,

$$1 + x^2 \leqslant (1 + 3y_i^2)(1 + 3y_j^2)(1 + 3\eta_{ij}^2), \tag{4.12}$$

so $D$ is bounded by the product of $(1 + 3y_i^2)^{\mu} |\psi_1(y_i, y_i^c)|$, $(1 + 3y_j^2)^{\mu} |\psi_2(y_j, y_j^c)|$ and $(1 + 3\eta_{ij}^2)^{\mu} |v_{ij}(\eta_{ij})|$. Then condition (A1) and Eq. (4.1) $D$ is $L^2$.

To complete the proof that $A$ is Schmidt class, Fourier transform expression (4.10). We have

$$\langle p q_1 q_2 | V_{ij} M^{-1} P_{\alpha} | p' q_1' q_2' \rangle = \widetilde{D}(p - p', q_1, q_2) \widetilde{\psi}_1^*(q_1') \widetilde{\psi}_2^*(q_2'). \tag{4.13}$$

Proceed, as with the estimate of $B$, and use the fact that $R(H; -l)V$ is a bounded operator times $R(H_0; -l)V$. Then

$$\|R(H; -l) V^{\alpha} M^{-1} P_{\alpha}\|_2 < \text{const} l^{-1/4}. \tag{4.14}$$

Combining Eqs. (4.6) and (4.14) shows that $A$ is Schmidt class if $\nu \in (\frac{3}{4}, 1)$. This completes the proof.

It is of some interest to see if the addition of three-body forces alters our integral bound on $\sigma_{\rm tot}^{\alpha}(E)$. Denote by $V_{ijk}$ the three-body interaction between particles $i,j$ and $k$. We require that the $V_{ijk}$ satisfy[5]

$$\int |v_{ijk}(\eta_{ij},\eta_{ik})|^{3+\epsilon} d\eta_{ij} d\eta_{ik} < \infty, \quad \epsilon > 0, \qquad \text{(B1)}$$

and

$$\int |v_{ijk}(\eta_{ij},\eta_{ik})(1+3\eta_{ij}^2)^{\mu}|^2 d\eta_{ij} < M_3 < \infty, \quad 1 \geqslant \mu > \tfrac{3}{4}, \qquad \text{(B2)}$$

for all $\eta_{ik}$ and all labels $i,j$ and $k$. The first condition ensures that the scattering theory of Sec. II remains valid. Property (B2) has a simple physical interpretation at large particle separation. When the distance between $i$ and $k$ is held fixed and $\mu$ is near $\frac{3}{4}$, then the three-body potential must decrease faster than $\eta_{ij}^{-3}$.

The only part of our proof that needs reconsideration is demonstrating that the operator $A$ is Schmidt class. Take particles $i$ and $k$ to be in cluster 1 and $j$ in cluster 2. The kernel for the $V_{ijk}$ contribution to $A$ is again given by expression (4.10), but with $D$ replaced by $D_3$ where

$$D_3(x,y_1,y_2) = v_{ijk}(\eta_{ij},\eta_{ik})(1+x^2)^{\mu}\psi_1(y_1)\psi_2(y_2). \qquad (4.15)$$

We must establish that $D_3$ is in $L_2$. Using Eq. (4.12) we see that $D_3$ is bounded by the product of $|v_{ijk}(\eta_{ij},\eta_{ik})(1+3\eta_{ij}^2)^{\mu}|$, $|\psi_1(y_i,y_i^c)(1+3y_i^2)^{\mu}|$ and $|\psi_2(y_j,y_j^c)(1+3y_j^2)^{\mu}|$. The measure for the $L^2$ integration may be taken as $d\eta_{ij}\, dy_1\, dy_2$. Condition (B2) ensures that the first factor integrated over $d\eta_{ij}$ is bounded by $M_3$. Condition (A1) implies that the last two factors are $L^2$ with respect to $dy_1$ and $dy_2$. Thus we have demonstrated:

*Lemma* 2: Let $v_{ij}$ satisfy (A1) and $v_{ijk}$ satisfy (B1) and (B2) for all $i,j$, and $k$. Then for each two-cluster channel $\alpha$ the operator $\Gamma$ is trace class when $\nu > \frac{3}{4}$ and $z$ is to the left of the spectrum of $H$.

This means, of course, that our discussion in Sec. II is not altered and that three-body forces of the type (B1), (B2) do not effect the integral bound obtained there.

In closing we note that for potentials in class (A1) similar trace class estimates for operators analogous to $\Gamma$ have been found by Simon.[12] The operator appearing in Simon's work has the resolvents of Eq. (4.2) replaced by spectral projections of $\widetilde{H}_{\alpha}$ on finite energy interval.


[1] J.M. Jauch, K.B. Sinha, and B.N. Misra, Helv. Phys. Acta **45**, 398 (1972).
[2] C. Chandler and A.G. Gibson, J. Math. Phys. **14**, 1328 (1973).
[3] Ph. Martin and B. Misra, in *Scattering Theory and Mathematics Physics* (Reidel, Dordrecht, 1973); Ph. Martin and B. Misra, J. Math. Phys. **14**, 997 (1973).
[4] M.N. Hack, Nuovo Cimento **13**, 231 (1959).
[5] W. Hunziker, in *Lectures in Theoretical Physics*, edited by A.O. Barut and W.E. Brittin (Gordon and Breach, New York, 1965), Vol. X-A.
[6] L.D. Faddeev, *Mathematical Aspects of the Three-Body Problem* (Davey, New York, 1965).
[7] J. Ginibre and M. Moulin, Ann. Inst. Henri Poincaré **21**, 29 (1974).
[8] M.A. Naimark and S.V. Fomin, Am. Math. Soc. Transl. Ser. 2, **5**, 35 (1957); L.H. Loomis, *Introduction to Abstract Harmonic Analysis* (Van Nostrand, New York, 1953).
[9] T. Ikebe, Pac. J. Math. **15**, 511 (1965).
[10] T.A. Osborn and D. Bollé, Phys. Rev. C **8**, 1198 (1973).
[11] A.J. O'Connor, Commun. Math. Phys. **32**, 319 (1973); J.M. Combes and L. Thomas, Commun. Math. Phys. **34**, 251 (1973).
[12] B. Simon, Commun. Math. Phys. **55**, 259 (1977).

# An algebraic approach to Coulomb scattering in $N$ dimensions

W. O. Rasmussen

*Champlain Regional College, Lennoxville Campus, Physics Department, Lennoxville, Quebec J1M 2A1, Canada*

S. Salamó

*Universidad Simón Bolívar, Departamento de Física, Apartado Postal 80659, Caracas 108, Venezuela*
(Received 20 June 1978; revised manuscript received 13 November 1978)

Using purely algebraic techniques, based on the larger symmetry group of the Kepler problem, the phase shifts and the scattering amplitude for Coulomb scattering in $N$ dimensions are derived.

## I. INTRODUCTION

The $N$-dimensional analog of the ordinary hydrogen atom has been discussed extensively in the physics literature.[1] The focus of these studies has generally been on the larger symmetry group and on the spectrum. The scattering problem in $N$ dimensions seems not to have been discussed so far. Some time ago Biedenharn and Brussard[2] and Zwanziger[3] pointed out that the nonrelativistic Coulomb phase shifts can be calculated solely with algebraic techniques. This work represented a novel introduction of group theory into the study of scattering problems in the sense that symmetry properties are employed as tools for treating the dynamics is of scattering.[4] We shall be guided by Zwanziger techniques in our strictly algebraic derivation of the phase shifts and the scattering amplitude.

We begin with a brief review of the $N$-dimensional Kepler problem. Next, the scattering states are defined and some needed matrix elements are derived. Finally, the phase shifts and the scattering amplitude are evaluated.

The simplest extension to $N$-dimensions of the three-dimensional Kepler Hamiltonian in the center of mass frame is

$$H = \frac{p^2}{2m} - \frac{\alpha}{r}, \quad \hbar = 1, \tag{1}$$

where $\mathbf{r} = (x_1, x_2, \dots x_N)$, $p = (p_1, p_2, \dots p_N)$, and

$$r^2 = \sum_{i=1}^{N} x_i^2 \quad \text{and}\, p_i = -i\frac{\partial}{\partial x_i}.$$

The usual canonical commutation relations holds, $[x_i, p_j] = i\delta_{ij}$, where $i,j = 1,2,\dots,N$ and all other commutators vanish.

The generators of the $N$-dimensional rotation group are defined to be

$$L_{ij} = x_i p_j - x_j p_i, \quad i,j = 1,2,\dots,N. \tag{2}$$

As expected, a generalized Runge–Lenz vector exists for this problem also and is given by

$$A_i = \frac{1}{2m}(L_{ij}p_j - p_j L_{ji}) - \frac{\alpha x_i}{r}$$

$$= \frac{1}{m}\left[x_i p^2 - p_i(\mathbf{x}\cdot\mathbf{p}) + i\left(\frac{N-1}{2}\right)p_i\right] - \frac{\alpha x_i}{r}. \tag{3}$$

These operators satisfy the following set of commutation relations:

$$[L_{ij}, L_{kl}] = i(\delta_{ik}L_{jl} + \delta_{jl}L_{ik} - \delta_{il}L_{jk} - \delta_{jk}L_{il}), \tag{4a}$$

$$[A_i, A_j] = iL_{ij}\left(-\frac{2H}{m}\right), \quad i,j = 1,\dots,N, \tag{4b}$$

$$[L_{ij}, A_k] = i(\delta_{ik}A_j - \delta_{jk}A_i), \tag{4c}$$

$$[L_{ij}, H] = [A_i, H] = 0. \tag{4d}$$

Since $H$ commutes with both the $A_i$'s and the $L_{ij}$'s, we can replace $H$ in these relations by its eigenvalues $E$ whenever we are working in fixed energy subspaces. Therefore, for each fixed energy subspace with $E<0$ (bound states), we define a new set of operators $L_{i,N+1}$ as follows

$$L_{i,N+1} \equiv -L_{N+1,i} \equiv \left(-\frac{m}{2E}\right)^{1/2} A_i. \tag{5}$$

Replacing the set $\{A_i\}$ by $\{L_{i,N+1}\}$ in Eqs. (4a), (4b) and (4c) we obtain the Lie algebra of the group SO($N+1$). As expected, we would have the usual SO(4) symmetry in three dimensions.

The generators $L_{ij}$, with $i,j = 1,\dots,N, N+1$, satisfy a representation relation, analogous to the three-dimensional counterpart $\mathbf{L}\cdot\mathbf{A} = 0$, namely,

$$L_{ij}L_{kl} - L_{ik}L_{jl} + L_{jk}L_{il} = 0,$$
$$\text{for } i \neq j \neq k \neq k,\ 1,2,\dots,N,\ N+1, \tag{6}$$

as can be checked with a simple calculation. As Louck[5] has shown, the above relation characterizes the generators of orbital angular momentum. Moreover, the number of independent commuting operators is reduced from $[(N-1)/2]^2$ for the case where $N+1$ is even to $N$ [for $N+1$ odd, is reduced from $N(N+1)/4$ to $N$]. The states in such an irreducible orbital angular momentum representation of SO($N+1$) may then be characterized uniquely by the set of eigenvalues of the second order Casimir operators $L_\alpha^2$,

$$L_\alpha^2 = \frac{1}{2}\sum_{i,j}^{\alpha} L_{ij}^2, \quad \alpha = 2,3,\dots,N,N+1, \tag{7}$$

 0022-2488/79/061064-04$01.00  

and the eigenvalues of $L_{12}$ in the subgroup chain $SO(N+1)\supset SO(N)\cdots\supset SO(2)$. The eigenvalues of these operators are given by:

$$L_\alpha^2|l_{N+1},l_N,\ldots,l_\alpha,\ldots,m\rangle = l_\alpha(l_\alpha+\alpha-2)|\cdots l_\alpha\cdots\rangle,$$
$$L_{12}|l_{N+1},l_N,\ldots\ldots,m\rangle = m|l_{n+1}\ldots\ldots,m\rangle, \tag{8a}$$

where the $l_\alpha$'s are integers satisfying

$$l_{\alpha+1}\geqslant l_\alpha\geqslant l_{\alpha-1}\geqslant\cdots\geqslant|m|. \tag{8b}$$

A concrete realization of these representations can be given in terms of the hyperspherical harmonics. The $l_\alpha$ then represent the degree of homogeneity of the harmonic polynomial in the first $\alpha$ dimensions.

The bound state energy spectrum can now be obtained easily if we notice that $L_{N+1}^2$ is given by

$$L_{N+1}^2 = L_N^2 - \frac{m}{2E}A^2 = -\left(\frac{N-1}{2}\right)^2 - \frac{\alpha^2 m}{2E}. \tag{9}$$

combining these relations with the eigenvalue equations (8) we find the well-known result,

$$E = \frac{\alpha^2 m}{2[l_{N+1}+(N-1)/2]^2}, \tag{10}$$

where $l_{N+1}$ takes on integer values from zero up.

Thus far, we have discussed the bound state case where $E$ is negative. The scattering states, however, are characterized by positive eigenvalues $E$ of the Hamiltonian

$$E\equiv\frac{k^2}{2m}>0. \tag{11}$$

We have defined the operator $L_{i,N+1}$ by Eq. (5), however this relation no longer defines a Hermitian operator. The appropriate modified definition is

$$L'_{i,N+1}\equiv\left(\frac{m}{2E}\right)^{1/2}A_i = \frac{m}{k}A_i = iL_{i,N+1}. \tag{12}$$

Therefore, the commutation relations (4a)–(4c) are replaced by the following ones:

$$[L_{ij},L_{kl}] = i(\delta_{ik}L_{jl}+\delta_{jl}L_{ik}-\delta_{il}L_{jk}-\delta_{jk}L_{il}), \tag{13a}$$
$$[L'_{i,N+1},L'_{j,N+1}] = -iL_{ij},\quad i,j=1,\ldots,N, \tag{13b}$$
$$[L_{ij},L'_{k,N+1}] = i(\delta_{ik}L'_{j,N+1}-\delta_{jk}L'_{i,N+1}),\quad i,j,k=1,\ldots,N. \tag{13c}$$

These relations define the Lie algebra of the noncompact group $SO(N,1)$. Again, we would just have the expected symmetry group $SO(3,1)$ in 3-space.

The $SO(N+1)$ Casimir operator $L_{N+1}^2$, Eq. (9), now is replaced by the $SO(N,1)$ Casimir operator $L_{N+1}'^2$,

$$L_{N+1}'^2 = \frac{m^2}{k^2}A^2 - L_N^2 = -\left(\frac{N-1}{2}\right)^2 - \frac{\alpha^2 m^2}{k^2}.$$

Then the eigenvalues $l'_{N+1}$ of $L_{N+1}'^2$ in the new subgroup chain description $SO(N,1)\supset SO(N)\cdots\supset SO(2)$ are thus just

$$l'_{N+1} = i\alpha\frac{m}{k} - \left(\frac{N-1}{2}\right). \tag{14}$$

This result shows that the continuum belongs to the principal series of the group $SO(N,1)$.[6]

## II. SCATTERING STATES AND MATRIX ELEMENTS

Ordinary, the Coulomb phase shifts are derived by constructing incoming and outgoing scattering states in parabolic coordinates. Our ability to separate the Hamiltonian in both spherical and parabolic coordinates is intimately connected to the higher symmetry of the Kepler problem. A separation in parabolic coordinates corresponds to a diagonalization of an alternate complete set of operators. In three dimensions, this means that we simultaneously diagonalize the following set of operators: $\{H, L^2, L_{12}\}$ for an spherical quantization and $\{H, A_3 = L_{34}, L_{12}\}$ in a parabolic one.[7] With this hint, we can consider in $N$ dimensions the complete set of $N$ commuting operators: $L_{N+1}'^2, L'_{N,N+1}, L_{N-1}^2, L_{12}$. Note that we have just replaced $L_N^2$ by $L'_{N,N+1}$.

The generator $L'_{N,N+1}$ unlike $L_N^2$ is diagonal on the special in and out scattering states which at $x_N\to+\infty$ or $x_N\to-\infty$ are asymptotically moving as plane waves $e^{ikx_N}$ of momentum $k\hat{x}_N$ parallel to the $x_N$ axis. Using Eq. (12) and (3), we find that for the asymptotic plane wave forms of the in and out scattering states

$$L'_{N,N+1}|\Psi_{\mathbf{k}\,\text{out}}^{\text{in}}\rangle = \frac{m}{k}A_N|\Psi_{\mathbf{k}\,\text{out}}^{\text{in}}\rangle = \left[\pm\frac{\alpha m}{k}+i\left(\frac{N-1}{2}\right)\right]|\Psi_{\mathbf{k}\,\text{out}}^{\text{in}}\rangle. \tag{15a}$$

In addition, the action of the other operators on the asymptotic plane wave states is

$$L_\alpha^2|\Psi_{\mathbf{k}\,\text{out}}^{\text{in}}\rangle = L_{12}|\Psi_{\mathbf{k}\,\text{out}}^{\text{in}}\rangle = 0,\quad \alpha = N-1,\ldots,3. \tag{15b}$$

Even though the in and out states take on these asymptotic plane wave forms only at the earliest times $t\to-\infty$ for the in state or the latest times $t\to-\infty$ for the out state, these eigenvalue equations at very early or late times are valid at all times, since the $SO(N,1)$ Lie algebra commutes with the Hamiltonian.

Since a partial wave expansion corresponds to expanding the in or out scattering states in terms of the orbital angular momentum states, Eq. (15b) tells us that only states of the form $|l'_{N+1}, l_N, 0,\ldots,0\rangle$ can appear in this expansion. Physically, this means that we do not have an angular dependence (except on $\theta_N$!) in the scattering amplitude. The in and out states can be expanded as

$$|\Psi_{\mathbf{k}\,\text{out}}^{\text{in}}\rangle = \sum_{l_N=0}^{\infty} a_{l_N\,\text{out}}^{\text{in}}|l'_{N+1},l_N,0,0,\ldots,0\rangle \equiv \sum_{\lambda=0}^{\infty} a_{\lambda_{\text{out}}}^{\text{in}}|k,\lambda\rangle \tag{16}$$

We have simplified the labeling of the state $|l'_{N+1},l_N,0,\ldots,0\rangle$ by replacing it with $|k,\lambda\rangle$. The label $k$ indicates that we are in a subspace of fixed energy $E=k^2/2m$ and the label $\lambda$ has replaced the orbital angular momentum parameter $l_N$.

In order to find the phase shifts, we need to evaluate the matrix elements of $L'_{N,N+1}$ between scattering states. These matrix elements can be obtained from those given by Louck for the $SO(N)$ groups in the following way. We take his matrix elements and make an analytical continuation in the $SO(N+1)$ representation parameter $l_{N+1}$ to the $SO(N,1)$ representation parameter $l'_{N+1}$ in them. We find[8]

$$L'_{N,N+1}|l'_{N+1},l'_N,0,...\rangle$$
$$= -i[(l'_N+1)(l'_N+N-2)\{[l'_N+(N-1)/2]^2+\alpha^2m^2/k^2\}\{(2l'_N+N)(2l'_N+N-2)\}^{-1}]^{1/2}\cdot|l'_{N+1},l'_N+1,0,...\rangle$$
$$+i\left[\frac{l'_N(l'_N+N-3)[[l'_N+(N-3)/2]^2+\alpha^2m^2/k^2]}{(2l'_N+N-2)(2l'_N+N-4)}\right]^{1/2}|l'_{N+1},l'_N-1,0,...\rangle. \quad (17)$$

With these results we can now evaluate the coefficients $a^{\text{in}}_{\lambda_{\text{out}}}$ of the partial wave expansion (16) for the scattering states. In fact, from Eq. (15) and (16) we have

$$L'_{N,N+1}|\Psi_{\mathbf{k}}{}^{\text{in}}_{\text{out}}\rangle = \left(\frac{N-1}{2}\pm iq\right)|\Psi_{\mathbf{k}}{}^{\text{in}}_{\text{out}}\rangle = \sum_{\lambda=0}^{\infty} a^{\text{in}}_{\lambda_{\text{out}}} L'_{N,N+1}|k,\lambda\rangle, \quad (18)$$

where we have defined $q$ to be $q\equiv\alpha m/k\equiv q^{\text{in}}$. Clearly if we find $a^{\text{in}}_\lambda$, then $a^{\text{out}}_\lambda$ can be obtained from $a^{\text{in}}_\lambda$ by replacing $q$ by $-q$ in it.

Using Eq. (17) together with Eq. (18), we can establish the following recursion relation for the partial wave coefficients $a^{\text{in}}_\lambda$:

$$\left[\frac{N-1}{2}-iq\right]a^{\text{in}}_\lambda = -\left(\frac{\lambda(\lambda+N-3)\{[\lambda+(N-3)/2]^2+q^2\}}{(2\lambda+N-2)(2\lambda+N-4)}\right)^{1/2}a^{\text{in}}_{\lambda-1}$$
$$+\left[\frac{(\lambda+1)(\lambda+N-2)\{[\lambda+(N-1)/2]^2+q^2\}}{(2\lambda+N-2)(2\lambda+N)}\right]^{1/2}a^{\text{in}}_{\lambda+1}. \quad (19)$$

The solution of this recursion relation is easily obtained and reads as follows

$$a^{\text{in}}_\lambda = \left(\frac{(2\lambda+N-2)(\lambda+N-3)}{\lambda!}\,\frac{[\lambda+(N-3)/2-iq]!}{[\lambda+(N-3)/2+iq]!}\right)^{1/2}a_0 \quad (20)$$

where $a_0$ is an arbitrary constant that can only depend on $k$ and must be real. As one expects this result agrees with the one given in Ref. (3) for the three-dimensional case.

From the definition of the $S$ matrix:

$$|\Psi^{\text{in}}_{\mathbf{k}}\rangle = S\,|\Psi^{\text{out}}_{\mathbf{k}}\rangle \quad (21)$$

and knowing that the problem is invariant under rotations, i.e., that $S$ is diagonal in the angular momentum representation: $S\,|k,\lambda\rangle = S_\lambda\,|k,\lambda\rangle$, with $S_\lambda = e^{2i\delta_\lambda}$, we find that.

$$a^{\text{in}}_\lambda = e^{2i\delta_\lambda}a^{\text{out}}_\lambda. \quad (22)$$

Combining this relation with the results obtained in Eq. (20) we obtain the desired phase shifts:

$$e^{2i\delta_\lambda} = \frac{[\lambda+(N-3)/2-iq]!}{[\lambda+(N-3)/2+iq]!} \quad (23)$$

We see that the analytical structure of this expression provides us with the correct energy spectrum for the $N$-dimensional analog of the hydrogen atom [see Eq. (10)]. Also, as one expects, this result agrees with the well-known three-dimensional result.

The next question that one can ask concerns the scattering amplitude. Recently Adawi[9] has considered the scattering amplitude for an arbitrary central potential in an $N$-dimensional space. From his results and specifying that we are dealing with a "Coulomb potential," one obtains for the partial wave expansion of the scattering amplitude:

$$f^{(N)}(\theta) = -i2^{\alpha-1}\Gamma(\alpha)\left[\frac{2}{\pi k^{2\alpha+1}}\right]^{1/2}$$
$$\times\sum_{\lambda=0}^{\infty}(\alpha+\lambda)C^\alpha_\lambda(\cos\theta)e^{2i\delta_\lambda}, \quad (24)$$

where $\alpha\equiv(N-2)/2$ and $C^\alpha_\lambda(\cos\theta)$ are the spherical harmonics of Gegenbauer. This expression reduces to the well-known expansion when $N$ is set equal to three, namely

$$f^{(3)}(\theta) = \frac{-i}{2k}\sum_{\lambda=0}^{\infty}(2\lambda+1)e^{2i\delta}\lambda P_\lambda(\cos\theta).$$

We need to point out that this expansion, as J.R. Taylor[10] has shown, does not converge in the usual sense, but does converge in the sense of distributions. Thus one can expect that the same problem will occur in our $N$ dimensional Coulomb problem. In other words, we must consider the expansion (24) as a formal one.

If we use the relation[11]

$$(1-x)^{-p} = \frac{2^{2\lambda-p}}{\pi^{1/2}}\frac{\Gamma(\lambda)\Gamma(\lambda-p+\frac{1}{2})}{\Gamma(p)}$$
$$\times\sum_{n=0}^{\infty}(n+\lambda)C^\lambda_n(x)\frac{\Gamma(p+n)}{\Gamma(2\lambda-p+n-1)} \quad (25)$$

and the phase shifts given in Eq. (23), we can carry out the sum in Eq. (24) and thus finally obtain the scattering amplitude

$$f^{(N)}(\theta) = -\frac{1}{k}\frac{1}{[2k]^{(N-1)/2}}\frac{1}{[\sin^2(\theta/2)]^{(N-1)/2}}$$

$$\times\frac{\Gamma((N-1/2)-iam/k)}{\Gamma(1+iam/k)}\exp\frac{iam}{k}\ln\sin^2\frac{\theta}{2}. \tag{26}$$

This expression reduces for $N=3$ to the usual Coulomb amplitude.

As a final remark, we would like to point out that we have obtained the result exhibited in Eq. (26) also in a different way i.e., by directly separating the Schrödinger equation in parabolic coordinates, as one usually does for the three-dimensional case. This result is not surprising as we pointed out earlier, since the fact that the equation can be separated in this set of coordinates is a consequence of the SO($N$ + 1) symmetry of this problem.

[1]See for example the references given by M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330 (1966). Also see D.R. Herrick, and O. Sinanŏglu, Rev. Mex. Física **22**, 1 (1973).

[2]L.C. Biedenharn and P.J. Brussard, "Coulomb Exitation," (Oxford U.P., New York, 1965).

[3]D. Zwanziger, J. Math. Phys. **8**, 1858 (1967).

[4]It is worth while to mention that the dynamical group of the hydrogen atom, the O(4,2) group has been used in order to evaluate the Coulomb amplitude, see A.O. Barut and W. Rasmussen, J. Phys. B. Atom. Molec. Phys. **6**, 1965 (1973). Also the O(4) dyon–dyon system has been investigated in the spirit of reference (3) by D. Zwanziger, Phys. Rev. **176**, 1480 (1968).

[5]J.D. Louck, "Theory of Angular Momentum in $N$-Dimensional Space", Los Alamos Scientific Lab. Rept. LA-2451 (1960); J.D. Louck and H.W. Galbraith, Rev. Mod. Phys. **44**, 540 (1972).

[6]Kurt Bernardo Wolf, J. Math. Phys. **12**, 197 (1971).

[7]D. Park, Z. Physik, **159**, 155 (1960). See also B.G. Wybourne, *Classical groups for Physicists* (Wiley, New York, 1974).

[8]This result can be obtained from Louck's[5] results given in Sec. 3, Eq. (3.7a), (3.7b), and (3.7c), or equivalently from the results given by S.C. Pang and K.T. Hecht, J. Math. Phys. **8**, 1233 (1967). In fact, if we consider Eqs. (5.44), (5.45) and (5.46) of this last reference, we obtain after a tedious calculation the matrix elements of $L_{N,N+1}$ for the SO($N$ + 1) group between the states labeled by $l_{N+1}$ and $l_N$:

$$L_{N,N+1}|l_{N+1},l_N,0,\ldots\rangle$$
$$= -i\left(\frac{(l_N+1)(l_N+N-2)(l_{N+1}-l_N)(l_{N+1}+l_N+N-1)}{(2l_N+N)(2l_N+N-2)}\right)^{1/2}$$
$$\times|l_{N+1},l_N+1,0,\ldots\rangle$$
$$+ i\left(\frac{l_N(l_N+N-3)(l_{N+1}-l_N+1)(l_{N+1}+l_N+N-2)}{(2l_N-N-2)(2l_N+N-4)}\right)^{1/2}$$
$$\times|l_{N+1},l_N-1,0,\ldots\rangle.$$

In this expression we make an analytical continuation on $l_{N+1}$ to the values given in Eq. (14), i.e., $l_{N+1}\to l'_{N+1}$. Since we have ambiguities in the square roots occuring in the above matrix elements, we choose a phase convention that agrees with that given by Naimark for the case of SO(3,1). See M.A. Naimark, "Linear Representation of the Lorentz Group," American Mathematical Society, Providence, Rhode Island (1957).

[9]I. Adawi, J. Math. Phys. **12**, 358 (1971).

[10]J.R. Taylor, Nuovo Cimento B **23**, 313 (1974).

[11]W. Magnus, F. Oberhettinger, and R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*. (Springer-Verlag, N.Y. 1966), 3rd ed.

# Group content of the Foldy–Wouthuysen transformation and the nonrelativistic limit for arbitrary spin

J. León, M. Quirós, and J. Ramirez Mittelbrunn

*Instituto de Estructura de la Materia, Serrano, 119. Madrid-6, Spain and Departamento de Física, Universidad de Alcalá de Henares, Alcalá de Henares, Spain*

(Received 19 September 1978; revised manuscript received 15 November 1978)

The relationship between Foldy–Wouthuysen and Lorentz transformations has been clarified throughout this paper. We propose a generalized FW transformation connecting two particular realizations of the $(m,j)$ representation of the Poincaré group: the covariant realization and a canonical realization acting on relativistic probability amplitudes. Fermions and bosons must be considered separately because the intrinsic parity of the particle–antiparticle systems is $(-1)^{2j}$. Thus for fermions we can directly take the $2(2j+1)$ - dimensional Joos–Weinberg covariant realization, while for bosons we must double it to get a reducible $4(2j+1)$ - dimensional realization where particles and antiparticles lie in orthogonal subspaces. In short, in momentum space the FW transformation is the matrix representing a Lorentz boost times the factor $(m/p_0)^{1/2}$, while in configuration space the FW transformation does not belong to the Poincaré group. The last part of the paper is devoted to getting quantum-mechanical representations of the Galileo group as a contraction of Poincaré group representations by using mathematical methods earlier developed by Mickelsson and Niederle. The relevance of our generalized FW transformation for getting a smooth, well defined, nonrelativistic limit is a remarkable result.

## 1. INTRODUCTION

The aim of this work is threefold: (i) to clarify the relationship between Foldy–Wouthuysen (FW) and Lorentz transformations; (ii) To study—using rigorous mathematical methods—the nonrelativistic limit for arbitrary spin particles by means of contracting representations from the Poincaré ($\mathscr{P}$) to the Galileo ($\mathscr{G}$) group; and (iii) To analyze the link among FW transformations, nonrelativistic limit, and probabilistic interpretation of wavefunctions.

Since Foldy and Wouthuysen built up—in their famous paper[1]—a unitary transformation passing from Dirac to canonical equation

$$i\frac{\partial}{\partial t}\phi = \beta\omega(-i\nabla)\phi(x)$$

for spin-$\frac{1}{2}$ particles, several generalizations of this transformation for higher spins have been proposed. Let us emphasize the main features of the original FW transformation: (a) It is unitary in the ordinary sense,[2] that is $U^{\dagger} = U^{-1}$; (b) it is related in a direct way to Lorentz boosts; and (c) it diagonalizes the Dirac Hamiltonian eliminating in this way odd operators. These three features have already been included in the different generalizations of FW transformations for higher spins. Thus, Ramanujan[3] and Tekumalla and Santhanaman[4] have defined unitary FW transformations starting from canonical wavefunctions. However, as it has been pointed out by Jayaramanan,[5] it is not possible to get local covariant wavefunctions, Eq. (25), by applying a unitary transformation to canonical wavefunctions, and so generalized FW transformations are not allowed to be unitary. In fact the unitarity of the FW transformation for spin 1/2 is, purely accidental, a consequence of the Lorentz invariant scalar product for Dirac wavefunctions, in the configuration space, whose metric operator is the unity.

On the other hand, a generalized FW transformation has been associated to Lorentz transformations passing from rest to laboratory covariant wavefunctions by Weaver, Hammer, and Good, in a classic paper,[6a] and, independently, by Matthews.[6b] However the work of these authors is lacking a clear group-theoretical interpretation. In fact the original motivation for FW transformation was to perform the nonrelativistic limit $c\to\infty$ starting from covariant wavefunctions. Nevertheless getting the canonical equation—and subsequently an Euclidean scalar product in configuration space—is not enough to carry out a nonrelativistic limit with a finite answer over group transformations. As we shall see later there is an infinity of $\mathscr{P}$ realizations obeying the canonical equation, but only one of these having an associated probabilistic interpretation, is suitable to perform the nonrelativistic limit. The distinction between fermions and bosons, implemented by the requirement of causality in second quantization, was pointed out by Nelson and Good,[7a] in a review paper, and by Hammer, McDonald, and Pursey,[7b] in a subsequent article. However these authors[7b] prove that Foldy's transformed of WHG wavefunctions do not obey the canonical equation for bosons. Thus, in order to overcome this difficulty, a different realization will be given for bosons, and FW transformations will be considered separately for fermions and for bosons.

Recently Krajcik and Nieto[8] gave an abstract definition of FW transformation as a pseudounitary transformation diagonalizing a pseudo-Hermitian Hamiltonian in a space

 0022-2488/79/061068-09$01.00  

with indefinite metric. These authors have studied the existence conditions proving that such a transformation does exist, and it is related to Lorentz transformations, for a class of first order relativistic equations without constrains. We shall not adhere to this approach but, instead, we shall find explicit solutions for generalized FW transformations with the Hamiltonian in diagonal form.

In the present work we study the problem from a group-theoretical point of view. Since Wigner's classic paper[9] unitary irreducible representations (UIR) of the Poincaré group are well known. In fact, a UIR is specified by two indices, coming from the two invariants of $\mathscr{P}$: $m$ and $j$ ($m$ being the mass and $j$ the spin of the associated particle). However in order to specify a particular realization we need to fix a corresponding functional Hilbert space by means of a wave equation and an invariant scalar product which makes the realization unitary. Moreover two different realizations of the same representation $(m, j)$ are related by a similarity transformation. In the following we shall consider finite mass and arbitrary spin representations.

Following the Mackey theory of induced representations[10,11] we can display two kinds of realizations for a given representation $(m, j)$ of $\mathscr{P}$:(a) Covariant realizations, characterized by a local transformation law as

$$U(a,\Lambda)\psi(x) = \mathscr{D}^{(j)}(\Lambda)\psi(\Lambda^{-1}(x-a))$$

and, (b) Wigner realizations, characterized by a nonlocal transformation law

$$U(a,\Lambda)\Phi(x) = \mathscr{D}^{(j)}(B_p^{-1}\Lambda B_{\Lambda^{-1}p})\Phi(\Lambda^{-1}(x-a)),$$

where $\mathbf{p} = -i\nabla$, $B_p^{-1}\Lambda B_{\Lambda^{-1}p}$ belongs to the little group of $(m,\mathbf{0})$ and $(a,\Lambda)$ stands for an element of $\mathscr{P}$. Only for the spin zero case, $\mathscr{D}^{(0)}(\Lambda) = 1$, both realizations coincide.

In other respects, neither covariant nor Wigner wavefunctions have a direct probabilistic interpretation. However, as we shall see along this work, by making use of Newton, Wigner, and Wightman's localizability theory[12] it is possible to construct functions $\phi(x)$—representing probability amplitudes for finding the particle at point $x$—from Wigner functions. Because the relevance of the probabilistic interpretation we shall introduce a third realization of the $(m, j)$ representation of $\mathscr{P}$ over amplitudes.

In this way we shall define the generalized FW transformation of a mass $m$ spin $j$ particle as that transformation mapping a covariant realization of $\mathscr{P}$ into a realization over amplitudes. With this definition, the FW transformation is unique and composed by two transformations: a Lorentz boost passing from the covariant to Wigner realization, and a second transformation mapping Wigner functions into probability amplitudes.

In this work we shall construct generalized FW transformations starting from the Joos and Weinberg[13] $2(2j+1)$-dimensional realization corresponding to $\mathscr{D}^{(j)}(\Lambda) = D^{(j,0)}(\Lambda) \oplus D^{(0,j)}(\Lambda)$. The doubling of the dimensionality is necessary when we wish to represent the full Poincaré group—in particular the parity—and not only the proper orthocronus group.

Starting from this point of view, it is easy to convince oneself that a clear distinction between fermions and bosons does emerge. This is due mainly to two facts: (a) The intrinsic parity particle antiparticle is $(-1)^{2j}$, as has been pointed out above, and (b) The parity operator is represented in the Joos–Weinberg realization by the $\beta$ matrix, with eigenvalues $\pm 1$, each one corresponding to a $(2j+1)$-dimensional subspace.[11] Thus the nonlocal condition $p_0 > 0$ ($p_0 < 0$) specifying particle (antiparticle) orbits can be replaced, for fermions, by the positive (negative) parity condition. In this way for half-integer spin representations in Weingberg realizations, fermions and antifermions do appear in orthogonal subspaces corresponding to the eigenvalues $+1$ and $-1$ of the parity matrix. On the other hand, for integer spin representations in Weinberg realization, bosons and antibosons are in the same subspace because the intrinsic parity is $+1$ in this case. We can summarize the distinction between fermions and bosons in the following way: To separate a particle and an antiparticle is equivalent to separating parity eigenstates for fermions while this is not true for bosons. As we shall see, this difficulty for bosons can be solved starting, from a $4(2j+1)$-dimensional covariant realization where particle and antiparticle states are separated from the very beginning.

Regarding the nonrelativistic limit, Inönü and Wigner[14] studied a limiting procedure between Lie algebras—called contraction—which was applied to get Galileo algebra from Poincaré algebra. Several problems where left open in Inönü–Wigner's work. In particular: (a) The contraction was not well defined for the group but only for the algebra. (b) The relation between $\mathscr{P}$ and $\mathscr{G}$ representations was not defined. A first step towards the enlargement of the definition to the group was given by Saletan[15] while a rigorous mathematical definition of group contraction is due to Mickelsson and Niederle.[16] These authors also gave the mathematical prescriptions for contracting representations. Finally a generalized contraction for Lie algebras has been studied by Doebner and Melsheimer,[17] where the power of the contraction parameter is no longer constrained to be equal to one. In this work, using Michelsson and Niederle prescriptions we are able to contract $\mathscr{P}$ representations into $\mathscr{G}$ representations for finite mass and arbitrary spin. The essential features of this contraction are already present for spinless particles and were studied in more detail in a previous paper.[18] The fundamental point for contracting representations is to find two representation spaces, for both groups, related by a similarity transformation depending on the contraction parameter ($c$ in this case) in such a way that, in the contraction limit ($c \to \infty$), the realization of a group goes to the realization of the other. In our case this role is played by probability amplitudes. We can compute the transformation mapping relativistic into nonrelativistic probability amplitudes. This transformation is the identity in momentum space while it is given, in the configuration space, by a rather complicated expression, Eq. (99).

Let us remark that starting from a true representation of $\mathscr{P}$ we can get nontrivial projective representations of $\mathscr{G}$ having a quantum-mechanical interpretation.[19]

## 2. REALIZATIONS

### A. (2j + 1) component representations

Let us consider here the restricted Poincaré group $\mathscr{P}^{\uparrow}_{+}$ whose unitary irreducible representations (UIR) are widely known since Wigner's work.[9] Thus a mass $m$, spin $j$ particle is represented by a $(2j+1)$ component wavefunction $a_\sigma(\mathbf{p})(\sigma = -j,...,j)$ transforming under elements $\{a,\Lambda\} \in \mathscr{P}^{\uparrow}_{+}$ as follows:

$$\{u(a,\Lambda)a\}_\sigma(\mathbf{p}) = e^{ipa} D^{(j)}_{\sigma\sigma'}(B_p^{-1}\Lambda B_{\Lambda^{-1}p}) a_{\sigma'}(\Lambda^{-1}\mathbf{p}), \quad (1)$$

where $D^{(j)}$ is the $(2j+1)$-dimension UIR of the three-space rotation group, $B_p$ is the boost transforming the 4-vector $\mathring{p} = (m,\mathbf{0})$ to $p = (\omega,\mathbf{p})$, with $\omega = (p^2+m^2)^{1/2}$, and the rotation $B_p^{-1}B_{\Lambda^{-1}p}$ belongs to the little group of $\mathring{p}$. Summation over repeated indices will be understood from now on.

The invariant scalar product for Wigner functions can thus be given as

$$(a,a') = \int \frac{d^3p}{\omega}\, a^*_\sigma(\mathbf{p}) a'_\sigma(\mathbf{p}). \quad (2)$$

Hence the components $\omega^{-1/2}a_\sigma(\mathbf{p})$ should be square-integrable functions of $\mathbf{p}$. Complex conjugated Wigner functions $a^*_\sigma$ will transform with the corresponding complex conjugate representation of the rotation group given by

$$D^{(j)*}(R) = CD^{(j)}(R)C^{-1}. \quad (3)$$

$C$ being a $(2j+1)^2$ matrix characterized by the properties $C^*C = (-)^{2j}$ and $C^+C = 1$, and whose explicit representations we do not need. A mass $m$, spin $j$ antiparticle will be represented in this way by the wavefunction $b^*_\sigma(\mathbf{p})$ transforming under $\{a,\Lambda\} \in \mathscr{P}^{\uparrow}_{+}$ as

$$[u(a,\Lambda)b^*]_\sigma(\mathbf{p}) = e^{-ipa}\{CD^{(j)}(B_p^{-1}\Lambda B_{\Lambda^{-1}p})C^{-1}\}_{\sigma\sigma'} b^*_{\sigma'}(\Lambda^{-1}\mathbf{p}). \quad (4)$$

Taking as the starting point the Wigner functions $a_\sigma(\mathbf{p})$, Weinberg[11] defines two covariant wavefunctions $\alpha_n(\mathbf{p})$ and $\beta_n(\mathbf{p})$

$$\alpha_n(\mathbf{p}) = u_n(\mathbf{p},\rho)a_\rho(\mathbf{p}), \quad (5a)$$

$$\beta_n(\mathbf{p}) = v_n(\mathbf{p},\rho)b^*_\rho(\mathbf{p}), \quad (5b)$$

the covariant spinors $u_n$, $v_n$ being constructured by the following boosting procedure:

$$u_n(p,\rho) = D_{nm}(B_p)u_m(\rho) \quad (6a)$$

$$v_n(p,\rho) = D_{nm}(B_p)v_m(\rho')C^{-1}_{\rho'\rho}, \quad (6b)$$

where $u_m(\rho)$ is the $m$th row of a Pauli spinor associated to the state with a third spin component, $\rho$, and $D_{nm}(\Lambda)$ being any Lorentz group representation such that $D(R)$ contains the component $D^{(j)}R$. This means that the representation applies to the spinors $u(\sigma)$ in the usual quantum-mechanical way

$$D_{nm}(R)u_m(\sigma) = u_n(\sigma')D^{(j)}_{\sigma'\sigma}(R), \quad (7a)$$

$$D_{nm}(R)v_m(\sigma) = v_n(\sigma')D^{(j)}_{\sigma'\sigma}(R). \quad (7b)$$

Using Eqs. (1), (4), and (5)–(7) we translate into covariant function language what was said about the trasnformation of Wigner functions

$$[u(a,\Lambda)\alpha]_n(\mathbf{p}) = e^{ipa}D_{nm}(\Lambda)\alpha_m(\Lambda^{-1}\mathbf{p}), \quad (8a)$$

$$[u(a,\Lambda)\beta]_n(\mathbf{p}) = e^{-ipa}D_{nm}(\Lambda)\beta_m(\Lambda^{-1}\mathbf{p}). \quad (8b)$$

Let us now restrict the above considerations to two important UIR of the Lorentz group.

(a) Representation $(j,0)$ characterized by the following choice for the algebra:

$$\mathbf{J}^{(j,0)} = \mathbf{J}, \quad \mathbf{K}^{(j,0)} = -i\mathbf{J}, \quad (9)$$

where $\mathbf{J}$ are the $(2j+1)$-dimensional matrices of the SU(2) algebra, and $D^{(j)}(\Lambda)$ will denote the matrix representation of the finite element $\Lambda$, in particular,

$$D^{(j)}(B_p) = \exp\left\{\frac{\mathbf{pJ}}{|\mathbf{p}|}\theta\right\}, \text{ with } |\mathbf{p}| = (mc)\sinh\theta. \quad (10)$$

The condition $D_{nm}(R) = D^{(j)}_{nm}(R)$ along with Eqs. (7) permits one to choose for the rotation basis

$$u_n(\sigma) = \delta_{n\sigma} = -v_n(\sigma). \quad (11)$$

In the configuration space we can build a covariant wavefunction

$$\phi_n(x) = (2\pi)^{-3/2}(mc)^{1/2}\int\frac{d^3p}{\omega} \times \{\alpha_n(\mathbf{p})e^{-ipx} + \beta_n(\mathbf{p})e^{ipx}\}, \quad (12)$$

where $\alpha_n$ and $\beta_n$ are the Weinberg functions (5) associated to the representation $(j,0)$. $\phi_n(x)$ transforms covariantly under $\mathscr{P}$

$$[u(a,\Lambda)\phi]_n(x) = D^{(j)}_{nm}(\Lambda)\phi_m(\Lambda^{-1}(x-a)). \quad (13)$$

(b) Representation $(0,j)$ characterized by the following choice:

$$\mathbf{J}^{(0,j)} = \mathbf{J}, \quad \mathbf{K}^{(0,j)} = i\mathbf{J}. \quad (14)$$

Let us denote by $\hat{D}^{(j)}(\Lambda)$ the matrix representing $\Lambda$

$$\hat{D}^{(j)}(\Lambda) = D^{(j)}(\Lambda^{-1})^+. \quad (15)$$

Thus for rotations both representations coincide, and the relations (7) and (10) are also valid in this case. However the corresponding definitions for covariant functions $\hat{\alpha}_n$ and $\hat{\beta}_n$ are given in terms of the $(0,j)$ covariant spinors

$$\hat{u}_n(\mathbf{p},\rho) = \hat{D}^{(j)}_{nm}(B_p)u_m(\rho), \quad (16a)$$

$$\hat{v}_n(p,\rho) = \hat{D}^{(j)}_{nm}(B_p)v_m(\sigma)C^{-1}_{\sigma\rho}, \quad (16b)$$

$\hat{D}$ being, according to (15),

$$\hat{D}^{(j)}(B_p) = \exp\left(\frac{\mathbf{pJ}}{|\mathbf{p}|}\theta\right) = D^{(j)}(B_p^{-1}). \quad (17)$$

In configuration space we can write in analogy to (12), a covariant wave function

$$\hat{\phi}_n(x) = (2\pi)^{-3/2}(mc)^{1/2} \times \int\frac{d^3p}{\omega}\{\hat{\alpha}_n(\mathbf{p})e^{-ipx} + (-)^{2j}\hat{\beta}_n(\mathbf{p})e^{ipx}\}. \quad (18)$$

It was necessary to insert the factor $(-)^{2j}$ in (18) in order to get the correct commutation relations between the fields. As we shall see in the following, this factor will be of the utmost importance in writing the FW transformation between antiparticles.

## B. 2(2$j$ + 1) component representations

Niederer and O'Raifertaigh[11] showed that a linear implementation of parity is possible if and only if there exists a pseudounitary representation $u(\Lambda)$. In other words, $u(\Lambda)$ should verify

$$u^{\dagger}(\Lambda)\eta u(\Lambda) = \eta \tag{19}$$

with $\eta^2 = 1$, $\eta^+ = \eta$, where $\eta$ is the parity matrix, that is,

$$\pi\psi(p) = \eta\psi(\hat{p}), \tag{20}$$

with $\hat{p} = (\omega, -\mathbf{p})$.

We shall next see that the pseudounitarity condition can be fulfilled by the direct sum $D^{(j,0)} \oplus D^{(0,j)}$. It is thus convenient to assemble both functions $\phi_n$ and $\hat{\phi}_n$ in a 2(2$j$ + 1)-component function, for instance

$$\psi(x) = 2^{-1/2}\begin{pmatrix}\phi(x)\\ \hat{\phi}(x)\end{pmatrix}. \tag{21}$$

This can be written as

$$\Psi(x) = (2\pi)^{-3/2}(mc)^{1/2} \times \int \frac{d^3p}{\omega}\{\Psi^{(+)}(\mathbf{p})e^{-ipx} + \Psi^{(-)}(\mathbf{p})e^{ipx}\}, \tag{22}$$

where

$$\Psi^{(+)}(\mathbf{p}) = \sum_{\sigma} U(\mathbf{p},\sigma)\, a(\mathbf{p},\sigma),$$

$$\Psi^{(-)}(\mathbf{p}) = \sum_{\sigma} V(\mathbf{p},\sigma)\, b^*(\mathbf{p},\sigma). \tag{23}$$

with the 2(2$j$ + 1)-dimensional spinors given by

$$U(\mathbf{p},\sigma) = \tilde{\mathscr{D}}^{(j)} b^*(B_p)U(\sigma),$$

$$V(\mathbf{p},\sigma) = \tilde{\mathscr{D}}^{(j)}(B_p)\, V(\sigma). \tag{24}$$

$\tilde{\mathscr{D}}^{(j)}$ is the representation $(j,0) \oplus (0,j)$, that is

$$\tilde{\mathscr{D}}^{(j)}(\Lambda) = \begin{pmatrix} D^{(j)}(\Lambda) & 0 \\ 0 & D^{(j)}(\Lambda^{-1})^{\dagger}\end{pmatrix} \tag{25}$$

which fullfills the pseudounitarity condition

$$\tilde{\mathscr{D}}^{(j)}(\Lambda)^{\dagger} = \tilde{\beta}\,\tilde{\mathscr{D}}^{(j)}(\Lambda^{-1})\,\tilde{\beta} \tag{26}$$

$$\text{with } \tilde{\beta} = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}. \tag{27}$$

Finally the four spinors $U$ and $V$ are simply given as

$$U(\sigma) = 2^{-1/2}\begin{pmatrix}u(\sigma)\\ u(\sigma)\end{pmatrix},\ V(\sigma) = -2^{-1/2} c_{\sigma\sigma'}\begin{pmatrix}u(\sigma')\\ (-)^{2j} u(\sigma')\end{pmatrix} = -2^{-1/2}\begin{pmatrix}v(\sigma)\\ (-)\, v(\sigma)\end{pmatrix} \tag{28}$$

Each branch of the hyperboloid $p^2 = m^2c^2$ transforms according to a separate representation of $\mathscr{P}$, therefore two independent invariants can be drawn out of $\Psi(x)$. Moreover $\Psi^{(+)}(\mathbf{p})$ and $\Psi^{(-)}(\mathbf{p})$ belong the same representation of SL(2,$C$), so that we can write (11)

$$(\Psi_1,\Psi_2) = \int \frac{d^3p}{\omega}\{\Psi_1^{(+)\dagger}\tilde{\beta}\,\Psi_2^{(+)} + \epsilon(-)^{2j}\,\Psi_1^{(-)\dagger}\tilde{\beta}\,\Psi_2^{(-)}\}. \tag{29}$$

Only in the case where $\epsilon = +1$ the invariant is positive definite. A further analysis of this invariant shows that for $\epsilon = (-)^{2j+1}$ it is local in configuration space, explicitly

$$(\Psi_1,\Psi_2) = i(2\pi)^{-3} \times \int d^3x\,\{\Psi_1^{\dagger}(x)\tilde{\beta}\,\dot{\Psi}_2(x) - \dot{\Psi}_1^{\dagger}(x)\tilde{\beta}\,\Psi_2(x)\}\,(mc)^{-1}, \tag{30}$$

while for $\epsilon = (-)^{2j}$ it is non-local. As a consequence only for fermions a local invariant does exist which is positive definite, unlike the case of bosons where we should decide between positivity or configuration space locality. The treatment for fermions will be clearly different of that for bosons, as will be patent in the next section.

# 3. FOLDY–WOUTHUYSEN TRANSFORMATIONS FOR ARBITRARY SPIN

In the above paragraphs we noticed the fact that representation $(j,0) \oplus (0,j)$ has room for both fermions and bosons. Here we shall be concerned with the possibility of separating particles and antiparticles [for instance, putting particles (antiparticles) in the (2$j$ + 1) upper (lower) dimensional subspace]. Let us introduce the representation $\mathscr{D}^{(j)}$ equivalent to $\tilde{\mathscr{D}}^{(j)}$

$$\mathscr{D}^{(j)}(\Lambda) = S\tilde{\mathscr{D}}^{(j)}S^{-1}, \tag{31}$$

where $S$ is the unitary transformation

$$S = 2^{-1/2}\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}. \tag{32}$$

In this new representation the wavefunction (22) reads

$$\psi(x) \equiv S\Psi(x) = (2\pi)^{-3/2}(mc)^{1/2} \times \int \frac{d^3p}{\omega}\{\psi^{(+)}(\mathbf{p})e^{-ipx} + \psi^{(-)}(\mathbf{p})e^{ipx}\}, \tag{33}$$

with

$$\psi_n^{(+)}(\mathbf{p}) = \omega_n^{(+)}(\mathbf{p},\sigma)a(\mathbf{p},\sigma),$$

$$\psi_n^{(-)}(\mathbf{p}) = \omega_n^{(-)}(\mathbf{p},\sigma)b^*(\mathbf{p},\sigma), \tag{34}$$

and

$$\omega^{(\pm)}(\mathbf{p},\sigma) = \mathscr{D}^{(j)}(B_p)u_j^{(\pm)}(\sigma). \tag{35}$$

The spinors $u_j^{(+)}$ are the same for bosons and fermions

$$u_j^{(+)}(\sigma) = \begin{pmatrix}u(\sigma)\\ 0\end{pmatrix}, \tag{36}$$

but

$$u_j^{(-)}(\sigma) = \begin{pmatrix}0\\ v(\sigma)\end{pmatrix}, \quad \text{for half-integer } j,$$

$$u_j^{(-)}(\sigma) = \begin{pmatrix} -v(\sigma) \\ 0 \end{pmatrix} \quad \text{for integer } j. \tag{37}$$

Let us now define $2(2j+1)$-dimensional Wigner functions as

$$\Phi_n^{(\pm)}(\mathbf{p}) = \mathscr{D}_{nm}^{(j)}(B_p^{-1})\psi_m^{(\pm)}(p). \tag{38}$$

these will transform under the action of $\mathscr{P}$ as

$$[u(a,\Lambda)\Phi^{(\pm)}]_n(\mathbf{p}) = e^{\pm iap}\mathscr{D}_{nm}^{(j)}[B_p^{-1}\Lambda B_{\Lambda^{-1}p}]\Phi_m^{(\pm)}(\Lambda^{-1}\mathbf{p}). \tag{39}$$

Again, $B_p^{-1}\Lambda B_{\Lambda^{-1}p}$ is an element of the little group associated to $\overset{\circ}{p}$.

It is worth while to remark here the resemblance between the $2(2j+1)$-component functions introduced in this section and the original $(2j+1)$-component Wigner functions. However in the $(2j+1)$ case, functions transform under $D^{(j)}$,(1), while antiparticles transform under $CD^{(j)}C^{-1}$, (4). This is not the case here where both particles and antiparticles transform stand in the same representation $\mathscr{D}^{(j)}$, because of the rotation $C$ performed in the rotation basis, Eq. (26).

We can write $2(2j+1)$ Wigner functions in configuration space as

$$\Phi(x) = (2\pi)^{-3/2}(mc)^{1/2}\int \frac{d^3p}{\omega}\{\Phi^{(+)}(\mathbf{p})\,e^{-ipx} + \Phi^{(-)}(\mathbf{p})e^{ipx}\}. \tag{40}$$

An analysis of Eq. (38) reveals that the functions $\Phi^{(+)}$ and $\Phi^{(-)}$ of (38) are orthogonal if and only if $j$ is half-integer. As a consequence of this, the functions $\Phi$ [Eq. (40)] describing fermions satisfy

$$i\partial_t\Phi(x) = \beta\omega(-i\nabla)\Phi(x) \tag{41}$$

which is not true for bosons, due to the lack of projection operators $(1\pm\beta)$ for integer $j$. As Eq. (41) will be essential in the next section we shall treat fermions and bosons separately.

## A. Regarding fermions

We shall generalize here the treatment developed by Foldy–Wouthuysen for spin 1/2, that is for wavefunctions verifying the Dirac equation and transforming according to the representation $(1/2,0)\oplus(0,1/2)$. Our first step concerns localizability.

Following Newton–Wigner–Wightman[12] we shall define localized fermions states as

$$\Phi_{\mathbf{x}}^{(\pm)}(\mathbf{p}) = e^{\pm ipx}\omega^{1/2}, \tag{42}$$

with $u^{(\pm)}$ as given by (36) and (37). Therefore the probability amplitude for finding fermions at $x$ is given by

$$\phi^{(\pm)}(x) = (2\pi)^{-3/2}\int \frac{d^3p}{\omega^{1/2}}e^{\mp ipx}\Phi^{(\pm)}(\mathbf{x}) = (2\pi)^{-3/2}\int d^3p\, e^{\mp ipx}\phi^{(\pm)}(\mathbf{p}), \tag{43}$$

where

$$\phi^{(\pm)}(\mathbf{p}) = \omega^{-1/2}\Phi^{(\pm)}(p) \tag{44}$$

is the probability amplitude in momentum space, which [with the aid of (37)] we see transformed as

$$[u(a,\Lambda)\phi^{(\pm)}]_n(p) = e^{\pm iap}\left(\frac{\omega(\Lambda^{-1}p)}{\omega(p)}\right)^{1/2}\mathscr{D}_{nm}^{(j)} \times [B_p^{-1}\Lambda B_{\Lambda^{-1}p}]\,\phi_m^{(\pm)}(\Lambda^{-1}p). \tag{45}$$

The configuration space amplitude $\phi(x) = \phi^{(+)}(x) + \phi^{(-)}(x)$ satisfies the canonical equation (41), as can be directly deduced from the fact

$$\beta\phi^{(\pm)}(p) = \pm\phi^{(\pm)}(p). \tag{46}$$

Let us now define the FW transformation $U^{(j)}$ as that mapping covariant functions (33) into probability amplitudes (43), that is

$$\psi(x) = U^{(j)}\phi(x). \tag{47}$$

Taking into account that $\phi^{(+)}(p)$ is orthogonal to $\phi^{(-)}(p)$, we obtain the explicit form of the FW transformation

$$U^{(j)} = (mc/\omega)^{1/2}\mathscr{D}^{(j)}(-i\beta\nabla), \tag{48}$$

where $\mathscr{D}^{(j)}(-i\beta\nabla)$ is the boost $\mathscr{D}^{(j)}(Bp)$ in configuration space, where according to (43) and (46) we substitute $\mathbf{p}$ by $-i\beta\nabla$. Let us see that for $J=\frac{1}{2}$ this definition coincides with that of Foldy. Using (11) and taking into account that

$$\mathbf{J}_{1/2} = \boldsymbol{\sigma}/2 \tag{49}$$

we get

$$D^{(1/2)}(Bp) = \frac{mc+\omega-\boldsymbol{\sigma}\mathbf{p}}{[2mc(\omega+mc)]^{1/2}} \tag{50}$$

and

$$\mathscr{D}^{(1/2)}(Bp) = \frac{mc+\omega+\boldsymbol{\alpha}\mathbf{p}}{[2mc(\omega+mc)]^{1/2}}, \quad \boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}. \tag{51}$$

Then inserting (51) in (48) we get the FW expression[1]

$$U^{(1/2)} = \frac{mc+\omega+\boldsymbol{\gamma}\mathbf{p}}{[2\omega(\omega+mc)]^{1/2}}, \quad (\mathbf{p} = -i\nabla). \tag{52}$$

To complete our approach and to prepare the nonrelativistic limit, let us write the positive definite scalar product [(29) and (30)] in terms of amplitudes. Taking into account (19) and the fact that $\eta=\beta$ up to a phase factor, we can use definition (38) to get

$$(\Psi_1,\Psi_2) = \int d^3p\,\{\phi_1^{(+)\dagger}\beta\phi_1^{(+)} - \phi_1^{(-)\dagger}\beta\phi_2^{(-)}\} \tag{53}$$

or, what is the same

$$(\Psi_1,\Psi_2) = \int d^3p\,\{\phi_1^{(+)\dagger}\phi_2^{(+)} + \phi_1^{(-)\dagger}\phi_2^{(-)}\}. \tag{54}$$

In configuration space this expression is simply

$$(\Psi_1,\Psi_2) = (2\pi)^3\int d^3x\,\phi_1^{\dagger}(x)\phi_2(x) \tag{55}$$

in complete analogy with the nonrelativistic scalar product.

## B. Regarding bosons

We have already mentioned the fact that neither covariant wavefunctions (33) nor Wigner functions (40) are suitable for defining the FW transformation for bosons. This can be easily understood by inspecting the spinor part [(36) and (37)] of Wigner functions. Particles and antiparticles belong to the same subspace, thus we have not at hand projectors equivalent to $(1 \pm \beta)/2$ to build up the canonical representation. This is closely related to the fact that the intrinsic parity of the boson–antiboson system should be $+1$ due to causality. In order to introduce a FW transformation we shall need to start from covariant functions with $4(2j+1)$ components. We first invert (33)

$$\psi^{(\pm)}(p) = [4(2\pi)^3 mc]^{-1/2} \int d^3x\, e^{\mp ipx} \{\omega\psi(x) \pm i\dot{\psi}(x)\}. \tag{56}$$

Now we take the $4(2j+1)$ covariant spinor

$$\psi(p) = \begin{pmatrix} \psi^{(+)}(p) \\ \psi^{(-)}(p) \end{pmatrix} \tag{57}$$

which transforms under $\mathscr{P}$ as

$$[U(a,\Lambda)\,\psi](p) = \begin{pmatrix} e^{ipa}\mathscr{D}^{(j)}(\Lambda) & 0 \\ 0 & e^{-ipa}\mathscr{D}^{(j)}(\Lambda) \end{pmatrix} \psi(\Lambda^{-1}p). \tag{58}$$

This is a completely reducible representation, being simply a multiple of the identity for $\mathscr{L}^{\uparrow}_{+}$ transformations. This is the price to be paid for the canonical representation. The $4(2j+1)$ Wigner functions can be defined following (38)

$$\Phi(p) = (\mathscr{D}^{(j)}(B_p^{-1}) \otimes I)\psi(p) \tag{59}$$

with transformations laws parallel to (39).

Let us now define the localized boson states as in the fermion case

$$\Phi_{\mathbf{x}}^{(\pm)}(p) = e^{\pm ipx}\omega^{1/2}. \tag{60}$$

Thus the probability amplitude for finding the boson at point $x$ will be

$$\phi_n^{(\pm)}(x) = (2\pi)^{-3/2} \int \frac{d^3p}{\omega^{1/2}} e^{\mp ipx} \Phi^{(\pm)}(p)$$

$$= (2\pi)^{-3/2} \int d^3p\, e^{\mp ipx} \phi^{(\pm)}(p), \tag{61}$$

where $\phi(p)$ is again the corresponding probability amplitude in momentum space

$$\phi^{(\pm)}(p) = \omega^{-1/2}\Phi^{(+)}(p). \tag{62}$$

In these formulas we made use of the projectors

$$P_\pm = (1 \pm \eta)/2, \tag{63}$$

where $\eta$ is the operator giving the sign of the energy

$$\eta = I \otimes \beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{64}$$

The $4(2j+1)$ amplitude in configuration space is now

$$\phi(x) = (2\pi)^{-3/2} \int d^3p\, \{e^{-ipx}\phi^{(+)}(p) + e^{ipx}\phi^{(-)}(p)\} \tag{65}$$

which obviously satisfies the canonical equation

$$i\partial_t\phi(x) = \eta\omega(-i\nabla)\phi(x). \tag{66}$$

On the other hand, covariant functions, (57) are in configuration space

$$\psi(x) = [(2\pi)^{-3}mc]^{1/2} \int \frac{d^3p}{\omega} \{\psi^{(+)}(p) \times e^{-ipx} + \psi^{(-)}(p)\, e^{ipx}\}, \tag{67}$$

where

$$\psi^{(\pm)}(p) = P_\pm \psi(p). \tag{68}$$

In this representation, the scalar product (29) is

$$(\psi_1,\psi_2) = \int \frac{d^3p}{\omega} \{\psi_1^{+}\beta'\,\psi_1^{(+)} + \psi_2^{(-)\prime}\beta'\,\psi_2^{(-)}\}, \tag{69}$$

with

$$\beta' = \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix} = \beta \otimes I. \tag{70}$$

Now using the pseudounitarity of the representation (58), that is,

$$\mathscr{D}^{(j)}(\Lambda)^{+} \otimes I = \beta'\{\mathscr{D}^{(j)}(\Lambda^{-1}) \otimes I\}\beta' \tag{71}$$

[which is a direct consequence of (26)], we can write the scalar product in terms of amplitudes as

$$(\psi_1,\psi_2) = \int d^3p\, \{\phi_1^{(+)\dagger}\phi_2^{(+)} + \phi_1^{(-)\dagger}\phi_2^{(-)}\}$$

$$= \int d^3x\, \phi_1^\dagger(x)\phi_2(x) \tag{72}$$

which looks like (55) but with a double number of components for the fields.

Returning to the FW transformation, we define it as that mapping $4(2j+1)$ covariant functions to amplitudes, that is,

$$\psi(x) = U^{(j)}\phi(x), \tag{73}$$

or explicitly

$$U^{(j)} = \left(\frac{mc}{\omega}\right)^{1/2} \begin{pmatrix} \mathscr{D}^{(j)}(i\nabla) & 0 \\ 0 & \mathscr{D}^{(j)}(-i\nabla) \end{pmatrix}$$

$$= \left(\frac{mc}{\omega}\right)^{1/2} (\mathscr{D}^{(j)}(\mathbf{p}) \otimes I)_{\mathbf{p}=i\eta\nabla}. \tag{74}$$

Let us remark that the parity matrix is $\beta'$ in this representation, so the boson–anti–boson intrinsic parity is $+1$ as requested. However, unlike the fermion case $\beta'$ does not coincide with the operator giving the sign of the energy $\eta$ (64).

The spinless boson is a specially simple case because the spin part of the Lorentz group is trivially represented by the identity $D^{00}(\Lambda) = 1$, therefore we do not need to double the representation as in Sec. 2.2 in order to get $2(2j+1)$ components. It is enough to follow the treatement introduced here in order to get a two-component canonical representation acting on (56) and (57) (where now $\psi^{(\pm)}$ has only one component). The representation thus obtained is the canonical representation due to Foldy.[20]

## 4. NONRELATIVISTIC LIMIT FOR ARBITRARY SPIN

### A. Contraction from the Poincaré group to the Galileo group

In a previous paper,[18] we worked out a rigorous method for contracting the Poincaré group to the Galileo group (in the limit $c \to \infty$). Let us briefly remind you of the main points used in that construction.[16]

Given two groups $G$ and $G'$, we say $G'$ is a global contraction of $G$ if we can construct a family of homeomorphisms $f_\epsilon$ of $G$ on $G'$ ($\epsilon \in (0,1]$) such that

$$x' \cdot y' = \lim_{\epsilon \to 0} f_\epsilon \, [f_\epsilon^{-1}(x') \cdot f_\epsilon^{-1}(y')] \quad \forall \; x',y',G'. \quad (75)$$

This means it is possible to deform continuously the group law in $G$ to the group law in $G'$.

In our particular problem $\mathscr{P} \to \mathscr{G}$, we can take $\epsilon = c^{-1}$ and we define the family $f_\epsilon$ in the following manner:

Given $g = e^{iaP} e^{i\boldsymbol{\beta}\mathbf{K}} R(\boldsymbol{\alpha}) \in \mathscr{P}$, we define $f_c(g)$ as

$$f_c(g) \equiv g' = e^{ibP'_0} e^{-i\mathbf{a}\mathbf{P}} e^{i\mathbf{V}\mathbf{K}'} R(\boldsymbol{\alpha}) \in \mathscr{G}, \quad (76)$$

where $b = a_0 c^{-1}, \mathbf{V} = c\boldsymbol{\beta}$, $P^\mu$, $\mathbf{K}$, $\mathbf{J}$ are the infinitesimal generators of $\mathscr{P}$ and $P'_0$, $\mathbf{P}$, $\mathbf{K}'$, and $\mathbf{J}$ those of $\mathscr{G}$ [$R(\boldsymbol{\alpha}) = \exp(i\boldsymbol{\alpha}\mathbf{J})$]. Then with the aid of the composition laws of $\mathscr{P}$ and $\mathscr{G}$ and the commutation rules of their Lie algebras, it is straightforward to check Eq. (75).

Along with group contraction we have an associated contraction of the algebras starting from

$$[J_i,J_j] = i\epsilon_{ijn} J_k, \quad [J_i,K_j] = i\epsilon_{ijn} K_n,$$
$$[K_i,K_j] = -i\epsilon_{ijn} J_n, \quad [J_i,P_j] = i\epsilon_{ijn} P_n,$$
$$[J_i,P_0] = 0, \quad [K_i,P_j] = i\delta_{ij} P_0, \quad [K_i,P_0] = iP_i \quad (77)$$

and defining the generalized Inönü–Wigner contraction[17,18]

$$\mathbf{K}' = c^{-1}\mathbf{K}, \quad P'_0 = cP_0 \quad (78)$$

we get the Galileo algebra in the limit $c \to \infty$. Explicitly

$$[J_i,J_j] = i\epsilon_{ijn} J_n, \quad [J_i,K'_j] = i\epsilon_{ijn} K'_k, \quad [K'_i,K'_j] = 0,$$
$$[J_i,P_j] = i\epsilon_{ijn} P_n, \quad [J_i,P'_0] = K'_i,P_j = 0, \quad [K'_i,P'_0] = iP_i, \quad (79)$$

which are the commutation rules for the Galileo group.

### B. Positive mass and arbitrary spin representations of the Galileo group

Nonrelativistic particles with positive mass and arbitrary spin are described by unitary irreducible projective representations of the Galileo group labeled by the values of the mass $m$, spin $\mathbf{S}' = \mathbf{J} - m^{-1}\mathbf{K}' \times \mathbf{P}$, and internal energy $u = \mathbf{P}'_0 - \mathbf{P}^2/2m$. They act in the space of square-integrable functions $\phi'_n(\mathbf{p})$ as follows[19]:

$$[u'(b,\mathbf{a},\mathbf{v},\alpha)\phi'](\mathbf{p})$$
$$= e^{ib(u + \mathbf{P}^2/2m)} e^{-i\mathbf{p}a} D^{(j)}[\boldsymbol{\alpha}] \, \phi'(R^{-1}(\mathbf{p} - m\mathbf{v})), \quad (80)$$

where $D^{(j)}(R)$ is the $(2j+1)$-dimensional representation of SU(2).

Introducing the configuration space functions $\phi'_n(\mathbf{x},t)$ by the transformation

$$\phi'(\mathbf{x},t) = (2\pi)^{-3/2} \int d^3p \, e^{-i(u + \mathbf{P}^2/2m)t} e^{i\mathbf{p}\mathbf{x}} \phi_n(\mathbf{p}), \quad (81)$$

we get the associated representation

$$[u'(b,\mathbf{a},\mathbf{v},\boldsymbol{\alpha})\phi'](\mathbf{x},t)$$
$$= e^{-i(mv^2/2)(t-b)} e^{im\mathbf{v}(\mathbf{x}-\mathbf{a})} D^{(j)}[\boldsymbol{\alpha}]\phi'$$
$$\times (R^{-1}(\mathbf{x} - \mathbf{a} - \mathbf{v}(t-b)), t-b). \quad (82)$$

The phase factor appearing in (85) corresponds to the nontrivial projective character of this representation. On the other hand, it is necessary to keep the Galileo invariance of the Schrödinger equation. Functions $\phi_n(\mathbf{p})$ and $\phi_n(\mathbf{x},t)$ not only transform in a covariant way but at the same time they represent probability amplitudes. This fact simplifies the interpretation of those representations. We only need to consider one realization of each representation [that given by (80) or (82)] unlike the Poincaré case where we have to choose between a covariant realization or a Wigner realization on amplitudes. This was shown in Secs. 2 and 3.

Antiparticles have room in the Galileo group being described by representations with mass $m$, spin $j$, and internal energy $u$, which can be deduced from (81), (82) by complex conjugation [keeping in mind that $D^{(j)*}(R) = CD^{(j)}(R)C^{-1}$]. Denoting by $\phi'^{(+)}(\phi'^{(-)})$ the wavefunctions associated with particles (antiparticles) we have the Schrödinger equation

$$i\partial_t \phi'^{(\pm)}(\mathbf{x},t) = \pm \, [\mathbf{P}^2/2m + u] \; \phi'^{(\pm)}(\mathbf{x},t). \quad (83)$$

### C. Contraction of representations

Let us now settle a bridge between the relativistic and Galilean descriptions of finite mass and arbitrary spin particles. Our first step will be to define the contraction for representations:[16]

Given a group $G$ and its contracted group $G'$, and given a representation of $G, D(g)$, (acting on $\mathscr{H}$) and a representation of $G', D'(g')$, (acting on $\mathscr{H}'$), we say $D'(g')$ is a contraction of $D(g)$ if there are a family of representations $D^\epsilon(g)$ equivalents to $D(g)$ (acting on the spaces $\mathscr{H}^\epsilon$) and a family $U_\epsilon$ of continuous linear mappings from $\mathscr{H}^\epsilon$ on $\mathscr{H}'$, such that

(i) $D^1(g) = D(g)$

(ii) $\lim_{\epsilon \to 0} U_\epsilon D^\epsilon [f_\epsilon^{-1}(g')] \, U_\epsilon^{-1} = D'(g')$. (84)

It is worthwhile to point out that $\widetilde{D}_\epsilon(g) = U_\epsilon D^\epsilon \times [f_\epsilon^{-1}(g')] \, U_\epsilon^{-1}$ is a realization of $D(g)$ on the space $\mathscr{H}'$, such that $\widetilde{D}_\epsilon(g) \to D'(g')$ as $\epsilon \to 0$. Thus we go from $D$ to $D'$ in two steps. First we find a representation $\widetilde{D}_\epsilon$ (unitary equivalent to $D$) acting on the same space as $D'$ (this is attained by the use of FW transformation as we shall see below). Second we go to the limit $\epsilon \to 0$ arriving to $D'$.

Let us start with half-integer $j$ representations in momentum space. We begin with the FW transformed representation (45). Taking into account (25), (31), and (32) we have

$$\mathscr{D}^{(j)}(\Lambda) = \frac{1}{2}\begin{pmatrix} D^{(j)}(\Lambda)+D^{(j)\dagger}(\Lambda^{-1}) & D^{(j)}(\Lambda)-D^{(j)\dagger}(\Lambda^{-1}) \\ -D^{(j)}(\Lambda)+D^{(j)\dagger}(\Lambda^{-1}) & D^{(j)}(\Lambda)+D^{(j)\dagger}(\Lambda^{-1}) \end{pmatrix}, \tag{85}$$

but $R_w = B_p^{-1}\Lambda B_{\Lambda^{-1}p}$ is a rotation, therefore $D^{(j)\dagger}(R_w^{-1}) = D^{(j)}(R_w)$, and

$$\mathscr{D}^{(j)}(R_w) = \begin{pmatrix} D^{(j)}(R_w) & 0 \\ 0 & D^{(j)}(R_w) \end{pmatrix}. \tag{86}$$

Thence (45) is a direct sum of two irreducible representations acting on $\phi^{(+)}(\mathbf{p})$ and $\phi^{(-)}(\mathbf{p})$, they can be given separately using the projectors $(1 \pm \beta)/2$. Choosing, for instance, that acting on $\phi^{(+)}$

$$[U(a,\Lambda)\phi^{(+)}](\mathbf{p}) = e^{iap}\left(\frac{\omega(\Lambda^{-1}p)}{\omega(p)}\right)^{1/2} D^{(j)}(R_w)\,\phi^{(+)}(\Lambda^{-1}\mathbf{p}). \tag{87}$$

Both (87) and (90) act on the same Hilbert space of $(2j+1)$ component square-integrable functions. We now choose a family of trivial projective representations equivalent to (87).

$$U_c(a,\Lambda) = e^{-imca_0}\,U(a,\Lambda). \tag{88}$$

The generator of time translations is in this family the kinetic energy. Moreover, going to the limit $c\rightarrow\infty$ as in (84), we get (80)

$$\lim_{c\rightarrow\infty} [U_c(f_c^{-1}(b,\mathbf{a},\mathbf{v},\boldsymbol{\alpha}))\,\phi^{(+)}](p) = \lim_{c\rightarrow\infty} e^{i(p^0-mc)cb}\,e^{-i\mathbf{a}\mathbf{p}}\left(\frac{\omega(\Lambda^{-1}p)}{\omega(p)}\right)^{1/2} \times D^{(j)}(B_p^{-1}\Lambda B_{\Lambda^{-1}p})\,\phi^{(+)}(\Lambda^{-1}p), \tag{89}$$

where according to (76)

$$\Lambda = f_c^{-1}(\mathbf{v},\boldsymbol{\alpha}) = e^{i\mathbf{v}\mathbf{k}/c}\,R(\boldsymbol{\alpha}). \tag{90}$$

Thus we have

$$(\Lambda^{-1}p)^0 = (1+v^2/2c^2)p^0 - \mathbf{v}\mathbf{p}/c + O(v^3/c^3),$$
$$(\Lambda^{-1}p) = R^{-1}(\mathbf{p}-m\mathbf{v}) + O(v/c), \tag{91}$$

and

$$\lim_{c\rightarrow\infty} D^{(j)}(B_p^{-1}\Lambda B_{\Lambda^{-1}p}) = \lim_{c\rightarrow\infty} D^{(j)}(B_p^{-1})\,D^{(j)}(\Lambda)\,D^{(j)}(B_{\Lambda^{-1}p}) = D^{(j)}(R), \tag{92}$$

where we used (91) and (10). Lastly

$$\lim_{c\rightarrow\infty}\{U_c[f_c^{-1}(b,\mathbf{a},\mathbf{v},\boldsymbol{\alpha})]\,\phi^{(+)}\}(p) = e^{i(\mathbf{p}^2/2m)b}\,e^{-i\mathbf{a}\mathbf{p}}\,D^{(j)}[\boldsymbol{\alpha}]\,\phi^{(+)}(R^{-1}(\mathbf{p}-m\mathbf{v})) \tag{93}$$

which is (80) with $U=0$.

The factor $e^{-imca_0}$ plays two roles in the limit $c\rightarrow\infty$. (i) To assure the convergence of this limit. (ii) To lead up to nontrivial projective representation of the Galileo group for which, instead of (79), we have

$$[K_i', P_j] = i\delta_{ij}\,m \tag{94}$$

as requested for quantum-mechanical representations.

If we choose, instead of (87), the representation acting on $\phi^{(-)}$, we must introduce a factor $e^{+imca_0}$ to get a finite limit. Then

$$\lim_{c\rightarrow\infty}[U_c[f_c^{-1}(b,a,\mathbf{v},\boldsymbol{\alpha})]\,\phi^{(-)}](\mathbf{p}) = e^{-i(\mathbf{p}^2/2m)b}\,e^{i\mathbf{a}\mathbf{p}}\,D^{(j)}[\boldsymbol{\alpha}]\,\phi^{(-)}(R^{-1}(\mathbf{p}-m\mathbf{v})) \tag{95}$$

as corresponds to antiparticles.

To complete our discussion, let us now go to configuration space. After (43) and (45) the realization of Poincaré group in terms of configuration space amplitudes is given by

$$[U(a,\Lambda)\,\phi^{(\pm)}](x) = \int d^3x'\,K^{(\pm)}(x,x';a,\Lambda)\,\phi^{(\pm)}(x'),\quad x_0' = x_0, \tag{96}$$

with

$$K_{mn}^{(\pm)}(x,x';a,\Lambda) = (2\pi)^{-3}\int d^3p\,e^{\mp ipx}\,e^{\pm i(\Lambda^{-1}p)x'} \times\left(\frac{\omega(\Lambda^{-1}p)}{\omega(p)}\right)^{1/2} D_{mn}^{(j)}(B_p^{-1}\Lambda B_{\Lambda^{-1}p}). \tag{97}$$

We now pass to the equivalent projective representation

$$U_c^{(\pm)}(a,\Lambda) = e^{\mp imca_0}U(a,\Lambda), \tag{98}$$

and introducing the transformation

$$V_c^{(\pm)} = \exp i(c\sqrt{\mathbf{p}^2+m^2c^2} - \mathbf{p}^2/2m)t,$$
$$\mathbf{p} = -i\nabla. \tag{99}$$

we get

$$\lim_{c\rightarrow\infty}[V_c^{(\pm)}U_c^{(\pm)}[f_c^{-1}(b,\mathbf{a},\mathbf{v},\boldsymbol{\alpha})]\,V_c^{(\pm)-1}\,\phi'^{(\pm)}(\mathbf{x},t)] = e^{\mp i\frac{mv^2}{2}(t-b)}\,e^{\pm im\mathbf{v}(\mathbf{x}-\mathbf{a})}\,D^{(j)}[\boldsymbol{\alpha}]\phi'^{(\pm)} \times(R^{-1}[\mathbf{x}-\mathbf{a}-\mathbf{v}(t-b)],t-b), \tag{100}$$

which are the Galilean representations associated to (83) with $u=0$. The role of (99) is clear: it maps the relativistics amplitude (43) on the nonrelativistic one with $u=0$.

The local exponent which appears in (100) is the only surviving remembrance of the nonlocal nature of the transformation law (96). This factor is needed for the invariance of the Schrödinger equation as we pointed out in the paragraph following equation (82).

For representations with integer $j$ we contract along the lines stated above, but taking into account that the representation defined on functions (62) trivially reduces to two representations with $(2j+1)$ dimensions when we use the projectors

$$P^{\pm} = \frac{1+\beta}{2} \otimes \frac{1 \pm \eta}{2}.$$

These representations act like those associated with fermions, (87) being the corresponding transformation law.

## 5. CONCLUSIONS

Let us briefly summarize our main results. The first achievement of this paper has been to settle the relationship between FW and Lorentz transformations for arbitrary spin. Due to the fact that the original FW tranformation was unitary and non-Hermitian, it could not be directly associated to a Lorentz boost, even if this was the intuitive physical picture for this transformation. In this work a definite mathematical meaning was given to it. We first introduce two different realizations of Poincaré group: the covariant realization—widely used due to its simple transformation properties—and a special case of Wigner realization, whose quantum-mechanical interpretation will be relevent for further applications. Then we define FW transformation as that mapping the first realization into the second . With this definition FW transformation in momentum space turns out to be a Lorentz boost—that carrying the particle to rest—times the trivial factor $(m/p_0)^{1/2}$, while this simple decomposition is no longer possible in configuration space. As a byproduct of this formalism an essential difference does emerge between bosons and fermions due to the different intrinsic parity for the corresponding particle–antiparticle system. As a consequence of this fact FW transformation should be considered separately for fermions and bosons.

The second part of the paper is devoted to the nonrelativistic limit. Using a contraction formalism previously developed by the authors,[18] we can establish a smooth connection between relativistic representations and quantum-mechanical Galilean representations. The role of FW transformations has been proved to be capital as a first step in this limiting procedure. Unlike other approaches,[21] our contraction procedure directly leads to the $(2j+1)$ physical representations of the Galileo group, free of redundant components. The heavy mathematical apparatus we used is a price to be paid for this remarkable result.

In this work we gave up the case of interacting particles. As it is well known, Foldy and Wouthuysen introduced[1] a sequence of unitary transformations which depend on the electromagnetic field in order to get the nonrelativistic Hamiltonian. However, as has been pointed out by Goldman,[22] this procedure runs into difficulties because the time dependence of the transformation shifts the expectation values of the Hamiltonian. To find a correct answer to the interaction problem several questions must be studied in detail: the compatibility of gauge invariance with Lorentz covariance of wavefunctions, and secondly the kind of interaction—minimal coupling or else—compatible with the implemented gauge. A detailed analysis of this situation deserves further attention and it will be the task in a forthcoming paper.[23]

Using procedures similar to those employed throughout this paper, we hope to disentangle the groupal content of other transformations, relevant to different physical situations, such a Melosh, Gomberoff *et al*,[24] etc. This task seems simpler in the first quantization language we used here. We intend to use the present formalism as the basis of a correspondence principle to be applied to such situations. The possibility of getting a precise prescription for writing the representations of the various contracted groups associated with physical situations is the main virtue of our method. In particular we are interested in high energy limits and their related covariance group, a first step in this direction has been accomplished in a previous work.[25]


[1] L.L. Foldy and S.A. Wouthuysen, Phys. Rev. **78**, 29 (1950).

[2] In a space whose metric is not Euclidean, $\eta \neq I$, an isometric transformation does not satisfy $U^\dagger = U^{-1}$ but $U^\dagger \eta U = \eta$.

[3] G.A. Ramanujan, Nuovo Cimento A **20**, 27 (1974).

[4] A.R. Tekumalla and T.S. Santhanaman, Nuovo Cimento Lett. **10**, 737 (1974).

[5] J. Jayaraman, J. Phys. A **8**, L1 (1975).

[6] (a) D.L. Weaver, C.L. Hammer, and R.H. Good, Phys. Rev. **135**, B241 (1964); (b) P.M. Mathews, Phys. Rev. **143**, 985 (1966).

[7] (a) T.J. Nelson and R.H. Good, Jr., Rev. Mod. Phys. **40**, 508 (1968); (b) C.L. Hammer, S.C. McDonald, and D.L. Pursey, Phys. Rev. **171**, 1349 (1968).

[8] R.A. Krajcik and M.M. Nieto, Phys. Rev. D **13**, 2245 (1976); **13**, 2250, (1976); **15**, 416 (1977); **15**, 426 (1977).

[9] E.P. Wigner, Ann. Math. **40**, 149 (1939).

[10] G.W. Mackey, *Induced Representation of Groups and Quantum Mechanics* (Benjamin, New York, 1968).

[11] U. H. Niederer and L.O'Raifeartaigh, Fortsch. Phys. **22**, 111 (1974); **22**, 131 (1974).

[12] T. Newton and E. P. Wigner, Rev. Mod. Phys. **21**, 400 (1949); A.S. Wightman, **34**, 845 (1962).

[13] H.Joos, Fortsch. Phys. **10**, 65 (1962). S. Weinberg, Phys. Rev. **133**, B1318 (1964).

[14] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. (USA) **39**, 510 (1953).

[15] E. Saletan, J. Math. Phys. **2**, 1 (1961).

[16] J. Mickelsson and J. Niederle, Commun. Math. Phys. **27**, 167 (1972).

[17] M. D. Doebner and O. Melsheimer, Nuovo Cimento A **49**, 306 (1967).

[18] J. León, M. Quirós, and J. Ramirez Mittelbrunn, "Probabilistic Interpretation, Group Contraction and the Galilei Limit of Relativistic Quantum Mechanics," Nuovo Cimento B **46**, 109 (1978).

[19] V. Bargamann, Ann. Math. (N.Y.) **59**, 1 (1954); E. Inönü and E.P. Wigner, Nuovo Cimento **9**, 705 (1952).

[20] L.L. Foldy, Phys. Rev. **102**, 568 (1956).

[21] L.P.S. Singh and C.R. Hagen, Phys. Rev. D **9**, 898 (1974); **9**, 910 (1974).

[22] T. Goldman, Phys. Rev. D **15**, 1063 (1977).

[23] J. León, M. Quirós, and J. Ramiŕez Mittelbrunn, "Gauge invariance, Lorentz covariance and the electromagnetic properties of elementary systems," I.E.M. Preprint (1978).

[24] H.J. Melosh, Ph.D. thesis, California Institute of Technology (1973); H.J. Melosh, Phys. Rev. D **9**, 1095 (1974); J.S. Bell, Acta Phys. Austriaca, Suppl. **13**, 395 (1974); W.F. Palmer and V. Rabl, Phys. Rev. D **10**, 2554 (1974); N.H. Fuchs, **11**, 1569 (1975); V. Aldaya and J. De Azcárraga, **14**, 1049 (1976).

[25] J. León, M. Quirós, and J. Ramírez Mittelbrunn, Phys. Lett. B **68**, 247 (1977); "Proceedings of Vth International Winter Meeting on Particle Physics" Candanchú (1977); Nuovo Cimento A **41**, 141 (1977).

# Internal Galilei group for a two-particle system[a)]

M. Daumens[b)] and M. Perroud[c)]

*Centre de Recherches Mathématiques, Université de Montréal, Montreal, Canada*
(Received 9 October 1978)

Two variable expansions of Galileian scattering amplitudes are proposed for a two-particle system with arbitrary spins. The barycentric decomposition of such a system permits, on the one hand, to realize the associated Hilbert space in the form $H = L^2(\mathbf{R}^3) \otimes H_{\rm INT}$, where $H_{\rm INT}$ is the Hilbert space of the states of the system expressed in their CM frame, and, on the other hand, to introduce what we call the internal Galilei group $G(3)_{\rm INT}$ as the maximal subgroup in $G(3)\times G(3)$ which respects the barycentric decomposition. For this group we have a natural unitary representation acting on $H_{\rm INT}$ induced by the PIUR of $G(3)\times G(3)$ acting on H. The reduction of this representation with respect to the chain of subgroups $G(3)_{\rm INT} \supset E(3)_{\rm INT} \supset {\rm SU}(2) \supset {\rm SU}(1)$ provides the usual one variable expansions while the reduction with respect to the chain $G(3)_{\rm INT} \supset G(3)_{0\ \rm INT} \supset {\rm SU}(2) \supset {\rm SU}(1)$ provides an example of a two variable expansion.

## INTRODUCTION

Several papers and a review article[1] were devoted to two-variable nonrelativistic expansions of the scattering amplitudes. Let us just recall that the basic idea consists of expressing the amplitudes as series and/or integrals in such a way that the kinematical variables, namely the energy and the scattering angle, only appear as arguments of known functions. The expansion in one variable (the scattering angle) is very familiar: it leads to a series of Legendre polynomials, the coefficients of which, $a_l(E)$, still depend on the energy. Provided that certain mathematical conditions are satisfied for the functions $a_l$, it is, of course, always possible to expand them into a series or an integral, for example in terms of Laguerre polynomials or of Bessel functions, in order to obtain the desired two-variable expansion. However, this is not the most fruitful way of proceeding because there is no physical motivation in the choice of such an expansion. The authors cited in Ref. 1 have introduced physical ingredients to justify several choices of expansions by means of group theoretical considerations, but they have restricted their study to the case of particles without spin. It is the aim of this paper to generalize their results to the general case of particles with arbitrary spins.

The motivation for such a study was largely explained in Ref. 1. The relativistic treatment is certainly the ultimate goal, and we intend to pursue it in a following article. The big advantage of the nonrelativistic scattering theory rests in the extreme simplicity of the many-body Galileian kinematics compared to the Einsteinian case. In particular, we have in mind the barycentric decompositions.

When we try to generalize the results of Ref. 1 to particles with spins, we are immediately faced with the problem of physically interpretating various spinlike indexes. To overcome this difficulty, it is necessary to have a deeper understanding of the group considerations which were introduced. In order to do that, we must adopt another point of view.

A scattering amplitude is more or less nothing else than the kernel of the transition operator $T$. For an elastic scattering process, this operator is defined on the Hilbert $\mathscr{H}$ space of two noninteracting particles. The kinematical assumptions are introduced by means of a projective unitary representation of the Galilei group G(3). This representation cannot be chosen arbitrarily: the two particles being noninteracting, it is obtained by a restriction to the diagonal subgroup of $G(3)\times G(3)$ of the tensor product of the irreducible representations which define each particle. [Clearly we have a unitary action of $G(3)\times G(3)$ on $\mathscr{H}$.] It is well known that this representation is reducible into a direct integral of irreducible ones essentially labelled by the internal energy and internal angular momentum. Corresponding to this decomposition, we have a decomposition of the Hilbert space and of the transition operator $T$ which is assumed to be an intertwining operator for the representations. The standard one-variable expansions of its kernel follow directly from this decomposition. If we want to introduce group theoretical considerations in order to motivate some choices of two-variable expansions, it is certainly not the kinematical group which will be useful since the internal energy is an invariant. In $G(3)\times G(3)$, however, there exist many subgroups, and in particular the one we call the internal Galilei group is quite relevant in such an approach. It is a subgroup closely related to the barycentric decomposition of a two-particle system. We introduce it, and mention some of its properties in Sec. 1; for reasons of clarity we have adopted the context of classical mechanics.

In Sec. 2 we make some minor additions to the standard results on projective unitary irreducible representations of the Galilei group. In particular we consider two realizations of the Hilbert space of these representations associated with

a)Work supported partly by the Echanges France-Québec.
b)Permanent address: Laboratoire de Physique Théorique, Université de Bordeaux 1, chemin du Solarium 33170 Gradignan, France.
c)Permanent address: Départment de Mathématiques, Ecole Polytechnique de Montréal.

  

two chains of subgroups, their relativistic analog being extensively used.

In Sec. 3 we deal with the reduction of the tensor product of two PIUR's of G(3) with respect to both the kinematical Galilei group and the internal Galilei group. We obtain several realizations of the Hilbert space of the representation, and we construct the unitary operators which connect them.

Finally in Sec. 4 we use the results of the previous section to get the two variable expansions of the scattering amplitudes, both in the $(l-s)$ coupling scheme and in the helicity one.

## 1. GALILEIAN GROUP ACTION ON CLASSICAL TWO-PARTICLE SYSTEMS

A very extensive treatment of the Galilei group G(3) can be found in a review articel by Lévy-Leblond[2] Let us just recall some fundamental facts in order to establish notation.

With respect to a standard parametrization, the group law of the extended Galilei group

$$\mathscr{G}(3) = \mathbb{R} \,\square\, G(3) \qquad [G(3) \simeq \mathbb{R}^7 \,\square\, SU(2)]$$

is given by

$$(\xi,\tau,\mathbf{a},\mathbf{v},A)(\xi',\tau',\mathbf{a}',\mathbf{v}',A') = (\xi+\xi'+\tfrac{1}{2}\tau'|\mathbf{v}|^2+\mathbf{v}\cdot R_A\mathbf{a}', \tau+\tau', \mathbf{a}+\tau'\mathbf{v}+R_A\mathbf{a}', \mathbf{v}+R_A\mathbf{v}', AA'), \tag{1.1}$$

where $R$: SU(2) $\rightarrow$ SO(3) denotes the usual Hamilton homomorphism.

Corresponding to this parametrization, a basis for the Lie algebra $\mathfrak{g}(3)$ is denoted by

$$\{\mathscr{M},\mathscr{H},\mathscr{P}_i,\mathscr{K}_i,\mathscr{J}_i\}, \quad i=1,2,3. \tag{1.2}$$

The nonzero commutation relations are

$$\begin{aligned} &[\mathscr{J}_i,\mathscr{J}_j] = \epsilon_{ijk}\mathscr{J}_k, \quad [\mathscr{J}_i,\mathscr{P}_j] = \epsilon_{ijk}\mathscr{P}_k, \\ &[\mathscr{J}_i,\mathscr{K}_j] = \epsilon_{ijk}\mathscr{K}_k, \\ &[\mathscr{K}_i,\mathscr{P}_j] = \delta_{ij}\mathscr{M}, \quad [\mathscr{K}_i,\mathscr{H}] = \mathscr{P}_i, \end{aligned} \tag{1.3}$$

and the invariants in the Lie algebra are generated by

$$\mathscr{M}, \quad \mathscr{U} = \mathscr{H} - \tfrac{1}{2}\mathscr{M}^{-1}|\mathscr{P}|^2, \quad |\mathscr{S}| = |\mathscr{J} - \mathscr{M}^{-1}\mathscr{K}\times\mathscr{P}|. \tag{1.4}$$

We shall have to consider the following two subgroups of $\mathscr{G}(3)$, each isomorphic to $R^3 \,\square\,$ SU(2).

The homogeneous Galilei subgroup

$$G(3)_0 = \{(0,0,\mathbf{0},\mathbf{v},A)\}$$

with invariants generated by

$$|\mathscr{K}| \quad \text{and} \quad \mathscr{K}\cdot\mathscr{J}. \tag{1.5}$$

The Euclidean subgroup

$$E(3) = \{(0,0,\mathbf{a},\mathbf{0},A)\}$$

with invariants generated by

$$|\mathscr{P}| \quad \text{and} \quad \mathscr{P}\cdot\mathscr{J}. \tag{1.6}$$

It is not our purpose to treat scattering in the context of classical mechanics: However, the reasons for the introduction of the various subgroups of $\mathscr{G}(3)\times\mathscr{G}(3)$ that we consider appear more directly in the classical framework than in its quantal counterpart.

The fundamental object of the mechanics of a Galileian classical system is a symplectic manifold on which the Galilei group acts via symplectomorphisms. Such a system is called elementary if the action of the group is transitive. According to an analysis of Souriau, who the first developed these ideas,[3] the symplectic manifolds associated to the elementary systems are all symplectomorphic to the orbits of the co-adjoint representation of the extended Galilei group. The orbit analysis furnishes a classification of all elementary Galileian classical systems. It turns out that this classification is very close to the classification of the elementary Galileian quantum systems furnished by the projective irreducible representations of G(3).

Denoting by $\{\mathscr{M}^*,\mathscr{H}^*,\mathscr{P}_i^*,\mathscr{K}_i^*,\mathscr{J}_i^*\}$, the dual basis of (1.2) and by

$$l^* = -m\mathscr{M}^* - E\mathscr{H}^* + \mathbf{p}\cdot\mathscr{P}^* - \boldsymbol{\xi}\cdot\mathscr{K}^* + \mathbf{j}\cdot\mathscr{J}^*,$$

an arbitrary element of the dual space $\mathfrak{g}(3)^*$, it is easy to verify that the co-adjoint action $\Delta(g) = \mathrm{Ad}(g^{-1})^*$ of $\mathscr{G}(3)$ is given by

$$\Delta(\xi,\tau,\mathbf{a},\mathbf{v},A): \begin{cases} m \rightarrow m, \\ E \rightarrow E + \tfrac{1}{2}m|\mathbf{v}|^2 + R_A\mathbf{p}\cdot\mathbf{v}, \\ \mathbf{p} \rightarrow R_A\mathbf{p} + m\mathbf{v}, \\ \boldsymbol{\xi} \rightarrow R_A\boldsymbol{\xi} - \tau(R_A\mathbf{p}+m\mathbf{v}) + m\mathbf{a}, \\ \mathbf{j} \rightarrow R_A\mathbf{j} + m\mathbf{a}\times\mathbf{v} + \mathbf{a}\times R_A\mathbf{p} + R_A\boldsymbol{\xi}\times\mathbf{v}. \end{cases} \tag{1.7}$$

Note the trivial action of the one-dimensional extension group. Given an element $l^*$ *with* $m\neq 0$, one sees that the invariant functions defined on its orbit are

$$m, \quad U = E - \frac{1}{2m}|\mathbf{p}|^2, \quad s = |\mathbf{s}| = \left|\mathbf{j} - \frac{1}{m}\boldsymbol{\xi}\times\mathbf{p}\right|. \tag{1.8}$$

Performing the change of variables

$$(m,E,\mathbf{p},\boldsymbol{\xi},\mathbf{j}) \rightarrow (m,U,\mathbf{p},\mathbf{q},\mathbf{s}), \quad \mathbf{q} = \frac{1}{m}\boldsymbol{\xi},$$

we get

$$\Delta(\xi,\tau,\mathbf{a},\mathbf{v},A): \begin{cases} m \rightarrow m, \\ U \rightarrow U, \\ \mathbf{p} \rightarrow R_A\mathbf{p} + m\mathbf{v}, \\ \mathbf{q} \rightarrow R_A\mathbf{q} - (\tau/m)(R_A\mathbf{p}+m\mathbf{v}) + \mathbf{a}, \\ \mathbf{s} \rightarrow R_A\mathbf{s}. \end{cases} \tag{1.9}$$

Consequently, when $m\neq 0$ and $s\neq 0$, which will be tacitly supposed from now on, the orbits $\mathscr{O}(m,s,U)$ are of dimension 8, diffeomorphic to $\mathbb{R}^3\times\mathbb{R}^3\times S^2$, and completely characterized by the three real numbers $m$, $s$ $(s>0)$, and $U$. The mechanical interpretation of these numbers are the mass (when

$m > 0$), the spin, and the internal energy of the elementary system (a particle). Taking into account the transformation formulas (1.9), it seems obvious that **p** and **s** have to be interpreted as the linear momentum and the intrinsic angular momentum, the interpretation of **q**, however, is not so obvious. It must be emphasized that the orbit $\mathscr{O}(m,s,U)$ is not the usual phase space $\mathscr{X}$ of the particle since no natural action of the Galilei group can be defined on such a space. On the extended phase space $\mathscr{X}\times\mathbb{R}$, a natural action is defined by

$$D(\tau,\mathbf{a},\mathbf{v},A): \begin{cases} \mathbf{p}\mapsto R_A\mathbf{p}+m\mathbf{v}, \\ \mathbf{x}\mapsto R_A\mathbf{x}+t\mathbf{v}+\mathbf{a}, \\ \mathbf{s}\mapsto R_A\mathbf{s}, \\ t\mapsto t+\tau, \end{cases} \tag{1.10}$$

where **x** is the position and $t$ the time.

The connection between $\mathscr{O}(m,s,U)$ and $\mathscr{X}\times\mathbb{R}$ requires a dynamical postulate. The extended phase space and the Galileian action (1.10) do not contain any dynamics as opposed to $\mathscr{O}(m,s,U)$ which is a model for the manifold of all motions of a free particle. More precisely, the dynamical group for a free particle

$$U_\lambda: (\mathbf{p},\mathbf{x},\mathbf{s},t)\mapsto(\mathbf{p},\mathbf{x}+(\lambda/m)\mathbf{p},\mathbf{s},t+\lambda), \quad \lambda\in\mathbb{R}, \tag{1.11}$$

foliates the extended phase space into orbits (motions) of dimension one. Since the phase space "at $t=0$," i.e., the section $\Sigma=\{(\mathbf{p},\mathbf{x},s,\ t=0)\}\subset\mathscr{X}\times\mathbb{R}$ intersects each motion exactly once, this section provides a realization of the motion manifold $\mathscr{X}\times\mathbb{R}/U$ and can obviously be identified with the orbit $\mathscr{O}(m,s,U)$. Furthermore, it can be seen that an action of G(3) can be induced on $\mathscr{X}\times\mathbb{R}/U$; in the special realization we consider, it is simply given by

$$\tilde{D}(\tau,\mathbf{a},\mathbf{v},A)=U_{-\tau}\circ D(\tau,\mathbf{a},\mathbf{v},A):\Sigma\to\Sigma,$$

and it is easily verified that it is identical with (1.9).

Each point $(m,U,\mathbf{p},\mathbf{q},\mathbf{s})$ in $\mathscr{O}(m,s,U)$ can therefore be interpreted either as a motion (not a state) of a free particle or as an initial state at $t=0$; the variable **q** represents the *position of the particle at the time* $t=0$. In the passive interpretation of a Galilei transformation, formulas (1.9) express the connection between the characterizations of free motion given by two Galileian observers.

Since our concern in this article is neither classical mechanics nor geometric quantization, we do not introduce any symplectic structure; only group actions are relevant for the following.

The symplectic manifold $\mathscr{M}$ associated with a nonelementary system, for example, a system of several particles not necessarily without interactions, can always be decomposed into orbits of the Galilei group

$$M\sim\bigcup_{\{S,U,\xi\}}\tilde{\mathscr{O}}_\xi(M,S,U), \tag{1.12}$$

where $M$ is the total mass of the system and $\xi$ is a label which removes possible degeneracies. These orbits $\tilde{\mathscr{O}}_\xi(M,S,U)$, generically of dimension ten, should not be confused with the orbits $\mathscr{O}(m,s,U)$ of the co-adjoint representation of $\mathscr{G}(3)$. In general, however, there exists a covariant mapping $\mu:\mathscr{M}\to\mathfrak{g}(3)^*$ which gives a meaning to the labels $M,S,U$.

Moreover, a fundamental feature in Galileian mechanics is the existence of a *barycentric decomposition*

$$\mathscr{M}\sim\mathscr{M}_{\rm CM}\times\mathscr{M}_{\rm INT}, \tag{1.13}$$

where $\mathscr{M}_{\rm CM}\sim\mathbb{R}^3\times\mathbb{R}^3$ is the set of pairs $(\mathbf{P},\mathbf{Q})$ on which the Galilei group acts as follows:

$$(\mathbf{P},\mathbf{Q})\mapsto(R_A\mathbf{P}+M\mathbf{v},R_A\mathbf{Q}-(\tau/M)(R\,\mathbf{P}+M\mathbf{v})+\mathbf{a}). \tag{1.14}$$

We recognize the transformation laws for the total linear momentum and the position of the barycenter at the time $t=0$. Owing to this decomposition (1.13), the decomposition (1.12) can be reduced to the corresponding decomposition of the "internal manifold" $\mathscr{M}_{\rm INT}$.

Now let us restrict ourselves to the case of a system of two noninteracting particles. The motion space is here simply the Cartesian product

$$\mathscr{O}(m_1,S_1,u_1)\times\mathscr{O}(m_2,s_2,u_2)$$

on which the group $\mathscr{G}(3)\times\mathscr{G}(3)$ acts in an obvious way. We introduce the barycentric decomposition by means of the bijective mapping $\Phi$: $\mathscr{O}(m_1,s_1,u_1)\times\mathscr{O}(m_2,s_2,u_2)\to\mathscr{M}_{\rm CM}\times\mathscr{M}_{\rm INT}$ defined by

$$\Phi:\langle(\mathbf{p}_1,\mathbf{q}_1,\mathbf{s}_1),(\mathbf{p}_2,\mathbf{q}_2,\mathbf{s}_2)\rangle\to\langle(\mathbf{P},\mathbf{Q}),(\mathbf{p},\mathbf{q},\mathbf{s}_1,\mathbf{s}_2)\rangle,$$

where

$$\begin{cases} \mathbf{P}=\mathbf{p}_1+\mathbf{p}_2 \\ \mathbf{Q}=(1/M)(m_1\mathbf{q}_1+m_2\mathbf{q}_2) \\ \mathbf{p}=(1/M)(m_2\mathbf{p}_1-m_1\mathbf{p}_2) \\ \mathbf{q}=(\mathbf{q}_1-\mathbf{q}_2) \end{cases} \qquad \begin{cases} \mathbf{p}_1=(m_1/M)\mathbf{P}+\mathbf{p} \\ \mathbf{p}_2=(m_2/M)\mathbf{P}-\mathbf{p} \\ \mathbf{q}_1=\mathbf{Q}+(m_2/M)\mathbf{q} \\ \mathbf{q}_2=\mathbf{Q}-(m_1/M)\mathbf{q} \end{cases} \tag{1.15}$$

and $M=m_1+m_2$ denotes the total mass of the system. This mapping is based on the dual mapping of the diagonal injection $l\to(l,l)$ of $\mathfrak{g}(3)$ into $\mathfrak{g}(3)\times\mathfrak{g}(3)$. Here $\mathscr{M}_{\rm INT}\sim\mathbb{R}^3\times\mathbb{R}^3\times S^2\times S^2$ and **p** and **q** are respectively interpreted as the relative linear momentum and the relative position at the time $t=0$.

Since $\Phi$ is a bijective mapping, it induces an action of the group $\mathscr{G}(3)\times\mathscr{G}(3)$ on $\mathscr{M}_{\rm CM}\times\mathscr{M}_{\rm INT}$. Taking into consideration formulas (1.9) and (1.15), it is easily verified that this action does not preserve the Cartesian decomposition (1.13). Some subgroups, however, do, and it can be checked that the maximal subgroup of $\mathscr{G}(3)\times\mathscr{G}(3)$ which preserves this decomposition is

$$\mathscr{B}=\{(\xi_1,\tau,\mathbf{a}_1,\mathbf{v}_1,A),(\xi_2,\tau,\mathbf{a}_2,\mathbf{v}_2,A)\}\subset\mathscr{G}(3)\times\mathscr{G}(3). \tag{1.16}$$

Indeed, when setting

$$\begin{aligned} \mathbf{a}_1&=\mathbf{a}+(m_2/M)\mathbf{b}, \quad \mathbf{v}_1=\mathbf{v}+(m_2/M)\mathbf{u}, \\ \mathbf{a}_2&=\mathbf{a}-(m_1/M)\mathbf{b}, \quad \mathbf{v}_2=\mathbf{v}-(m_1/M)\mathbf{u}, \end{aligned} \tag{1.17}$$

we get

$$\begin{cases} \mathbf{P}\mapsto R_A\mathbf{P}+M\mathbf{v}, \\ \mathbf{Q}\mapsto R_A\mathbf{Q}-(\tau/M)(R_A\mathbf{p}+M\mathbf{v})+\mathbf{a}, \\ \mathbf{p}\mapsto R_A\mathbf{p}+\mu\mathbf{u}, \\ \mathbf{q}\mapsto R_A\mathbf{q}-(\tau/\mu)(R_A\mathbf{q}+\mu\mathbf{u})+\mathbf{b}, \\ \mathbf{s}_i\mapsto R_A\mathbf{s}_i, \quad i=1,2, \end{cases} \tag{1.18}$$

where $\mu = m_1 m_2/M$ denotes the reduced mass of the system.

This maximal subgroup itself contains two subgroups of particular interest:

1) *The diagonal subgroup*:

$$\mathscr{G}(3)_d = \{(\xi,\tau,\mathbf{a},\mathbf{v},A),(\xi,\tau,\mathbf{a},\mathbf{v},A)\}, \tag{1.19}$$

which is the extended kinematical Galilei group of the system. It is the group with respect to which we perform the foliation (1.12) or the decomposition (1.13). Its Lie algrbra is simply the diagonal subalgebra of $\mathfrak{g}(3)\times\mathfrak{g}(3)$.

2) *The "internal subgroup"*:

$$\mathscr{I} = \{(\xi_1,\tau,(m_2/M)\mathbf{b},(m_2/M)\mathbf{u},A), (\xi_2,\tau,-(m_1/M)\mathbf{b},-(m_1/M)\mathbf{u},A)\}, \tag{1.20}$$

Note that this subgroup depends essentially on the masses $m_1$ and $m_2$. It could be defined as the maximal subgroup of $\mathscr{B}$ which stabilizes the pair $\{\mathbf{P}=\mathbf{0},\mathbf{Q}=\mathbf{0}\}$. This is a 12-dimensional group which can be seen as a central extension of the Galilei group G(3) by a two-dimensional Abelian group. We define the *internal Galilei group of the two-particle system* as the subgroup

$$G(3)_{\rm INT} = \{(0,\tau,(m_2/M)\mathbf{b},(m_2/M)\mathbf{u},A), (0,\tau,-(m_1/M)\mathbf{b},-(m_1/M)\mathbf{u},A)\}\subset\mathscr{I}, \tag{1.21}$$

the Lie algebra of which is generated by the elements

$$\mathscr{H} = \mathscr{H}_1 + \mathscr{H}_2,\quad \mathscr{P}_{\rm INT} = \frac{m_2}{M}\mathscr{P}_1 - \frac{m_1}{M}\mathscr{P}_2,$$
$$\mathscr{K}_{\rm INT} = \frac{m_2}{M}\mathscr{K}_1 - \frac{m_1}{M}\mathscr{K}_2,$$
$$\left(\mathscr{Q}_{\rm INT} = -\frac{1}{\mu}\mathscr{K}_{\rm INT} = -\frac{1}{m_1}\mathscr{K}_1 + \frac{1}{m_2}\mathscr{K}_2 = \mathscr{Q}_1 - \mathscr{Q}_2\right),\quad \mathscr{J} = \mathscr{J}_1 + \mathscr{J}_2. \tag{1.22}$$

[Indexes 1 and 2 refer to the algebra $\mathfrak{g}(3)\times\{0\}$ and $\{0\}\times\mathfrak{g}(3)$ in $\mathfrak{g}(3)\times\mathfrak{g}(3)$.]

For each given motion of the two-particle system, it is possible to choose a Galilei frame (with respect to the kinematical group) in which $\mathbf{P}=\mathbf{0}$, $\mathbf{Q}=\mathbf{0}$ (*the CM frame of motion*). In such a frame, the internal Gaililei group acts purely on the internal variables $\mathbf{p}$, $\mathbf{q}$, $\mathbf{s}_1$, and $\mathbf{s}_2$ through formulas (1.18). (Note that in the CM frame, $\mathbf{p}=\mathbf{p}_1=-\mathbf{p}_2$, which gives a useful interpretation of $\mathbf{p}$.) In this manner, it acts transitively on the set of all motions expressed in their CM frames. It conserves the total mass $M$ of the system, but it generates all the values of the internal angular momentum and of the internal energy of the system. As a matter of fact, the total angular momentum and the total energy

$$\mathbf{J} = \mathbf{j}_1 + \mathbf{j}_2 = \mathbf{l}_{\rm CM} + \mathbf{S}_{\rm INT} = \mathbf{Q}\times\mathbf{P} + (\mathbf{q}\times\mathbf{p} + \mathbf{s}_1 + \mathbf{s}_2), \tag{1.23}$$
$$E = E_1 + E_2 = K_{\rm CM} + U_{\rm INT} = \frac{1}{2M}|\mathbf{P}|^2 + \left(\frac{1}{2\mu}|\mathbf{p}|^2 + u_1 + u_2\right)$$

reduce, in the CM frame, to $\mathbf{S}_{\rm INT}$ and $U_{\rm INT}$. Taking into account the action of $G(3)_{\rm INT}$ on $\mathbf{p}$, $\mathbf{q}$, $\mathbf{s}_1$, and $\mathbf{s}_2$, we see that

$$0\leqslant S_{\rm INT} = |\mathbf{S}_{\rm INT}| < \infty,\quad u_1 + u_2 \leqslant U_{\rm INT} < \infty. \tag{1.24}$$

Hence, the action of $G(3)_{\rm INT}$ provides a tool to classify the orbits $\tilde{\mathscr{O}}(M,S_{\rm INT},U_{\rm INT})$ of the kinematical group $G(3)_d$ in the product manifold $\mathscr{O}(m,s_1,u_1)\times\mathscr{O}(m_2,s_2,u_2)$. Since we are not interested in classical mechanics, we do not perform this classification here [by finding the "degeneracy label" of formula (1.12)]; this tool, however, will be used in Sec. 4 in the quantum mechanical analog, which consists of reducing the tensor product of two irreducible unitary representations of $\mathscr{G}(3)$ into its irreducible components.

## 2. THE PIUR'S OF THE GALILEI GROUP G(3)

A Hilbert space on which a projective unitary representation of the Galilei group acts is associated with each Galileian quantum system. Such a system is called elementary if the representation is irreducible. The projective irreducible unitary representations (PIUR's) of G(3) have all been classified.[1] The ones which are not equivelent to true representations are characterized by the pairs $(m,s)$ $m\in\mathbb{R}_*$, $2s\in\mathbb{N}$, and up to unitary equivalence, they can all be obtained by restriction to G(3) of irreducible unitary representations of the extended Galilei group $\mathscr{G}(3)$ which belong to the classes $(m,s,u)$, $u\in\mathbb{R}$. A standard realization of such representations is given by

$$(U^{(msu)}(\xi,\tau,\mathbf{a},\mathbf{v},A)f)(\mathbf{p},\sigma) = \exp\left[-im\xi - i\left(\frac{1}{2m}|\mathbf{p}|^2 + u\right)\tau + i\mathbf{p}\cdot\mathbf{a}\right] \times \sum_{\sigma'=-s}^{s} D^s_{\sigma\sigma'}(A)f(R_A^{-1}(\mathbf{p}-m\mathbf{v}),\sigma'), \tag{2.1}$$

where

$$f\in L^2(\mathbb{R}^3\times\Sigma),\quad \Sigma = \{-s_1-s+1,\ldots,s-1,s\},$$
$$(f|f) = \sum_{\sigma=-s_1}^{s}\int_{\mathbb{R}^3} d^3p\,|f(\mathbf{p},\sigma)|^2$$

and $D^s$ is a spin-s representation of SU(2).

The corresponding representation $\nu$ of the Lie algebra [notation: $\nu(\mathscr{L}) = iL$] gives, for the basis (1.2),

$$M = -m,\quad H = -\frac{1}{2m}|\mathbf{p}|^2 - u,\quad P_k = p_k,$$
$$K_k = im\frac{\partial}{\partial p_k},\quad \left(Q_k = -M^{-1}K_k = i\frac{\partial}{\partial p_k}\right), \tag{2.2}$$
$$J_k = S_k + i\left(p_l\frac{\partial}{\partial p_j} - p_j\frac{\partial}{\partial p_l}\right),\quad (k,l,j) = \text{Cycl}\,(1,2,3).$$

Since the mapping $\mu: L^2(\mathbb{R}^3\times\Sigma)\rightarrow\mathfrak{g}(3)^*$ defined by $\langle\mu(f),\mathscr{L}\rangle = (f|(1/i)\nu(\mathscr{L})f)$, $\forall\,\mathscr{L}\in\mathfrak{g}(3)$, is covariant: $\Delta(g)\circ\mu = \mu\circ|U^{(msu)}(g)$, $\mathfrak{g}(3)^*$ of the Lie algebra. Consequently, the physical interpretation of the functions in $L^2(\mathbb{R}^3\times\Sigma)$ and the observables (2.2) associated to the Lie algebra has to follow the same pattern as in classical mechanics: A function $f\in L^2(\mathbb{R}^3\times\Sigma)$ represents either a *quantal mo-*

*tion* (not a state) of a free particle of mass $m$ ($m > 0$), spin $s$, and internal energy $u$ or an initial state (at the time $t = 0$) for the evolution law generated by the Hamiltonian $H$. Consequently, the observable $\mathbf{Q} = \{Q_1, Q_2, Q_3\}$ represents the *position operator at the time* $t = 0$.

Note that $L^2(\mathbb{R}^3 \times \Sigma)$ is the spectral representation space for the operators $P_i$, $i = 1,2,3$, $\mathbf{S}^2$, and $S_3$ (and $M$):

$$(P_i f)(\mathbf{p},\sigma) = p_i f(\mathbf{p},\sigma), \quad (S^2 f)(\mathbf{p},\sigma) = s(s+1) f(\mathbf{p},\sigma),$$
$$(S_3 f)(\mathbf{p},\sigma) = \sigma f(\mathbf{p},\sigma).$$

This is the so called "canonical realization" of the $(m,s)$-PIUR class of G(3). Another equivalent realization known as the "helicity realization" is obtained when, instead of the spin operator $S_3$, it is the helicity operator

$$\Lambda = \frac{\mathbf{J}\cdot\mathbf{P}}{|\mathbf{P}|}, \tag{2.3}$$

which is diagonalized. The relation between these two realizations is given by the unitary operator on $L^2(\mathbb{R}^3 \times \Sigma]$

$$\tilde{f}(\mathbf{p},\lambda) = (T_1 f)(\mathbf{p},\lambda) = \sum_{\sigma=-s}^{s} D^s_{\lambda\sigma}(h(\mathbf{p})^{-1}) f(\mathbf{p},\sigma), \tag{2.4}$$

$$f(\mathbf{p},\sigma) = (T_1^{-1}\tilde{f})(\mathbf{p},\sigma) = \sum_{\lambda=-s}^{s} D^s_{\sigma\lambda}(h(\mathbf{p}))\tilde{f}(\mathbf{p},\lambda),$$

where $h(\mathbf{p})$ is defined in (A1). It is easy to verify that the representation

$$\tilde{U}^{(msu)}(\cdots) = T_1 U^{(msu)}(\cdots) T_1^{-1}$$

acts on $L^2(\mathbb{R}^3 \times \Sigma)$ as follows:

$$(\tilde{U}^{(msu)}(\xi,\tau,\mathbf{a},\mathbf{v},A)\tilde{f})(\mathbf{p},\lambda)$$
$$= \exp\{-im\xi - i[(1/2m)|\mathbf{p}|^2 + u]\tau + i\mathbf{p}\cdot\mathbf{a}\}$$
$$\times \sum_{\lambda'=-s}^{s} D^s_{\lambda\lambda'}(h(\mathbf{p})^{-1} A h(R_A^{-1}(\mathbf{p}-m\mathbf{v})))\tilde{f}$$
$$(R_A^{-1}(\mathbf{p}-m\mathbf{v}),\lambda'). \tag{2.5}$$

In Sec. 3, we need two particular realizations of the representation class $(m,s)$. They are related to two group–subgroup reductions, and they provide the spectral representation space for two different complete sets of operators.

## A. The Euclidean subgroup E(3)

We consider here the chain of subgroups $G(3) \supset E(3) \supset SU(2) \supset SU(1)$.

By considering the restriction to E(3) of the representation $\tilde{U}^{(msu)}$ in (2.5), we obtain

$$(\tilde{U}^{(msu)}\downarrow_{E(3)}(\mathbf{a},A)\tilde{f})(\mathbf{p},\lambda)$$
$$= \exp(i\mathbf{p}\cdot\mathbf{a}) \times \chi_\lambda(h(\mathbf{p})^{-1} A h(R_A^{-1}\mathbf{p}))\tilde{f}(R_A^{-1}\mathbf{p},\lambda)).$$

{since $B = h(\mathbf{p})^{-1} A h(R_A^{-1}\mathbf{p}) \in SU(1)$, $D_{\lambda\lambda'}(B) = \delta_{\lambda\lambda'}\chi_\lambda(B)$ [cf. (A1)].}

By virtue of (A1), this representation is reducible following the decomposition formula (A12). Then, by using the unitary transformation defined by (A15), we complete the reduction of the chain of subgroups considered. We obtain the Hilbert space $\mathscr{H}_{(E(3))}$ of functions $\hat{f}((ms),p,\lambda,j,n)$ satisfying (A18) and defined on the spectra of the operators (in addition, of course, to $M$ and $\mathbf{S}^2$) $|\mathbf{P}|$, helicity $\Lambda$, total angular momentum $\mathbf{J}$, and its projection $J_3$. For completeness, let us rewrite transformations (A15) in this context:

$$\hat{f}((ms),p,\lambda,j,n)$$
$$= (T_2\tilde{f})((ms),p,\lambda,j,n)$$
$$= \left(\frac{2j+1}{4\pi}\right)^{1/2} \int_{S^2} d\Omega(\omega) D^j_{n\lambda}(h(\omega))\tilde{f}(p\omega,\lambda), \tag{2.6}$$

$$\tilde{f}(\mathbf{p},\lambda) = (T_2^{-1}\hat{f})(\mathbf{p},\lambda)$$
$$= \sum_{j=|\lambda|}^{\infty} \sum_{n=-j}^{j} \hat{f}((ms),p,\lambda,j,n)$$
$$\times \left(\frac{2j+1}{4\pi}\right)^{1/2} D^j_{n\lambda}(h(\omega))^*.$$

## B. The homogeneous Galilei subgroup $G(3)_0$

Here the chain of subgroups considered is $G(3) \supset G(3)_0 \supset SU(2) \supset SU(1)$.

The restriction ot $G(3)_0$ of the representation $U^{(msu)}$ gives

$$(U^{(msu)}\downarrow_{G(3)_0}(\mathbf{v},A)f)(\mathbf{p},\sigma)$$
$$= \sum_{\sigma'=-s}^{s} D^s_{\sigma\sigma'}(A) f(R_A^{-1}(\mathbf{p}-m\mathbf{v}),\sigma').$$

Up to the factor $m$ before $\mathbf{v}$, we recognize the quasiregular representation of E(3) defined by (A11). As indicated in the Appendix, the complete reduction of the chain of subgroups is achieved by the unitary transformation (A17). In order to avoid notational confusion we rewrite (A17) in the form

$$\check{f}((ms),\rho,\nu,j,n)$$
$$= (Tf)((ms),\rho,\nu,j,n)$$
$$= \left[\frac{1}{2\pi^2(2s+1)}\right]^{1/2} \sum_{\sigma=-s}^{s} \int_{\mathbb{R}^3} d^3p\, D^{\rho\nu}_{jns\sigma}(\mathbf{p},1) f(\mathbf{p},\sigma), \tag{2.7}$$

$$f(\mathbf{p},\sigma) = (T^{-1}\check{f})(\mathbf{p},\sigma)$$
$$= \left[\frac{1}{2\pi^2(2s+1)}\right]^{1/2} \sum_{\nu=-s}^{s} \sum_{j=|\nu|}^{\infty} \sum_{n=-j}^{j} \int_0^\infty d\rho\,\rho^2$$
$$\times D^{\rho\nu}_{jns\sigma}(\mathbf{p},1)^* \check{f}((ms),\rho,\nu,j,n).$$

In this case, we obtain the Hilbert space $\mathscr{H}_{(G(3)_0)}$ of functions $\check{f}((ms),\rho,\nu,j,n)$ satisfying (A18) and defined on the spectra of the operators $(M,\mathbf{S}^2)$, $|\mathbf{K}|$, $|\mathbf{K}|^{-1}\mathbf{J}\cdot\mathbf{K}$, $\mathbf{J}^2$, and $J_3$. According to the interpretation previously given for the operator $\mathbf{Q}$, the operator $|\mathbf{Q}| = (1/m)|\mathbf{K}|$ has to be interpreted as the distance between the particle and the origin of the frame at the time $t=0$, while the operator $|\mathbf{K}|^{-1}\mathbf{J}\cdot\mathbf{K}$ is a kind of helicity, i.e., the projection on the position vector at the time $t=0$ of the spin of the particle.

The unitary transformation

$$T_3 : \mathscr{H}_{(\mathrm{E}(3))} \rightarrow \mathscr{H}_{(G(3)_0)}$$

is given by

$$T_3 = T T_1^{-1} T_2^{-1},$$

where $T_1$, $T_2$, and $T$ are respectively defined by (2.4), (2.6) and (2.7). After some manipulations involving in particular (A5), (A6), and (A7) we get

$$\begin{aligned}\check{f}((ms),\rho,\nu,j,n) &= (T_3\hat{f})((ms),\rho,\nu,j,n)\\ &= \left(\frac{2}{\pi(2s+1)(2j+1)}\right)^{1/2} \sum_{\lambda=-s}^{s} \int_0^\infty dp\, p^2 d^{\rho\nu}_{js\lambda}(p)\\ &\quad \times \hat{f}((ms),p,\lambda,j,n)\end{aligned}$$

$$\begin{aligned}\hat{f}((ms),p,\lambda,j,n) &= (T_3^{-1}\check{f})((ms),p,\lambda,j,n)\\ &= \left(\frac{2}{\pi(2s+1)(2j+1)}\right)^{1/2} \sum_{\nu=-s}^{s} \int_0^\infty d\rho\, \rho^2 d^{\rho\nu}_{js\lambda}(p)^*\\ &\quad \times \check{f}((ms),\rho,\nu,j,n).\end{aligned}$$

*Remark*: On the space $\mathscr{H}_{(\mathrm{E}(3))}$ the group E(3) acts according to (A19)

$$\begin{aligned}(\hat{U}(\mathbf{a},A)\hat{f})((ms),p,\lambda,j,n) = \sum_{j'=|\lambda|}^{\infty} \sum_{n'=-j'}^{j'} D^{p\lambda}_{jnj'n'}(\mathbf{a},A)\hat{f}((ms),p,\lambda,j',n'),\end{aligned}$$

while on the space $\mathscr{H}_{(\mathrm{G}(3)_0)}$, the group $\mathrm{G}(3)_0$ acts as

$$\begin{aligned}(\check{U}(\mathbf{v},A)\check{f})((ms),\rho,\nu,j,n) = \sum_{j'=|\lambda|}^{\infty} \sum_{n'=-j'}^{j'} D^{\rho\nu}_{jnj'n'}(m\mathbf{v},A)\check{f}((ms),\rho,\nu,j',n').\end{aligned}$$

The extension of these representations to a representation of G(3) equivalent to $U^{(msu)}$ is of course possible, but it involves in general integral operators which we do not write down here.

## 3. TWO-PARTICLE GALILEIAN QUANTUM SYSTEMS

For a nonelementary system, the projective unitary representation of G(3) which is defined on the Hilbert space associated to the system is not irreducible. However it is possible to decompose the Hilbert space $\mathscr{H}$ into a direct integral of spaces $\mathscr{H}_{M,j,U,\xi}$ on which the action of G(3) [or $\mathscr{G}(3)$] is given by the representations $U^{(MjU)}$:

$$\mathscr{H} \simeq \int^{\oplus} d\mu(j,U,\xi)\mathscr{H}_{M,j,U,\xi}. \tag{3.1}$$

Here $M$ denotes the total mass of the system and $\xi$ denotes some label which removes the degeneracies. Moreover, as in classical mechanics, there always exists a barycentric decomposition so that

$$\mathscr{H} \simeq \mathscr{H}_{\mathrm{CM}} \otimes \mathscr{H}_{\mathrm{INT}} \tag{3.2}$$

where $\mathscr{H}_{\mathrm{CM}} \simeq L^2(\mathbb{R}^3)$. The action of G(3) on this space $\mathscr{H}_{\mathrm{CM}}$ is unitarily equivalent to

$$\begin{aligned}(U|_{\mathrm{CM}}(\tau,\mathbf{a},\mathbf{v},A)f)(\mathbf{P}) = \exp\left(-\frac{i\tau}{2M}|\mathbf{P}|^2 + i\mathbf{P}\cdot\mathbf{a}\right)f(R_A^{-1}(\mathbf{P}-M\mathbf{v})).\end{aligned} \tag{3.3}$$

When the space $\mathscr{H}$ is realized in the form (3.2), the decomposition (3.1) reduces to a corresponding decomposition of the "internal space" $\mathscr{H}_{\mathrm{INT}}$.

Now let us consider the special case of a system consisting of two noninteracting particles. Then the Hilbert spaces is the tensor product $\mathscr{H} = \mathscr{H}_{m_1,s_1,u_1} \otimes \mathscr{H}_{m_2,s_2,u_2}$ and the action of the kinematical group is simply defined as the restriction to the diagonal subgroup $\mathscr{G}(3)_d$ of the representation $U^{(m_1s_1u_1)} \otimes U^{(m_2s_2u_2)}$ of $\mathscr{G}(3)\times\mathscr{G}(3)$. For the realization (2.1), we have

$$\begin{aligned}\mathscr{H} &= L^2(\mathbb{R}^3\times\Sigma_1)\otimes(\mathbb{R}^3\times\Sigma_2)\\ &\simeq L^2(\mathbb{R}^3\times R^3\times\Sigma_1\times\Sigma_2)\end{aligned} \tag{3.4}$$

and the barycentric decomposition (3.2) is performed by means of the unitary transformation of $\mathscr{H}$,

$$\begin{aligned}\tilde{f}(\mathbf{P},\mathbf{p},\sigma_1,\sigma_2) &= (T_B f)(\mathbf{P},\mathbf{p},\sigma_1,\sigma_2)\\ &= f\left(\frac{m_1}{M}\mathbf{P}+\mathbf{p},\frac{m_2}{M}\mathbf{P}-\mathbf{p},\sigma_1,\sigma_2\right)\end{aligned} \tag{3.5}$$

induced by the change of variables [cf. (1.15)]

$$B\begin{cases}\mathbf{P} = \mathbf{p}_1+\mathbf{p}_2,\\ \mathbf{p} = \dfrac{1}{M}(m_2\mathbf{p}_1 - m_1\mathbf{p}_2),\end{cases} \qquad B^{-1}\begin{cases}\mathbf{p}_1 = \dfrac{m_1}{M}\mathbf{P}+\mathbf{p},\\ \mathbf{p}_2 = \dfrac{m_2}{M}\mathbf{P}-\mathbf{p}.\end{cases}$$

Indeed, the space of functions $\tilde{f}(\mathbf{P},\mathbf{p},\sigma_1,\sigma_2)$ is naturally isometric to the tensor product $L^2(\mathbb{R}^3)\otimes\mathscr{H}_{\mathrm{INT}}$ with $\mathscr{H}_{\mathrm{INT}} \simeq L^2(\mathbb{R}^3\times\Sigma_1\times\Sigma_2)$.

In the case of the Hilbert space (3.4), the decomposition (3.1) is well known[2,4]:

$$\mathscr{H} \simeq \bigoplus_{\xi} \bigoplus_{j=0}^{\infty} \int_0^{\infty\,\oplus} p^2\, dp\, \mathscr{H}_{M,j,U,\xi}, \tag{3.6}$$

where $p = |\mathbf{p}|$ and $U = U_{\mathrm{INT}} = (1/2\mu)p^2 + u_1 + u_2$ [cf. (1.23)]. There exist two standard choices for the degeneracy label $\xi$ depending on the manner of "coupling" representations of SU(2) which occur in the reduction process.

### A. $(l-s)$ or canonical coupling

Here $\xi = (l,s)$ and the unitary transformation which realizes the decomposition (3.6) is given by ($\mathbf{p} = p\boldsymbol{\omega}, \boldsymbol{\omega}\in S^2$)

$$\begin{aligned}(C_c\tilde{f})(\mathbf{P},p,j,m,l,s) &= \sum_{\sigma}\sum_{n}\sum_{\sigma_1\sigma_2} (s\sigma ln|jm)(s_1\sigma_1 s_2\sigma_2|s\sigma)\\ &\quad\times \int_{S^2} d\Omega(\boldsymbol{\omega}) Y^l_n(\boldsymbol{\omega})^*\tilde{f}(\mathbf{P},\mathbf{p},\sigma_1,\sigma_2).\end{aligned} \tag{3.7}$$

## B. Helicity coupling

Here $\xi = (\lambda_1,\lambda_2)$ but first let us introduce the following unitary transformation on $\mathscr{H}$:

$$(B\tilde{f})(\mathbf{P},\mathbf{p},\lambda_1,\lambda_2) = \sum_{\sigma_1}\sum_{\sigma_2} D^{s_1}_{\lambda_1\sigma_1}(h(\mathbf{p})^{-1}) D^{s_2}_{\lambda_2\sigma_2}(h(\mathbf{p})^{-1})\tilde{f}(\mathbf{P},\mathbf{p},\sigma_1,\sigma_2).$$

Then

$$(C_h\tilde{f})(\mathbf{P},p,j,m,\lambda_1,\lambda_2) = \frac{\sqrt{2j+1}}{4\pi}\int_{S^2} d\Omega(\omega) D^{j}_{m\lambda_1+\lambda_2}(h(\omega))\tilde{f}(\mathbf{P},\mathbf{p},\lambda_1,\lambda_2).$$

Let us mention that $\lambda_1$ and $-\lambda_2$ are the helicities of the two particles since $\mathbf{p}$ represents the momentum of the first particle in the CM frame of the two-particle system.

Let us examine in more detail the decomposition (3.6) in relation with the barycentric decomposition $\mathscr{H}\simeq L^2(\mathbb{R}^3)\otimes\mathscr{H}_{\rm INT}$. Since the group $\mathscr{G}(3)\times\mathscr{G}(3)$ acts on $\mathscr{H}$ via the representation $U^{(m_1s_1u_1)}\otimes U^{(m_2s_2u_2)}$, we have a well defined action of its subgroups $\mathscr{B}$, $\mathscr{I}$, and $G(3)_{\rm INT}$. Moreover, the action of the group $\mathscr{B}$ respects the barycentric decomposition given by (3.5). Indeed, by setting

$$g = \left(\left(\xi_1,\tau,\mathbf{a}+\frac{m_2}{M}\mathbf{b},\mathbf{v}+\frac{m_2}{M}\mathbf{u},A\right),\left(\xi_2,\tau,\mathbf{a}-\frac{m_1}{M}\mathbf{b},\mathbf{v}-\frac{m_1}{M}\mathbf{u},A\right)\right)\in\mathscr{B}, \tag{3.8}$$

it is easy to verify that the representation of $\mathscr{B}$ on $\mathscr{H}$ is

$$(\mathscr{U}(g)\tilde{f})(\mathbf{P},\mathbf{p},\sigma_1,\sigma_2) = \exp\left[-im_1\xi_1 - im_2\xi_2 - i\left(\frac{1}{2M}|\mathbf{P}|^2+\frac{1}{2\mu}|\mathbf{p}|^2 + u_1 + u_2\right)\tau + i\mathbf{P}\cdot\mathbf{a} + i\mathbf{p}\cdot\mathbf{b}\right]\sum_{\sigma_1'}\sum_{\sigma_2'} D^{s_1}_{\sigma_1\sigma_1'}(A)$$
$$\times D^{s_2}_{\sigma_2\sigma_2'}(A)\tilde{f}(R_A^{-1}(\mathbf{P}-M\mathbf{v}),\ R_A^{-1}(\mathbf{p}-\mu\mathbf{u}),\sigma_1',\sigma_2'). \tag{3.9}$$

This representation can be written in the form

$$(\mathscr{U}(g)F)(\mathbf{P}) = \mathscr{L}(\Lambda(\mathbf{P})^{-1}g\Lambda(g^{-1}\mathbf{P}))F(g^{-1}\mathbf{P}),$$

where

$F:\mathbb{R}^3\to\mathscr{H}_{\rm INT}\quad [F(\mathbf{P})(\mathbf{p},\sigma_1,\sigma_2)] = \tilde{f}[\mathbf{P},\mathbf{p},\sigma_1,\sigma_2)]$,

$\Lambda(\mathbf{P}):\mathbf{0}\to\mathbf{P},\Lambda(\mathbf{P}) = ((0,0,\mathbf{0},\mathbf{P}/M,1),(0,0,\mathbf{0},\mathbf{P}/M,1))\in\mathscr{G}(3)_d$,

$g\mathbf{P} = R_A\mathbf{P} + M\mathbf{v}$,

and $\mathscr{L}$ is the following unitary representation of the group $\mathscr{I}$ defined on the space $\mathscr{H}_{\rm INT}$:

$$(\mathscr{L}(\xi_1,\xi_2,\tau,\mathbf{b},\mathbf{u},A)\phi)(\mathbf{p},\sigma_1,\sigma_2) = \exp\left[-im_1\xi_1 - im_2\xi - i\left(\frac{1}{2\mu}|\mathbf{p}|^2 + u_1 + u_2\right)\tau + i\mathbf{p}\cdot\mathbf{b}\right]$$
$$\times\sum_{\sigma_1'}\sum_{\sigma_2'} D^{s_1}_{\sigma_1\sigma_1'}(A)D^{s_2}_{\sigma_2\sigma_2'}(A)\phi(R_A^{-1}(\mathbf{p}-\mu\mathbf{u}),\sigma_1',\sigma_2'). \tag{3.10}$$

Hence the representation $\mathscr{U}$ of $\mathscr{B}$ appears as a representation induced by the representation $\mathscr{L}$ of the internal subgroup $\mathscr{I}$. A projective representation $\tilde{\mathscr{L}}$ of the Internal Galilei group $G(3)_{\rm INT}$ is simply obtained from $\mathscr{L}$ by setting $\xi_1 = \xi_2 = 0$ in formula (3.10). It is not irreducible but of course

$$\tilde{\mathscr{L}} \simeq \bigoplus_{s=|s_1-s_2|}^{s_1+s_2} U^{(\mu s u_1 + u_2)} \tag{3.11}$$

and this reduction can be realized by the Clebsch–Gordan transformation

$$\tilde{\phi}(\mathbf{p},s,\sigma) = (C\phi)(\mathbf{p},s,\sigma) = \sum_{\sigma_1\sigma_2}(s_1\sigma_1\, s_2\sigma_2\,|\, s\sigma)\phi(\mathbf{p},\sigma_1,\sigma_2). \tag{3.12}$$

The decomposition (3.6) is performed by expressing the internal space $\mathscr{H}_{\rm INT}$ as a direct integral of spaces irreducible with respect to the diagonal subgroup $\mathscr{G}(3)_d$. By virtue of (3.9) this group acts on $\mathscr{H}_{\rm INT}$ only through its subgroup $\mathscr{A}$ consisting of the elements of SU(2) and of the time translations; according to the definitions (1.19) and (1.21) we have

$$\mathscr{A} = \mathscr{G}(3)_d \cap G(3)_{\rm INT}.$$

Hence, since $\mathscr{A}\subset G(3)_{\rm INT}$, all we have to do is to compute the reduction into irreducible components of the restriction to $\mathscr{A}$ of the representation $\tilde{\mathscr{L}}$ of $G(3)_{\rm INT}$. In other words, we have to find a realization of $\mathscr{H}_{\rm INT}$ in which the internal energy operator $H_{\rm INT}$ (i.e., the generator of the time translations) and the internal angular momentum $\mathbf{J}^2_{\rm INT}$ are diagonal.

In the previous section we have considered two particular realizations of the Hilbert space of the PIUR's of G(3).

The first one was related to the chain of subgroups which in the present context will be written as

$$G(3)_{\rm INT}\supset E(3)_{\rm INT}\supset SU(2)\supset SU(1).$$

The subgroup $E(3)_{\rm INT}$ does not contain $\mathscr{A}$; however it is clear from (1.1) that the time translations commute with the entire subgroup $E(3)_{\rm INT}$. Consequently, in the realization defined by (2.6) both $H_{\rm INT}$ and $J^2_{\rm INT}$ are diagonal. Actually, this realization, together with the reduction (3.12) give, up to some appropriate change in the degeneracy label $\xi$, either the realization (3.7) or the realization (3.8).

The second one, related to the chain of subgroups

$$G(3)_{\rm INT}\supset G(3)_{0\ \rm INT}\supset SU(2)\supset SU(1).$$

does not lead to the decomposition (3.6) because the internal energy operator $H_{\rm INT}$ is not left invariant by the subgroup $G(3)_{0\ \rm INT}$. Indeed we have

$$\tilde{\mathscr{L}}(0,\mathbf{0},\mathbf{u},1)^\dagger H_{\rm INT}\tilde{\mathscr{L}}(0,\mathbf{0},\mathbf{u},1) = H_{\rm INT} + \tfrac{1}{2}\mu|\mathbf{u}|^2.$$

However, in spite of, or rather thanks to this lack of invariance, the realization of $\mathscr{H}$ associated with the realization (2.7) of $\mathscr{H}_{\rm INT}$ will be used to obtain a two-variable expansion of the scattering amplitude. By virtue of (2.7), the unitary transformation

$$S:\mathscr{H}\to\check{\mathscr{H}},$$

where $\check{\mathscr{H}}$ is the space of functions $\check{f}(\mathbf{P},s,\rho,\nu,j,m)$ such that

$$(\check{f}|\check{f}) = \sum_{s=|s_1-s_2|}^{s_1+s_2} \sum_{\nu=-s}^{s} \sum_{j=|\nu|}^{\infty} \sum_{m=-j}^{j} \int_0^\infty \rho^2\, d\rho \times \int_{\mathbb{R}^3} d^3P\, |\check{f}(\mathbf{P},s,\rho,\nu,j,m)|^2 < \infty$$

is given by

$$(S\tilde{f})(\mathbf{P},s,\rho,\nu,j,m) = \frac{1}{[2\pi^2(2s+1)]^{1/2}} \times \sum_{\sigma=-s}^{s} \int_{\mathbb{R}^3} d^3p\, D^{\rho\nu}_{jm\,s\sigma}(\mathbf{p},1)\tilde{f}(\mathbf{P},\mathbf{p},s,\sigma).$$

$$(S^{-1}\check{f})(\mathbf{P},\mathbf{p},s,\sigma,) = \frac{1}{[2\pi^2(2s+1)]^{1/2}} \sum_{\nu=-s}^{s} \sum_{j=|\nu|}^{\infty} \sum_{m=-j}^{j} \int_0^\infty d\rho\, \rho^2 \times D^{\rho\nu}_{jm\,s\sigma}(\mathbf{p},1)^* \check{f}(\mathbf{P},s,\rho,\nu,j,m). \tag{3.13}$$

The space $\check{\mathscr{H}}$ is the spectral representation space for the total linear momentum, the length $\mathbf{S}^2$ of the vector sum of the two spins, the internal angular momentum $J^2_{\mathrm{INT}}$, its projection $J_{3\,\mathrm{INT}}$, the distance at $t=0$ between the two particles and finally the projection of $J_{\mathrm{INT}}$ on the relative vector position at $t=0$.

The unitary operators which connect this realization of the space $\mathscr{H}$ with either the $(l-s)$ coupling one or the helicity coupling one are given by

(1) *$(l-s)$ coupling*:

$$(T_c\bar{f})(\mathbf{P},s,\rho,\nu,j,m) = \sum_\lambda \sum_l \left(\frac{2(2l+1)}{\pi(2s+1)(2j+1)^2}\right)^{1/2} (s\lambda l 0|j\lambda) \times \int_0^\infty dp\, p^2\, d^{\rho\nu}_{js\lambda}(p)\bar{f}(\mathbf{P},p,j,m,l,s),$$

$$(T_c^{-1}\check{f})(\mathbf{P},p,j,m,l,s) = \sum_\nu \sum_\lambda \left(\frac{2(2l+1)}{\pi(2s+1)(2j+1)^2}\right)^{1/2} (s\lambda l 0|j\lambda) \times \int_0^\infty d\rho\, \rho^2\, d^{\rho\nu}_{js\lambda}(p)^*\check{f}(\mathbf{P},s,\rho,\nu,j,m), \tag{3.14}$$

(2) *Helicity coupling*:

$$(T_h\bar{f})(\mathbf{P},s,\rho,\nu,j,m) = \left(\frac{2}{\pi(2s+1)(2j+1)}\right)^{1/2} \sum_{\lambda_1\lambda_2\lambda} (s_1\lambda_1 s_2\lambda_2|s\lambda) \times \int_0^\infty dp\, p^2\, d^{\rho\nu}_{js\lambda}(p)\bar{f}(\mathbf{P},p,j,m,\lambda_1,\lambda_2), \tag{3.15}$$

$$(T_h^{-1}\check{f})(\mathbf{P},p,j,m,\lambda_1,\lambda_2) = \left(\frac{2}{\pi(2s+1)(2j+1)}\right)^{1/2} \sum_s \sum_\nu (s_1\lambda_1 s_2\lambda_2|s\lambda) \times \int_0^\infty d\rho\, \rho^2\, d^{\rho\nu}_{js\lambda}(p)^*\check{f}(\mathbf{P},s,\rho,\nu,j,m).$$

These transformations are obtained after some manipulations by using formulas (2.6), (2.7), (3.6), (3.7), and (3.12).

Let us note that in both cases, the transformations in one direction involve an integration over the spectrum of the internal energy operator (which is a kinematical invariant) while in the other direction they involve an integration over the spectrum of the operator interpreted as the distance at $t=0$ between the two particles.

## 4. TWO-VARIABLE EXPANSIONS OF SCATTERING AMPLITUDES

The phenomenological Galileian elastic scattering theory deals with a unitary operator (the scattering operator)

$$S: \mathscr{H} \to \mathscr{H}$$

defined on the Hilbert space (3.4) of two noninteracting particles. The kinematical Galileian invariance in scattering is introduced by means of the intertwining condition

$$\mathrm{U}(g)S = S\mathrm{U}(g), \quad \forall\ g \in \mathrm{G}(3),$$

where $\mathrm{U}(g)$ denotes the projective unitary representation $U^{(m_1,s_1,u_1)} \otimes U^{(m_2,s_2,u_2)}$. According to the Schur lemma and to the decomposition (3.6) of the space $\mathscr{H}$ we must have for both realizations (3.7) and (3.9),

$$(Sf)(\mathbf{P},p,j,m,\xi) = \sum_{\xi'} S(p,j,\xi,\xi')f(\mathbf{P},p,j,m,\xi') \tag{4.1}$$

with $|S(p,j,\xi,\xi')| \leqslant 1$. The interpretation of functions $f$ as describing a motion of two free particles is well adapted to this scattering theory: These motions are the asymptotically free limits of the true motion of the interacting particles. This formalism can be extended to non elastic scattering but for this more general case the Hilbert space (3.4) is only a part of the space which has to be considered and consequently in its decomposition (3.1), the degneracy labels $\xi$ in (4.1) have to be completed in order to take into account the various channels. However in this section we limit ourselves to reactions of the type $1+2 \to 3+4$. Recall that the interwining condition implies by virtue of (3.1) the conservation of the total mass $M$ of the system.

For applications, it is not the scattering operator which is relevant but the transition operator $T$ defined by

$$T = i(S-1).$$

Therefore, we shall be concerned with the "matrix elements"

$$T(p,j,\xi,\xi') = i(S(p,j,\xi,\xi') - \delta_{\xi\xi'}).$$

The barycentric decomposition (3.2) is particularly useful since, because of (4.1), we have

$$T = 1 \otimes \widetilde{T}.$$

We can therefore restrict ourselves to functions belonging to the internal space $\mathscr{H}_{\rm INT}$ and from now on, when referring to formulas (3.7), (3.8), (3.13), (3.14), and (3.15) it will be always understood that it is their restriction to $\mathscr{H}_{\rm INT}$ which has to be considered. Hence, for the transition operator $T$, and with this convention, formula (4.1) becomes

$$(\widetilde{T}f)(p,j,m,\xi) = \sum_{\xi'} T(p,j,\xi,\xi')f(p,j,m,\xi'). \tag{4.2}$$

This equation holds both for the canonical and for the helicity coupling realization of $\mathscr{H}_{\rm INT}$. When considering the original realization $\mathscr{H}_{\rm INT} = L^2(\mathbb{R}^3 \times \Sigma_1 \times \Sigma_2)$, it is not difficult to establish that the corresponding operator

$$T' = C_\xi^{-1}\widetilde{T}C_\xi, \tag{4.3}$$

where $C_\xi$ is defined either by (3.7) or by (3.8), is of the following form,

$$(T'f)(\mathbf{P},\sigma_1,\sigma_2) = \sum_{\sigma_1'}\sum_{\sigma_2'}\int_{S^2} d\Omega(\omega')\, T(p,\omega,\omega',\sigma_1,\sigma_2,\sigma_1',\sigma_2') \times f(\mathbf{p}_1',\sigma_1',\sigma_2') \tag{4.4}$$

with $\mathbf{p} = p\omega$, $\mathbf{p}' = p\omega'$ ($|\mathbf{p}| = |\mathbf{p}'| = p$).

Following a standard analysis which we do not reproduce here, the kernel $T(p,\omega,\omega',\sigma_1,\sigma_2,\sigma_1',\sigma_2')$ is related to the scattering amplitudes by

$$a(p,\omega,\sigma_1,\sigma_2,\sigma_1',\sigma_2') = T(p,\omega,\mathbf{e}_3,\sigma_1,\sigma_2,\sigma_1',\sigma_2'), \tag{4.5}$$

where $\omega' = \mathbf{e}_3$ is the direction of the incident beam and $\omega = \omega(\varphi,\theta)$ is the direction of a scattered particle traditionally expressed in the CM frame of a pair of incident and target particles.

When the scattering amplitudes are expressed, by using (4.3), in terms of the matrix elements $T(p,j,\xi,\xi')$ defined by (4.2), we obtain the two following well known one-variable expansions:

(1) *Canonical coupling* [cf. (3.7)]:

$$a(p,\omega,\sigma_1,\sigma_2,\sigma_1',\sigma_2') = \sum_s\sum_{s'}(s_1\sigma_1 s_2\sigma_2|s\sigma)(s_1\sigma_1' s_2\sigma_2'|s'\sigma') \times \sum_l\sum_{l'}\sum_j (s\sigma\, lm|\, j\sigma')(s'\sigma' l\, 0|j\sigma')T(p,j,l,s,l',s') \times\left(\frac{2l'+1}{4\pi}\right)^{1/2} Y^l_m(\omega). \tag{4.6}$$

(2) *Helicity coupling* [cf. (3.8)]:

$$a(p,\omega,\lambda_1,\lambda_2,\lambda_1',\lambda_2') = \sum_j\left(\frac{2j+1}{4\pi}\right)T(p,j,\lambda_1,\lambda_2,\lambda_1',\lambda_2') \times D^j_{\lambda_1+\lambda_2\,\lambda_1'\lambda_2'}(h(\omega))^*. \tag{4.7}$$

For the case of spin-zero particles, both formulas reduce to the familiar expansion

$$a(p,\varphi,\theta) = \sum_l\left(\frac{2l+1}{4\pi}\right)T(p,l)P_l(\cos\theta). \tag{4.8}$$

In these formulas, the specific dymanics of the scattering process is entirely contained in the matrix elements $T(p,j,\xi,\xi')$. The direction of scattering $\omega$ only appears as the argument of known special functions related to the group SU(2), but this is not the case for the internal kinetic energy $E = (1/2\mu)p^2$ which appears (via $p$) in the matrix elements $T(p,j,\xi,\xi')$. As mentioned in the Introduction, this situation presents some disadvantages and it would be a significant improvement to provide a formalism which treat the internal kinetic energy on the same footing as the direction of scattering. It is the idea of the two-variable expansions to construct satisfactory representations of the functions $p \mapsto T(p,j,\xi,\xi')$ in terms of known functions depending on a set of interpretable indices.

It is clear that such two-variable expansions cannot be obtained on the basis of purely kinematical considerations. Indeed, while it is true that the definition (4.2) of the matrix elements $T(p,j,\xi,\xi')$, based on the decomposition of the space $\mathscr{H}_{\rm INT}$ into subspaces irreducible with respect to the kinematical Galilei group, is responsible for the one-variable expansion of the scattering amplitudes in terms of functions depending on the direction variable $\omega$, it is also true that this definition cannot lead to a similar treatment for the variable $p$ since this latter is an invariant for the action of this group.

However, the transition operator $\widetilde{T}$ acts on the internal Hilbert space $\mathscr{H}_{\rm INT}$ and on this space we have a unitary representation of the internal Galilei group whose generators all have a physical meaning. Moreover, let us recall that the decomposition of this space into subspaces irreducible with respect to the kinematical group was performed by means of the realization of $\mathscr{H}_{\rm INT}$ associated with the group–subgroup chain $\mathrm{G(3)_{INT}} \supset \mathrm{E(3)_{INT}} \supset \mathrm{SU(2)} \supset \mathrm{SU(1)}$. In this realization, the internal energy operator and the internal angular momentum, which are two kinematical invariants, are represented by diagonal operators. The decomposition (4.2) of the transition operator $\widetilde{T}$ and the subsequent one-variable expansions (4.6) and (4.7) follow directly from that: As functions of the direction variable $\omega$, the scattering amplitudes are expanded in series of orthogonal eigenfunctions of the internal angular momentum operator. A similar treatment will be possible for the $p$ variable only if we start from a realization of $\mathscr{H}_{\rm INT}$ in which the internal kinetic energy is not a diagonal operator. In the other realization of this space that we considered in Sec. 3, associated with the group-subgroup chain $\mathrm{G(3)_{INT}} \supset \mathrm{G(3)_{0\ INT}} \supset \mathrm{SU(2)} \supset \mathrm{SU(1)}$, this is precisely the case. Moreover, since the internal angular momentum is still represented by a diagonal operator, nothing will be changed in the $\omega$ dependence of the scattering amplitudes.

In this realization, the transition operator is given by

$$T'' = T_\xi \widetilde{T} T_\xi^{-1},$$

where $T_\xi$ denotes either the unitary transformation $T_C$ or $T_h$

defined respectively by (3.14) or (3.15), and it can be written in the (possibly formal) form

$$(T''\check{f})(s,\rho,j,m) = \sum_{s'}\sum_{\nu'}\int_0^\infty d\rho'\,\rho'^2 T(\rho,\rho',j,s,\nu,s',\nu')\check{f}(s',\rho',j,m). \quad (4.9)$$

In the case $\xi = (l,s)$, it can be directly verified that

$$T(\rho,\rho',j,s,\nu,s',\nu') = \frac{2}{\pi(2j+1)^2}\sum_\lambda\sum_{\lambda'}\sum_l\sum_{l'}\left(\frac{(2l+1)(2l'+1)}{(2s+1)(2s'+1)}\right)^{1/2}$$
$$\times(s\lambda l0|j\lambda)(s'\lambda' l'0|j\lambda')\int_0^\infty dp\,p^2 d^{\rho\nu}_{js\lambda}(p)$$
$$\times T(p,j,l,s,l',s')d^{\rho'\nu'}_{js'\lambda'}(p)^*.$$

Conversely, by computing $\widetilde{T}$ as $T_c^{-1}T''T_c$, we get

$$T(p,j,l,s,l',s') = \frac{2}{\pi(2j+1)^2}\sum_\nu\sum_{\nu'}\sum_\lambda\sum_{\lambda'}\left(\frac{(2l+1)(2l'+1)}{(2s+1)(2s'+1)}\right)^{1/2}$$
$$\times(s\lambda l0|j\lambda)(s'\lambda' l'0|j\lambda')\int_0^\infty d\rho\,\rho^2\int_0^\infty dp'\,p'^2$$
$$\times\int_0^\infty d\rho'\,\rho'^2 d^{\rho\nu}_{js\lambda}(p)^* T(\rho,\rho',j,s,\nu,s',\nu')d^{\rho'\nu'}_{js'\lambda'}(p'). \quad (4.10)$$

Similar formulas hold of course for $\xi = (\lambda_1,\lambda_2)$. The important point to note in this formula is that the variable $p$ only appears in the special function $d^{\rho\nu}_{js\lambda}(p)$.

For later use, we define the following function

$$A(\rho,\nu,j,s,s',\lambda') = \frac{1}{2\pi^2\sqrt{(2s+1)(2s'+1)}}\sum_{\nu'}\int_0^\infty d\rho'\,p'^2$$
$$\times\int_0^\infty d\rho\,\rho^2 T(\rho,\rho',j,s,\nu,s',\nu')d^{\rho'\nu'}_{js'\lambda'}(p'). \quad (4.11)$$

Then, formula (4.10) can be written as

$$T(p,j,l,s,l',s') = \frac{4\pi}{(2j+1)^2}\sqrt{(2l+1)(2l'+1)}\sum_\nu\sum_\lambda\sum_{\lambda'}(s\lambda l0|j\lambda)$$
$$\times(s'\lambda' l'0|j\lambda')\int_0^\infty d\rho\,\rho^2 d^{\rho\nu}_{js\lambda}(p)A(\rho,\nu,j,s,s',\lambda'). \quad (4.12)$$

Finally we obtain the two-variable expansions of the scattering amplitudes either by replacing $T(p,j,l,s,l',s')$ in (4.6) by its expression (4.12) or by computing directly the transition operation $T'$ as $C^{-1}S^{-1}T''SC$ for the canonical coupling, where $C$ is defined by (3.12) and $S$ by (3.13). For the helicity coupling scheme, we have to compute the kernel of $BC^{-1}S^{-1}T''SCB^{-1}$, where $B$ is defined by (3.8), we get:

(1) *Canonical coupling*:

$$a(p,\omega,\sigma_1,\sigma_2,\sigma_1',\sigma_2') = \sum_s\sum_{s'}(s_1\sigma_1s_2\sigma_2|s\sigma)(s_1\sigma_1's_2\sigma_2'|s'\sigma')\sum_\nu\sum_j\int_0^\infty d\rho\,\rho^2$$
$$\times D^{\rho\nu}_{j\sigma's\sigma}(\mathbf{p},1)^*A(\rho,\nu,j,s,s',\sigma'). \quad (4.13)$$

(2) *Helicity coupling*:

$$a(p,\omega,\lambda_1,\lambda_2,\lambda_1',\lambda_2') = \sum_s\sum_{s'}(s_1\lambda_1s_2\lambda_2|s\lambda)(s_1\lambda_1's_2\lambda_2'|s'\lambda')\sum_\nu\sum_j\int_0^\infty d\rho$$
$$\times\rho^2 D^{\rho\nu}_{j\lambda's\lambda}(\mathbf{p},h(\mathbf{p}))^*A(\rho,\nu,j,s,s',\lambda'). \quad (4.14)$$

Let us notice that, by using (3.14), formula (4.12) can be inverted to give

$$A(\rho,\nu,j,s,s',\lambda') = \frac{1}{2\pi^2}\sum_l\sum_{l'}\sum_\lambda\frac{\sqrt{(2l+1)(2l'+1)}}{(2s+1)(2s'+1)}(s'\lambda' l'0\,|\,j\lambda')$$
$$\times(s\lambda l0\,|\,j\lambda)\int_0^\infty dp\,p^2 d^{\rho\nu}_{js\lambda}(p)^*T(p,j,l,s,l',s')$$

but it is obvious that some conditions have to be imposed on the function $A(\rho,\nu,j,s,s',\lambda')$ in order to obtain not just a formal expression.

It is possible to exploit the remarkable structure of the $d^{\rho\nu}_{jj'm}(p)$ functions which is given explicitly in (A9) in order to express the two-variable expansions (4.13) and (4.14) directly in terms of spherical Bessel functions. Before doing this, it is judicious to introduce the following function

$$B(\rho,l,j,s,s',\lambda') = \sqrt{4\pi}\left(\frac{(2l+1)(2s+1)}{2j+1}\right)^{1/2} i^l$$
$$\times\sum_\nu(s\nu l0\,|\,j\nu)A(\rho,\nu,j,s,s',\lambda')$$

which is the generalization to the cases of arbitrary spins, of the so-called "Galilei amplitudes" introduced in Ref. 1.

After some manipulations we obtain:

(1) *Canonical coupling*:

$$a(p,\omega,\sigma_1,\sigma_2,\sigma_1',\sigma_2') = \sum_s\sum_{s'}(s_1\sigma_1s_2\sigma_2\,|\,s\sigma)(s_1\sigma_1's_2\sigma_2'\,|\,s'\sigma')$$
$$\times\sum_l\sum_j(s\sigma lm\,|\,j\sigma')$$
$$\times Y^l_m(\omega)\int_0^\infty d\rho\,\rho^2 j_l(\rho p)B(\rho,l,j,s,s',\sigma').$$

$B(\rho,l,j,s,s',\sigma')$

$$= \frac{2l+1}{2j+1} \sum_{\sigma_1\sigma_2} \sum_{\sigma_1'\sigma_2'} (s_1\sigma_1 s_2\sigma_2 \mid s\sigma)$$

$$\times (s_1\sigma_1' s_2\sigma_2' \mid s'\sigma')(s\sigma lm \mid j\sigma') \int_{S^2} d\Omega(\omega)$$

$$\times \int_0^\infty dp\, p^2\, Y^l_m(\omega)^* j_l(\rho p) a(p,\omega,\sigma_1,\sigma_2,\sigma_1',\sigma_2').$$

(2) *Helicity coupling*:

$a(p,\omega,\lambda_1,\lambda_2,\lambda_1',\lambda_2')$

$$= \sum_s \sum_{s'} (s_1\lambda_1 s_2\lambda_2 \mid s\lambda)$$

$$\times (s_1\lambda_1' s_2\lambda_2' \mid s'\lambda') \sum_l \sum_j (s\lambda l0 \mid j\lambda)$$

$$\times \left(\frac{2l+1}{2j+1}\right)^{1/2} \left(\frac{2j+1}{4\pi}\right)^{1/2} D^j_{\lambda'\lambda}(h(\omega))^*$$

$$\times \int_0^\infty d\rho\, \rho^2 j_l(\rho p) B(\rho,l,j,s,s',\lambda').$$

$B(\rho,l,j,s,s',\lambda')$

$$= \left(\frac{2l+1}{2j+1}\right)^{3/2} \sum_{\lambda_1\lambda_2} \sum_{\lambda_1'\lambda_2'} (s_1\lambda_1 s_2\lambda_2 \mid s\lambda)(s_1\lambda_1' s_2\lambda_2' \mid s'\lambda')$$

$$\times (s\lambda l0 \mid j\lambda) \int_{S^2} d\Omega(\omega) \int_0^\infty dp\, p^2 \left(\frac{2j+1}{4\pi}\right)^{1/2}$$

$$\times D^j_{\lambda'\lambda}(h(\omega)) j_l(\rho p) a(p,\omega,\lambda_1,\lambda_2,\lambda_1',\lambda_2').$$

By taking into account the known behavior of the functions $j_l(x)$ near $x = 0$, it is easily seen that these expansions have the correct threshold behavior ($p \to 0$).

## CONCLUSION

The two variable expansions of scattering amplitudes we propose here are based on the decomposition of the "internal space" of the two particles into invariant subspaces associated with the chain of subgroups $G(3)_{INT} \supset G(3)_{0\,INT} \supset SU(2) \supset SU(1)$. It would be possible to obtain other expansions by choosing other chains, for example, $G(3)_{INT} \supset G(3)_{0\,INT} \supset E(2) \supset SU(1)$ or $G(3)_{INT} \supset G(3)_{0\,INT} \supset \mathbb{R}^3$. These two examples have already been considered in Ref. 1 for scalar particles and from a slightly different point of view. The first one is related to the eikonal expansion.

We have shown that the internal Galilei group can be naturally introduced by means of the barycentric decomposition concept. Except for the rotations and for the time translations, which are shared with the kinematical group, this group has no further connections with kinematical properties of the system. It acts on the internal properties of the system: relative space translations and relative Galileian boosts. Note that this internal group concept associated with the barycentric decomposition could be easily generalized to an $N$-particle system.

This method of approaching the two variable expansions has been applied in relativistic theory and will be the subject of a future paper. The main and new problem which arises in this case is concerned with some technical difficulties due to the lack of good properties in the barycentric decomposition of a relativistic system when the particles have different masses.

## ACKNOWLEDGMENTS

We thank P. Winternitz for discussions on this subject and related topics. We thank also the Centre de Recherches Mathématique of the Université de Montréal for its hospitality.

## APPENDIX

We give in this Appendix some results about the quasi-regular representations of the Euclidean group E(3). It does not contain anything new except for the introduction of the $d^{p\lambda}_{jj'm}$ functions which are used in the main text so as to make the Galileian treatment as close as possible to the relativistic one. For references about representation theory of E(3), see W. Miller Jr.[5]

The faithful irreducible unitary representations of E(3) are characterized by the pairs $(p,\lambda)$, $p \in \mathbb{R}_*^+$, $2\lambda \in \mathbb{Z}$. A standard realization is

$$(U^{(p\lambda)}(\mathbf{a},A)f)(\omega) = \exp\{ip\omega\cdot\mathbf{a}\}\chi_\lambda(h(\omega)^{-1}Ah(R_A^{-1}\omega))f(R_A^{-1}\omega), \quad \text{(A1)}$$

where

$$f \in L^2_\Omega(S^2),\ (f \mid g) = \int_{S^2} d\Omega(\omega) f(\omega)^* g(\omega),$$

$$d\Omega(\omega) = \sin\theta\, d\varphi\, d\theta,$$

$h(\omega) \in SU(2)$ is such that $R_{h(\omega)} : \mathbf{e}_3 \to \omega$,[5]

$\chi_\lambda$ is a character of SU(1).

An orthonormal basis of $L^2_\Omega(S^2)$ of eigenvectors of the operators $J^2$ and $J_3$ is given by

$$F^\lambda_{jm}(\omega) = \left(\frac{2j+1}{4\pi}\right)^{1/2} D^j_{m\lambda}(h(\omega))^*, \quad \text{(A2)}$$

where $D^j_{m\lambda}$ is a Wigner $D$ function and $|\lambda| \leqslant j < \infty$, $-j \leqslant m \leqslant j$.

The matrix elements of $U(\mathbf{a},A)$ are defined by

$$D^{p\lambda}_{jmj'm'}(\mathbf{a},A) = (F^\lambda_{jm} \mid U^{(p\lambda)}(\mathbf{a},A)F^\lambda_{j'm'}). \quad \text{(A3)}$$

In particular

$$D^{p\lambda}_{jmj'm'}(0,A) = \delta_{jj'} D^j_{mm'}(A). \quad \text{(A4)}$$

Any element of E(3) can be factorized in such a way that only translations along the third axis are involved,

$$(\mathbf{a},A) = (0,h(\mathbf{a}))(a\mathbf{e}_3,1)(0,h(\mathbf{a})^{-1}A), \quad \text{(A5)}$$

where $a = |\mathbf{a}|$ and $h(\mathbf{a}) \equiv h(a^{-1}\mathbf{a})$. Hence the only new matrix element to be considered is $D^{p\lambda}_{jmj'm'}(a\mathbf{e}_3,1)$. By using the definition (A3) with the scalar product defined by (A1), one shows that

$$D^{p\lambda}_{jmj'm'}(a\mathbf{e}_3,1) = 0 \quad \text{if } m \neq m'.$$

Then, by defining the reduced matrix element $d^{p\lambda}_{jj'm}(a)$ of E(3) as following

$$D^{p\lambda}_{jmj'm'}(a\mathbf{e}_3,1) = \delta_{mm'} d^{p\lambda}_{jj'm}(a), \tag{A6}$$

we have

$$D^{p\lambda}_{jmj'm'}(\mathbf{a},A) = \sum_n D^{j}_{mn}(h(\mathbf{a})) d^{p\lambda}_{jj'n}(a) D^{j'}_{nm'}(h(\mathbf{a})^{-1}A), \tag{A7}$$

where the summation over $n$ runs from $-\min(j,j')$ to $+\min(j,j')$.

The reduced matrix elements are explicitly given by

$$d^{p\lambda}_{jj'm}(a) = \tfrac{1}{2}\sqrt{(2j+1)(2j'+1)} \int_0^\pi d\theta \sin\theta e^{ipa\cos\theta} \times d^{j}_{m\lambda}(\cos\theta) d^{j'}_{m\lambda}(\cos\theta)^*. \tag{A8}$$

By using the familiar expansion in spherical Bessel functions

$$e^{i\mathbf{a}\cdot\mathbf{b}} = \sum_{l=0}^{\infty} (2l+1) i^l j_l(ab) P_l(\mathbf{v}\cdot\boldsymbol{\omega}), \quad \mathbf{a} = a\mathbf{v}, \quad \mathbf{b} = b\boldsymbol{\omega},$$

the integration in Eq. (A8) can be easily performed and we get

$$d^{p\lambda}_{jj'm}(a) = \left(\frac{2j'+1}{2j+1}\right)^{1/2} \sum_l (2l+1) i^l j_l(pa) \times (l0j'm \mid jm)(l0j'\lambda \mid j\lambda). \tag{A9}$$

This simple expression exhibits the symmetry of these functions with respect to the permutations $p \leftrightarrow a$, $j \leftrightarrow j'$, and $\lambda \leftrightarrow m$.

Let us mention the following "orthonormality" property

$$\sum_m \int_0^\infty da\, a^2 d^{p\lambda}_{j_1j_2m}(a)^* d^{p'\lambda'}_{j_1j_2m}(a) = \frac{\pi}{2}(2j_1+1)(2j_2+1)\delta^{\lambda\lambda'} \frac{\delta(p-p')}{p^2} \tag{A10}$$

and, by virtue of the symmetry $p \leftrightarrow a$, a similar "closure" property.

By *quasiregular representation* of E(3), we mean a unitary representation of the form

$$(U^{(s)}(\mathbf{a},A)f)(\mathbf{x},\sigma) = \sum_{\sigma'=-s}^{s} D^{s}_{\sigma\sigma'}(A) f(R_A^{-1}(\mathbf{x}-\mathbf{a}),\sigma'), \tag{A11}$$

where $f \in L^2(\mathbb{R}^3 \times \Sigma)$, $\Sigma = \{-s, -s+1, \ldots, s-1, s\}$, $2s \in \mathbb{N}$,

$$(f \mid f) = \sum_{\sigma=-s}^{s} \int_{\mathbb{R}^3} d^3x\, |f(\mathbf{x},\sigma)|^2 < \infty.$$

This representation can be seen as induced by the $D^s$ representation of the maximal compact subgroup SU(2) of E(3). It is reducible, more precisely we have

$$U^{(s)} \sim \bigoplus_{\lambda=-s}^{s} \int_0^{\infty\,\oplus} dp\, p^2 U^{(p\lambda)}. \tag{A12}$$

Actually, when performing at first a fourier transformation

$$\tilde{f}(\mathbf{p},\sigma) = (\mathscr{F}f)(\mathbf{p},\sigma) = (2\pi)^{-3/2} \int_{\mathbb{R}^3} d^3x\, e^{i\mathbf{p}\cdot\mathbf{x}} F(\mathbf{x},\sigma),$$

we obtain an equivalent representation

$$\tilde{U}^{(s)}(\ldots) = \mathscr{F} U^{(s)}(\ldots) \mathscr{F}^{-1}$$

such that

$$(\tilde{U}^{(s)}(\mathbf{a},A)\tilde{f})(\mathbf{p},\sigma) = \exp\{i\mathbf{p}\cdot\mathbf{a}\} \sum_{\sigma'=-s}^{s} D^{s}_{\sigma\sigma'}(A) \tilde{f}(R_A^{-1}\mathbf{p},\sigma'). \tag{A13}$$

Then, by using the unitary transformation ($p = |\mathbf{p}|$, $\mathbf{p} = p\boldsymbol{\omega}$)

$$\mathbf{f}(p,\lambda,\boldsymbol{\omega}) = (T_2\tilde{f})(p,\lambda,\boldsymbol{\omega}) = \sum_{\sigma=-s}^{s} D^{s}_{\lambda\sigma}(h(\boldsymbol{\omega})^{-1}) \tilde{f}(p\boldsymbol{\omega},\sigma) \tag{A14}$$

we easily establish the decomposition (A11) by introducing the spherical coordinates of $\mathbb{R}^3$.

The spectral representation space $\mathscr{H}^{(s)}$ for the operators $|\mathbf{P}|$, $\mathbf{J}\cdot\mathbf{P}$, $\mathbf{J}^2$, and $J_3$ associated to the representation $U^{(s)}$ is simply obtained starting from the space $L^2(\mathbb{R}^+ \times S^2 \times \Sigma)$ of the functions $\mathbf{f}(p,\lambda,\boldsymbol{\omega})$ by introducing the unitary transformation

$$\hat{f}(p,\lambda,j,m) = (T_3\bar{f})(p,\lambda\boldsymbol{\omega}) = \left(\frac{2j+1}{4\pi}\right)^{1/2} \int_{S^2} d\Omega(\boldsymbol{\omega}) D^{j}_{m\lambda}(h(\boldsymbol{\omega})) \bar{f}(p,\lambda,\boldsymbol{\omega}),$$

$$\bar{f}(p,\lambda,\boldsymbol{\omega}) = (T_3^{-1}\hat{f})(p,\lambda,\boldsymbol{\omega}) = \sum_{j=|\lambda|}^{\infty} \sum_{m=-j}^{j} \hat{f}(p,\lambda,j,m) \times \left(\frac{2j+1}{4\pi}\right)^{1/2} D^{j}_{m\lambda}(h(\boldsymbol{\omega}))^*. \tag{A15}$$

Then

$$\begin{aligned}
(|\mathbf{P}|\hat{f})(p,\lambda,j,m) &= p\hat{f}(p,\lambda,j,m);\\
(\mathbf{J}\cdot\mathbf{P}\hat{f})(p,\lambda,j,m) &= p\lambda\hat{f}(p,\lambda,j,m),\\
(\mathbf{J}^2\hat{f})(p,\lambda,j,m) &= j(j+1)\hat{f}(p,\lambda,j,m);\\
(\mathbf{J}_3\hat{f})(p,\lambda,j,m) &= m\hat{f}(p,\lambda,j,m).
\end{aligned} \tag{A16}$$

When starting from the space $L^2(\mathbb{R}^3 \times \Sigma)$ of the functions $f(\mathbf{x},\sigma)$, we have to consider the unitary transformation

$$T = T_3 T_2 \mathscr{F}.$$

After some manipulations [using (A5) and (A6)], we can write $T$ and its inverse in the form

$$\hat{f}(p,\lambda,j,m) = (Tf)(p,\lambda,j,m) = \frac{1}{[2\pi^2(2s+1)]^{1/2}} \sum_{\sigma=-s}^{s} \int_{\mathbb{R}^3} d^3x\, D^{p\lambda}_{jms\sigma}(\mathbf{x},1) f(\mathbf{x},\sigma) \tag{A17}$$

$$f(\mathbf{x},\sigma) = (T^{-1}\hat{f})(\mathbf{x},\sigma)$$
$$= \frac{1}{[2\pi^2(2s+1)]^{1/2}} \sum_{\sigma=-s}^{s} \sum_{j=|\lambda|}^{\infty} \sum_{m=-j}^{j} \int_0^\infty dp\, p^2$$
$$\times D^{p\lambda}_{jms\sigma}(\mathbf{x},1)^* \hat{f}(p,\lambda,j,m).$$

Note that

$$(\hat{f}\,|\,\hat{f}) = \sum_{\sigma=-s}^{s} \sum_{j=|\lambda|}^{\infty} \sum_{m=-j}^{j} \int_0^\infty dp\, p^2 |\hat{f}(p,\lambda,j,m)|^2 \quad \text{(A18)}$$

and that the representation

$$\hat{U}^{(s)}(\cdots) = TU^{(s)}(\cdots)T^{-1}$$

acts as follows,

$$(\hat{U}^{(s)}(\mathbf{a},A)\hat{f})(p,\lambda,j,m)$$
$$= \sum_{j'=|\lambda|}^{\infty} \sum_{m'=-j'}^{j'} D^{p\lambda}_{jmj'm'}(\mathbf{a},A)\hat{f}(p,\lambda,j',m'). \quad \text{(A19)}$$


[1]E.G. Kalnins, J. Patera, R.T. Sharp, and P. Winternitz, in *Group Theory and its Applications*, edited by E. Loebl (Academic, New York, 1975), Vol. 3; Phys. Rev. D **8**, 2552 (1973); **10**, 3527 (1973).

[2]J.M. Lévy-Leblond, in *Group Theory and its Applications*, edited by E. Loebl,(Academic, New York, 1971), Vol. 2.

[3]J.M. Souriau, *Structure des systémes dynamiques* (Dunod, Paris, 1970).

[4]J. Voisin, J. Math. Phys. **6**, 1822 (1965).

[5]W. Miller, Jr., Commun. Pure Appl. Math. **18**, 527 (1964).

# Positivity conditions on correlation functions that imply Debye screening

Paul Federbush

*Department of Mathematics, The University of Michigan, Ann Arbor, Michigan 48109*
(Received 13 July 1978)

In the classical statistical mechanics setting, a set of positivity conditions on certain two-point correlation functions is exhibited that implies Debye screening for a large class of Coulomb-like models. For example, for the model treated by Brydges, for which he has rigorously proved shielding, in a range of parameters where $\langle \phi^s(x)J(y)\rangle \geq 0$ for all $x$ and $y$ and all $s$ odd, there is screening. (Alternative conditions require positivity for only two correlation functions.) Strong estimates are obtained for the rate of exponential falloff.

Currently there is much interest in acquiring understanding of Debye screening. An outstanding problem is the question of proving shielding in quantum statistical mechanics (assuming it is valid). Toward this end, the short range difficulties of the $1/r$ potential have been controlled.[1] Brydges has recently proved screening for classical Coulomb systems.[2] We feel the present approach may provide insight into the problem, dealing with shielding charges in a suggestive way. We will deduce exponential falloff of the two point correlation function, for a large class of models, under the assumption of positivity for certain two-point correlation functions. This work may be interesting for a number of reasons. First, it is possible a proof of these positivity conditions will be forthcoming, giving a new proof of shielding. Second, we obtain strong statements on the rate of exponential falloff. Third, the positivity conditions may be tested for in numerical experiments (or theoretically) to provide a good estimate of the range of parameters for which shielding holds.

We study a classical statistical mechanical system of several species of "charged" particles, species $i$ with charge $q_i$ and fugacity $z_i$. See Ref. 3, for example, for the basic definitions. The partition function is written as

$$Z = \sum \prod_i \left( \frac{z_i^{n_i}}{n_i!} \int dx^{n_i} \right) e^{-\beta V}, \tag{1}$$

with

$$V = \tfrac{1}{2} \int :J\left( \frac{\gamma}{r} + v \right) J:, \tag{2}$$

$v$ the short range potential, $\gamma/r$ the Coulomb term, and

$$J = \sum q_i \rho_i, \tag{3}$$

$\rho_i$ the density of species $i$. We assume for convenience

$$\sum z_i q_i = 0. \tag{4}$$

We define $\phi$ as

$$\phi = \int \left( \frac{\gamma}{r} + v \right) J. \tag{5}$$

It is helpful to define the notation:

$$Z = I(e^{-\beta V}), \tag{6}$$

$$[A\,] = I(e^{-\beta V}A\,), \tag{7}$$

$$\langle A \rangle = \frac{[A\,]}{Z}, \tag{8}$$

to discuss ensemble averages of a function $A$. We also define

$$w = (-\Delta\,)v. \tag{9}$$

We are now prepared to state several conditions from which we will prove a number of results.

*Condition I*: There exists $f(z,x) \equiv f(z)$ (i.e., the $x$ dependence is suppressed) satisfying

$$\left( -\Delta + 4\pi\gamma \sum_i z_i q_i^2 \beta + \sum_i z_i q_i^2 \beta w * \right) f(z) = 4\pi\gamma\delta(z-x) + w(z-x), \tag{10}$$

with $f \geqslant 0$ all $z$ (and $f$ falling off exponentially).

Later we will show such an $f$ exists in a number of interesting cases, and find the falloff explicitly in these cases.

*Condition II*: $\langle J(y)\,\phi(x)\rangle \geqslant 0 \quad$ all $x,y$, $\qquad$ (11)

*Condition III*.1: $\langle J(y)\,\phi^s(x)\rangle \geqslant 0 \quad$ all $x,y$, $\;s$ odd, $\qquad$ (12)

*Condition III*.2:

$$\left\langle J(y)\left( -J(x) - \sum_i z_i q_i^2 \beta \phi(x) \right)\right\rangle \geqslant 0 \quad \text{all } x \neq y, \tag{13}$$

*Condition III*.3:

$$\left\langle J(y) \sum_i z_i q_i \left( -e^{-\beta q_i \phi(x)} + 1 - \beta q_i \phi(x) \right)\right\rangle \geqslant 0 \quad \text{all } x,y. \tag{14}$$

A system is charge symmetric if the species occur in pairs with equal fugacities and equal and opposite charges.

*Theorem* 1: For a system satisfying Conditions I, II, and III.3, or for a system satisfying Conditions I, II, and III.2, or for a charge symmetric system satisfying Conditions I, II, and III.1, one has

$$0 \leqslant \langle (y)\,\phi(x)\rangle \leqslant cf(y,x). \tag{15}$$

 0022-2488/79/061090-03$01.00  

From the conditions on $f$, this implies the exponential falloff of the two point correlation function; the system shields.

For the system studied by Brydges, suitably scaled, (10) becomes

$$(-\Delta + 2zq^2\beta)f = 4\pi\delta, \tag{16}$$

an equation on a unit lattice, $\Delta$ the discrete Laplacian. [we have picked his $l = 1$, and $\gamma = (1/4\pi)$.]

*Theorem* 2: For this system one may pick

$$0 \leqslant f(x,y) = g(x-y) = g(x_1,x_2,x_3), \tag{17}$$

where $g$ is the fundamental solution of (16), some of whose properties are given in the Appendix.

We note that it is not known that Conditions II or III hold in the region where Brydges has proved shielding, but we believe they do.

We next consider a system with

$$v = -\gamma \frac{e^{-\alpha r}}{r}. \tag{18}$$

This choice eliminates the singularity of the total potential at $r=0$. Equation (10) in momentum space becomes

$$\left(k^2 + \frac{4\pi\alpha^2\tilde{\gamma}}{k^2+\alpha^2}\right)\tilde{f} = \alpha^2\tilde{\gamma}\frac{4\pi}{k^2+\alpha^2}, \tag{19}$$

with

$$\tilde{\gamma} = \sum z_i q_i^2 \beta\gamma. \tag{20}$$

We pick

$$\tilde{f} = \frac{4\pi\alpha^2\tilde{\gamma}}{r_2 - r_1}\left(\frac{1}{k^2+r_1} - \frac{1}{k^2+r_2}\right), \tag{21}$$

$$r_2 = \frac{\alpha^2 + (\alpha^4 - 16\pi\alpha^2\tilde{\gamma})^{1/2}}{2}, \tag{22}$$

$$r_1 = \frac{\alpha^2 - (\alpha^4 - 16\pi\alpha^2\tilde{\gamma})^{1/2}}{2}. \tag{23}$$

*Theorem* 3: For this system, with

$$\alpha^2 > 16\pi\tilde{\gamma} > 0, \tag{24}$$

one may pick $f$ satisfying

$$0 \leqslant f(x,y) \leqslant \frac{c\exp(-\sqrt{r_1}\,|x-y|)}{|x-y|}. \tag{25}$$

We thus have two natural systems satisfying Condition I.

Before turning to a proof of Theorem 1 we first consider a simpler statistical mechanics model, a Gaussian distribution of continuous charges on a unit rectangular lattice.

$$Z = \prod_i \left(\int dJ_i e^{-\alpha/2J_i^2}\right)\exp\left[\left(\frac{-\beta}{2}\right)\sum_{i,j} J_i\left(\frac{\gamma}{r} + v\right)J_j\right]. \tag{26}$$

There is a $J_i$ for each lattice site $i$. $1/r$ denotes the Green's function for a discrete Laplacian $\Delta$. With a notation similar to Eqs. (6)–(8) one has the pull-through formula (an integration by parts),

$$I(J_iB) = \frac{1}{\alpha} I\left(\frac{\partial B}{\partial J_i}\right). \tag{27}$$

As in (5) one defines

$$\phi = \sum \left(\frac{\gamma}{r} + v\right)_i J_i. \tag{28}$$

We want to study $\langle J_y\phi_x\rangle$, $y$ and $x$ lattice sites, and write

$$\langle J_y\phi_x\rangle = \left\langle J_y\left\{\phi_x - \sum f_i J_i + \sum f_i J_i\right\}\right\rangle \tag{29}$$

(the dependence of $f$ on $x$ is suppressed). We apply (27) to the third term in braces,

$$= \frac{1}{\alpha} f_y + \left\langle J_y\left\{\phi_x - \sum f_i J_i - \frac{\beta}{\alpha}\sum f_i\left(\frac{\gamma}{r} + v\right)_{ij} J_j\right\}\right\rangle. \tag{30}$$

If we can find an $f$ satisfying

$$(\alpha + \beta\gamma(1/r)* + \beta v*)f = \alpha\left(\frac{\gamma}{r} + v\right), \tag{31}$$

an exact analogy of (10), we get

$$\langle J_y\phi_x\rangle = \frac{1}{\alpha} f_y, \tag{32}$$

(31) and (32) directly yield

$$\langle J_i J_j\rangle = \left(\frac{1}{\alpha + \beta\gamma/r + \beta v}\right)_{ij}. \tag{33}$$

Any correlation function can be calculated by this result and Wick's theorem for this Gaussian model. It follows that for such a Gaussian model if $\alpha$, $\beta$, $\gamma$ and $v$ are picked to ensure

$$\langle J_i\phi_j\rangle \geqslant 0 \quad \text{all } i \text{ and } j, \tag{34}$$

and (as is automatic)

$$\langle \phi_i\phi_i\rangle \geqslant 0 \quad \text{all } i, \tag{35}$$

that it satisfies Conditions II, III.1, III.2, and III.3. It is easy to show that a Gaussian model approximating the models of Theorem 2 or Theorem 3, having the same $f$ in a discretized form, satisfies (34). That is, our Conditions II and III hold for the continuous approximations of our models, essentially the $\beta \to 0$ limit (with $\beta z$ fixed).

Returning to a proof of Theorem 1, we first exhibit an analog of the pull-through formula (27) in the classical statistical mechanics setting. We let $B$ be a functional of the $\{\rho_i(x)\}$. Then

$$I(\rho_i(x)B) = z_i I(B^{(\cdot ix)}), \tag{36}$$

where $B^{(\cdot ix)}$ is $B$ with $\rho_i(y)$ replaced by $\rho_i(y) + \delta(y-x)$. As in (29) we write

$$\langle J(y)\phi(x)\rangle = \left\langle J(y)\left\{\phi(x) - \int f(z)J(z) + \int f(z)J(z)\right\}\right\rangle \tag{37}$$

and apply (36) to the last term in braces getting

$$\langle J(y)\phi(x)\rangle = C + E + M, \tag{38}$$

where $C$ will be a set of terms identically *canceling*, $M$ will be the *main* term, and $E$ will serve the place of an *error*.

$$C=0=\left\langle J(y)\int\left[\frac{\gamma}{r}+v-f-\sum z_i q_i^2\beta \times\int f\left(\frac{\gamma}{r}+v\right)\right]J\right\rangle. \quad (39)$$

Applying the Laplacian to the coefficient of $J$ one gets exactly Eq. (10) in Condition I,

$$M=\sum z_i q_i^2 f(y,x)\langle e^{-\beta q_i\phi(y)}\rangle. \quad (40)$$

The expectations in (40) are all positive, their values determine the constant in (15) of Theorem 1.

$$E=\sum z_i q_i\int f(z)\langle J(y)(e^{-\beta q_i\phi(z)}-1+\beta q_i\phi(z))\rangle. \quad (41)$$

By Condition III.3 we assume $E\leqslant 0$. Thus from (38) we have

$$0\leqslant\langle J(y)\,\phi(x)\rangle=M+E\leqslant M, \quad (42)$$

The first inequality is Condition II; this is Eq. (15). The statement of III.1 trivially yields III.3 for symmetric systems. The inequalities of (13) and (14) are identical by a simple application of (36).

Proceeding to collect some final points we note that it was only necessary to ensure $C\leqslant 0$, not that $C=0$, and this freedom may be helpful in some situations. It is amusing to attempt to eliminate the need for an error term in (38) by allowing the $f$ in (37) to be a functional of the $J$'s. This was attempted in Ref. 4 with very limited success; other expressions like (38) were obtained with smaller error terms.

## ACKNOWLEDGMENT

This work was supported in part by NSF Grant PHY 77-02187.

## APPENDIX

We look at some properties of the Green's function for the discrete Helmholtz equation on a unit rectangular lattice,

$$(-\Delta+a)g=4\pi\delta.$$

(I) In one dimension,

$$g(x)=c'e^{-k|x|},$$

with

$$e^k+e^{-k}-2=a.$$

(II) In three dimensions $g(x_1,x_2,x_3)$, symmetric in its arguments, satisfies

$$\sum_{x_2,x_3} g(x_1,x_2,x_3)=c'e^{-k|x_1|},$$

for the same $k$ as in (I).

(III) In three dimensions

$$g\leqslant c''\exp\left(-\sum_i\frac{|x_i|}{2}\ln\frac{N_i+r_i}{N_i-r_i}\right),$$

where

$$r_i=|x_i|\Big/\Sigma|x_i|,$$

and

$$\frac{4(\Sigma N_i)^2+(6+a)^2 r_i^2}{(6+a)^2}=N_i^2,\quad N_i>0.$$

(II) follows easily from (I). (I) is verified by directly substituting into the difference equation. (III) is derived from a random walk expression for the Green's function.[5]

[1]D. Brydges, P. Federbush, Commun. Math. Phys. **49**, 233 (1976); Commun. Math. Phys. **53**, 19 (1977).
[2]D. Brydges, Commun. Math. Phys. **58**, 313 (1978).
[3]D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[4]P. Federbush, "A Functional Relationship for the Two Point Correlation Function and Approximate Debye Screening in Classical Statistical Mechanics" (unpublished preprint).
[5]D. Brydges and P. Federbush, Commun. Math. Phys. **62**, 79 (1978).

# Generalized susceptibility of a solitary wave[a)]

K. C. Lee[b)] and S. E. Trullinger

*Department of Physics, University of Southern California, Los Angeles, California 90007*
(Received 31 August 1978)

We define a generalized susceptibility for solitary wave solutions of the nonlinear Klein–Gordon equation and obtain its expression in terms of the complete set of functions which arise in the linear stability analysis of the solitary wave. Explicit expressions are presented for the susceptibility of the sine-Gordon soliton and the $\phi^4$ kink. Plots are presented for the long-wave dynamic polarizability of the $\phi^4$ kink which have application to the response of ferroelectric domain walls to an oscillating external electric field.

## I. INTRODUCTION

Solitary wave solutions of nonlinear wave equations have received considerable attention recently because of their wide applicability in several branches of physics,[1] particularly in condensed matter systems.[2] The sine–Gordon (SG) soliton and $\phi^4$ kink are two cases of particular interest, for example, as models for charge carriers[3] in weakly-pinned charge-density wave condensates and domain walls[4] in ferroelectrics, respectively. Solitary waves propagate through the system without distortion of shape and exhibit remarkable stability and other particle-like properties. This has prompted their use as models for extended particles in nonlinear quantum field theories.[5,6]

There have been some recent advances in understanding the behavior of solitary waves under the influence of external perturbations.[7,8] These studies indicate that solitary waves of the nonlinear Klein–Gordon type (e.g., SG and $\phi^4$) may be viewed as stable extended particles which are deformable under the influence of external forces. In this paper we study the response of a nonlinear Klein–Gordon solitary wave when it is subject to an external disturbance and damping. The linear response of the field when a solitary wave is present is compared to the response in the absence of a solitary wave in order to extract the intrinsic response of the solitary wave.

In Sec. II, we define the generalized susceptibility[9] for the linear response. We derive an expression for the susceptibility in terms of eigenfunctions of a Schrödinger-like equation which arises in the stability analysis for the solitary wave solution of the nonlinear Klein–Gordon equation. In Sec. III we calculate the susceptibilities for both the SG soliton and $\phi^4$ kink, obtaining explicit expressions for these two examples. In Sec. IV we present plots of the $\phi^4$ kink long-wavelength susceptibility for representative values of the damping constant. These results have application to the electric polarizability of domain walls[4] in ferroelectrics. Finally, in Sec. V we summarize our results and present some general comments.

[a)]Research supported by the National Science Foundation under grant No. DMR77-08845 and by the SNU-AID Graduate Basic Sciences Program.

[b)]Permanent address: Department of Physics, Seoul National University, 151 Seoul, Korea.

## II. GENERAL FORMALISM

We consider a Lagrangian density, $\mathscr{L}(\phi)$ of the nonlinear Klein–Gordon type for a real scalar field in one space and one time dimension. In dimensionless variables $(z,\tau)$, $\mathscr{L}(\phi)$ has the form

$$\mathscr{L}(\phi)=\frac{1}{2}\left(\frac{\partial\phi}{\partial\tau}\right)^2-\frac{1}{2}\left(\frac{\partial\phi}{\partial z}\right)^2-U(\phi), \qquad (2.1)$$

where $U(\phi)$ has at least two degenerate minima, say at $\phi_1$ and $\phi_2$, such that $U(\phi_1)=U(\phi_2)=0$. The equation of motion for the field $\phi$ is given by

$$\frac{\partial^2\phi}{\partial\tau^2}-\frac{\partial^2\phi}{\partial z^2}+U'(\phi)=0. \qquad (2.2)$$

The static solitary-wave (kink) solutions, $\phi_s(z)$, are obtained by integrating Eq. (2.2) (with $\partial\phi/\partial\tau=0$) subject to the boundary conditions $\phi_s(+\infty)=\phi_1$, $\phi_s(-\infty)=\phi_2$, or $\phi_s(-\infty)=\phi_1$, $\phi_s(+\infty)=\phi_2$. The traveling solutions may be obtained by boosting $\phi_s(z)$ to a frame moving with velocity $v$ ($|v|<1$).

The stability of the solitary-wave solutions may be investigated by determining the nature of small deviations from the static waveform $\phi_s(z)$. This is accomplished by considering solutions to Eq. (2.2) of the form

$$\phi(z,\tau)=\phi_s(z)+\psi(z)e^{i\omega\tau}, \qquad (2.3)$$

where $|\psi(z)|$ is assumed to be small. Substitution of (2.3) into (2.2) and subsequent linearization in $\psi$ yields the following equation for $\psi$,

$$\left(-\frac{d^2}{dz^2}+U''(\phi_s(z))\right)\psi=\omega^2\psi. \qquad (2.4)$$

The ground state of this Schrödinger-like equation always occurs at $\omega^2=0$, implying that all eigenvalues $\omega^2$ satisfy $\omega^2\geqslant 0$, and hence the solitary wave is stable against small oscillations.

The "bound" state at $\omega^2=0$ corresponds to the so-called "translation mode"[5-8] of the solitary wave since $\psi=\phi_s'(z)$ satisfies Eq. (2.4) with $\omega^2=0$. In general, there also exists a continuous spectrum of eigenvalues corresponding to "continuum" states with the dispersion relation

$$\omega_k^2=m^2+k^2, \qquad (2.5)$$

where $m^2$ is defined by

$$m^2=U''(\phi_s(z=\pm\infty)). \qquad (2.6)$$

 0022-2488/79/061093-06$01.00  

In addition, there may exist other bound states (with $0<\omega^2<m^2$) which correspond to localized "internal" oscillations of the solitary wave.

One can form a complete orthonormal set, $\{\psi_i\}$, from the eigenfunctions of Eq. (2.4). The orthonormality condition is expressed by

$$\int_{-\infty}^{+\infty} dz \psi_i^*(z)\psi_j(z) = \delta_{ij}, \tag{2.7}$$

where $i$ and $j$ are discrete labels for bound states and continuous labels ($k$) for the continuum states. For $i=k$ and $j=k'$, $\delta_{ij}=\delta(k-k')$. The completeness relation is expressed by

$$\int_i \psi_i^*(z)\psi_i(z') = \delta(z-z'), \tag{2.8}$$

where the symbol $\int_i$ denotes summation over discrete bound states plus integration over continuum states. We shall use this complete orthonormal set for the expansion of external disturbances and for the response of the field to such disturbances.

We consider the effect of a spatially and temporally varying external "force" $F(z,\tau)$ together with damping on the solitary-wave solution $\phi_s(z)$. The equation of motion (2.2) for the total field is modified to read

$$\frac{\partial^2\phi}{\partial\tau^2} - \frac{\partial^2\phi}{\partial z^2} + U'(\phi) + \Gamma\frac{\partial\phi}{\partial\tau} = F(z,\tau), \tag{2.9}$$

where $\Gamma$ is a damping constant. We seek a solution of Eq. (2.9) of the form

$$\phi(z,\tau) = \phi_s(z) + \Delta\phi(z,\tau), \tag{2.10}$$

i.e., we assume that a response $\Delta\phi$ to the disturbance is superimposed on the initially static solitary wave. We assume that $|\Delta\phi|$ is small[10] if $F(z,\tau)$ is small in magnitude. We then substitute Eq. (2.10) into Eq. (2.9) and linearize in $\Delta\phi$ to obtain

$$\frac{\partial^2\Delta\phi}{\partial\tau^2} - \frac{\partial^2\Delta\phi}{\partial z^2} + U''(\phi_s(z))\Delta\phi + \Gamma\frac{\partial\Delta\phi}{\partial\tau} = F(z,\tau). \tag{2.11}$$

In general the solution of the linear equation (2.11) can be expressed formally in terms of a linear integral operator whose kernel we denote by $K(z,z';\tau-\tau')$. We can then write

$$\Delta\phi(z,\tau) = \int_{-\infty}^{\tau} d\tau' \int_{-\infty}^{+\infty} dz' K(z,z';\tau-\tau')F(z',\tau'). \tag{2.12}$$

The upper limit of the $\tau'$ integration in (2.12) is bounded by $\tau$ because of causality. It is convenient, therefore, to define

$$K(z,z';\tau)=0 \quad (\tau<0), \tag{2.13}$$

and rewrite (2.12) as

$$\Delta\phi(z,\tau) = \int_{-\infty}^{+\infty} d\tau' \int_{-\infty}^{+\infty} dz' K(z,z';\tau')F(z',\tau-\tau'). \tag{2.14}$$

We now introduce Fourier time transforms of $\Delta\phi$, $F$, and $K$,

$$\Delta\hat{\phi}(z,\Omega) = \int_{-\infty}^{+\infty} d\tau e^{-i\Omega\tau}\Delta\phi(z,\tau), \tag{2.15a}$$

$$\Delta\hat{F}(z,\Omega) = \int_{-\infty}^{+\infty} d\tau e^{-i\Omega\tau}F(z,\tau), \tag{2.15b}$$

and

$$\alpha(z,z';\Omega) = \int_{-\infty}^{+\infty} d\tau e^{-i\Omega\tau}K(z,z';\tau). \tag{2.15c}$$

It then follows from Eq. (2.14) that

$$\Delta\hat{\phi}(z,\Omega) = \int_{-\infty}^{+\infty} dz'\alpha(z,z';\Omega)\hat{F}(z',\Omega). \tag{2.16}$$

In order to express $\alpha(z,z';\Omega)$ in terms of the eigenfunctions $\{\psi_i\}$ of Eq. (2.4), we first expand $F(z,\tau)$ and $\Delta\phi(z,\tau)$ in terms of $\{\psi_i\}$:

$$F(z,\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\Omega e^{i\Omega\tau}\hat{F}(z,\Omega) \tag{2.17a}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\Omega e^{i\Omega\tau}\int_i f_i(\Omega)\psi_i(z) \tag{2.17b}$$

and

$$\Delta\phi(z,\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\Omega e^{i\Omega\tau}\Delta\hat{\phi}(z,\Omega) \tag{2.18a}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\Omega e^{i\Omega\tau}\int_i \Delta\hat{\phi}_i(\Omega)\psi_i(z). \tag{2.18b}$$

Substituting Eqs. (2.17b) and (2.18b) into Eq. (2.11) and using Eq. (2.4), we obtain

$$\int_{-\infty}^{+\infty}\frac{d\Omega'}{2\pi}e^{i\Omega'\tau}\int_i \Delta\hat{\phi}_i(\Omega')(\omega_i^2-\Omega'^2+i\Omega'\Gamma)\psi_i(z)$$
$$= \int_{-\infty}^{+\infty}\frac{d\Omega'}{2\pi}e^{i\Omega'\tau}\int_i f_i(\Omega')\psi_i(z). \tag{2.19}$$

Multiplying both sides of Eq. (2.19) by $e^{-i\Omega\tau}\psi_j^*(z)$ and integrating over $z$ and $\tau$, we obtain

$$\Delta\hat{\phi}_j(\Omega) = \frac{f_j(\Omega)}{(\omega_j^2-\Omega^2)+i\Gamma\Omega}. \tag{2.20}$$

Substitution of Eq. (2.20) into Eq. (2.18b) yields

$$\Delta\phi(z,\tau) = \int_{-\infty}^{+\infty}\frac{d\Omega}{2\pi}e^{i\Omega\tau}\int_j \frac{f_j(\Omega)\psi_j(z)}{(\omega_j^2-\Omega^2)+i\Gamma\Omega}$$
$$= \int_{-\infty}^{+\infty}\frac{d\Omega}{2\pi}e^{i\Omega\tau}\int_j\int_{-\infty}^{+\infty}dz'$$
$$\times\frac{\hat{F}(z',\Omega)\psi_j^*(z')\psi_j(z)}{(\omega_j^2-\Omega^2)+i\Gamma\Omega}, \tag{2.21}$$

where the last equality is obtained using the inverse transform of Eq. (2.17b),

$$f_j(\Omega) = \int_{-\infty}^{+\infty} dz'\,\hat{F}(z',\Omega)\psi_j^*(z'). \tag{2.22}$$

The Fourier transform of Eq. (2.21) with respect to $\tau$ then gives

$$\Delta\hat{\phi}(z,\Omega) = \int_{-\infty}^{+\infty} dz' \left\{ \int_j \frac{\psi_j(z)\psi_j^*(z')}{(\omega_j^2 - \Omega^2) + i\Gamma\Omega} \right\} \hat{F}(z',\Omega). \tag{2.23}$$

By comparing Eq. (2.23) with Eq. (2.16), we finally obtain $\alpha(z,z';\Omega)$ in terms of $\{\psi_j\}$,

$$\alpha(z,z';\Omega) = \int_j \frac{\psi_j(z)\psi_j^*(z')}{(\omega_j^2 - \Omega^2) + i\Gamma\Omega}. \tag{2.24}$$

It is often of interest to consider only that portion of the response to the perturbation corresponding to the deformation of the solitary waveform. For this purpose we drop the term with $\omega_j^2 = 0$ from the $j$ sum in Eq. (2.24), since this term represents the contribution of the translation mode without deformation. We denote this modified response function by $\alpha'(z,z';\Omega)$, i.e.,

$$\alpha'(z,z';\Omega) = \int_j' \frac{\psi_j(z)\psi_j^*(z')}{(\omega_j^2 - \Omega^2) + i\Gamma\Omega}, \tag{2.25}$$

where the prime on the generalized summation symbol indicates that the term with $\omega_j^2 = 0$ is to be excluded. In general, $\psi_i$ for the continuum states is nonvanishing in the limit as $|z| \to \infty$. Therefore, $\alpha'(z,z';\Omega)$ contains a portion which measures the response of the field far from the center of the solitary wave. Thus, we must also subtract this background contribution in order to obtain the *intrinsic* response of the solitary wave. The subtraction procedure employed guarantees the convergence of several integrals encountered in obtaining the generalized susceptibility of a solitary wave,

$$\hat{\alpha}(p,k;\Omega) \equiv \int_{-\infty}^{+\infty} dz\, e^{-ipz} \alpha'(z,k;\Omega), \tag{2.26}$$

where

$$\alpha'(z,k;\Omega) = \int_{-\infty}^{+\infty} dz'\, e^{-ikz'} \alpha'(z,z';\Omega) - \lim_{|z|\to\infty} \int_{-\infty}^{+\infty} dz'\, e^{-ikz'} \alpha'(z,z';\Omega). \tag{2.27}$$

This generalized susceptibility, $\hat{\alpha}(p,k;\Omega)$, provides a measure of the $p$-Fourier component of the solitary wave response to a perturbation in the form of a monochomatic wave obeying the dispersion relation $\Omega = \Omega(k)$. If the field carries charge and the perturbing wave is an electric field, the dynamic polarizability of the solitary wave may be defined as the $p = 0$ component of $\hat{\alpha}(p,k;\Omega)$,

$$\alpha_0(k;\Omega) \equiv \hat{\alpha}(0,k;\Omega). \tag{2.28}$$

## III. GENERALIZED SUSCEPTIBILITIES OF THE SINE–GORDON SOLITON AND "$\phi^4$" KINK

In this section we calculate the generalized susceptibility for two example solitary waves, namely the sine–Gordon soliton[1] and the $\phi^4$ kink.[4–6] For both of these examples the eigenfunctions $\{\psi_i\}$ are known analytically, enabling us to obtain explicit expressions for the susceptibility.

If we choose

$$U(\phi) = 1 - \cos\phi, \tag{3.1}$$

then the sine–Gordon soliton is obtained as

$$\phi_s(z) = 4 \tan^{-1} e^z, \tag{3.2}$$

and

$$U''(\phi_s) = 1 - 2\,\mathrm{sech}^2 z. \tag{3.3}$$

The normalized eigenfunctions are[7]

$$\psi_0 = \frac{1}{\sqrt{2}} \mathrm{sech} z, \quad \omega_0^2 = 0, \tag{3.4}$$

for the translation mode, and

$$\psi_k = \frac{1}{\sqrt{2\pi}} \frac{1}{\omega_k} e^{ikz}(k + i \tanh z), \tag{3.5}$$

with

$$\omega_k^2 = 1 + k^2 \tag{3.6}$$

for the continuum states. Substituting Eqs. (3.5) and (3.6) into Eq. (2.25), we obtain

$$\alpha'(z,z';\Omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dk \frac{e^{ik(z-z')}(k + i\tanh z)(k - i\tanh z')}{(k^2+1)(k^2+1-\Omega^2+i\Gamma\Omega)}. \tag{3.7}$$

The integral in Eq. (3.7) can be evaluated by closing the contour in the upper half of the complex $k$ plane for $z > z'$ and in the lower half plane for $z < z'$. We find

$$\alpha'(z,z';\Omega) = \frac{1}{2(Q^2+1)} \Big( e^{-(z-z')}(1 + \tanh z)(1 - \tanh z') + \frac{1}{iQ} e^{iQ(z-z')}(iQ - \tanh z)(iQ + \tanh z') \Big) \quad (z > z'), \tag{3.8a}$$

and

$$\alpha'(z,z';\Omega) = \frac{1}{2(Q^2+1)} \Big( e^{-(z'-z)}(1 + \tanh z')(1 - \tanh z) + \frac{1}{iQ} e^{iQ(z'-z)}(iQ - \tanh z')(iQ + \tanh z) \Big) \quad (z < z'). \tag{3.8b}$$

In Eqs. (3.8a) and (3.8b), $Q$ is defined to be the branch of $(\Omega^2 - 1 - i\Gamma\Omega)^{1/2}$ having a positive imaginary part.

If we choose

$$U(\phi) = \tfrac{1}{2}(1 - \phi^2)^2, \tag{3.9}$$

then the $\phi^4$ kink is obtained as

$$\phi_s(z) = \tanh z \tag{3.10}$$

and

$$U''(\phi_s) = 4 - 6\,\text{sech}^2 z. \tag{3.11}$$

Unnormalized eigenfunctions $\{\psi_i\}$ for this case are given elsewhere,[11] and upon normalization we obtain

$$\psi_0(z) = \frac{\sqrt{3}}{2}\,\text{sech}^2 z, \quad \omega_0^2 = 0, \tag{3.12}$$

for the translation mode,

$$\psi_1(z) = \left(\frac{3}{2}\right)^{1/2} \text{sech}z\,\tanh z, \quad \omega_1^2 = 3, \tag{3.13}$$

for the additional bound state corresponding to localized internal vibration of the kink, and

$$\psi_k(z) = \frac{e^{ikz}}{[2\pi(1+k^2)(4+k^2)]^{1/2}} \times [3\tanh^2 z - 3ik\tanh z - (1+k^2)], \tag{3.14}$$

with $\omega_k^2 = 4 + k^2$, for the continuum states.

The susceptibility $\alpha'(z,z';\Omega)$ of the $\phi^4$ kink may be calculated from Eq. (2.25) in a manner similar to that above for the sine–Gordon soliton. We find

$$\alpha'(z,z';\Omega) = -\frac{3}{2}\frac{1}{Q'^2+1}\,\text{sech}z\,\tanh z\,\text{sech}z'\,\tanh z' - \frac{3}{2}\frac{1}{Q'^2+1}e^{-(z-z')}(\tanh^2 z + \tanh z)(\tanh^2 z' - \tanh z')$$
$$+ \frac{3}{4}\frac{1}{Q'^2+4}e^{-2(z-z')}(\tanh z + 1)^2(\tanh z' - 1)^2 + \frac{i}{2}\frac{1}{Q'(Q'^2+1)(Q'^2+4)}e^{iQ'(z-z')}$$
$$\times[3\tanh^2 z - 3iQ'\tanh z - (Q'^2+1)][3\tanh^2 z' + 3iQ'\tanh z' - (Q'^2+1)] \quad (z > z'), \tag{3.15a}$$

and

$$\alpha'(z,z';\Omega) = -\frac{3}{2}\frac{1}{Q'^2+1}\,\text{sech}z'\,\tanh z'\,\text{sech}z\,\tanh z - \frac{3}{2}\frac{1}{Q'^2+1}e^{-(z'-z)}(\tanh^2 z' + \tanh z')(\tanh^2 z - \tanh z)$$
$$+ \frac{3}{4}\frac{1}{Q'^2+4}e^{-2(z'-z)}(\tanh z' + 1)^2(\tanh z - 1)^2 + \frac{i}{2}\frac{1}{Q'(Q'^2+1)(Q'^2+4)}e^{iQ'(z'-z)}$$
$$\times[3\tanh^2 z' - 3iQ'\tanh z' - (Q'^2+1)][3\tanh^2 z + 3iQ'\tanh z - (Q'^2+1)] \quad (z < z'). \tag{3.15b}$$

In Eqs. (3.15a) and (3.15b), $Q'$ is defined to be the branch of $(\Omega^2 - i\Gamma\Omega)^{1/2}$ having a positive imaginary part.

Except for the contribution of the first term on the right-hand side of Eqs. (3.15), the integrations over $z'$ appearing in the expression (2.27) for $\alpha'(z,k;\Omega)$ are of the following form for both the sine–Gordon soliton and $\phi^4$ kink:

$$\int_{-\infty}^{z} dz'\, e^{i\lambda_+(z-z')}(A\sigma^2 + B\sigma + C)\,(A\tanh^2 z' - B\tanh z' + C) + \int_z^{\infty} dz'\, e^{-i\lambda_-(z-z')}(A\sigma^2 - B\sigma + C)$$
$$\times\,(A\tanh^2 z' + B\tanh z' + C),$$

where $\sigma = \tanh z$ and $\lambda_+, \lambda_-, A, B, C$ are all complex constants. These integrals are evaluated in terms of the hypergeometric function,[12] $F$, and the final integrals over $z$ in Eq. (2.26) may be expressed in terms of the generalized hypergeometric function,[12] ${}_3F_2$. After some tedious algebra, we find the following result for the sine–Gordon and $\phi^4$ cases,

$$\hat{\alpha}(p,k;\Omega) = \frac{3}{2}\Delta\frac{\pi^2 kp}{1+Q'^2}\,\text{sech}\frac{\pi}{2}p\,\text{sech}\frac{\pi}{2}k + \sum_{l=1}^{M} I_l(D_l,a_0(l),a_1(l),a_2(l)), \tag{3.16}$$

where $\Delta = 0$, $M = 2$ for the sine–Gordon case and $\Delta = 1$, $M = 3$ for the $\phi^4$ case. The quantities $I_l$ are given by

$$I_l = \frac{D_l}{2i(k+p)}\sum_{m=0}^{2}\sum_{n=0}^{2} a_m(l)a_n(l)\sum_{r=\pm1}\frac{2ir}{\lambda_l^{(r)}}\Bigg[m\frac{\Gamma(m-(i/2)r(p+k))}{\Gamma(m+n+1)}\Gamma(n+1+(i/2)r(p+k))$$
$$\times{}_3F_2\left(n,1,m-\frac{i}{2}r(p+k);-\frac{i}{2}\lambda_l^{(r)}+1;m+n+1\right) - n\frac{\Gamma(m+1-(i/2)r(p+k))}{\Gamma(m+n+1)}\Gamma\left(n+\frac{i}{2}r(p+k)\right)$$
$$\times{}_3F_2\left(n,1,m+1-\frac{i}{2}r(p+k);-\frac{i}{2}\lambda_l^{(r)}+1;m+n+1\right) + \frac{n}{1-(i/2)\lambda_l^{(r)}}$$
$$\times\frac{\Gamma(m+1-(i/2)r(p+k))\Gamma(n+1+(i/2)r(p+k))}{\Gamma(m+n+2)}{}_3F_2\left(n+1,2,m+1-\frac{i}{2}r(p+k);2-\frac{i}{2}\lambda_l^{(r)};m+n+2\right)\Bigg], \tag{3.17}$$

where, for the sine–Gordon case,

$$D_1 = \frac{1}{2(1+Q^2)}, \quad D_2 = \frac{1}{2iQ(1+Q^2)}, \tag{3.18a}$$

$$a_0(1) = a_2(1) = a_2(2) = 0, \quad a_1(1) = -a_1(2) = 2,$$
$$a_0(2) = 1 + iQ, \tag{3.18b}$$

$$\lambda_1^{(\pm 1)} = i \pm k, \quad \lambda_2^{(\pm 1)} = Q \pm k, \tag{3.18c}$$

and for the $\phi^4$ case,

$$D_1 = -\frac{3}{2(1+Q'^2)}, \quad D_2 = \frac{3}{4(4+Q'^2)},$$
$$D_3 = \frac{i}{2Q'(1+Q'^2)(4+Q'^2)}, \tag{3.19a}$$

$$a_0(1) = a_0(2) = a_1(2) = a_2(2) = 0,$$
$$a_1(1) = -2, \quad a_2(1) = 4, \quad a_1(3) = -6(2+iQ'), \tag{3.19b}$$
$$a_2(3) = 12, \quad a_0(3) = 2 - Q'^2 + 3iQ',$$

$$\lambda_1^{(\pm)} = i \pm k, \quad \lambda_2^{(\pm)} = 2i \pm k, \quad \lambda_3^{(\pm)} = Q' \pm k. \tag{3.19c}$$

Equations (3.16)–(3.19) embody our results for the generalized susceptibility for all wave vectors $p$ and $k$. It should be noted that because of the factor $(k+p)^{-1}$ appearing in $I_l$, the limiting form of $I_l$ must be used when $k + p = 0$.

For the sine–Gordon soliton, the long-wavelength ($k = 0$) dynamic polarizability, $\alpha_0(0;\Omega)$, defined by Eq. (2.28), is found to agree with the quantity $-\alpha(\Omega)$ determined previously in Ref. 8, and we do not repeat the expression here. Instead we focus in the next section on the nature of the dynamic polarizability of the $\phi^4$ kink.

## IV. DYNAMIC POLARIZABILITY OF THE $\phi^4$ KINK

The general expression (3.16) for $\hat{\alpha}(p,k;\Omega)$ simplifies a great deal when we take the long-wavelength limit ($p \to 0, k \to 0$) to obtain the $k = 0$ dynamic polarizability $\alpha_0(k=0;\Omega) = \hat{\alpha}(0,0;\Omega)$. For the $\phi^4$ kink we obtain

$$\alpha_0(0;\Omega) = 3(Q'^2+4)^{-1} - [2Q'^2(1+Q'^2)(4+Q'^2)]^{-1}$$
$$\times \Bigg[ \tfrac{1}{6}a_2^2(3) + a_1(3)a_2(3) + a_1^2(3) - 2a_0(3)a_2(3)$$
$$+ \sum_{n=1}^{\infty} \frac{n!}{(1-\frac{1}{2}iQ')_n} \bigg( \frac{a_0(3)a_2(3) + a_0(3)a_1(3)}{n}$$
$$+ \frac{a_1^2(3)}{n+1} + \frac{2a_1(3)a_2(3) - a_2^2(3)}{n+2} + \frac{2a_2(3)}{n+3} \bigg) \Bigg], \tag{4.1}$$

where $(z)_n$ denotes the Pochammer symbol,

$$(z)_n = z(z+1)\cdots(z+n-1). \tag{4.2}$$

The $k = 0$ polarizability corresponds to the response of $\phi$ to a spatially uniform external oscillating field. As we can see from Eq. (3.16), the internal vibration mode of the $\phi^4$ kink does not respond to a spatially uniform disturbance [the first term in (3.16) vanishes when $k = 0$]. This is due to the fact that the eigenfunction [$\psi_1$, see Eq. (3.13)] for this mode is an odd function of $z$.

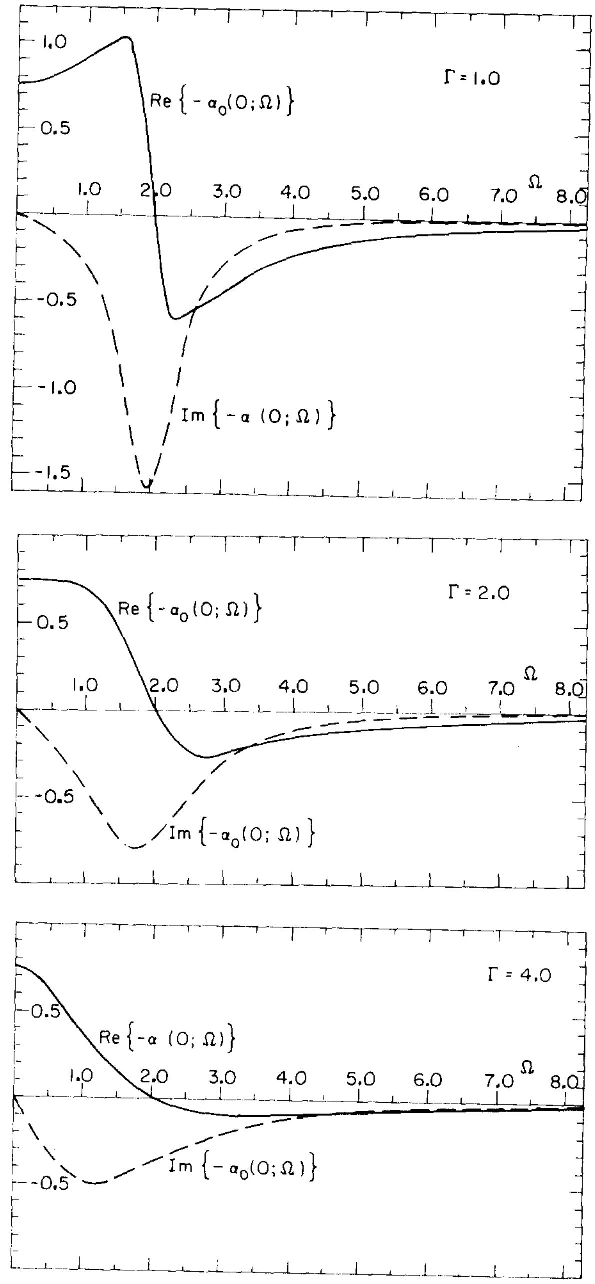


FIG. 1. Dynamic polarizability of the $\phi^4$ kink vs. frequency. The real and imaginary parts of $-\alpha_0(0;\Omega)$ are shown as solid and dashed curves, respectively, for three example values (1.0,2.0,4.0) of the damping constant $\Gamma$. There is only one absorption peak (dip in $-\alpha$) near the fundamental harmonic oscillation frequency, $\Omega = 2$, since the internal oscillation at $\Omega = \sqrt{3}$ is not excited by a uniform field.

Equation (4.1) may be rewritten in terms of $Q'$ as

$$\alpha_0(0;\Omega) = \frac{3}{Q'^2(1+Q'^2)(4+Q'^2)} \Bigg[ 4 + 3Q'^2 + Q'^4$$
$$- \sum_{n=1}^{\infty} \frac{n!}{(1-\frac{1}{2}iQ')_n} \bigg( \frac{3Q'^2 - iQ'(2-Q'^2)}{n}$$

$$+\frac{6(4-Q'^2+4iQ')}{n+1}-\frac{24(3+iQ')}{n+2}$$

$$+\frac{48}{n+3}\Big)\Big]. \tag{4.3}$$

For small $\Omega$, we find

$$-\alpha_0(0;\Omega)\cong\tfrac{3}{4}-\tfrac{3}{16}i\Gamma\Omega+\mathscr{O}(\Omega^2). \tag{4.4}$$

In Fig. 1 we have plotted our numerical results for $-\alpha_0(0;\Omega)$ obtained from Eq. (4.3) for three representative values of the damping constant $\Gamma$. The real part of $-\alpha_0(0;\Omega)$ starts at 0.75 (independent of $\Gamma$) at $\Omega=0$ and for $\Gamma<\Gamma_0(\lesssim 2.0)$ it rises to a maximum, then falls to a minimum after passing through zero near $\Omega=2$, and finally rises to its asymptotic value of zero at $\Omega=\infty$. If $\Gamma>\Gamma_0$, $\mathrm{Re}[-\alpha_0(0;\Omega)]$ decreases from 0.75 immediately as $\Omega$ increases from zero. The imaginary part of $-\alpha_0(0;\Omega)$ is always zero at $\Omega=0$, decreases to a minimum and then rises to its asymptotic value of zero. The form of this response is very reminiscent of that for a harmonic oscillator and is quite similar to the behavior of the response of the sine–Gordon soliton as described in Ref. 8.

## V. SUMMARY

In this paper we have defined a generalized susceptibility for a solitary wave solution of the nonlinear Klein–Gordon equation, and we have derived its expression in terms of the eigenfunctions of the Schrödinger-like equation which arises in a stability analysis of the solitary wave. Two examples have been worked out explicitly; the sine–Gordon soliton polarizability reduces to that obtained in Ref. 8 in the long-wavelength limit; for the $\phi^4$ kink we have numerically evaluated our general expression to obtain the dynamic polarizability for example values of the damping constant $\Gamma$. We find that if the perturbing field is uniform in space, the internal vibration mode of the kink is not excited, due to its odd parity. However, if the disturbance is inhomogeneous then there is a nonvanishing contribution of this mode to the generalized susceptibility $\hat{\alpha}(p,k;\Omega)$ when $p\neq 0$. The maximum response of the vibration mode can be obtained by applying a standing wave with its node coincident with the center ($\phi=0$) of the $\phi^4$ kink. If we view the $\phi^4$ kinks as models[4] of domain walls in ferroelectrics, the $\mathrm{Im}[\alpha_0(k;\Omega(k))]$ provides a measure of the absorption of energy from a monochromatic electric field with dispersion $\Omega=\Omega(k)$.

After this work was completed, we became aware of recent work by Theodorakopoulos and coworkers[13] in which they have also studied the response of the $\phi^4$ kink to an oscillating field. These authors also employ the complete orthonormal set of eigenfunctions used here. These functions have also been used recently[14,15] to study the interaction of the $\phi^4$ kink with the linear oscillations.

## ACKNOWLEDGMENT

We are grateful to Dr. Theodorakopoulos for communicating his results prior to publication.

[1]A.C. Scott, F.Y.F. Chu, and D.W. Mclaughlin, Proc. IEEE **61**, 1443 (1973).
[2]See, for example, *Solitons and Condensed Matter Physics*, Springer Series in Solid State Sciences, Vol. 8, edited by A.R. Bishop and T. Schneider (Springer-Verlag, Berlin, 1978).
[3]M.J. Rice, A.R. Bishop, J.A. Krumhansl, and S.E. Trullinger, Phys. Rev. Lett. **36**, 432 (1976).
[4]J.A. Krumhansl and J.R. Schrieffer, Phys. Rev. B **11**, 3535 (1975).
[5]R. Rajaraman, Phys. Rep. C **21**, 229 (1975).
[6]R. Jackiw, Rev. Mod. Phys. **49**, 681 (1977).
[7]M.B. Fogel, S.E. Trullinger, A.R. Bishop, and J.A. Krumhansl, Phys. Rev. Lett. **36**, 1411 (1976); Phys. Rev. B **15**, 1578 (1977).
[8]M.B. Fogel, S.E. Trullinger, and A.R. Bishop, Phys. Lett. A **59**, 81 (1976).
[9]See, for example, L.D. Landau and E.M. Lifschitz, *Statistical Physics* (Addison–Wesley, Reading, Massachusetts, 1958), p. 391.
[10]If $F(z,\tau)$ contains a term which is constant in time, then $\Delta\phi$ is small only after the removal of secularities due to the translation of the solitary wave. See Ref. 7.
[11]J. Goldstone and R. Jackiw, Phys. Rev. D **11**, 1486 (1975).
[12]I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965).
[13]N. Theodorakopoulos, S. Hanna, and R. Klein, in *Solitons and Condensed Matter Physics*, Springer Series in Solid State Sciences, Vol. 8, edited by A.R. Bishop and T. Schneider (Springer-Verlag, Berlin, 1978).
[14]W. Hasenfratz and R. Klein, Physica A **89**, 191 (1977).
[15]Y. Wada and J.R. Schrieffer, Phys. Rev. B **18**, 3897 (1978).

# Resolvent integration techniques for generalized transport equations[a)]

R. L. Bowden

*Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia*

William Greenberg

*Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia*

P. F. Zweifel

*Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia*
(Received 24 July, 1978)

A generalized class of "transport type" equations is studied, including most of the known exactly solvable models; in particular, the transport operator $K$ is a scalar type spectral operator. A spectral resolution for $K$ is obtained by contour integration techniques applied to bounded functions of $K$. Explicit formulas are developed for the solutions of full and half range problems. The theory is applied to anisotropic neutron transport, yielding results which are proved to be equivalent to those of Mika.

## I. INTRODUCTION

In 1973 Larsen and Habetler[1] introduced a technique for solving the one-speed neutron transport equation based on contour integration of the resolvent of the transport operator about its spectrum. (Although the transport operator, $K^{-1}$ in their notation, is unbounded, its bounded inverse $K$ can be treated by resolvent integration. This leads to an "eigenfunction expansion" in the sense of Titchmarsh, for $K$. To return to the neutron transport equation which involves $K^{-1}$ rather than $K$, it is necessary to develop a functional calculus for $K$, after the manner von Neumann introduced into quantum mechanics[2]; this was accomplished in a later paper.[3])

The Larsen–Habetler method has been extended in the past two or three years to a number of more general forms of the neutron transport equation. Instead of listing these references here, the reader is referred to a recent comprehensive review article.[4] One special case should be noted, however, namely the so-called "critical" situation (which in one-speed theory corresponds to the situation $c = 1$). The orthodox resolvent integration technique cannot be applied in such a case, because $K^{-1}$ is not invertible on its range. A modified and somewhat cumbersome technique can be used, however.[5] The idea is to restrict $K^{-1}$ to a domain on which it is invertible, and proceed after the manner of Ref. 1, later extending results to the whole space.

Since the (linearized) equations describing electron oscillations in plasma and the kinetics of rarefied gases are similar in form to the neutron transport equation, it seems that resolvent integration techniques might be valuable in solving these equations also. However, they both pose difficult problems. For example, the unbounded operator describing gas kinetics is not invertible, and even if it is restricted to a domain in which it is invertible, its inverse is still unbounded. Thus it is not possible to integrate around the spectrum. The linearized Vlasov equation describing plasma oscillations is also unbounded, and although it is in general invertible, its inverse also is unbounded, so again straightforward resolvent integration techniques fail. (The relevant equations for these two physical problems are discussed in Ref. 6, Chap. 10.)

Very recently, a method suggested by Larsen[7] has been successfully applied to the Vlasov equation,[8] the gas kinetics equation (for a BGK model)[9,10] and also to conservative neutron transport.[11] For the first two cases this method was crucial to the solution; in the neutron transport case it merely simplified the previous somewhat cumbersome method described above. The basic idea was to transform the transport operator $K \to S = (K - \xi I)^{-1}$, where $\xi$ is in the resolvent set of $K$. Then since $S$ is a bounded operator with "thin" spectrum, the orthodox contour integration method can be applied to $S$ to develop an eigenfunction expansion. Then a functional calculus is obtained along the lines of Ref. 3, so that the equation involving $K = S^{-1} + \xi I$ can be solved. (In more mathematical terms, a "constructive existence theorem" can be proved.)

In the present paper, we extend this technique, as developed in Refs. 8–11 to a general class of transport type equations of the form

$$\frac{\partial}{\partial x}\psi(x,\mu) = -(K\psi)(x,\mu) + q(x,\mu), \tag{1a}$$

where

$$(Kf)(\mu) = k(\mu)f(\mu) + \sum_{n=1}^{N} g_n(\mu)\int_A J_n(s)f(s)\,ds. \tag{1b}$$

Here $f \in \mathscr{L}^p(A,\sigma) = \mathscr{B}$, $A \subset \mathbb{R}$ is a directed Liaponov contour and $k(\mu)$ is a real valued, $\sigma$-measurable function on $A$. The functions $k$, $g_n$, and $J_n$ are assumed to obey certain continuity and differentiability conditions which we enumerate later. In order to place Eq. (1) in perspective, we observe that the three equations discussed in Refs. 8–11 correspond to the following values of $k$, $g_n$, and $J_n$:

[a)]Research supported by the National Science Foundation under Grant ENG 75-15882.

  

1. One-speed conservative neutron transport (Ref. 11)

$$k(\mu) = \frac{1}{\mu},$$
$$g_n(\mu) = -\frac{1}{\mu},$$
$$J_n(\mu) = \tfrac{1}{2},$$
$$A = [-1,1],$$
$$N = 1.$$

2. BGK model for gas kinetics (Ref. 9),

$$k(\mu) = \frac{1}{\mu},$$
$$g_n(\mu) = -\frac{1}{\mu},$$
$$J_n(\mu) = \frac{1}{\sqrt{\pi}} e^{-\mu^2},$$
$$A = \mathbb{R},$$
$$N = 1.$$

3. Linearized Vlasov equation (Ref. 8),

$$k(\mu) = \mu,$$
$$g_n(\mu) = \eta(\mu),$$
$$J_n(\mu) = 1,$$
$$A = \mathbb{R},$$
$$N = 1.$$

[$\eta(\mu)$ is proportional to the derivative of the equilibrium electron distribution.]

4. One-speed neutron transport, anisotropic scattering (Ref. 6, p. 87),

$$k(\mu) = \frac{1}{\mu},$$
$$g_n(\mu) = \frac{-(2n-1)}{2}\frac{c}{\mu} f_{n-1} P_{n-1}(\mu),$$
$$J_n(\mu) = P_{n-1}(\mu),$$
$$A = [-1,1],$$
$$N = N_0.$$

(The $f_n$ are the Legendre moments of the scattering kernel and $P_n$ are Legendre polynomials.)

Other equations of transport type can be expected to occur in various areas, gas dynamics, radiative, and electron transport, etc., which basically involve the linearized Boltzmann equation. The solution to such equations can then be read off from our results. Although the smoothness restrictions which we place on coefficients in our transport-type equation are merely sufficient conditions, we believe that they are sufficiently general to encompass virtually all cases which may arise from physical application.

The plane of our paper is as follows. In Sec. II we compute $S = (K-\xi)^{-1}$ and the resolvent of $S$, $(zI - S)^{-1}$. We also obtain the spectrum of $S$. Then in Sec. III we perform the integration about the continuous spectrum of $S$, and in Sec. IV integrate about the point spectrum. These two results together give a "full-range" eigenfunction expansion for $S$. A similar "half-range" expansion is obtained in Sec. V. Then, in order to translate these into eigenfunction expansions for $K$, which are needed to solve Eq. 11, we first need to extend the results to Banach space. The analysis of Secs. III–V has been restricted to a dense subspace of Hölder continuous functions (since it was necessary to evaluate boundary values of Cauchy integrals). Once the extension is carried out in Sec. VI, a functional calculus for $S$ can be developed; this is done in VII and so, as was explained earlier, an eigenfunction expansion for $K = S^{-1} + \xi I$ is thereby obtained. Sec. VIII presents some applications to boundary value problems.

## II. RESOLVENTS

We wish to consider the integrodifferential Eq. (1). The operator $K$ [Eq. (1b)] may be written in the obvious notation

$$Kf = kf + \mathbf{g}\cdot\int_A \mathbf{J}(s) f(s)\, ds, \tag{2}$$

and where there is no confusion, we shall abbreviate $\int_A \mathbf{J}(f) f(s)\, ds = \mathbf{J}(s)$. $K$ is not assumed to be bounded; in any case we shall for the most part restrict its domain to $D(K) = \{f \mid Kf \in \mathscr{B},\ f \text{ Hölder continuous on compacts}\}$. Finally, we shall let $(\mathbf{J}\otimes\mathbf{g})_{mn}(\mu) = J_m(\mu) g_n(\mu)$, and shall write $\mathscr{B}$ for the product of $N$ copies of the Banach space $\mathscr{B}$. By a solution of Eq. (1), we demand a continuously differential function $\psi:\mathbb{R}\to\mathscr{B}$ satisfying specified boundary conditions (to be discussed later), where the inhomogeneous source term $q:\mathbb{R}\to\mathscr{B}$ is assumed to satisfy a uniform Hölder condition (on every compact subset of $A$).

*Lemma* 1: If there exists $\xi\in\mathbb{C}/\mathbb{R}$ such that the following are satisfied:

(i) $T_\xi = \int_A \frac{\mathbf{J}\otimes\mathbf{g}(s)}{k(s)-\xi}\, ds + I$ invertible on $\mathbb{C}^N$,

(ii) $\frac{1}{k-\xi}\mathbf{g}\in\mathscr{B}$,

(iii) $\mathscr{P}_\xi : f \to \mathbf{J}\left(\frac{1}{k-s} f\right) \in \mathscr{B}^*$,

then $S_\xi = (K - \xi I)^{-1}$ exists as a bounded operator on $\mathscr{B}$.

*Proof*: Letting $(K-\xi I)f = (k-\xi)f + \mathbf{g}\cdot\mathbf{J}(f) = h$, we obtain

$$f = \frac{1}{k-\xi} h - \frac{\mathbf{g}\cdot\mathbf{J}(f)}{k-\xi}$$

for $f$ in the (dense) domain of $K$. This is valid in $\mathscr{B}$ by (ii) and the fact that $[1/(k-\xi)] : f \to [1/(k-\xi)] f \in L(\mathscr{B})$, the bounded operators on $\mathscr{B}$, since $\operatorname{ess\,sup}|1/(k-\xi)| \leqslant |\operatorname{Im}\xi|^{-1}$. Then, computing $\mathbf{J}(f)$ and utilizing (iii), after some rearranging we have

$$\sum_{n=1}^{N} J_n(f) \int_A \frac{J_m(s) g_n(s)\, ds}{k(s)-\xi} + \delta_{mn} = J_m\left(\frac{1}{k-\xi} h\right),$$

which may be written

$$\int_A \frac{\mathbf{J}\otimes\mathbf{g}(s)\, ds}{k(s)-\xi} + I\,\mathbf{J}(f) = \mathbf{J}\left(\frac{1}{k-\xi} h\right).$$

Inverting

$$\Lambda(\xi) = \int_A \frac{\mathbf{J} \otimes \mathbf{g}(s)\, ds}{k(s) - \xi} + I,$$

by virtue of (i), we obtain

$$(K - \xi I)^{-1} f = \left[ \frac{1}{k - \xi} \right] f - \frac{1}{k - \xi} \mathbf{g} \cdot \Lambda^{-1}(\xi) \mathbf{J}\left( \frac{1}{k - \xi} f \right). \tag{3}$$

We have thus proved as well,

*Lemma* 2: With $\xi$ defined as in the previous lemma, and

$$\Lambda(z) = \int_A \frac{\mathbf{J}(s) \otimes \mathbf{g}(s)\, ds}{k(s) - z} + I, \tag{4}$$

then

$$S_\xi f = (K - \xi I)^{-1} f = \left[ \frac{1}{k - \xi} \right] f - \frac{1}{k - \xi} \mathbf{g} \cdot \Lambda^{-1}(\xi) \times \mathbf{J}\left( \frac{1}{k - \xi} f \right). \tag{5}$$

*Lemma* 3: The resolvent of $S_\xi$ is given by

$$(S_\xi - zI)^{-1} f = \left[ \frac{k - \xi}{1 - z(k - \xi)} \right] f + \frac{1}{1 - z(k - \xi)} \times \mathbf{g} \cdot \Lambda^{-1}\left( \xi + \frac{1}{z} \right) \mathbf{J}\left( \frac{1}{1 - z(k - \xi)} f \right). \tag{6}$$

*Proof*: We compute

$$(S_\xi - zI) f = h,$$

as in Lemma 1, obtaining

$$f = \frac{k - \xi}{1 - z(k - \xi)} h + \frac{1}{1 - z(k - \xi)} \mathbf{g} \cdot \Lambda^{-1}(\xi) \times \mathbf{J}\left( \frac{1}{k - \xi} f \right);$$

thus,

$$\mathbf{J}\left( \frac{1}{k - \xi} f \right) = \mathbf{J}\left( \frac{1}{1 - z(k - \xi)} h \right) + \sum_{i=1}^{N} \mathbf{J}\left( \frac{1}{(k - \xi)(1 - z(k - \xi))} g_i \right) (\Lambda^{-1}(\xi) \mathbf{J})_i \left( \frac{1}{k - \xi} f \right)$$
$$= \left( I - \int_A \frac{\mathbf{J} \otimes \mathbf{g}(s)\, ds}{(k(s) - \xi)(1 - z(k(s) - \xi))} \Lambda^{-1}(\xi) \right)^{-1} \times \mathbf{J}\left( \frac{1}{1 - z(k - \xi)} h \right).$$

We may then rewrite the expression for $f$ as

$$f = \frac{k - \xi}{1 - z(k - \xi)} h + \frac{1}{1 - z(k - \xi)} \times \mathbf{g} \cdot \left( \Lambda(\xi) - \int_A \frac{\mathbf{J} \otimes \mathbf{g}(s)\, ds}{(k(s) - \xi)(1 - z(k(s) - \xi))} \right)^{-1} \times \mathbf{J}\left( \frac{h}{k - z(k - \xi)} \right).$$

Let

$$\Lambda_\xi(z) = \Lambda(\xi) - \int_A \frac{\mathbf{J} \otimes \mathbf{g}(s)\, ds}{(k(s) - \xi)(1 - z(k(s) - \xi))}.$$

Then we may write

$$\Lambda_\xi(z) = I + \int_A \mathbf{J} \otimes \mathbf{g}(s) \left( \frac{1}{k(s) - \xi} - \frac{1}{(k(s) - \xi)(1 - z(k(s) - \xi))} \right) ds,$$

and utilizing

$$\frac{1}{k - \xi} - \frac{1}{(k - \xi)(1 - z(k - \xi))} = \frac{1}{k - \xi - (1/z)},$$

obtain

$$\Lambda_\xi(z) = \Lambda(\xi + 1/z),$$

which completes the proof.

*Lemma* 4: Let $\Omega_\xi(z) = \det \Lambda(\xi + 1/z)$ and $N_p$ be the set of zeroes of $\Omega_\xi$. Let $Q = \{ z = 1/(\omega - \xi) \in \mathbb{C} \mid \omega \in \operatorname{Ran} k \}$. Then

$$\sigma_p(S_\xi) = N_p, \quad \sigma_c(S_\xi) = Q.$$

*Proof*: This is an immediate consequence of Weyl's theorem and Eqs. (5) and (6).

## III. CONTINUOUS SPECTRUM

*Definition*: We shall call the triple $\{k, g, J\}$ of transport type if $k$ is one–one and differentiable and if for $\beta = 1$ or else for $\beta = -1$, $s \to k(s)^{-\beta}$ is continuous and each of the functions $t \to k^{-1}(t^\beta)$, $s \to J(s)/k(s)^{-1/2(1-\beta)} k'(s)$, and $s \to J(s) \otimes g(s)/k(s)^{-1/2(1-\beta)} k'(s)$ is Hölder continuous on compact subsets of $A$.

Here and throughout we write $k(s)^{-1}$ for $1/k(s)$ and $k^{-1}(s)$ for the inverse function evaluated at $s$, e.g., $k^{-1}(s) = s$ if $k(s) = s$.

*Lemma* 5: Assume $\{k, g, J\}$ is of transport type, and for $f \in \mathscr{B}$ Hölder continuous with compact support, define

$$M_f(z) = \Lambda^{-1}(\xi + 1/z) \mathbf{J}\left( \frac{1}{1 - z(k - \xi)} f \right). \tag{7a}$$

Then the boundary values $M^\pm$ and $\Lambda^\pm$ are given by

$$M_f^\pm\left( \frac{1}{k(\omega) - \xi} \right) = \Lambda_\xi^{-1}(k(\omega)) \pm \int_A \frac{\mathbf{J}(s) f(s)\, ds}{k(\omega) - k(s)} (k(\omega) - \xi) \pm \frac{i\pi \mathbf{J}(\omega) f(\omega)(k(\omega) - \xi)}{k'(\omega)}, \tag{7b}$$

$$\Lambda_\xi^{\pm}(k(\omega)) = I + \int_\Lambda \frac{\mathbf{J}\otimes\mathbf{g}(s)\,ds}{k(s)-k(\omega)} \pm \frac{i\pi\mathbf{J}\otimes\mathbf{g}(\omega)}{k'(\omega)}$$

$$= \lambda(k(\omega)) \pm \frac{i\pi\mathbf{J}\otimes\mathbf{g}(\omega)}{k'(\omega)}, \tag{7c}$$

where + and − refer to nontangential approach of $z$ to the contour from the right and left, respectively.

*Proof*: With the substitution $t = k(s)^\beta$, the integral to be evaluated for $\Lambda^{\pm}M^{\pm}$ is

$$-\int \frac{\mathbf{J}(k^{-1}(t^\beta))f(k^{-1}(t^\beta))}{t^{1/2(1-\beta)}(t-[(1+z\xi)/z]^\beta)k'(k^{-1}(t))}\,dt.$$

The continuity of $s \to k(s)^{-\beta}$ assures that the integration may be restricted to a compact set. Then the Plemelj formulas may be applied by virtue of the required Hölder continuity, and (7b) follows. The computation of $\Lambda^{\pm}$ is similar.

Using Lemma 5, we may compute

$$\Lambda_\xi^{+}(k(\omega))\mathbf{M}^{+}\left(\frac{1}{k(\omega)-\xi}\right) - \Lambda_\xi^{-}(k(\omega))\mathbf{M}^{-}\left(\frac{1}{k(\omega)-\xi}\right)$$

$$= \frac{-2\pi i\mathbf{J}(\omega)f(\omega)(k(\omega)-\xi)}{k'(\omega)} \tag{8a}$$

$$= \lambda(k(\omega))\left\{\mathbf{M}^{+}\left(\frac{1}{k(\omega)-\xi}\right) - \mathbf{M}^{-}\left(\frac{1}{k(\omega)-\xi}\right)\right\} + \pi i\frac{\mathbf{J}\otimes\mathbf{g}(\omega)}{k'(\omega)}\left[\mathbf{M}^{+}\left(\frac{1}{k(\omega)-\xi}\right) + \mathbf{M}^{-}\left(\frac{1}{k(\omega)-\xi}\right)\right] \tag{8b}$$

Taking the inner product of Eqs. (8) with $J(\omega)$, denoting

$$\mathbf{J}(\omega)\cdot\mathbf{J}(\omega) = J^2(\omega),$$

and using the identity $\mathbf{J}\cdot\mathbf{J}\otimes\mathbf{g}\mathbf{M} = J^2\mathbf{g}\cdot\mathbf{M}$, we find

$$f(\omega) + \frac{1}{k(\omega)-\xi}\mathbf{g}(\omega)\cdot\tfrac{1}{2}\left[\mathbf{M}^{+}\left(\frac{1}{k(\omega)-\xi}\right) + \mathbf{M}^{-}\left(\frac{1}{k(\omega)-\xi}\right)\right]$$

$$= -\frac{1}{2\pi i}\frac{k'(\omega)}{J^2(\omega)}\frac{1}{k(\omega)-\xi}\mathbf{J}(\omega)\cdot\lambda(k(\omega))\left[\mathbf{M}^{+}\left(\frac{1}{k(\omega)-\xi}\right) - \mathbf{M}^{-}\left(\frac{1}{k(\omega)-\xi}\right)\right]. \tag{9}$$

Further, we may compute the difference in boundary values of $M$ from Lemma 5.

Utilizing

$$\Lambda^{+}(z)^{-1} \pm \Lambda^{-}(z)^{-1} = \Lambda^{-}(z)^{-1}\{\Lambda^{-}(z) \pm \Lambda^{+}(z)\}\Lambda^{+}(z)^{-1},$$

we obtain

$$M^{+}\left(\frac{1}{k(\omega)-\xi}\right) - M^{-}\left(\frac{1}{k(\omega)-\xi}\right)$$

$$= -\Lambda_\xi^{-}(k(\omega))^{-1}\mathbf{J}\otimes\mathbf{g}(\omega)\Lambda_\xi^{+}(k(\omega))^{-1}2\pi i\frac{(k(\omega)-\xi)}{k'(\omega)}\int_\Lambda\frac{\mathbf{J}(s)f(s)\,ds}{k(\omega)-k(s)} - \Lambda_\xi^{-}(k(\omega))^{-1}\lambda(k(\omega))\Lambda_\xi^{+}(k(\omega))^{-1}$$

$$\times 2\pi i\frac{(k(\omega)-\xi)}{k'(\omega)}\mathbf{J}(\omega)f(\omega). \tag{10}$$

We will now integrate the resolvent of $S_\xi$ on a contour $\Gamma$ about its spectrum,

$$f(\omega) = \frac{1}{2\pi i}\oint_{\Gamma(Q)}(zI-S_\xi)^{-1}f(\omega)\,dz + \frac{1}{2\pi i}\oint_{\Gamma(N_p)}(zI-S_\xi)^{-1}f(\omega)\,dz.$$

Denoting the two integrals by $f_1(\omega), f_2(\omega)$, respectively, we have

$$f_1(\omega) = f(\omega) + \frac{1}{2\pi i}\mathbf{g}(\omega)\cdot\oint_{\Gamma(Q)}\frac{1}{z(k(\omega)-\xi)-1}\mathbf{M}(z)\,dz$$

$$= f(\omega) + \frac{1}{2\pi i}\mathbf{g}(\omega)\cdot\int_R\frac{1}{(x-\xi)(x-k(\omega))}\left[\mathbf{M}^{+}\left(\frac{1}{x-\xi}\right) - \mathbf{M}^{-}\left(\frac{1}{x-\xi}\right)\right]dx + \frac{1}{2(k(\omega)-\xi)}$$

$$\times\mathbf{g}(\omega)\cdot\left[\mathbf{M}^{+}\left(\frac{1}{(k(\omega)-\xi)}\right) + \mathbf{M}^{-}\left(\frac{1}{(k(\omega)-\xi)}\right)\right],$$

where $R = \operatorname{Ran}k$ and we have utilized the analyticity and Hölder continuity of $\mathbf{M}(z)$. Using Eq. (9), this becomes

$$f_1(\omega) = \frac{1}{2\pi i}\,\mathbf{g}(\omega)\cdot\int_R \frac{1}{(x-\xi)(x-k(\omega))}\left[\mathbf{M}^+\left(\frac{1}{x-\xi}\right) - \mathbf{M}^+\left(\frac{1}{x-\xi}\right)\right]dx - \frac{1}{2\pi i}\frac{k'(\omega)}{J^2(\omega)}\frac{1}{(k(\omega)-\xi)}$$

$$\times \mathbf{J}(\omega)\cdot\lambda(k(\omega))\left[\mathbf{M}^+\left(\frac{1}{(k(\omega)-\xi)}\right) - \mathbf{M}^-\left(\frac{1}{(k(\omega)-\xi)}\right)\right].$$

Let us define

$$\mathbf{\Phi}_x(\omega) = P\frac{\mathbf{g}(\omega)}{x-k(\omega)} - \frac{k'(\omega)}{J^2(\omega)}\delta(x-k(\omega))\cdot\lambda^t(x)\mathbf{J}(\omega), \quad x\in R,\ \omega\in A \tag{11a}$$

$$\mathbf{A}(x) = -\Lambda^-(x)^{-1}\frac{\mathbf{J}\otimes\mathbf{g}(k^{-1}(x))}{k'(k^{-1}(x))}\Lambda^+(x)^{-1}\int_A\frac{\mathbf{J}(x)f(s)\,ds}{x-k(s)} - \Lambda^-(x)^{-1}\lambda(x)\cdot\Lambda^+(x)^{-1}\mathbf{J}(k^{-1}(x))f(k^{-1}(x))/k'(k^{-1}(x)), \tag{11b}$$

where $t$ indicates the transpose. We have proved

*Theorem* 6: Suppose $K$ satisfies the hypothesis of Lemma 5, and $f\in\mathscr{B}$ is Hölder continuous with compact support. Then

$$f_1(\omega) = \int_R \mathbf{\Phi}_x(\omega)\cdot\mathbf{A}(x)\,dx, \tag{11c}$$

where $\Phi_x$ and $\mathbf{A}$ are defined by Eqs. (11a) and (11b).

## IV. POINT SPECTRUM

The contributions to $f(\omega)$ of zeroes of

$$\Omega(z) = \det\Lambda\left(\frac{1}{z}+\xi\right)$$

is a routine exercise in residue theory. We shall write

$$N_0 = \{z\in\mathbb{C}\mid \Omega(z)=0,\ z\notin Q\},\qquad N_Q = \{z\in\mathbb{C}\mid \Omega(z)=0, z\in Q\}.$$

We will for simplicity assume that the zeroes in $N_0$ have multiplicity one, although for later applications, zeroes in $N_Q$ of multiplicity one and two will be considered. Multiplicity of any order can be computed simply by using the residue formula for higher order poles. Finally, we assume

$$\frac{1}{k-\alpha_z}g\in\mathscr{B},\quad \text{for } \alpha_z = k\left(\frac{1}{z}+\xi\right),\ z\in N_Q. \tag{12}$$

*Theorem* 7: If $K$ satisfies the hypothesis of Lemma 5 and Eq. (12), and $f\in\mathscr{B}$, then $f_2(\omega)$ is a sum of contributions,

$$f_2(\omega) = \sum_{z\in N_0\cup N_Q} f_2^z(\omega), \tag{13}$$

where

(i) if $z\in N_0$ (multiplicity one),

$$f_2^z(\omega) = -\frac{1}{z(k(\omega)-\xi)-1}\,\mathbf{g}(\omega)\cdot\left(\frac{d\Omega}{dz}(z)\right)^{-1}\Lambda_c\left(\frac{1}{z}+\xi\right)\times\int_A\frac{\mathbf{J}(s)f(s)\,ds}{z(k(s)-\xi)-1}, \tag{14a}$$

(ii) if $z\in N_Q$ (multiplicity one),

$$f_2^z(\omega) = \tfrac{1}{2}\frac{1}{z(k(\omega)-\xi)-1}\,\mathbf{g}(\omega)\cdot\left\{\left[\left(\frac{d\Omega^+}{dz}(z)\right)^{-1}\Lambda_c^+\left(\frac{1}{z}+\xi\right)+\left(\frac{d\Omega^-}{dz}(z)\right)^{-1}\Lambda_c^-\left(\frac{1}{z}+\xi\right)\right]\int\frac{\mathbf{J}(s)f(s)\,ds}{z(k(s)-\xi)-1}\right.$$

$$\left. + \pi i\left[\left(\frac{d\Omega^+}{dz}\right)(z)^{-1}\Lambda_c^+\left(\frac{1}{z}+\xi\right)-\left(\frac{d\Omega^-}{dz}(z)\right)^{-1}\Lambda_c^-\left(\frac{1}{z}+\xi\right)\right]\mathbf{J}(\omega)f(\omega)\frac{(k(\omega)-\xi)}{k'(\omega)}\right\}, \tag{14b}$$

(iii) if $z\in N_Q$ (multiplicity two)

$$f_2^z(\omega) = \frac{1}{\Omega''(z)^+}\frac{d}{dz}\left[\frac{1}{z(k(\omega)-\xi)-1}\,\mathbf{g}(\omega)\cdot\Lambda_c^+\left(\frac{1}{z}+\xi\right)\left(\int_A\frac{\mathbf{J}(s)f(s)\,ds}{z(k(s)-\xi)-1}+\frac{\pi i\mathbf{J}(\omega)f(\omega)(k(\omega)-\xi)}{k'(\omega)}\right)\right]$$

$$+\frac{1}{\Omega''(z)}\frac{d}{dz}\left[\frac{1}{z(k(\omega)-\xi)-1}\,\mathbf{g}(\omega)\cdot\Lambda_c^-\left(\frac{1}{z}+\xi\right)\left(\int_A\frac{\mathbf{J}(s)f(s)\,ds}{z(k(s)-\xi)-1}-\pi i\frac{\mathbf{J}(\omega)f(\omega)(k(\omega)-\xi)}{k'(\omega)}\right)\right]$$

$$-\frac{2}{3}\frac{\Omega'''(z)}{(\Omega''(z))^2}\frac{1}{z(k(\omega)-\xi)-1}\mathbf{g}(\omega)\cdot\left\{\left[\Lambda_c^+\left(\frac{1}{z}+\xi\right)+\Lambda_c^-\left(\frac{1}{z}+\xi\right)\right]\int_A\frac{\mathbf{J}(s)f(s)\,ds}{z(k(s)-\xi)-1}\right.$$

$$\left.+\pi i\left[\Lambda_c^+\left(\frac{1}{z}+\xi\right)-\Lambda_c^-\left(\frac{1}{z}+\xi\right)\right]\frac{\mathbf{J}(\omega)f(\omega)(k(\omega)-\xi)}{k'(\omega)}\right\}. \tag{14c}$$

In these formulas, $\Lambda_c$ indicates the cofactor matrix, $\Lambda^{-1}=\Omega\Lambda_c$.

## V. CONSTRUCTION OF SOLUTIONS

In this section, we wish to establish a norm on $\mathscr{B}$ which will enable Eqs. (11) to be extended to the full Banach space. Because of certain technical difficulties in treating the problem when there are eigenvalues imbedded in the continuous spectrum, we will consider two cases.

Case (a): $\det\Omega(z)\neq 0$ for $z\subset R\subset\mathbb{R}$.

Let us define $F$: $\mathscr{B}\rightarrow\mathscr{B}'$ by

$$F(f)(x)=\mathbf{A}(x),$$

where $\mathbf{A}(x)$ is given by Eq. (11b). We wish to choose spaces $\mathscr{B}$ and $\mathscr{B}'$ such that $F$ will be an invertible bounded transformation. We shall consider separately the terms

$$F_1(f)(x)=\Lambda^-(x)^{-1}\frac{\mathbf{J}\otimes\mathbf{g}(k^{-1}(x))}{k'(k^{-1}(x))}\Lambda^+(x)^{-1}\int_A\frac{J(s)f(s)\,ds}{x-k(s)} \tag{15a}$$

and

$$F_2(f)(x)=\Lambda^-(x)^{-1}\lambda(x)\Lambda^+(x)^{-1}\mathbf{J}(k^{-1}(x))f(k^{-1}(1)). \tag{15b}$$

The $L_p$ estimates of the contributions to $\mathbf{A}(x)$ due to the zeros of $\Omega$ are trivial [by virtue of assumption (a) above] and are omitted.

Suppose $f$ is Hölder continuous with compact support. Then, as a function of $\hat{x}=x^\beta$, $\hat{x}^{-\frac{1}{2}(1-\beta)}F_1(f)(\hat{x}^\beta)$ is Hölder continuous, and may be estimated in $L_p(\hat{x})$ norm $\|\ \|_{L_p(\hat{x})}$ by

$$\|\hat{x}^{-(1-\beta)/2}F_1(f)(\hat{x}^\beta)\|_{L_p(\hat{x})}$$
$$\leqslant\left\|\frac{\|\Lambda^-(\hat{x}^\beta)^{-1}\mathbf{J}\otimes\mathbf{g}(k^{-1}(\hat{x}^\beta))\Lambda^+(x^\beta)^{-1}\|}{k'(k^{-1}(\hat{x}^\beta))}\right\|_\infty$$
$$\cdot C_p\left\|\frac{\mathbf{J}(k^{-1}(t^\beta))f(k^{-1}(t^\beta))}{k'(k^{-1}(t^\beta))t^{(1-\beta)/2}}\right\|_{L_p^N(t^\beta)}$$

where $C_p$ depends upon $p$ only, and

$$\|\Lambda(\hat{x})\|_{L_p^N(\hat{x})}=\left\{\sum_{i=1}^N\int|\mathbf{A}_i(\hat{x})|^p d\hat{x}\right\}^{1/p}, \tag{16a}$$

$$\|\Lambda(\hat{x})\|_\infty=\sup_{\hat{x}\subset\mathbb{R}}\sup_{|ij|}|\Lambda_{ij}(\hat{x})|. \tag{16b}$$

Similarly, $\hat{x}^{-(1-\beta)/2}F_2(f)(\hat{x}^\beta)$ may be estimated by

$$\|\hat{x}^{-(1-\beta)/2}F_2(f)(\hat{x}^\beta)\|_{L_p^N(\hat{x})}$$
$$\leqslant\|\Lambda^-(\hat{x}^\beta)^{-1}\lambda(\hat{x}^\beta)\Lambda^+(\hat{x}^\beta)^{-1}\hat{x}^{(1-\beta)}k'(k^{-1}(\hat{x}^\beta))\|_\infty$$
$$\cdot\left\|\frac{\mathbf{J}(k^{-1}(\hat{x}^\beta))f(k^{-1}(\hat{x}^\beta))}{k'(k^{-1}(\hat{x}^\beta))\hat{x}^{(1-\beta)/2}}\right\|_{L_p^N(\hat{x})}.$$

Let $\mathscr{B}$ be the Banach space of real-valued measurable functions $f$ on $A$ such that

$$\|f\|_{\mathscr{B}}=\left\|\frac{\mathbf{J}(k^{-1}(\hat{x}^\beta))f(k^{-1}(\hat{x}^\beta))}{k'(k^{-1}(\hat{x}^\beta))\hat{x}^{(1-\beta)/2}}\right\|_{L_p^N(\hat{x})}<\infty,$$

and $\mathscr{B}'$ the Banach space of real-valued measurable functions $A$ on $R$ such that

$$\|A\|_{\mathscr{B}'}\equiv\|\hat{x}^{-(1-\beta)/2}\mathbf{A}(\hat{x}^\beta)\|_{L_p^N(\hat{x})}<\infty.$$

Call the triple $\{k,g,J\}$ smooth if

$$\Lambda^-(s^\beta)^{-1}\frac{\mathbf{J}\otimes\mathbf{g}(k^{-1}(s_\beta))}{k'(k^{-1}(s^\beta))},$$

$\Lambda^+(s^\beta)^{-1}$, and the functions $\Lambda^\pm(s^\beta)^{-1}\{s^{1-\beta}k'(k^{-1}(s^\beta))\}^\alpha$, $\alpha=\pm 1$, are bounded as $s\rightarrow\infty$.

*Lemma* 8: If $\{k,g,J\}$ is smooth of transport type, then, $F$:$\mathscr{B}\rightarrow\mathscr{B}'$ extends to a bounded linear transformation.

Likewise, we define

$$F_1'(A)(\omega)=\mathbf{g}(\omega)\cdot\int\frac{\mathbf{A}(x)\,dx}{x-k(\omega)} \tag{17a}$$

and

$$F_2'(A)(\omega)=-\frac{1}{J^2(\omega)}\Lambda'(k(\omega))\mathbf{J}(\omega)\cdot\mathbf{A}(k(\omega)). \tag{17b}$$

Let $k(\omega)=t^\beta$ and $x=\hat{x}^\beta$. Then

$$\frac{F_1'(A)(k^{-1}(t^\beta))\mathbf{J}(k^{-1}(t^\beta))}{k'(k^{-1}(t^\beta))t^{(1-\beta)/2}}$$
$$=\frac{\mathbf{J}(k^{-1}(t^\beta))\mathbf{g}(k^{-1}(t^\beta))}{k'(k^{-1}(t^\beta))}\cdot\int\frac{\mathbf{A}(\hat{x}^\beta)d\hat{x}}{(\hat{x}-t)\hat{x}^{(1-\beta)/2}}$$

and

$$\|F_1'(A)\|_{\mathscr{B}}\leqslant\left\|\frac{\|\mathbf{J}(k^{-1}(t^\beta))\otimes\mathbf{g}(k^{-1}(t^\beta))\|}{k'(k^{-1}(t^\beta))}\right\|_\infty\|A\|_{\mathscr{B}'}.$$

The second term may be estimated by

$$\|F_2'(A)\|_{\mathscr{B}}\leqslant\left\|\frac{\|\mathbf{J}(k^{-1}(t^\beta))\cdot\lambda'(t^\beta)\mathbf{J}(k^{-1}(t^\beta))\|}{J^2(k^{-1}(t^\beta))t^{(1-\beta)}k'(k^{-1}(t^\beta))}\right\|_\infty\|A\|_{\mathscr{B}'}.$$

*Lemma* 9: If $\{k,g,J\}$ is smooth of transport type, then $F'$:$\mathscr{B}'\rightarrow\mathscr{B}$ extends to a bounded linear transformation and $F'=F^{-1}$.

We may now obtain a resolution of the identity corresponding to $K$. For $\lambda\in R$ let us define

$$E(\lambda)f(\omega)=\int_{-\infty}^{\lambda}\mathbf{\Phi}_x(\omega)\cdot\mathbf{A}(x)\,dx, \tag{18}$$

where $\mathbf{A}$ and $\mathbf{\Phi}$ are given by Eqs. (11). To obtain the discrete eigenprojections, we note that $\Lambda_c(1/z+\xi)$ may be written

$$A_c\left(\frac{1}{z}+\xi\right)=\gamma_z\otimes\alpha_z, \tag{19}$$

for $z\in N_p$, where

$$A\left(\frac{1}{z}+\xi\right)\gamma_z=0.$$

Then $f_2^z(\omega)$ may be expanded as

$$f_2^z(\omega)=\Phi_z(\omega)A_z, \tag{20a}$$

with

$$\Phi_z(\omega)=\frac{\mathbf{g}(\omega)\cdot\gamma_z}{z(k(\omega)-\xi)-1}, \tag{20b}$$

an eigenvector of $K$, and

$$A_z=-\frac{1}{\Omega'(z)}\alpha_z\cdot\int\frac{\mathbf{J}(s)f(s)\,ds}{z(k(s)-\xi)-1}. \tag{20c}$$

Defining $E(\lambda)$ for $1/(\lambda-\xi)=z\in N_p$ by

$$E(\lambda)f(\omega)=\Phi_z(\omega)A_z, \tag{21}$$

we may follow Ref. 3 to prove that the family of projections $E(\lambda)$ is a resolution of the identity, and

$$K=\int_R\lambda\,dE(\lambda)+\sum_{(1/(\lambda-\xi)\in N_p}\lambda E(\lambda). \tag{22}$$

We state this as

*Theorem* 10: The family $E(\lambda)$ is a resolution of the identity for $K$. The solution of Eq. (1) satisfying the boundedness condition $\lim_{x\to\infty}\|\psi(x)\|_{\mathscr{B}}=0$ is given by

$$\psi(x,\nu)=\int_{-\infty}^{x}d\xi\int_0^\infty e^{-(x-\xi)\lambda}d(E(\lambda)q)(\xi,\nu)$$

$$-\int_x^\infty d\xi\int_{-\infty}^{0}e^{-(x-\xi)\lambda}d(E(\lambda)q)(\xi,\nu)$$

$$+\sum_{1/(\lambda-\xi)\in N_p}e^{x(1-\lambda)}E(\lambda)q. \tag{23}$$

Case (b): $\det\Omega(z)=0$ for $z\in R$.

When there are eigenvalues imbedded in the continuous spectrum, the $L^p$ estimates of the previous paragraphs are not valid. It is nevertheless possible to verify that the expression in Theorem 10 is indeed the solution of the boundary value problem. To see this we may rederive the eigenfunction expression, Eq. (11c), for $Sf$, with $f\in\mathscr{B}$ Hölder continuous, obtaining

$$Sf=\int\frac{1}{x-\xi}\Phi_x(\omega)\cdot\mathbf{A}_f(x)\,dx+\sum_{Sf}, \tag{24a}$$

where $\mathbf{A}_f$ is used to denote the transform $A$ given by Eq. (11b), and $\Sigma_{Sf}$ denotes the discrete terms, Eq. (13). Now let us choose $h\in\mathscr{D}(S^{-1})$, whence

$$h=\int\Phi_x(\omega)\cdot\mathbf{A}_h(x)\,dx+\sum_g=Sf, \tag{24b}$$

and therefore

$$\frac{1}{x-\xi}A_f(x)=A_h(x). \tag{25}$$

We have used a Liouville theorem argument to go from Eqs. (24) to Eq. (25). For analogous use of this argument, see Refs. 12 and 13.

Thus, we may write

$$S^{-1}h=\int(x-\xi)\Phi_x(\omega)\cdot\mathbf{A}_h(x)\,dx+\sum_{S^{-1}h}-l_h. \tag{26}$$

We have then immediately $Kh$, and we may substitute the expression in Theorem 10 into Eq. (1) to obtain

*Corollary* 11: The solution of Eq. (11) satisfying the boundedness conditions $\lim_{x\to\infty}\|\psi(x)\|_{\mathscr{B}}=0$ is given by Eq. (23).

## VI. HALF RANGE

The eigenfunction expansion developed in the previous three sections can be used to solve so called "full range" problems involving Eq. (1). The terminology "full range" means we are interested in solutions for $x\in\mathbb{R}$, i.e., infinite media problems. Of more practical interest is the case $x\in\mathbb{R}^+$, i.e., half-space problems; typically one needs an eigenfunction expansion on the so-called "half-range," $\mu\in A^+=\{\mu\in A\,|\mu\geqslant 0\}$. (A detailed discussion of this point may be found in Ref. 1, for one-speed neutron transport.)

The idea, as introduced in Ref. 1, is to define a map $E$ with certain properties which guarantee that the "half-range" expansion of $f$ is given by the full range expansion of $Ef$. We define $E$: $D(E)\to D(K)$, with

$$D(E)=\{f\in L_p(A^+,\sigma)\mid f\text{ is Hölder continuous with compact support}\}$$

as

$$E(f)(\mu)=f(\mu),\quad\mu\in A^+$$
$$=\mathbf{g}(\mu)\cdot X^{-1}(\mu)\int_{A^+}\frac{Y^{-1}(-s)\mathbf{J}(s)f(s)\,ds}{k(s)-z},\quad\mu\in A^-. \tag{27}$$

Here the matrices $X$ and $Y$ are supposed to provide the Wiener–Hopf factorization of the matrix $A$, i.e.,

$$A(z)=Y(-z)X(z),$$

where $X$ and $Y$ are analytic in $z$ for $\mathrm{Re}z+\mathrm{Im}z<0$ and $\lim_{|z|\to\infty}Y(-z)$ and $\lim_{|z|\to\infty}X(z)$ exist. The sufficient conditions that such a factorization exist have been discussed by Mullikin[14] and Victory[15] (see also Ref. 4, Sec. V and VI). The existence of such factorization is crucial to the analysis of the present section. (For a slightly different approach, see Ref. 16.)

We now state

*Theorem* 12: Let $E$ be defined by Eq. (25). Then $(S_\xi-zI)^{-1}Ef$ is analytic in $z$ for $\mathrm{Re}z+\mathrm{Im}z<0$.

*Proof*: Writing

$(zI - S_\xi)^{-1}Ef$

$$= \frac{k(\mu)-\xi}{z(k(\mu)-\xi)-1}\left(Ef(\mu) - \frac{\mathbf{g}(\mu)}{k(\mu)-\xi}\cdot\mathbf{G}(z)\right),$$

where $\mathbf{G}$ is given by

$$\mathbf{G}(z) = \Lambda^{-1}\left(\xi + \frac{1}{z}\right)\int_A \frac{\mathbf{J}(s)Ef(x)\,ds}{1 - z(k(s)-\xi)}$$

$$= \Lambda^{-1}\left(\xi + \frac{1}{z}\right)\mathbf{T}\left(\xi + \frac{1}{z}\right),$$

we demand

$$\frac{\mathbf{g}(\mu)}{k(\mu)-\xi}\cdot\mathbf{G}^{+}\left(\frac{1}{k(\mu)-\xi}\right) = \frac{\mathbf{g}(\mu)}{k(\mu)-\xi}$$

$$\cdot\mathbf{G}^{-}\left(\frac{1}{k(\mu)-\xi}\right) = Ef(\mu),$$

for $\mu\in\Lambda^{-}$. In fact, let us assume

$$G^{+}\left(\frac{1}{k(\mu)-\xi}\right) = G^{-}\left(\frac{1}{k(\mu)-\xi}\right)$$

on $A$; whence

$$Y^{+}(-k(\mu))^{-1}\mathbf{T}^{+}(k(\mu)) = Y^{-}(-k(\mu))^{-1}\mathbf{T}^{-}(k(\mu)). \tag{28}$$

Define

$$\mathbf{Q}(z) = Y\left(-\xi - \frac{1}{z}\right)^{-1}\mathbf{T}\left(\xi + \frac{1}{z}\right)$$

$$- \int_{A^{\cdot}} \frac{Y^{-1}(-k(s))\mathbf{J}(s)f(s)\,ds}{1 - z(k(s)-\xi)}.$$

Then

$$Q^{+}\left(\frac{1}{\nu-\xi}\right) = Q^{-}\left(\frac{1}{\nu-\xi}\right)$$

for $\nu\in\mathbb{R}^{-}$ by virtue of Eq. (28), and on $\mathbb{R}^{+}$ by a direct computation. Since $Q$ is bounded near $N_p$ and $Q(t)\to 0$ at infinity, we conclude

$$Y\left(-\xi - \frac{1}{z}\right)\int_{A^{\cdot}} \frac{Y^{-1}(-k(s))\mathbf{J}(s)f(s)\,ds}{1 - z(k(s)-\xi)}$$

$$= \int_A \frac{\mathbf{J}(s)\,Ef(s)\,ds}{1 - z(k(s)-\xi)}.$$

By evaluating at limits at $z = 1/[k(\mu)-\xi]$, taking a scalar product with $\mathbf{J}(z)$, and computing $\mathbf{Y}^{+} - \mathbf{Y}^{-} = (\Lambda^{+} - \Lambda^{-})\mathbf{X}^{-1}$, this implies that

$Ef(\mu)$

$$= \frac{\mathbf{g}(\mu)}{k(\mu)-\xi}\cdot X^{-1}(k(\mu))\int_{A^{\cdot}} \frac{Y^{-1}(-k(s))\mathbf{J}(s)f(s)\,ds}{1 - (k(s)-\xi)/(k(\mu)-\xi)},$$

$$\mu\in\Lambda^{-}.$$

Finally, a straightforward computation gives

$$Ef(\mu) = \frac{\mathbf{g}(\mu)}{k(\mu)-\xi}\cdot\mathbf{G}\left(\frac{1}{k(\mu)-\xi}\right)$$

For $\omega\in\Lambda^{+}$, the half range expansion of $f\in L^p$ is given by

$$f(\omega) = \int_{R^{\cdot}} \mathbf{\Phi}_x(\omega)\cdot\mathbf{A}_x\,dx + \sum_f^{+}, \tag{29a}$$

where $\mathbf{\Phi}_x$ is defined by Eq. (11a), and

$$\mathbf{A}(x) = -\Lambda^{-}(x)^{-1}\mathbf{J}\otimes\mathbf{g}(k^{-1}(x))\Lambda^{+}(x)^{-1}Y(-x)$$

$$\times\int_{A^{\cdot}} \frac{Y(-k(s))\mathbf{J}(s)f(s)\,ds}{x - k(x)}$$

$$-\Lambda^{-}(x)^{-1}\lambda(x)\Lambda^{+}(x)^{-1}\mathbf{J}(k^{-1}(x))f(k^{-1}(x)). \tag{29b}$$

(Note this is a "half range expansion", as the negative spectrum does not enter.)

The contribution of isolated eigenvalues $\Sigma_f^{+}$ from the appropriate half space is carried out as in Sec. IV. We omit details.

## VII. ANISOTROPIC NEUTRON TRANSPORT

In this section we present a quick illustration of the full and half-range expansions obtained above. The illustration that we have in mind is the neutron transport equation with anisotropic scattering. In particular, if we assume the scattering function can be expanded as a finite series of Legendre polynomials, we obtain the triple $(k,g,J)$ as indicated in example 4 of Sec. I.

Therefore, from Eq. (4), we can write

$$\left[\Lambda\left(\frac{1}{\omega}\right)\right]_{lm} = \delta_{lm} + \frac{c}{2}\,\omega\int_{-1}^{1}\frac{(2m+1)f_m P_m(s)P_l(s)}{s-\omega}\,ds.$$

We have then, from Eq. (11),

$$f_1(\omega) = \int_{-1}^{+1}\mathbf{\Phi}_y(\omega)\cdot\mathbf{A}(y)dy,$$

where for this problem,

$$[\mathbf{\Phi}_y(\omega)]_l = P\,\frac{(c/2)(2l+1)yf_lP_l(\omega)}{y-\omega}$$

$$+\tfrac{1}{2}\sum_n \frac{[\Lambda_{nl}^{+}(1/y) + \Lambda_{nl}^{-}(1/y)]}{\Sigma_k P_K^2(\omega)}P_n(\omega)\delta(y-\omega),$$

and

$$[\Lambda(y)]_l = -\frac{1}{y}\sum_{\mu,k,m}\Lambda_{l\mu}^{-}\left(\frac{1}{y}\right)^{-1}\frac{c}{2}(2k+1)$$

$$\times f_kP_k(y)P_\mu(y)\Lambda_{km}^{+}\left(\frac{1}{y}\right)^{-1}\int_{-1}^{1}\frac{P_m(s)f(s)}{1/y - 1/s}\,ds$$

$$+\tfrac{1}{2}\sum_n\left[\Lambda_{lm}^{-}\left(\frac{1}{y}\right)^{-1} + \Lambda_{ln}^{+}\left(\frac{1}{y}\right)^{-1}\right]P_nf(y).$$

However let us note the identity[17]

$$\Lambda(1/\omega)\mathbf{h}(\omega) = \Omega(1/\omega)\mathbf{P}(\omega),$$

where

$$[\mathbf{P}(\omega)]_n = P_n(\omega)$$

and[18]

$$\Omega(1/\omega) = \det\Lambda(1/\omega) = 1 + \omega\int_{-1}^{+1}\frac{K(\omega,x)}{x-\omega}\,dx.$$

Here

$$K(\omega,x) = \frac{c}{2}\sum_{n=0}^{N}(2n+1)f_nP_n(x)h_n(\omega),$$

and

$$[\mathbf{h}(\omega)]_n = h_n(\omega),$$

where the polynomials $h_n(\omega)$ are defined recursively by[19]

$$(n+1)h_{n+1}(\omega) + nh_{n-1}(\omega) = \omega[(2n+1) - cf_n]h_n(\omega),$$

$$h_0(\omega) = 1, \quad h_1(\omega) = (1-c)\omega.$$

Use of this identity and a modest amount of algebra allows us to write

$$[\mathbf{A}(y)]_l = \frac{h_l(y)}{\Omega^+(1/y)\Omega^-(1/y)}\int_{-1}^{+1}\frac{sK(y,s)f(s)}{y-s}\,ds + \bar{\lambda}(y)f(y),$$

where

$$\bar{\lambda}(y) = \tfrac{1}{2}[\Omega^+(1/y) + \Omega^-(1/y)].$$

Using this result and Eqs. (11) we obtain

$$\int_{-1}^{+1}\mathbf{\Phi}_y(\omega)\cdot\mathbf{A}(y)dy = \int_{-1}^{+1}\varphi_y(\omega)\mathscr{A}(y)\,dy,$$

where

$$\varphi_y(\omega) = \frac{Pyk(\omega,y)}{y-\omega} + \bar{\lambda}(y)\delta(y-\omega) \tag{30a}$$

and

$$\mathscr{A}(y) = [y(\Omega^+(1/y)\Omega^-(1/y)]^{-1}\int_{-1}^{+1}sf(s)\varphi_y(s)\,ds. \tag{30b}$$

This is the same result obtained by Mika using the singular eigenfunction technique.[19]

For half-range problems it is necessary to use the factorization

$$\Lambda(1/z) = Y(-1/z)X(1/x).$$

For the problem under consideration, this factorization has been shown by Mullikin[20] for those cases when $c<1$. More precisely, the matrices $X(1/\omega)$ and $Y(1/\omega)$ can be written in the form[20]

$$[Y(1/\omega)]_{nk} = \delta_{nk} + \frac{c\omega}{2}\int_0^1\frac{(2k+1)f_k\Psi_k(s)P_n(s)}{s+\omega}\,ds, \tag{31a}$$

and

$$[X(1/\omega)]_{nk} = \delta_{nk} + \frac{(-1)^{n+k-1}c\omega}{2}\times\int_0^1\frac{(2k+1)f_k\Psi_k(s)P_n(s)}{s+\omega}\,ds, \tag{31b}$$

where the functions $\Psi_k$ satisfy the nonlinear equations

$$\Psi_l(s) = P_l(s) + \frac{cs}{2}\sum_{k=0}^{N}(-1)^{l+k}(2k+1)f_k\times\int_0^1\frac{\Psi_k(s)\Psi_k(t)P_l(t)}{t+s}\,dt, \tag{31c}$$

plus certain analyticity constraints. Mullikin has shown the existence of the solution to Eqs. (31) and recent results indicate that Eq. (31c) or a variant of it is a likely candidate for solution by iteration.

Using the same techniques as in the full-range we find from Eq. (29b) that

$$[\mathbf{A}(y)]_l = \frac{h_l(y)}{\Omega^+(1/y)\Omega^-(1/y)}\int_0^1\frac{sG(y,s)f(s)}{y-s}\,ds + \bar{\lambda}(y)f(y),$$

where

$$G(y,s) = \frac{c}{2}\sum_{n,l}(2n+1)f_nh_n(y)Y_{nl}(-1/y)\times Y_{lm}^{-1}(-1/s)P_m(s).$$

Substituting this into Eq. (29a), we will obtain the contribution to the half-range expansion from the continuum to be

$$\bar{f}_1(\omega) = \int_0^1\mathbf{\Phi}_y(\omega)\cdot\mathbf{A}(y)\,dy = \int_0^1\varphi_y(\omega)\bar{\mathscr{A}}(y)\,dy,$$

where $\varphi_y(\omega)$ is given by Eq. (30) and

$$\bar{\mathscr{A}}(y) = [y\Omega^+(1/y)\Omega^-(1/y)]^{-1}\int_0^1\frac{sy\,G(y,s)f(s)\,ds}{y-s} + \bar{\lambda}(y)f(y).$$

The contribution to the expansion from the discrete roots can be easily, if laboriously, worked out, using Theorem 7(i). Again, Mika's results[19] are reproduced. (Recall[21] that for the subcritical case considered here the discrete roots fall on the real line outside $[-1,1]$.) The singular eigenfunction method has the advantage that the matrix $\Lambda$ need not be factored; the method presented here however, is somewhat simpler, granted the matrix factorization known, and gives the results in somewhat simpler form. In any event, this section has been included as an illustration of our technique and to connect our results with the (seemingly) different formulas in the literature, rather than to obtain new results on this particular application.

[1]E.W. Larsen and G.J. Habetler, Commun. Pure Appl. Math **26**, 525 (1973).
[2]J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U. P., Princeton, N.J. 1955).
[3]E.W. Larsen, S. Sancaktar, and P.F. Zweifel, J. Math. Phys. **16**, 1117 (1976).
[4]W. Greenberg and P.F. Zweifel, Trans. Theor. Stat. Phys. **5**, 219 (1976).
[5]W. Greenberg and P.F. Zweifel, J. Math. Phys. **17**, 163 (1976); R.L. Bowden, W. Greenberg, and P.F. Zweifel, SIAM J. Appl. Math. **32**, 765 (1977).
[6]K.M. Case and P.F. Zweifel, *Linear Transport Theory* (Addison–Wesley, Reading, Mass., 1967).
[7]E.W. Larsen, private communications.
[8]M.D. Arthur, W. Greenberg, and P.F. Zweifel, Phys. Fluids **20**, 1296 (1977).
[9]R.L. Bowden and L. Garbanati, Trans. Theor. Stat. Phys. **1**, 1 (1978).
[10]W.L. Cameron, "Constructive Solutions of the Linearized Boltzmann Equation. Coupled BGK Model Equation," preprint.
[11]R.L. Bowden, W.L. Cameron, and P.F. Zweifel, J. Math. Phys. **18**, 201 (1977).
[12]W. Greenberg and S. Sancaktar, J. Math. Phys. **17**, 2092 (1976).
[13]Ref. 6, p. 74.
[14]T.W. Mullikin, Trans. Theor. Stat. Phys. **3**, 215 (1973).
[15]H.D. Victory, Trans. Theor. Stat. Phys. **5**, 87 (1976).
[16]R.J. Hangelbroek and C.G. Lekkerkerker, SIAM J. Math. Anal. **8**, 458 (1977).
[17]R.L. Bowden, F.J. McCrosson, and E.A. Rhodes, J. Math. Phys. **9**, 753 (1968).
[18]A. Leonard and T.W. Mullikin, J. Math. Phys. **5**, 399 (1964).
[19]J.R. Mika, Nucl. Sci. Eng. **11**, 415 (1961).
[20]T.W. Mullikin, Ap. J. **39**, 1267 (1964).
[21]K.M. Case, J. Math. Phys. **15**, 974 (1974).

# The Coulomb and Coulomb-like off-shell Jost functions

H. van Haeringen

*Natuurkundig Laboratorium der Vrije Universiteit, Amsterdam, The Netherlands and Institute for Theoretical Physics, P.O. Box 800, University of Groningen, The Netherlands*

(Received 29 August 1978)

The off-shell Jost functions are studied for a potential which is the sum of the Coulomb potential and an arbitrary local short-range central potential. We derive their singular on-shell behavior and their connection with the pure Coulomb off-shell Jost functions. For the latter we derive a large variety of interesting explicit analytic expressions which are useful for various purposes.

## 1. INTRODUCTION

In this paper we investigate the off-shell Jost functions $f_{cl}(k,q)$ for the Coulomb potential and the off-shell Jost functions $f_l(k,q)$ for a Coulomb plus short-range potential, $V = V_c + V_s$, where $V_s$ is assumed to be local and central. As is now well known, these off-shell Jost functions are particularly interesting in connection with the transition matrices.

In Sec. 2 we show that $f_{cl}(k,q)$ is a basic constituent of $f_l(k,q)$. In particular, we prove that $f_l(k,q)$ has exactly the same singularity in $q = k$ as $f_{cl}(k,q)$. In order to obtain the most convenient formula for $f_{cl}(k,q)$, a regrouping of certain hypergeometric function expressions has to be performed. By doing this we supply the supplementary proof of the simple formula for $f_{cl}(k,q)$ that we have given before.[1] This formula contains Jacobi polynomials and certain polynomials of two variables, $A_l$.

In Sec. 3 we derive a large number of interesting expressions for these polynomials $A_l$. Each of these is useful for different purposes, as is clearly illustrated at the end of Sec. 3. We shall use the notation of Ref. 1.

## 2. THE OFF-SHELL JOST FUNCTIONS

In this section we will express the off-shell Jost function $f_l(k,q)$ for a Coulomblike potential in terms of the Coulomb off-shell Jost function $f_{cl}(k,q)$. By using this expression the on-shell behavior at $q = k$ is easily obtained. Further, we shall sketch the derivation of a simple closed expression for $f_{cl}(k,q)$.

We start by noting that[2]

$$f_l(k,q) = 1 + \tfrac{1}{2}\pi q(q/k)^l f_l(k)\, {}_0\langle ql\downarrow|\, V_l\, |kl+\rangle. \quad (2.1)$$

Here $f_l(k)$ is the Jost function and $|kl+\rangle$ the "outgoing" scattering state, with energy $k^2$, for the potential $V_l = V_{cl} + V_{sl}$. We use the Coulomb analog of Eq. (2.1) and apply the two-potential formalism. In this way we get the convenient expression,

$$f_l^{-1}(k) f_l(k,q) = f_{cl}^{-1}(k) f_{cl}(k,q) + {}_c\langle kl-|\, V_{sl} G_l |X_l\rangle. \quad (2.2)$$

Here $G_l$ is the partial-wave Green operator for $V_l$, and $|X_l\rangle$ is defined by

$$|X_l\rangle = \tfrac{1}{2}\pi k G_{0l}^{-1}\,[(q/k)^{l+1}\,|ql\uparrow\rangle_0 - |kl\uparrow\rangle_0].$$

By inserting

$$\langle p|ql\uparrow\rangle_0 = 2(\pi q)^{-1}(p/q)^l(p^2-q^2)^{-1},$$

we obtain a simple expression for $|X_l\rangle$ in the momentum representation,

$$\langle p|X_l\rangle = (p/k)^l(k^2-q^2)/(p^2-q^2). \quad (2.3)$$

Equation (2.2) is very interesting, since it clearly shows that $f_l(k,q)$ has exactly the same singularity in $q = k$ as $f_{cl}(k,q)$. As a matter of fact, by using Eq. (2.3) we have

$$\lim_{q\to k}\omega|X_l\rangle = 0, \quad k>0,$$

and therefore,

$$\lim_{q\to k}\omega f_l(k,q) = f_l(k), \quad k>0. \quad (2.4)$$

Here

$$\omega \equiv \left(\frac{q-k}{q+k}\right)^{i\gamma}\frac{e^{\pi\gamma/2}}{\Gamma(1+i\gamma)} = \frac{f_{c0}(k)}{f_{c0}(k,q)}.$$

Now we are going to summarily derive explicit expressions for $f_{cl}(k,q)$ [cf. Eqs. (4) and (7) of Ref. 1]. In order to evaluate ${}_0\langle ql\downarrow|V_{cl}|kl+\rangle_c$, which occurs in

$$f_{cl}(k,q) = 1 + \tfrac{1}{2}\pi q(q/k)^l f_{cl}(k)\, {}_0\langle ql\downarrow|V_{cl}|kl+\rangle_c,$$

we use the well-known expressions,

$${}_0\langle ql\downarrow|r\rangle = (2/\pi)^{1/2} i^{-l} h_l^{(+)}(qr),$$

$$= (-)^l(2/\pi)^{1/2}(qr)^{-1}e^{iqr}{}_2F_0(-l,l+1;(2iqr)^{-1})$$

and

$$\langle r|kl+\rangle_c = (2/\pi)^{1/2} l!\,[f_{cl}(k)(2l+1)!]^{-1}(2ikr)^l e^{-ikr} \times {}_1F_1(l+1-i\gamma;2l+2;2ikr).$$

By using Ref. 3, p. 278, one obtains

$${}_0\langle ql\downarrow|V_{cl}|kl+\rangle_c = 2i\gamma l!\,[\pi q f_{cl}(k)(2l+1)!]^{-1} \times \sum_{m=0}^{l}(m+1)_l(k/q)^m z^{l+1-m} \times {}_2F_1(l+1+i\gamma,l+1-m;2l+2;z), \quad (2.5)$$

where $z = 2k/(q+k)$. The important step now is to separate off that part which contains the branch-point singularity in $q = k$. To this end we apply two transformations to the hypergeometric function ${}_2F_1$ on the right-hand side of Eq. (2.5) and find (Ref. 3, p. 47),

$${}_2F_1(l+1+i\gamma,l-m+1;2l+2;z) = (1-z)^{m-i\gamma}\left(\frac{\Gamma(2l+2)\Gamma(i\gamma-m)}{\Gamma(l+1-m)\Gamma(l+1+i\gamma)}\right)z^{-2l-1}$$

 0022-2488/79/061109-06$01.00  

$$\times {}_2F_1(m-l,-i\gamma-l;1+m-i\gamma;1-z)$$
$$+\left(\frac{\Gamma(2l+2)\Gamma(m-i\gamma)}{\Gamma(l+1+m)\Gamma(l+1-i\gamma)}\right)z^{-2l-1}$$
$$\times {}_2F_1(-m-l,i\gamma-l;1-m+i\gamma;1-z).$$

The hypergeometric series for the ${}_2F_1$'s on the right-hand side break off. Therefore, these ${}_2F_1$'s can be rewritten in terms of Jacobi polynomials. One has, with $z=2/(1+x)$,

$$P_{l+m}^{(i\gamma-m,-i\gamma-m)}(x)$$
$$=\binom{l+i\gamma}{l+m}z^{-m-l}\,{}_2F_1(-m-l,i\gamma-l;1+i\gamma-m;1-z),$$

and so

$$P_{l-m}^{(-i\gamma+m,i\gamma+m)}(x)$$
$$=\binom{l-i\gamma}{l-m}z^{m-l}\,{}_2F_1(m-l,-i\gamma-l;1-i\gamma+m;1-z).$$

When we insert all this in Eq. (2.5) we get a complicated expression. In order to simplify this expression we introduce the polynomials $A_l$,

$$A_l(q^2/k^2;\gamma^2)$$
$$\equiv\sum_{m=0}^{l}\binom{l+m}{l}(-)^m\left(\frac{q}{k}\right)^{l-m}P_{l+m}^{(i\gamma-m,-i\gamma-m)}\left(\frac{q}{k}\right). \tag{2.6}$$

Furthermore, we shall now prove that

$$\sum_{m=0}^{l}\binom{l+m}{l}\left(\frac{k^2-q^2}{4kq}\right)^m P_{l-m}^{(m-i\gamma,m+i\gamma)}\left(\frac{q}{k}\right)$$
$$=P_l^{(-i\gamma,i\gamma)}\left(\frac{q^2+k^2}{2qk}\right). \tag{2.7}$$

For this proof we use

$$P_n^{(\alpha,\beta)}(\zeta)$$
$$=\binom{n+\alpha}{n}{}_2F_1(-n,n+1+\alpha+\beta;1+\alpha;\tfrac12-\tfrac12\zeta), \tag{2.8}$$

and the well-known integral representation,

$${}_2F_1(a,b;c;\zeta)=\left(\frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\right)$$
$$\times\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-t\zeta)^{-a}dt. \tag{2.9}$$

The left-hand side of Eq. (2.7) then becomes

$$\Gamma(l+1-i\gamma)[\Gamma(-i\gamma-l)\Gamma^2(l+1)]^{-1}$$
$$\times\int_0^1 t^l(1-t)^{-i\gamma-l-1}[1-\tfrac12 t(1-q/k)]^l$$
$$\times\sum_{m=0}^{l}\binom{l}{m}[t^{-1}-\tfrac12(1-q/k)]^{-m}\left(\frac{k^2-q^2}{4kq}\right)^m dt.$$

By performing the summation and using again Eqs. (2.8) and (2.9) we obtain the desired expression, i.e.,

$$\binom{l-i\gamma}{l}{}_2F_1\left(-l,l+1;1-i\gamma;\frac{-(q-k)^2}{4kq}\right)$$
$$=P_l^{(-i\gamma,i\gamma)}\left(\frac{q^2+k^2}{2kq}\right).$$

This completes the proof of Eq. (2.7).

By inserting the above expressions in Eq. (2.5) and using Eqs. (2.6) and (2.7) we obtain

$${}_0\langle ql\downarrow|V_{cl}|kl+\rangle_c$$
$$=2c_{l\gamma}[\pi q f_{cl}(k)]^{-1}[-x^{-l}A_l(x^2;\gamma^2)$$
$$+f_{c0}(k,q)P_l^{(-i\gamma,i\gamma)}(u)], \tag{2.10}$$

cf. Eq. (7) of Ref. 1. Here $x=q/k$, $u=(q^2+k^2)/(2kq)$,

$$f_{c0}(k,q)=\left(\frac{q+k}{q-k}\right)^{i\gamma},$$

and

$$c_{l\gamma}\equiv\frac{\Gamma^2(l+1)\Gamma(1+i\gamma)\Gamma(1-i\gamma)}{\Gamma(l+1+i\gamma)\Gamma(l+1-i\gamma)}. \tag{2.11}$$

In Sec. 3 we will derive a large number of useful expressions for the polynomials $A_l$.

## 3. THE TWO-VARIABLE POLYNOMIALS $A_l$

In this section we shall derive a number of interesting explicit expressions for the polynomials $A_l$ that occur in the formula (2.10) for the Coulomb off-shell Jost functions $f_{cl}(k,q)$.

To start with, we have

$$A_l\equiv A_l(x^2;\gamma^2)=\sum_{n=0}^{l}\binom{l+n}{l}(-)^n x^{l-n}P_{l+n}^{(i\gamma-n,-i\gamma-n)}(x) \tag{3.1}$$

[Eq. (2.6)], where $x=q/k$. Substitution of

$$P_{l+n}^{(i\gamma-n,-i\gamma-n)}(x)$$
$$=\Gamma(l+1+i\gamma)[\Gamma(l+n+1)\Gamma(l-n+1)\Gamma(i\gamma-l)]^{-1}$$
$$\times\int_0^1 t^{l-n}(1-t)^{i\gamma-l-1}[1-\tfrac12 t(1-x)]^{l+n}dt$$

yields

$$A_l=\Gamma(l+1+i\gamma)[\Gamma^2(l+1)\Gamma(i\gamma-l)]^{-1}$$
$$\times\sum_{n=0}^{l}\binom{l}{n}\int_0^1(-tx)^{-n}[1-\tfrac12 t(1-x)]^n(tx)^l$$
$$\times[1-\tfrac12 t(1-x)]^l(1-t)^{+i\gamma-l-1}dt.$$

The summation is easily carried out. We then get

$$A_l=4^{-l}i\gamma c_{l\gamma}^{-1}\int_0^1(1-t)^{i\gamma-l-1}[(2-t)^2-x^2t^2]^l dt, \tag{3.2a}$$

where $c_{l\gamma}$ is given by Eq. (2.11). The polynomials $A_l$ can also be expressed in terms of Gegenbauer polynomials $C_n^{-l}$. Indeed, by introducing $\tau=1-t$ we get from Eq. (3.2a),

$$A_l=\left(\frac{1-x^2}{4}\right)^l i\gamma c_{l\gamma}^{-1}\int_0^1\tau^{i\gamma-l-1}$$
$$\times\left(1+\tau^2-2\tau\frac{x^2+1}{x^2-1}\right)^l d\tau. \tag{3.2b}$$

It is well known that

$$(1-2\tau\zeta+\tau^2)^\lambda=\sum_{n=0}^{\infty}C_n^{-\lambda}(\zeta)\tau^n,\quad |\tau|<1,\quad \lambda\neq 0.$$

Because of

$$C_n^{-l}\equiv 0, \quad n=2l+1,2l+2,\cdots,$$

we can apply the above expansion to Eq. (3.2b), the result being,

$$A_l=\left(\frac{1-x^2}{4}\right)^l i\gamma c_{l\gamma}^{-1}\sum_{n=0}^{2l}\frac{1}{n-l+i\gamma}C_n^{-l}\left(\frac{x^2+1}{x^2-1}\right). \quad (3.3a)$$

By using

$$C_{l-n}^{-l}(\zeta)\equiv C_{l+n}^{-l}(\zeta), \quad -l\leqslant n\leqslant l,$$

we recast the above sum in the more convenient form,

$$\sum_{n=0}^{2l}\frac{1}{n-l+i\gamma}C_n^{-l}\left(\frac{x^2+1}{x^2-1}\right)$$
$$=-i\gamma\sum_{n=0}^{l}\frac{\epsilon_n}{n^2+\gamma^2}C_{l\pm n}^{-l}\left(\frac{x^2+1}{x^2-1}\right).$$

Here $\epsilon_n$ is the Neumann symbol,

$$\epsilon_n=\begin{cases}1, & n=0,\\ 2, & n=1,2,3,\cdots.\end{cases}$$

In this way we obtain from Eq. (3.3a),

$$A_l=\left(\frac{1-x^2}{4}\right)^l c_{l\gamma}^{-1}\gamma^2\sum_{n=0}^{l}\frac{\epsilon_n}{n^2+\gamma^2}C_{l\pm n}^{-l}\left(\frac{x^2+1}{x^2-1}\right). \quad (3.3b)$$

This expression can be rewritten in terms of the Jacobi polynomials $P_l^{(n,-n)}$. By using

$$C_n^{\lambda}\left(\frac{x^2+1}{x^2-1}\right)$$
$$=(\lambda)_n(n!)^{-1}\left(\frac{x+1}{x-1}\right)^n {}_2F_1\left(\lambda,-n;1-\lambda-n;\left(\frac{x-1}{x+1}\right)^2\right),$$
$$=(\lambda)_n(n!)^{-1}\left(\frac{x-1}{x+1}\right)^n {}_2F_1\left(\lambda,-n;1-\lambda-n;\left(\frac{x+1}{x-1}\right)^2\right),$$

we derive the interesting relation,

$$\left(\frac{4x}{1-x^2}\right)^{-l}C_{l-n}^{-l}\left(\frac{x^2+1}{x^2-1}\right)$$
$$=\frac{\Gamma^2(l+1)}{\Gamma(l+n+1)\Gamma(l-n+1)}\left(\frac{1-x}{1+x}\right)^n$$
$$\times P_l^{(n,-n)}(\tfrac{1}{2}x+\tfrac{1}{2}x^{-1}), \quad |n|\leqslant l. \quad (3.3c)$$

By inserting this in (3.3b) we get

$$A_l=x^l c_{l\gamma}^{-1}\gamma^2\sum_{n=0}^{l}\frac{\epsilon_n}{n^2+\gamma^2}\frac{\Gamma^2(l+1)}{\Gamma(l+n+1)\Gamma(l-n+1)}$$
$$\times\left(\frac{1-x}{1+x}\right)^n P_l^{(n,-n)}(\tfrac{1}{2}x+\tfrac{1}{2}x^{-1}). \quad (3.4a)$$

Further, we have

$$P_l^{(n,-n)}(z)=(l+1)_n(z+1)^{n/2}(z-1)^{-n/2}\mathfrak{P}_l^{-n}(z),$$

where $\mathfrak{P}_l^{-n}$ is Legendre's function of the first kind. Substitution of this expression yields

$$A_l=x^l c_{l\gamma}^{-1}\gamma^2\sum_{n=0}^{l}\frac{\epsilon_n}{n^2+\gamma^2}\frac{\Gamma(l+1)}{\Gamma(l-n+1)}\mathfrak{P}_l^{-n}(\tfrac{1}{2}x+\tfrac{1}{2}x^{-1}),$$
$$0<x<1. \quad (3.4b)$$

When $x>1$ the Legendre function here has to be multiplied by $(-)^n$.

From Eq. (3.2a) one can find an expression containing either ${}_2F_1(\cdots;\tfrac{1}{2})$ or ${}_2F_1(\cdots;-1)$ or ${}_2F_1(\cdots;2)$. It turns out that the formula with ${}_2F_1(\cdots;2)$ is the more convenient one. We obtain this formula by using the binomial expansion, which yields

$$A_l=4^{-l}i\gamma c_{l\gamma}^{-1}\int_0^1(1-t)^{i\gamma-l-1}$$
$$\times\sum_{m=0}^{l}\binom{l}{m}(2-t)^{2m}(-x^2t^2)^{l-m}.$$

By again using the binomial expansion,

$$(2-t)^{2m}=\sum_{n=0}^{2m}\binom{2m}{n}2^n(-t)^{2m-n},$$

the integration can be performed, with the result,

$$\int_0^1(1-t)^{i\gamma-l-1}t^{2l-n}dt$$
$$=\Gamma(2l-n+1)\Gamma(i\gamma-l)/\Gamma(i\gamma+l-n+1).$$

In this way we get

$$A_l=4^{-l}i\gamma c_{l\gamma}^{-1}(2l)![\Gamma(i\gamma-l)/\Gamma(l+1+i\gamma)]$$
$$\times\sum_{m=0}^{l}\binom{l}{m}(-x^2)^{l-m}\sum_{n=0}^{2m}(-2m)_n(-l-i\gamma)_n$$
$$\times 2^n/[n!(-2l)_n].$$

The sum $\Sigma_n$ is a terminating hypergeometric series for which we write ${}_2F_1(-2m,-l-i\gamma;-2l;2)$. One should be careful here, since the third parameter, $-2l$, is a nonpositive integer. By using expression (2.11) for $c_{l\gamma}$ we obtain

$$A_l=4^{-l}\binom{2l}{l}\sum_{m=0}^{l}\binom{l}{m}(-)^m x^{2l-2m}$$
$$\times{}_2F_1(-2m,-l-i\gamma;-2l;2).$$

We note that $A_l$ is a function of $\gamma^2$ rather than of $\gamma$, as can be seen from Eq. (3.3b). So we have, by replacing $m$ by $l-n$,

$$A_l=4^{-l}\binom{2l}{l}\sum_{n=0}^{l}\binom{l}{n}(-)^{l-n}x^{2n}$$
$$\times{}_2F_1(2n-2l,-l\pm i\gamma;-2l;2). \quad (3.5a)$$

The hypergeometric function ${}_2F_1(\cdots;2)$ can be expressed in terms of a Jacobi polynomial with argument 0. By using Ref. 3, p. 212, we have

$$A_l=\frac{(-)^l}{l!}\sum_{n=0}^{l}\frac{(2n)!}{n!}\frac{(2l-2n)!}{(l-n)!}(-\tfrac{1}{4}x^2)^n$$
$$\times P_{2l-2n}^{(2n-l+i\gamma,2n-l-i\gamma)}(0), \quad (3.5b)$$

or

$$A_l=\frac{1}{l!}(\tfrac{1}{4}x^2)^l\sum_{n=0}^{l}\frac{(2n)!}{n!}\frac{(2l-2n)!}{(l-n)!}(-\tfrac{1}{4}x^2)^{-n}$$

$$\times P_{2n}^{(l-2n+i\gamma,l-2n-i\gamma)}(0)\,. \qquad (3.5c)$$

Now we come to the derivation of the most elegant formula for $A_l$, i.e., a generalized hypergeometric function ${}_3F_2$ with argument $1-x^2$. From Eq. (3.2a) we have

$$A_l=i\gamma c_{l\gamma}^{-1}\int_0^1(1-t)^{i\gamma-l-1}[1-t+\tfrac{1}{4}(1-x^2)t^2]^l dt.$$

After substitution of

$$[1-t+\tfrac{1}{4}(1-x^2)t^2]^l=\sum_{n=0}^{l}\binom{l}{n}(1-t)^{l-n}t^{2n}2^{-2n}$$

$$\times(1-x^2)^n,$$

we can perform the integration, the result being

$$\int_0^1(1-t)^{i\gamma-1-n}t^{2n}dt=\Gamma(i\gamma-n)\Gamma(2n+1)/$$

$$\Gamma(i\gamma+n+1).$$

In this way we obtain

$$A_l=c_{l\gamma}^{-1}\sum_{n=0}^{l}\frac{(2n)!}{n!}\frac{(-l)_n}{(1+i\gamma)_n(1-i\gamma)_n}2^{-2n}(1-x^2)^n. \qquad (3.6a)$$

By using the doubling formula for the gamma function we have

$$(2n)!=(\tfrac{1}{2})_n2^{2n}n!,$$

and so

$$A_l=c_{l\gamma}^{-1}\,{}_3F_2(-l,1,\tfrac{1}{2};1+i\gamma,1-i\gamma;1-x^2). \qquad (3.6b)$$

An alternative expression is

$$A_l=\sum_{n=0}^{l}\frac{\Gamma(l+1+i\gamma)}{\Gamma(n+1+i\gamma)}\frac{\Gamma(l+1-i\gamma)}{\Gamma(n+1-i\gamma)}$$

$$\times\frac{(\tfrac{1}{2})_n(x^2-1)^n}{\Gamma(l+1)\Gamma(l+1-n)}, \qquad (3.6c)$$

where we have inserted Eq. (2.11). Furthermore, we have the terminating hypergeometric series,

$$A_l=\frac{(\tfrac{1}{2})_l}{l!}(x^2-1)^l\sum_{n=0}^{l}\frac{(i\gamma-l)_n(-i\gamma-l)_n}{(\tfrac{1}{2}-l)_n}\frac{(1-x^2)^{-n}}{n!}. \qquad (3.6d)$$

From Eq. (3.6c) one can derive an expression involving a ${}_3F_2$ with argument 1. By inserting

$$(x^2-1)^n=\sum_{m=0}^{n}\binom{n}{m}x^{2m}(-)^{n-m}$$

in (3.6c) and introducing the new summation variable $\nu=n-m$, we have

$$\sum_{n=0}^{l}\sum_{m=0}^{n}\cdots=\sum_{m=0}^{l}\sum_{\nu=0}^{l-m}\cdots.$$

It turns out that the sum $\Sigma_\nu$ is a ${}_3F_2(\cdots;1)$, and thus we obtain

$$A_l=\sum_{n=0}^{l}\frac{x^{2n}(\tfrac{1}{2})_n}{\Gamma(l+1)\Gamma(l+1-n)}$$

$$\times\frac{\Gamma(l+1+i\gamma)\Gamma(l+1-i\gamma)}{\Gamma(n+1+i\gamma)\Gamma(n+1-i\gamma)}$$

$$\times{}_3F_2(n-l,n+1,n+\tfrac{1}{2};n+1+i\gamma,n+1-i\gamma;1). \qquad (3.7)$$

We transform this ${}_3F_2$ into a ${}_3F_2$ with different parameters by applying a generalization of Dixon's theorem, see Slater (Ref. 4, p. 52),

$${}_3F_2(n-l,n+1,n+\tfrac{1}{2};n+1+i\gamma,n+1-i\gamma;1)$$

$$=\Gamma\begin{bmatrix}l-n+\tfrac{1}{2}, & n+1+i\gamma, & n+1-i\gamma\\ \tfrac{1}{2}, & l+1, & n+1\end{bmatrix}$$

$$\times\ {}_3F_2(i\gamma,-i\gamma,l-n+\tfrac{1}{2};\tfrac{1}{2},l+1;1).$$

Then we have from (3.7),

$$A_l=\frac{\Gamma(l+1+i\gamma)\Gamma(l+1-i\gamma)}{\Gamma^2(l+1)}$$

$$\times\sum_{n=0}^{l}\frac{x^{2n}}{n!}\frac{(\tfrac{1}{2})_n\Gamma(l-n+\tfrac{1}{2})}{\Gamma(\tfrac{1}{2})\Gamma(l-n+1)}$$

$$\times{}_3F_2(i\gamma,-i\gamma,l-n+\tfrac{1}{2};\tfrac{1}{2},l+1;1). \qquad (3.8a)$$

Note that the hypergeometric series for this ${}_3F_2$ breaks off when $i\gamma=0,-1,-2,\cdots$. The case $i\gamma=0$ corresponds to no Coulomb interaction at all. On the other hand, $i\gamma=-1,-2,-3,\cdots$ occurs for the Coulomb bound states.

It is not difficult to derive from Eq. (3.8a) the corresponding series with *decreasing* powers of $x$. This expression has almost exactly the same form as (3.8a), namely,

$$A_l=\frac{\Gamma(l+1+i\gamma)\Gamma(l+1-i\gamma)}{\Gamma^2(l+1)}\sum_{n=0}^{l}\frac{x^{2l-2n}}{n!}$$

$$\times\frac{(\tfrac{1}{2})_n\Gamma(l-n+\tfrac{1}{2})}{\Gamma(\tfrac{1}{2})\Gamma(l-n+1)}$$

$$\times{}_3F_2(i\gamma,-i\gamma,n+\tfrac{1}{2};\tfrac{1}{2},l+1;1). \qquad (3.8b)$$

By comparing this expression with Eq. (3.5c) we get the interesting equality

$${}_3F_2(i\gamma,-i\gamma,n+\tfrac{1}{2};\tfrac{1}{2},l+1;1)$$

$$=\frac{(-4)^n(l-n)!l!n!}{\Gamma(l+1+i\gamma)\Gamma(l+1-i\gamma)}P_{2n}^{(l-2n+i\gamma,l-2n-i\gamma)}(0).$$

In the particular case when $l=2n$ this expression can be simplified. By using (e.g., Ref. 3, p. 167)

$$\Gamma(1+\mu)P_\nu^{-\mu}(0) = {}_2F_1(-\nu,\nu+1;\mu+1;\tfrac{1}{2})$$

$$= \Gamma(1+\tfrac{1}{2}\mu)\Gamma(\tfrac{1}{2}+\tfrac{1}{2}\mu) \times[\Gamma(1+\tfrac{1}{2}\mu+\tfrac{1}{2}\nu)\Gamma(\tfrac{1}{2}+\tfrac{1}{2}\mu-\tfrac{1}{2}\nu)]^{-1},$$

we get

$$P_{2n}^{(i\gamma,-i\gamma)}(0) = 2^{2n}\Gamma(\tfrac{1}{2}+\tfrac{1}{2}i\gamma+n)[\Gamma(\tfrac{1}{2}+\tfrac{1}{2}i\gamma-n)\Gamma(2n+1)]^{-1} = \frac{(-)^n n!}{(\frac{1}{2})_n}\binom{n-\frac{1}{2}+\frac{1}{2}i\gamma}{n}\binom{n-\frac{1}{2}-\frac{1}{2}i\gamma}{n}, \quad (3.8c)$$

and so

$${}_3F_2(i\gamma,-i\gamma,n+\tfrac{1}{2};\tfrac{1}{2},2n+1;1) = \frac{\pi\Gamma^2(n+1)}{\Gamma(\frac{1}{2}+\frac{1}{2}i\gamma)\Gamma(\frac{1}{2}-\frac{1}{2}i\gamma)\Gamma(n+1+\frac{1}{2}i\gamma)\Gamma(n+1-\frac{1}{2}i\gamma)},$$

cf. Eqs. (2.3) and (3.13) of Ref. 4.

One can see from Eq. (3.6c) in particular that the degree of the polynomial $A_l \equiv A_l(x^2;\gamma^2)$ is $l$, both in $x^2$ and in $\gamma^2$,

$$A_l = \sum_{n=0}^{l} x^{2l-2n} D_n^{(l)}(\gamma^2), \quad (3.9a)$$

$$A_l = \sum_{n=0}^{l} \gamma^{2l-2n} F_n^{(l)}(x^2). \quad (3.9b)$$

Here $D_n^{(l)}$ and $F_n^{(l)}$ are certain polynomials of degree $n$. It turns out that Eq. (3.9b) is less suitable for practical applications, so we shall mainly restrict ourselves to the expansion in the $D_n^{(l)}$'s. One can also write $A_l$ as

$$A_l = \sum_{n=0}^{l}\sum_{m=0}^{n} x^{2l-2n}\gamma^{2m}a_{n,m}^{(l)}. \quad (3.10)$$

Here the coefficients $a_{n,m}^{(l)}$ are *real positive* numbers, as can be proven with the help of Eq. (3.8).

It is of interest to discuss a number of special cases. In the first place we consider the zero-energy case, $k=0$. Recalling $x \equiv q/k$ and $\gamma \equiv -s/k$, we have from Eq. (3.6c),

$$A_l \simeq \gamma^{2l}(l!)^{-2}\,{}_3F_0(-l,1,\tfrac{1}{2};-x^2/\gamma^2), \quad k\to 0,$$

and so

$$\lim_{k\to 0}\gamma^{-2l}A_l = (l!)^{-2}\,{}_3F_0(-l,1,\tfrac{1}{2};-q^2/s^2). \quad (3.11)$$

On the other hand, for $k\to\infty$ we have $x\to 0$ and $\gamma\to 0$. In this case we get from Eq. (3.8),

$$A_l(0;0) = a_{l,0}^{(l)} = \frac{(\frac{1}{2})_l}{l!} = 4^{-l}\binom{2l}{l}. \quad (3.12)$$

For $x=1$ one easily derives from Eq. (3.6b)

$$A_l(1;\gamma^2) = c_{l\gamma}^{-1} = \binom{l+i\gamma}{l}\binom{l-i\gamma}{l}. \quad (3.13)$$

The numbers $a_{n,m}^{(l)}$ $(n,m=0,1,\ldots,l)$ can be considered as a matrix, which is triangular because of

$$a_{n,m}^{(l)} = 0, \quad n<m.$$

The matrix elements on the principal axis are given by

$$a_{n,n}^{(l)} = \frac{4^{n-l}(2l-2n)!}{l!n!(l-n)!}. \quad (3.14)$$

In particular for $n=l$ one has

$$a_{l,l}^{(l)} = F_0^{(l)} = (l!)^{-2}. \quad (3.15)$$

Equation (3.14) is obtained by considering

$$D_n^{(l)}(\gamma^2) = \sum_{m=0}^{n}\gamma^{2m}a_{n,m}^{(l)}$$

and

$$D_n^{(l)}(\gamma^2) = (-)^n 4^{n-l}\frac{(2n)!(2l-2n)!}{l!n!(l-n)!}P_{2n}^{(l-2n+i\gamma,l-2n-i\gamma)}(0)$$

$$= (-)^n 4^{-l}\binom{2l}{l}\binom{l}{n}\,{}_2F_1(-2n,-l\pm i\gamma;-2l;2). \quad (3.16)$$

It is interesting to note the connection of $D_n^{(l)}$ with certain known polynomials, namely Krawtchouk's polynomials $k_n(z)$, which depend in addition on a positive variable $p<1$ and a positive integer $N$. These polynomials are associated with the binomial distribution in probability theory. According to Refs. 5 and 6 one has, with $p=\frac{1}{2}$ and $N=2l$,

$$k_{2n}(i\gamma+l) = 4^{-n}\binom{i\gamma+l}{2n}\,{}_2F_1(-2n,i\gamma-l;1+l+i\gamma-2n;-1) = P_{2n}^{(l-2n\pm i\gamma,l-2n\mp i\gamma)}(0). \quad (3.17)$$

Since $k_n(z)$ is defined for an integer variable $z$ only, $D_n^{(l)}$ may be considered as a generalization of $k_{2n}$.

For $\gamma=0$ we get from Eqs. (3.4a) and (3.6b),

$$A_l(x^2;0) = x^l P_l(\tfrac{1}{2}x+\tfrac{1}{2}x^{-1}) = {}_2F_1(-l,\tfrac{1}{2};1;1-x^2). \quad (3.18)$$

By using these expressions we obtain

$$a_{n,0}^{(l)} = a_{l-n,0}^{(l)} = D_n^{(l)}(0) = 4^{-l}\binom{2n}{n}\binom{2l-2n}{l-n}. \quad (3.19)$$

Further we derive from Eqs. (3.8c) and (3.15),

$$D_n^{(2n)}(\gamma^2) = \binom{n-\frac{1}{2}+\frac{1}{2}i\gamma}{n}\binom{n-\frac{1}{2}-\frac{1}{2}i\gamma}{n}, \quad (3.20)$$

which again shows the dependence on $\gamma^2$ rather than on $\gamma$. For $x=0$ we have from Eq. (3.15),

$$A_l(0;\gamma^2) = D_l^{(l)}(\gamma^2) = (-)^l\binom{2l}{l}P_{2l}^{(i\gamma-l,-i\gamma-l)}(0). \quad (3.21)$$

In order to obtain explicit expressions for $A_0, A_1,\cdots$, Eq. (3.6) is very useful. We first recast Eq. (3.6c) in a more explicit form,

$$A_l = \binom{l+i\gamma}{l}\binom{l-i\gamma}{l}\sum_{n=0}^{l}\frac{(-l)_n(\frac{1}{2})_n}{(1+i\gamma)_n(1-i\gamma)_n}(1-x^2)^n$$

$$= (l!)^{-2}\sum_{\nu=0}^{l}(-x^2)^\nu\sum_{n=\nu}^{l}\binom{n}{\nu}(-l)_n(\tfrac{1}{2})_n\prod_{m=n+1}^{l}(m^2+\gamma^2). \quad (3.22)$$

Therefore, we have

$$D_{l-\nu}^{(l)}(\gamma^2) = (l!)^{-1} \sum_{n=\nu}^{l} \binom{n}{\nu} \frac{(-)^{n+\nu}(\frac{1}{2})_n}{\Gamma(l+1-n)} \prod_{m=n+1}^{l} (m^2+\gamma^2). \tag{3.23}$$

In particular,

$$D_0^{(l)} = \frac{\Gamma(l+\frac{1}{2})}{l!\Gamma(\frac{1}{2})},$$
$$D_1^{(l)} = \frac{\Gamma(l-\frac{1}{2})}{l!\Gamma(\frac{1}{2})}(\gamma^2 + \tfrac{1}{2}l), \tag{3.24}$$
$$D_2^{(l)} = \frac{\Gamma(l-\frac{3}{2})}{l!\Gamma(\frac{1}{2})} \tfrac{1}{2}[\gamma^4 + \gamma^2(3l-2) + \tfrac{3}{4}l(l-1)],$$
$$D_3^{(l)} = \frac{\Gamma(l-\frac{5}{2})}{l!\Gamma(\frac{1}{2})} \frac{1}{6}\Big[\gamma^6 + \gamma^4(\tfrac{15}{2}l - 10) + \tfrac{1}{4}\gamma^2(45l^2 - 105l + 46) + \tfrac{15}{8}l(l-1)(l-2)\Big].$$

Finally, we give the first four polynomials $A_l$ in explicit form,

$$A_0 = 1,$$
$$A_1 = \tfrac{1}{2}(x^2 + 1 + 2\gamma^2),$$
$$A_2 = \tfrac{1}{8}[3x^4 + 2x^2(1+\gamma^2) + 3 + 8\gamma^2 + 2\gamma^4],$$
$$A_3 = \tfrac{1}{48}[15x^6 + 3x^4(3+2\gamma^2) + x^2(9 + 14\gamma^2 + 2\gamma^4) + \tfrac{1}{3}(45 + 136\gamma^2 + 50\gamma^4 + 4\gamma^6)]. \tag{3.25}$$

## ACKNOWLEDGMENTS

I would like to thank Dr. T.H. Koornwinder and Dr. N.M. Temme for a number of interesting discussions. This investigation forms a part of the research program of the Foundation for Fundamental Research of Matter (FOM), which is financially supported by the Netherlands Organization for Pure Scientific Research (ZWO).

[1]H. van Haeringen, J. Math. Phys. **19**, 1379 (1978).
[2]M.G. Fuda, Phys. Rev. C **14**, 37 (1976).
[3]W. Magnus, F. Oberhettinger, and R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, New York, 1966).
[4]L.J. Slater, *Generalized Hypergeometric Functions* (Cambridge U.P., London, 1966).
[5]*Higher Transcendental Functions*, Bateman Manuscript Project, edited by A. Erdélyi (McGraw-Hill, New York, 1953), Vol. 2, pp. 221, 224.
[6]G. Szegö, *Orthogonal Polynomials* (American Mathematical Society, Providence, Rhode Island, 1975), p. 35.

# Petrov type *N* vacuum metrics and homothetic motions

W. D. Halford

*Department of Mathematics, Massey University, Palmerston North, New Zealand*
(Received 27 July 1978)

Petrov type *N* vacuum spaces which admit an expanding and/or twisting principal null congruence and a homothetic motion are considered. It is shown that there are no such spaces which admit two Killing vectors, or one Killing vector of special type. If there are no Killing vectors present, the form of the homothetic Killing vector is restricted to one possibility.

## 1. INTRODUCTION

A *conformal motion* of a space with metric tensor g is defined by $\mathcal{L}_\chi g = \varphi(x)g$ where $\mathcal{L}$ is the lie derivative in the direction of the *conformal Killing vector* $\chi$ (infinitesimal generator of the motion) and $\varphi$ is a function of the local coordinates $(x)$. If $\varphi =$ nonzero constant, the motion is *homothetic* and $\chi$ is a *homothetic Killing vector* (HKV). If $\varphi = 0$, the motion is *isometric* and $\chi$ is a *Killing vector* (KV). An important result is that any space admits at most one independent HKV.[1]

Petrov type N solutions of Einstein's vacuum field equations are among the most interesting, physically and mathematically, of all empty-space metrics. For example, Collinson[2] proved that the only curvature collineations[3] admitted by a vacuum space—time not of Petrov type N are conformal motions, but that type N vacuum spaces *do* admit curvature collineations which are *not* conformal motions (i.e., they admit more general types of symmetry). Collinson and French[4] proved that a conformal motion of nonflat empty space—time must be homothetic unless the space—time is Petrov type N with hypersurface-orthogonal geodesic rays. The only type N vacuum fields which admit (proper) conformal motions are the *pp* waves.[5] McIntosh[6] has given an example of a *pp* wave admitting a particular homothetic Killing vector.

In this paper we are concerned with Petrov type N vacuum spaces with expanding and/or twisting principal ray congruences (i.e., not *pp* waves). Such spaces may have no symmetries. If there are any isometries present, they depend upon whether the congruence is (i) twisting, or (ii) twist-free. In case (i) Collinson[7] has shown that there exists at most one KV in the space. An example with one KV is the metric found by Hauser.[8] In case (ii) Held[9] proved that the space admits at most two KVs. Our interest centers on type N vacuum metrics which admit a homothetic Killing vector and possibly one or two KVs.

## 2. VACUUM SPACES WITH TWO KILLING VECTORS

In another paper,[10] referred to as Paper I, we listed all nonflat algebraically special vacuum metrics with nonzero complex divergence which admit a HKV and 2, 3, or 4 KVs. None of these were of Petrov type N. Hence we have

*Theorem* 1: There are no Petrov type N vacuum spaces with nonzero expansion and/or twist which admit 2 KVs and one HKV.

## 3. VACUUM SPACES WITH ONE KILLING VECTOR

In order to obtain our next results we use the formalism developed by Debney, Kerr, and Schild and by Kerr and Debney,[11] extended in Paper I. In the $(\zeta, \bar{\zeta}, u, v)$ coordinate system the space is type N when

$$\bar{D}\partial_u D\Omega = 0 = \mu, \tag{1}$$

and the only surviving field equation is

$$\bar{D}\bar{D}D\Omega = DD\bar{D}\bar{\Omega}, \tag{2}$$

where the bar denotes complex conjugation, the operator $D$ is defined by

$$D = \partial_\zeta - \Omega\partial_u,$$

and $\Omega = \Omega(\zeta, \bar{\zeta}, u)$. The "complex mass" function $\mu$ vanishes for type N vacuum.

If there is one Killing vector $K$ it assumes one of the canonical forms

(i) $K = F(\zeta, \bar{\zeta})\partial_u$, (ii) $K = \partial_\zeta + \partial_{\bar{\zeta}}$.

Kerr and Debney[11] showed that if $K$ takes the form (i), then $\ddot{\Omega} = 0$ (dots denote differentiation with respect to $u$). This implies $\Omega = uf(\zeta, \bar{\zeta}) + g(\zeta, \bar{\zeta})$ so that $\partial_u\partial_u D\Omega = 0$. Coupled with (1) this implies flat space. Hence we have

*Theorem* 2: There are no Petrov type N vacuum spaces with nonzero expansion and/or twist which admit a KV of the form

$$K = F(\zeta, \bar{\zeta})\partial_u.$$

In Paper I we showed that the general form of a HKV admitted by an algebraically special vacuum space with nonzero complex divergence is (in these coordinates)

$$H = \alpha\partial_\zeta + \bar{\alpha}\partial_{\bar{\zeta}} + \mathrm{Re}(\alpha_\zeta)(u\partial_u - v\partial_v) + R\partial_u + a(u\partial_u + v\partial_v),$$

where $\alpha = \alpha(\zeta)$, $R = R(\zeta, \bar{\zeta}) = \bar{R}$, $a$ is a nonzero real constant, and $\alpha_\zeta \equiv \partial_\zeta\alpha$. If $H$ is to be in the space along with the KV

$$K = \partial_\zeta + \partial_{\bar{\zeta}} = \partial_x \quad (\zeta = x + iy), \tag{3}$$

we must determine the form of $H$ more precisely from the commutation relation

$$[K, H] = \lambda K, \quad \lambda \text{ real constant.}$$

This requires $\alpha = \lambda\zeta + e$, $R = R(y)$, where $\lambda$ and $e$ are constants, with $e$ complex. We can simplify the form of $H$ by means of a coordinate transformation

$$\zeta \to \zeta' = \lambda\zeta + e \equiv \Phi(\zeta), \quad u \to u' = u + S(y),$$

 0022-2488/79/061115-03$01.00  

and by using the appropriate transformations on $\alpha$ and $R$ (see Paper I):

$$\alpha \to \alpha' = \Phi_\zeta \alpha,$$

$$R \to R' = |\Phi_\zeta| [R - (\mathrm{Re}(\alpha_\zeta) + a)S + HS],$$

where $S = S(\zeta, \bar\zeta) = \bar S$. By solving $HS - (\lambda + a)S - R = 0$ we send $R$ to zero and so put $H$ into the form

$$H = \zeta\partial_\zeta + \bar\zeta\partial_{\bar\zeta} + (a+1)u\partial_u + (a-1)v\partial_v$$
$$= x\partial_x + y\partial_y + (a+1)u\partial_u + (a-1)v\partial_v, \tag{4}$$

where we have dropped the primes on the new variables and absorbed the constant $\lambda$ into $H$ in the process. These transformations do not alter the form of the Killing vector $K$.

The nontrivial Killing equations and homothetic Killing equations and their first order integrability conditions (see Paper I) reduce to

$$K(\Omega - u\dot\Omega) = 0, \tag{5}$$

$$K\dot\Omega = 0, \tag{6}$$

$$K(\bar D\dot\Omega) = 0, \tag{7}$$

$$K\Delta = 0, \tag{8}$$

$$K\ddot\Omega = 0, \tag{9}$$

with $K$ in form (3), and

$$(H - a)(\Omega - u\dot\Omega) = 0, \tag{10}$$

$$(H+1)\dot\Omega = 0, \tag{11}$$

$$(H+2)(\bar D\dot\Omega) = 0, \tag{12}$$

$$(H + 1 - a)\Delta = 0, \tag{13}$$

$$(H + 2 + a)\ddot\Omega = 0, \tag{14}$$

with $H$ in form (4). Equations (5) and (6) imply

$$\Omega = \Omega(y, u), \tag{15}$$

and this with Eqs. (10) and (11) gives two possibilities:

$$\text{(i) } a = -1, \quad \Omega = y^{-1} f(u), \tag{16}$$

$$\text{(ii) } a \neq -1, \quad \Omega = u^{a/(a+1)} g(y^{a+1}/u). \tag{17}$$

The functions $f$ and $g$ are to be determined from Eqs. (1) and (2). In both cases the integrability conditions (7), (9), (12), and (14) are satisfied identically. The function $\Delta = \Delta(x, y, u)$ is defined by

$$\Delta = i\,\mathrm{Im}(\bar D\Omega).$$

Substituting this expression into (8) and (13) places an additional constraint on $\Omega$. The space is twist-free if and only if $\Delta = 0$.

Case (i): Substituting (16) into (1) and (2) gives

$$i\dot E + \bar f \ddot E = 0, \tag{18}$$

and

$$3E - \dot E(2i\bar f - \bar E) = 3\bar E + \dot{\bar E}(2if + E), \tag{19}$$

where

$$E(u) = if - 2f\dot f,$$

and the dot denotes differentiation with respect to $u$. The complexity of this set of equations is such that so far we have found no solutions $f(u)$ for which the space is nonflat ($\ddot E \neq 0$ for type N). A possibility is that the LHS of Eq. (19) is a real constant $k$, when (18) and (19) give

$$2\dot z^3 \dddot z + \dot z^2 \ddot z^2 - 3z\ddot z^3 = 0, \tag{20}$$

where $z = E + k$. Equation (20) has the solution

$$Az = p - Cp^{3/2}, \quad Au + B = 3Cp^{1/2} - \log p, \tag{21}$$

expressed parametrically, where $p$ is the parameter and $A$, $B$, $C$ are arbitrary constants. Having found $z$ and therefore $E$, Eq. (18) must be satisfied for values of $f(u)$ determined from $E$.

Case (ii): The substitution of (17) into Eqs. (1) and (2) yields complex equations which have defied any attempts at obtaining solutions so far.

Hauser's solution: The twisting type N solution of Hauser[8] admits one KV of the form (3) and also admits a HKV of the form (4). In Hauser's coordinates these vectors are

$$K = \partial_\sigma, \quad H = 2\sigma\partial_\sigma + 2\rho\partial_\rho + 4\xi\partial_\xi + 3u\partial_u.$$

It remains to extract his solution from Eqs. (1), (2), and (15).

## 4. VACUUM SPACES WITH NO ISOMETRIES

In case the space admits just one HKV (and no KVs) we shall consider each of the canonical forms

$$\text{(i) } H = u\partial_u + v\partial_v, \quad \text{(ii) } H = \partial_\zeta + \partial_{\bar\zeta} + u\partial_u + v\partial_v.$$

The nontrivial homothetic Killing equations and their first order integrability conditions reduce to

$$(H-1)(\Omega - u\dot\Omega) = 0, \tag{22}$$

$$H\dot\Omega = 0, \tag{23}$$

$$H(\bar D\dot\Omega) = 0, \tag{24}$$

$$(H-1)\Delta = 0, \tag{25}$$

$$(H+1)\ddot\Omega = 0. \tag{26}$$

Case (i): Equations (22) and (23) with $H$ in the form (i) imply $\Omega = uh(\zeta, \bar\zeta)$. Then $\partial_u\partial_u D\Omega = 0$, and this together with (1) ensures that the space is flat. Hence we have

*Theorem* 3: There are no Petrov type N vacuum spaces with nonzero expansion and/or twist which admit only an HKV of the form

$$H = u\partial_u + v\partial_v.$$

Case (ii): It is possible that type N spaces of the kind under consideration exist which admit just one HKV,

$$H = \partial_\zeta + \partial_{\bar\zeta} + u\partial_u + v\partial_v.$$

However, the complexity of the high order nonlinear partial differential equation coming from (1) and (2) is such that no solutions have yet been obtained.

## ACKNOWLEDGMENTS

I am grateful to Professor R. P. Kerr for his advice, and I thank Dr. C. B. G. McIntosh for pointing out that Hauser's solution admits a homothetic motion.

[1]K. Yano, J. Indian Math. Soc. **15**, 105 (1951). See also T. Suguri and S. Ueno, Tensor (N. S.) **24**, 253 (1972); D.M. Eardley, Commun. Math. Phys. **37**, 287 (1974); C.B.G. McIntosh, Gen. Rel. Grav. **7**, 199 (1976).
[2]C.D. Collinson, J. Math. Phys. **11**, 818 (1970).
[3]G.H. Katzin, J. Levine, and W.R. Davis, J. Math. Phys. **10**, 617 (1969).
[4]C.D. Collinson and D.C. French, J. Math. Phys. **8**, 701 (1967).
[5]H.W. Brinkmann, Math. Ann. **91**, 269 (1924); **94**, 119 (1925); J. Ehlers and W. Kundt, in *Gravitation: An Introduction to Current Research* (Wiley, New York, 1962), p. 80; C.D. Collinson, see Ref. 2; A. Thompson, Phys. Rev. Lett. **34**, 507 (1975).
[6]C.B.G. McIntosh, Gen. Rel. Grav. **7**, 215 (1976).
[7]C.D. Collinson, J. Phys. A **2**, 621 (1969).
[8]I. Hauser, Phys. Rev. Lett. **33**, 1112 (1974).
[9]A. Held, J. Math. Phys. **17**, 39 (1976).
[10]W.D. Halford and R.P. Kerr, "Einstein spaces and Lomothetic motions. I.," to appear in J. Math. Phys. (1979). See also W.D. Halford, Ph.D. thesis (unpublished, 1977).
[11]G.C. Debney, R.P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969); R.P. Kerr and G.C. Debney, J. Math. Phys. **11**, 2807 (1970).

# Circular motion for a time-asymmetric relativistic two-body problem

Donald E. Fahnline

*The Pennsylvania State University, Altoona, Pennsylvania 16603*
(Received 9 October 1978)

A formalism developed in a previous paper yields necessary and sufficient conditions and a solution for the circular motion case of the time-asymmetric relativistic two-body problem in which one particle responds to the retarded Liénard–Wiechert field of a second, while the second responds to the advanced field of the first. The necessary conditions contradict Künzle's exceptional circular motion solution with zero angular momentum; consequently, zero angular momentum implies one-dimensional motion. The limit in which the mass of one of the particles becomes infinite commutes with the nonrelativistic limit and reduces the solution properly to the circular motion solution of the relativistic one-body Coulomb problem.

## INTRODUCTION

This paper discusses the circular motion case of a relativistic action-at-a-distance two-body problem due to Fokker[1]: one spinless electrically charged particle responds without self-action to the retarded Liénard–Wiechert field of a second, while the second responds similarly to the advanced field of the first.

In her study of the problem, Bruhns[2] showed that circular motion solutions exist similar to those found by Schild[3] for the Wheeler–Feynman time-symmetric problem. Künzle[4] found explicit solutions for the equal mass case, numerical solutions for arbitrary mass ratios, and an exceptional solution with zero total angular momentum. The existence of this exceptional solution prevented him from concluding that zero total angular momentum implies one-dimensional motion. Künzle's reduction of his circular motion equations in the limit where one of the masses becomes infinite did not produce the circular motion relations appropriate for the one-body relativistic Coulomb problem, and he found that this limit does not commute with the nonrelativistic limit.

The present paper uses a previously developed formalism,[5] which is reviewed in the first section, to establish necessary and sufficient conditions for circular motion and to obtain the general circular motion solution. That Künzle's exceptional solution does not satisfy the necessary conditions leads to proofs that zero angular momentum implies one-dimensional motion and that a possible singularity pointed out by Bruhns does not occur. Finally, the $m_2 \to \infty$ limit of the circular motion solution does produce the circular motion relations of the Coulomb problem, and this limit does commute with the nonrelativistic limit.

## REVIEW

The symbol $x_n^{\mu}$ $(n = 1,2)$ represents the Minkowski space coordinates of particle $n$, $\tau_n$ represents its proper time, and $v_n^{\mu} \equiv dx_n^{\mu}/d\tau_n$ represents its proper velocity obeying $v_n^{\mu} v_{n\mu} = -c^2$. The metric tensor is $g^{ii} = -g^{00} = 1, g^{\mu\nu} = 0$ for $\mu \neq \nu$.

A previous paper,[5] which should be read for the details of this brief outline, establishes the existence of a center of motion frame where the conserved total 4-momentum $P^{\mu}$ (assumed timelike and future pointing), the center of motion $x^{\mu}$, and the conserved total angular momentum $J^{\mu\nu}$ have zero components except for $P^0 = mc > 0$, $x^0 = c\tau_x$, and $J^{12} = -J^{21} \equiv J_3 \equiv J \geqslant 0$. In this frame the center of mass $z^{\mu}$, which is distinct from the center of motion, moves in a circle of radius $J/mc$ about the origin:

$$z^0 = x^0 = c\tau_x, \quad \mathbf{z} = \mathbf{J} \times \mathbf{r}/mcr, \tag{1}$$

where the separation between the particles $r^{\mu} \equiv x_1^{\mu} - x_2^{\mu}$ obeys the constraint $r^{\mu} r_{\mu} = 0$, $r \equiv |\mathbf{r}| > 0$, and $J_i \equiv \frac{1}{2}\epsilon_{ijk} J_{jk}$.

In terms of $z^{\mu}$ and $r^{\mu}$ the particle positions are

$$x_n^{\mu} = z^{\mu} - (-1)^n (m_f c\rho_f + g) r^{\mu}/mc\rho_x, \tag{2}$$

where the particle label $f = 2$ when $n = 1$ and vice versa, $m_n$ is the rest mass particle $n$, $g$ is the coupling constant in Gaussian units, $\rho_n \equiv -v_n^{\mu} r_{\mu}/c > 0$, and $\rho_x \equiv -v_x^{\mu} r_{\mu}/c$. In the center of motion frame $\rho_x$ reduces to $r$, the particles move in the plane perpendicular to $\mathbf{J}$, and $\mathbf{r}$ passes through the center of mass perpendicular to $\mathbf{z}$ as shown in Fig. 1.

Assuming $r^0 > 0$ for definiteness in the constraint $r^{\mu} r_{\mu} = 0$ yields

$$\sigma \equiv c\dot{\tau}_1 \rho_1 = c\dot{\tau}_2 \rho_2 > 0, \tag{3}$$

where a dot above a variable indicates differentiation with respect to an arbitrary scalar $s$ such that $\dot{x}_n^{\mu} \dot{x}_{n\mu} < 0$ and $\dot{x}_n^0 > 0$. It is natural and convenient to choose $s = \tau_x$, which determines $\sigma$ to be

$$\sigma = mc^2/(m_1 c\rho_1^{-1} + m_2 c\rho_2^{-1}). \tag{4}$$

The $\rho$'s are related by

$$mc\rho_x = m_1 c\rho_1 + m_2 c\rho_2 + 2g, \tag{5}$$

$$m^2c^2J^2/\eta = m^2c^2 - mc\rho_x(m_1 c\rho_1^{-1} + m_2 c\rho_2^{-1}), \tag{6}$$


0022-2488/79/061118-03$01.00

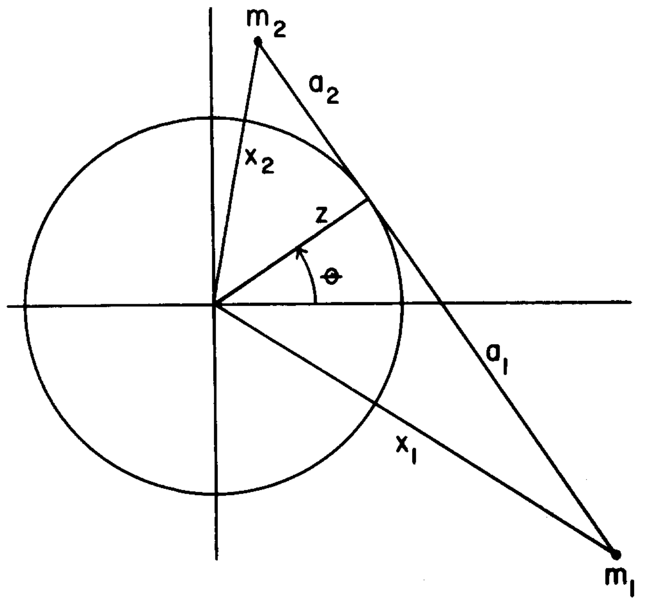


FIG. 1. Geometry of the time-asymmetric problem in the center of motion frame: $z = J/mc$, $a_1 = (m_2c\rho_2 + g)/mc$, $a_2 = (m_1c\rho_1 + g)/mc$.

and

$$m^2c^2J^2/\eta^2 = \dot r^\mu \dot r_\mu/\sigma^2 = 2\psi - \rho_1^{-2} - \rho_2^{-2} \geqslant 0, \qquad (7)$$

where $\eta = m_1c\rho_1 m_2c\rho_2 - g^2$ and $\psi \equiv -v_1^\mu v_{2\mu}/\rho_1\rho_2c^2 > 0$.

The problem now reduces to solving

$$m_1c\dot\rho_1/\sigma = m_1c/\rho_1 - m_1c\rho_1\psi - g/\rho_2^2, \qquad (8)$$

$$m_2c\dot\rho_2/\sigma = -m_2c/\rho_2 + m_2c\rho_2\psi + g/\rho_1^2, \qquad (9)$$

and

$$\dot\theta/\sigma = (\dot r^\mu \dot r_\mu)^{1/2}/\sigma\rho_x = mcJ/\rho_x\eta, \qquad (10)$$

where $\theta$ is the angle between $\mathbf{z}$ and the $x$ axis in the center of motion frame.

## CONDITIONS FOR CIRCULAR MOTION

Circular motion in the center of motion frame requires that $|\mathbf{x}_1|$ and $|\mathbf{x}_2|$ be constant. But (1) and (2) yield

$$|\mathbf{x}_n|^2 = |\mathbf{z}|^2 + (m_fc\rho_f + g)^2/m^2c^2, \qquad (11)$$

which shows that the $\rho_n$ must be constant:

$$\dot\rho_1 = \dot\rho_2 = 0. \qquad (12)$$

Using (12) in the derivatives of the time-components of (1) and (2) implies

$$\dot x_1^0 = \dot x_2^0 = \dot z^0 = \dot x^0 = c. \qquad (13)$$

Hence, (3) yields

$$\gamma_n \equiv (1 - |d\mathbf{x}_n/dt_n|^2/c^2)^{-1/2} = dt_n/d\tau_n = c\rho_n/\sigma, \qquad (14)$$

and (4) yields

$$m = m_1\gamma_1^{-1} + m_2\gamma_2^{-1} < m_1 + m_2, \qquad (15)$$

since $\gamma_n > 1$. Equation (13) also shows that $\dot r^\mu$ is spacelike for circular motion:

$$\dot r^\mu \dot r_\mu = |\dot{\mathbf{r}}|^2 > 0. \qquad (16)$$

Multiplying (8) by $m_2c\rho_2$ and (9) by $m_1c\rho_1$, adding and subtracting the resulting equations, and using (7) and (12) yield

$$m_1c\rho_2(m_2c\rho_2 + g) = m_2c\rho_1(m_1c\rho_1 + g) \qquad (17)$$

and

$$m_1c\rho_1m_2c\rho_2\dot r^\mu \dot r_\mu = -g\sigma^2(m_1c\rho_1^{-1} + m_2c\rho_2^{-1}). \qquad (18)$$

Equations (16) and (18) show that circular motion requires

$$g < 0. \qquad (19)$$

Equation (17) implies that if $m_1c\rho_1 + g \leqslant 0$, then $m_2c\rho_2 + g \leqslant 0$. Equation (5) would then give $mc\rho_x \leqslant 0$, which is contrary to the assumptions $\rho_x > 0$ and $m > 0$. This contradiction shows that

$$m_nc\rho_n + g > 0. \qquad (20)$$

Equations (19) and (20) give

$$\eta > 0; \qquad (21)$$

hence $\dot\theta$ is positive as well as constant.

Künzle's[4] exceptional circular motion solution with $J = 0$, $\eta = 0$, and $g > 0$ fails to satisfy either (19) or (21); this has an important implication for the $J = 0$ case. Although (10) shows that one-dimensional motion along a straight line through the origin of the center of motion frame requires $J = 0$, Künzle[4] found that the converse has difficulties: Setting $J = 0$ in (10) requires only that either $\dot\theta = 0$ or $\eta = 0$. However, if $J = 0$ and $\dot\theta \neq 0$ at any time, (10) and the continuity of $\dot\theta$ imply that $\eta = 0$ for an extended period of time. This restriction with (5), (6), and (7) forces $\rho_1$ and $\rho_2$ to be constants, and the motion must be circular. But circular motion requires condition (21), which contradicts $\eta = 0$. Hence $J = 0$ implies one-dimensional motion. This discussion also completes the answer to Bruhns' question about the possibility that $\eta = 0$ for nonlinear motion[2]: If $J = 0$, the motion is one-dimensional. If $J > 0$, then (7) requires $\eta \neq 0$.

Sufficient conditions for circular motion are simply $\dot\rho_1 = \dot\rho_2 = 0$ initially and $g < 0$. These conditions require by virtually the same arguments as before that (17), (18), (20), and (21) be true initially. Using (21) in (6) yields $\eta > J^2 > 0$ initially and, therefore, always. Hence there are no singularities in the pair of coupled ordinary differential equations resulting from using (7) to eliminate $\psi$ from (8) and (9):

$$m_1c\dot\rho_1/\sigma = \tfrac{1}{2}m_1c\rho_1(\rho_1^{-2} - \rho_2^{-2} - m^2c^2J^2\eta^{-2}) - g\rho_2^{-2}$$

and

$$m_2c\dot\rho_2/\sigma = \tfrac{1}{2}m_2c\rho_2(\rho_1^{-2} - \rho_2^{-2} - m^2c^2J^2\eta^{-2}) + g\rho_1^{-2},$$

where $\sigma$ is given by (4). The solution of these equations for the given initial conditions requires $\dot\rho_1 = \dot\rho_2 = 0$ always. Since $g < 0$, the particles cannot be at rest and must move in circles in the center of motion frame.

## CIRCULAR MOTION SOLUTION

For arbitrary given values of $m_1$, $m_2$, and $g < 0$, the most convenient parameter for expressing the circular motion solution of the time-asymmetric problem is one of the $\rho_n$'s, say $\rho_1$. Any value of $\rho_1$ such that $m_1c\rho_1 + g > 0$ determines via

(17) a single value of $\rho_2$ satisfying $m_2c\rho_2 + g > 0$:

$$\rho_2 = -g/2m_2c + [g^2/4m_2^2c^2 + \rho_1(\rho_1 + g/m_1c)]^{1/2}. \quad (22)$$

Using (22) in (5) gives the value of $mc\rho_x$. Solving (6), (7), and (18) first for $m^2c^2$ and then for $J^2$ yields their values in terms of those already found:

$$m^2c^2 = (mc\rho_x - g\eta/m_1c\rho_1m_2c\rho_2)(m_1c\rho_1^{-1} + m_2c\rho_2^{-1}) \quad (23)$$

and

$$J^2 = -g\eta^2(m_1c\rho_1^{-1} + m_2c\rho_2^{-1})/m^2c^2m_1c\rho_1m_2c\rho_2. \quad (24)$$

These values obey the inequalities $0 < m < (m_1 + m_2)$ and $0 < J^2 < \eta/2$.

Values for $\rho_x, \sigma, \gamma_n, |\mathbf{x}_n|, t_n, \tau_n$, and $\dot\theta$ now follow immediately from (22)–(24) and the equations in the preceding two sections. Finally, the angle $\alpha$ between $\mathbf{x}_1$ and $\mathbf{x}_2$ in the center of motion frame follows from

$$\cos\alpha = \mathbf{x}_1\cdot\mathbf{x}_2/|\mathbf{x}_1|\,|\mathbf{x}_2| = [J^2 - (m_1c\rho_1 + g)(m_2c\rho_2 + g)]/m^2c^2|\mathbf{x}_1|\,|\mathbf{x}_2|. \quad (25)$$

## THE ONE-BODY LIMIT

In the $m_2 \to \infty$ limit the time-asymmetric relativistic two-body problem should reduce to the relativistic Coulomb problem in which particle 1 responds to the static Coulomb field due to particle 2 at rest at the origin of the center of motion frame. Künzle[4] has already examined this limit for the circular motion case; but he does not obtain the expected Coulomb problem relations, and his $m_2 \to \infty$ limit does not commute with the nonrelativistic limit. His results may be obtained in the present context by holding $m_1$, $g$, and $m_2\rho_2$ fixed as the limit is taken, but this forces $\rho_2 \to 0$ and $\rho_1 \to -g/m_1c$.

Holding $m_1$, $g$, and $\rho_1$ fixed in the results of the last section as $m_2 \to \infty$ leaves the size of the orbits independent of the limit and produces the Coulomb problem relations:

$$\rho_x = \rho_2 = (\rho_1^2 + g\rho_1/m_1c)^{1/2}, \quad (26)$$

$$E/c \equiv mc - m_2c = \tfrac{1}{2}[m_1c(\rho_1\rho_2^{-1} + \rho_2\rho_1^{-1}) + g\rho_2^{-1}], \quad (27)$$

$$J^2 = -gm_1c\rho_1, \quad (28)$$

$$\dot\theta = J/m_1\rho_1\rho_2, \quad (29)$$

$$\mathbf{x}_2 = 0, \quad (30)$$

$$\gamma_2 = 1. \quad (31)$$

Furthermore, this limit does not commute with the nonrelativistic limit.

## DISCUSSION

This paper has given necessary and sufficient conditions and a general solution for the circular motion case of the time-asymmetric relativistic two-body problem. The solution reduces properly to the circular motion solution of the Kepler problem in the nonrelativistic limit and to that of the one-body relativistic Coulomb problem in the limit where one of the masses becomes infinite. One consequence of the necessary conditions is that zero total angular momentum implies one-dimensional motion along a line through the origin of the center of motion frame.

The circular motion solution in effect inverts the equations found by Bruhns[2] so that the masses of the two particles may be considered part of the given information. The similarity of her equations to Schild's[3] equations for the time-symmetric problem suggests that an inversion of the latter may be possible.


[1] A.D. Fokker, Physica **9**, 33 (1929).
[2] B. Bruhns, Phys. Rev. D **8**, 2370 (1973).
[3] A. Schild, Phys. Rev. **131**, 2762 (1963).
[4] H.P. Künzle, Int. J. Theor. Phys. **11**, 395 (1974).
[5] D.E. Fahnline, J. Math. Phys. **18**, 1006 (1977).

# Time delay in $N$-body scattering[a),b)]

D. Bollé[c)]

*Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium*

T. A. Osborn

*Cyclotron Laboratory, Department of Physics, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2*

(Received 10 July 1978)

We extend the theory of time delay to $N$-body scattering. The known results relating time delay to the $S$ matrix in the two-body and three-body problem suggest that these relationships are universal. Within the context of two-Hilbert space $N$-body scattering theory an abstract definition of time delay is provided. For all scattering processes initiated by the collision of two clusters a simple proof is constructed establishing the connection of time delay to the on-shell $S$ matrix and its energy derivatives. The definition of time delay and method of proof given here are compared with earlier approaches used in the three-body problem.

## I. INTRODUCTION

This paper gives a derivation of the relation between time delay and the $S$ matrix for the $N$-body scattering problem. Roughly speaking, time delay is the total spatial retardation of a scattering state induced by the collision process. This concept is defined by placing the center of a $3(N-1)$-dimensional sphere of radius $r$ about the $N$-body center-of-mass position. The duration time the asymptotic wavepacket (evolving without the inclusion of the intercluster interactions that cause multichannel scattering) resides in the sphere is subtracted from the corresponding time for the fully interacting wavepacket. Then the radius $r$ is taken into $\infty$. The resulting difference is the time delay. It is a function of the initial scattering channel, the incident asymptotic wavepacket, and all the interactions present in the $N$-body system. The basic problem in this theory is to find a universally valid definition of time delay and to relate this observable to the $S$ matrix.

Although time delay and its relation to the $S$ matrix has been extensively studied in the two-body problem[1-5] only recently has the theory been developed for the three-body problem.[6,7] The derivation given here extends the theory to $N$-body collisions for systems of particles interacting through short-range (non-Coulomb) forces. This derivation is simpler and at the same time more complete than the one used in the three-body case.

Our proof is carried out in the two-Hilbert space formalism appropriate for nonrelativistic $N$-body scattering.[8-11] It provides us with a simple and rigorous description of the $N$-body problem. The notation we adopt for the two-Hilbert-space theory is given in Sec. II. The definition of time delay is found in Sec. III. The basic result of this section is the transformation of the statement of the problem into a form in which the exact wavefunctions are replaced by asymptotic forms. The argument given here is a generalization of the one found by Martin[5] for the two-body case. This is a major simplification because the complicated exact $N$-body wavefunction has been taken out of the problem, leaving only simple asymptotic wavepackets. Section IV uses the generalized function properties of the momentum space matrix elements of the sphere to complete the derivation. This method is taken over from our earlier work on the three-body problem.[6] Section V contains our conclusions and the Appendix includes proofs of Fourier transform properties of the sphere that our derivation requires.

## II. $N$-BODY SCATTERING THEORY

In this section we define the two-Hilbert space multichannel theory that describes the scattering solutions of the $N$-body problem. The basic two-Hilbert space theory outlined here is the same as found in the work of Chandler and Gibson.[11] However, we must alter their notation in order to exhibit explicitly the individual asymptotic channels. These channel spaces and their associated matrix elements are required for the argument given in Sec. IV. The basic reason for employing the two-Hilbert space theory is that it gives us a general framework to describe on an equal footing both time-independent and time-dependent phenomena. In this framework one has simple and rigorous statements of probability conservation, channel orthogonality, and completeness. The associated operational calculus that accompanies this description of scattering is easy to use, carries an obvious physical content, and makes the ensuing proof possible. This theory is adequate for all collision processes not involving long-range (Coulomb-like) forces.

The $N$-body problem we discuss is that of $N$ distinguishable spinless particles which interact through pairwise potentials. We have ignored all spin and isospin labels. These degrees of freedom do not impose any difficulties in scattering theory. The physical behavior of the system is determined by its Hamiltonian. Let $H_0$ be the $N$-body kinetic energy operator in the center-of-mass frame and $V_{ij}$ the interaction between particles $i$ and $j$. The full Hamiltonian is

a)Work supported in part by a grant from the National Research Council of Canada, and by a NATO Research Grant.

b)One of the authors (D.B.) dedicates this paper to Professor L.P. Bouckaert on the occasion of his seventieth birthday.

c)Bevoegdverklaard Navorser N.F.W.O., Belgium.

  

$$H = H_0 + \sum_{i>j} V_{ij}. \tag{II.1}$$

This self-adjoint operator[10] acts on the Hilbert space, $\mathscr{H}$, of square integrable functions in $R^{3(N-1)}$. Here and in the following we systematically remove the total center-of-mass motion from the problem.

Each distinct asymptotic channel will be labeled by $\alpha$, and $\mathscr{A}$ will represent the collection of all channels $\alpha$. The symbol $\alpha$ denotes both a partition, $A$, of the $N$ particles into $N_\alpha$ clusters and the specification of the eigenfunction of each cluster. Set $A = \{al : l = 1, N_\alpha\}$. Here $al$ is the set containing the labels of the $nl$ particles in the $l$th cluster of partition $A$. The internal cluster wavefunctions are taken to be $\Psi^\alpha_{al}$. These bound state wavefunctions are normalized to unity. Consider cluster $al$. Let $h_0(al)$ denote the internal kinetic energy and $v(al)$ the total internal potential. The cluster Hamiltonian is

$$h(al) = h_0(al) + v(al), \tag{II.2}$$

and its eigenfunction $\Psi^\alpha_{al}$ satisfies

$$h(al)\Psi^\alpha_{al} = -\epsilon^\alpha_{al}\Psi^\alpha_{al}. \tag{II.3}$$

The cluster binding energy is $\epsilon^\alpha_{al}$. The total binding energy of all clusters in channel $\alpha$ is the sum

$$\epsilon^\alpha = \sum \epsilon^\alpha_{al}, \tag{II.4}$$

where the index $al$ runs over all clusters in $A$.

In terms of these cluster properties let us construct the channel space $\mathscr{H}_\alpha$ and its associated channel Hamiltonian $\widetilde{H}_\alpha$. Consider each cluster to be a point particle and let $H_0[A]$ denote the relative motion kinetic energy operator for these $N_\alpha$ bodies. Then $\widetilde{H}_\alpha$ is given by

$$\widetilde{H}_\alpha = H_0[A] - \epsilon^\alpha. \tag{II.5}$$

This Hamiltonian gives the energy available to the $N_\alpha$ clusters when they are mutually outside of each others force fields and freely moving. The space $\widetilde{H}_\alpha$ acts in is denoted by $\mathscr{H}_\alpha = L^2(R^{3(N_\alpha - 1)})$. Since $\widetilde{H}_\alpha$ is just the kinetic energy operator displaced by a fixed energy it will only have a continuous spectrum. The label $\alpha = 0$ is reserved for the channel where all particles are free. Clearly, $\widetilde{H}_0 = H_0$ and $\epsilon^0 = 0$.

The last preliminary definition needed is that of the injection operator $J_\alpha$ which maps $\mathscr{H}_\alpha$ into $\mathscr{H}$. Let $f_\alpha$ be any function in $\mathscr{H}_\alpha$. Set $F_\alpha$ to be

$$F_\alpha = \Big(\prod_{al} \Psi^\alpha_{al}\Big) f_\alpha, \tag{II.6}$$

then $J_\alpha$ is defined as

$$F_\alpha = J_\alpha f_\alpha. \tag{II.7}$$

Operator $J_\alpha$ is called the identification operator for channel $\alpha$. If $\alpha = 0$, then $J_0$ is the identity on $\mathscr{H}_0$.

The basic object in multichannel scattering theory is the wave operator that maps $\mathscr{H}_\alpha$ into $\mathscr{H}$ by

$$\Omega^{(\pm)}_\alpha = \underset{t\to\mp\infty}{\text{s-lim}}\, e^{iHt} J_\alpha \exp(-i\widetilde{H}_\alpha t). \tag{II.8}$$

The wave operators possess four basic properties. The first of these is

$$\int_0^{\mp\infty} dt\, \|[\Omega^{(\pm)}_\alpha - e^{iHt}J_\alpha \exp(-i\widetilde{H}_\alpha t)]f_\alpha\| < \infty. \tag{II.9}$$

Property (II.8) and (II.9) are somewhat redundant, since (II.9) implies (II.8). The second property is channel othogonality for wave operators with the same boundary condition,

$$\Omega^{(+)\dagger}_\alpha \Omega^{(+)}_\beta = \delta_{\alpha\beta} I_\alpha. \tag{II.10}$$

Here $I_\alpha$ is the identity on $\mathscr{H}_\alpha$ and the dagger indicates the adjoint. Furthermore the range of any $\Omega^{(\pm)}_\alpha$ is orthogonal to all bounds state eigenfunctions of $H$. Energy conservation becomes the intertwining relation

$$H\Omega^{(\pm)}_\alpha = \Omega^{(+)}_\alpha \widetilde{H}_\alpha. \tag{II.11}$$

The last wave operator property is completeness. This statement has two possible forms. First, define $R^{(\pm)}_\alpha = \Omega^{(\pm)}_\alpha \Omega^{(\pm)\dagger}_\alpha$. Property (II.10) implies $R^{(\pm)}_\alpha$ is a projection operator onto the range of $\Omega^{(\pm)}_\alpha$. It also tells us that for different $\alpha$ and $\beta$, then $R^{(\pm)}_\alpha$ and $R^{(\pm)}_\beta$ are orthogonal. Weak asymptotic completeness is the statement that

$$\sum_\alpha R^{(+)}_\alpha = \sum_\alpha R^{(-)}_\alpha. \tag{II.12}$$

This feature clearly is an underlying reflection of time reversal invariance. Now let $B$ denote the projection operator onto the subspace of bound state eigenfunctions of $H$. Then strong asymptotic completeness states that

$$\sum_\alpha R^{(\pm)}_\alpha + B = I, \tag{II.13}$$

where $I$ is the identity on $\mathscr{H}$. Clearly (II.13) implies (II.12).

The form of two-Hilbert space theory outlined here has gradually been developed from the time-dependent multichannel scattering theory set up by Jauch.[8] The existence and channel orthogonality properties have been proven by Hack[9] and Hunziker[10] for short-range potentials $V_{ij}$, that are multiplication operators in coordinate space by $L^2$ functions.[12] The intertwining property is an immediate consequence of Eq. (II.8). The difficult statements to prove are the completeness properties (II.12) and (II.13). For $N = 3$, Faddeev[13] has proved strong completeness. Recently, several new proofs were found for the three-body case.[14] For $N > 4$ completeness has not yet been proven. Note however, the different status of (II.12) and (II.13). The weak completeness seems to be required by any physically reasonable theory—or time reversal invariance will be violated. The strong completeness is far more difficult to establish since it implies that $H$ has no essentially singular spectrum. In the derivation of our time-delay results in Secs. III and IV, we require only the weak completeness. For the duration of the paper we take it as an hypothesis that (II.9)-(II.12) are valid for all $N$. We shall also assume that the total number of channels in $\mathscr{A}$ is finite. Even though we have no long-range Coulomb interaction it is possible to have an infinite number of channels. One mechanism that can generate an infinite number of independent bound states for a cluster is the Efimov effect.[15] If we have a pair potential $V_{ij}$ with a zero energy bound state in the two-body system, then the Efimov result predicts that there must be an infinite number of three-body bound states, and thus an infinite number of channels.

We conclude this section by a few comments on the many-body $S$ matrix. Channel components of the $s$ matrix are operators from $\mathscr{H}_\beta$ to $\mathscr{H}_\alpha$ defined by

$$S_{\alpha\beta}=\Omega_\alpha^{(-)\dagger}\Omega_\beta^{(+)}. \tag{II.14}$$

From Eq. (II.14) and the intertwining property (II.11) one finds

$$S_{\alpha\beta}\widetilde{H}_\beta=\widetilde{H}_\alpha S_{\alpha\beta}. \tag{II.15}$$

This means $S_{\alpha\beta}$ conserves energy. Unitarity of the $S$ matrix is

$$\sum_\gamma S_{\alpha\gamma}S^\dagger_{\beta\gamma}=\sum_\gamma S^\dagger_{\gamma\alpha}S_{\gamma\beta}=\delta_{\alpha\beta}I_\alpha. \tag{II.16}$$

It is not difficult to show that this is valid if and only if weak asymptotic completeness holds.[11]

## III. THE $S$ MATRIX TRANSFORMATION

The purpose of this section is to give a universally valid definition of $N$-body time delay for an arbitrary two-cluster incident channel. This definition is then transformed so that the problem may be stated exclusively in terms of asymptotic wavefunctions.

Let $f_\alpha\in\mathscr{H}_\alpha$ be the wavepacket specifying the incoming wave in channel $\alpha$. The time evolution of this asymptotic state is

$$\Phi_\alpha(t)=J_\alpha e^{-i\widetilde{H}_\alpha t}f_\alpha. \tag{III.1}$$

The corresponding exact $\Psi(t)$, satisfying

$$i\frac{\partial\Psi(t)}{\partial t}=H\Psi(t), \tag{III.2}$$

that converges to $\Phi_\alpha(t)$ as $t\to-\infty$ is

$$\Psi(t)=\Omega_\alpha^{(+)}e^{-i\widetilde{H}_\alpha t}f_\alpha. \tag{III.3}$$

In this notation the wave-operator property (II.9) reads

$$\int_{-\infty}^0 dt\,\|\Psi(t)-\Phi_\alpha(t)\|<\infty. \tag{III.4}$$

The state $\Psi(t)$ may be equally well characterized by the outgoing waves $\Phi'_\beta(t)$ that approximate $\Psi(t)$ for large positive times. These outgoing waves are given by

$$\Phi'_\beta(t)=J_\beta e^{-i\widetilde{H}_\beta t}f'_\beta, \tag{III.5}$$

where

$$f'_\beta=S_{\beta\alpha}f_\alpha, \tag{III.6}$$

The convergence criteria for $\Phi'_\beta(t)$ may be conveniently stated as

$$\int_0^\infty dt\,\|R_\beta^{(-)}\Psi(t)-\Phi'_\beta(t)\|<\infty. \tag{III.7}$$

This is equivalent to Eq. (II.9) because

$$\begin{aligned}\|R_\beta^{(-)}\Psi(t)-\Phi'_\beta(t)\|&=\|e^{-iHt}\Omega_\beta^{(-)}\Omega_\beta^{(-)\dagger}\Omega_\alpha^{(+)}f_\alpha-J_\beta e^{-i\widetilde{H}_\beta t}S_{\beta\alpha}f_\alpha\|\\&=\|(\Omega_\beta^{(-)}-e^{iHt}J_\beta e^{-i\widetilde{H}_\beta t})f'_\beta\|.\end{aligned} \tag{III.8}$$

One constructs the definition of time delay in terms of $\Psi(t)$ and the asymptotic waves $\Phi_\alpha(t)$ and $\Phi'_\beta(t)$. First we denote by $P(r)$ a projection operator in $\mathscr{H}$ onto functions with support inside a $3(N-1)$-dimensional sphere of radius $r$. Represent by $m_i$ and $\mathbf{r}_i$ the mass and position of particle $i$ in the center-of-mass frame. The radius $r$ is defined by

$$r^2=M^{-1}\sum_i^N m_i\bar{r}_i^2, \tag{III.9}$$

where $M$ is the total mass of all $N$ particles. The effect of $P(r)$ on a function $f$ in $\mathscr{H}$ is multiplication by one if the vector argument of $f$ is inside the sphere and zero otherwise.

The scattering processes of greatest interest in particle and nuclear physics are those initiated by the collision of two fragments. We will restrict our definition and analysis to these channels. One possible definition of time delay for a two-fragment channel $\alpha$ is

$$T_\alpha^{\text{in}}(f_\alpha;r)=\int_{-\infty}^\infty dt\,[\|P(r)\Psi(t)\|^2-\|P(r)\Phi_\alpha(t)\|^2]. \tag{III.10}$$

Clearly for finite $r$ the integrals have the interpretation of the total sojourn time $\Psi(t)$ is inside the sphere $P(r)$ minus the time the asymptotic wave $\Phi_\alpha(t)$ is in the sphere.

We note that the time integrals in (III.10) will exist, since for all $t$ the integrand is bounded by $2\|f_\alpha\|^2$ and for large $t$ the wavepackets disperse out of the finite volume of the sphere $P(r)$. For example, let $\beta$ be any channel with $n+1$ free clusters and asymptotic Hamiltonian $H_0[A]$. Take the Jacobian coordinate variables describing the relative cluster positions to be $\mathbf{x}=\{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n\}$. The Hamiltonian is $H_0[A]=-\hbar^2\Sigma_j(2\mu_j)^{-1}\nabla_j^2$, where $\mu_j$ are reduced masses. Thus the coordinate space representation of this free time evolution operator is[16]

$$\exp(-iH_0[A]t)f(\mathbf{x})=\left(\frac{m_\beta}{2\pi it}\right)^{3n/2}\int d\mathbf{x}'\prod_{j=1}^n\exp[i\mu_j(\mathbf{x}_j-\mathbf{x}'_j)^2/2t]f(\mathbf{x}'). \tag{III.11}$$

Here $m_\beta$ is a function of $\mu_j$. Let $A(r)$ be the volume of the sphere in $3(N-1)$ dimensions. Then Eq. (III.11) implies

$$\|P(r)J_\beta\exp(-i\widetilde{H}_\beta t)f_\beta\|^2\leqslant\left|\frac{m_\beta}{2\pi t}\right|^{3n}A(r)\|f_\beta\|_1^2. \tag{III.12}$$

Here $\|f_\beta\|_1^2$ denotes the $L^1$ norm of $f_\beta(\mathbf{x}_\beta)$. When estimate (III.12) is combined with the appropriate form of bound (II.9), it is easy to see that the integral (III.10) is convergent for finite $r$.

However the exact wave $\Psi(t)$ is as closely associated with the outgoing wave $\Phi'_\beta(t)$ as it is with the incoming waves $\Phi_\alpha(t)$. Thus an equally reasonable definition of time delay is

$$T_\alpha^{\text{out}}(f_\alpha;r)=\int_{-\infty}^\infty dt\left[\|P(r)\Psi(t)\|^2-\sum_\beta\|P(r)\Phi'_\beta(t)\|^2\right]. \tag{III.13}$$

The time delay we shall study in detail is one that is fully symmetric with respect to the asymptotic states that converge to $\Psi(t)$ for $t\to\pm\infty$. This symmetric definition is

$$T_\alpha(f_\alpha;r)=\tfrac{1}{2}[T_\alpha^{\rm in}(f_\alpha;r)+T_\alpha^{\rm out}(f_\alpha;r)]. \qquad \text{(III.14)}$$

The sensitive issue concerning these three definitions of time delay is whether or not the $r\to\infty$ limit exists and is finite. Our analysis shows that only $T_\alpha(f_\alpha;r)$ does have a finite limit. We note that in the two-body scattering case, where there are several rigorous treatments[3-5] of time delay, all these definitions have limits when $r\to\infty$ and are equivalent. We discuss this point in greater detail in Sec. V.

From here on we assume that the incoming wavepacket $f_\alpha$ belong to the space $\mathscr{F}$ that is defined as the set of all functions in momentum space with compact support and with continuous derivatives of all orders. Of course, $\mathscr{F}$ is a subset of the asymptotic space $\mathscr{H}_\alpha(\mathbf{Q}_\alpha)$. Furthermore, as we will discuss later on in detail, we assume that the $S$ matrix preserves the space $\mathscr{F}$.

The first simplication is to find an equivalent form for $T_\alpha(f_\alpha;r)$ in which the exact waves $\Psi(t)$ are systematically replaced by the asymptotic waves $\Phi_\alpha(t)$ and $\Phi'_\beta(t)$. The identities given in the following lemma allow this simplication.

*Lemma* 1: For any $f_\alpha\in\mathscr{F}$ let $\Psi(t)$, $\Phi_\alpha(t)$, and $\Phi'_\beta(t)$ be defined as above. Assume that $\Psi^\alpha_{al}$ and $\Psi^\beta_{bl}$ are bounded almost everywhere in momentum space. Set $\Delta^\pm(f_\alpha;r)$ to be

$$\Delta^+(f_\alpha;r)=\int_0^\infty dt\left[\|P(r)\Psi(t)\|^2-\sum_\beta\|P(r)\Phi'_\beta(t)\|^2\right], \qquad \text{(III.15)}$$

$$\Delta^-(f_\alpha;r)=\int_{-\infty}^0 dt\,[\|P(r)\Psi(t)\|^2-\|P(r)\Phi_\alpha(t)\|^2], \qquad \text{(III.16)}$$

then

$$\lim_{r\to\infty}\Delta^\pm(f_\alpha;r)=0. \qquad \text{(III.17)}$$

*Proof*: Consider $\Delta^+(f_\alpha;r)$ first. If we note that $\Psi(t)$ only has values in the $\Sigma_\beta R_\beta^{(-)}$ subspace of $\mathscr{H}$, we can split up the integrand of Eq. (III.15) in a sum of direct terms and cross terms, viz.,

$$\left[\sum_\beta\|P(r)R_\beta^{(-)}\Psi(t)\|^2-\sum_\beta\|P(r)\Phi'_\beta(t)\|^2\right]+\left(\sum_\beta R_\beta^{(-)}\Psi(t),P(r)\sum_{\gamma\neq\beta}R_\gamma^{(-)}\Psi(t)\right). \qquad \text{(III.18)}$$

Because the sum over $\beta$ is finite it suffices to consider only one term in the integrand of (III.15). For each $\beta$ we have the following bound for the direct term

$$\begin{aligned}&|\,\|P(r)R_\beta^{(-)}\Psi(t)\|^2-\|P(r)\Phi'_\beta(t)\|^2|\\&\leqslant[\|R_\beta^{(-)}\Psi(t)\|+\|\Phi'_\beta(t)\|][\|R_\beta^{(-)}\Psi(t)-\Phi'_\beta(t)\|]\\&\leqslant 2\|f_\alpha\|\|R_\beta^{(-)}\Psi(t)-\Phi'_\beta(t)\|.\end{aligned} \qquad \text{(III.19)}$$

The first inequality follows from adding and subtracting $(R_\beta^{(-)}\Psi(t),P(r)\Phi'_\beta(t))$ to the difference on the left-hand side. The second inequality is implied by the isometric property of the wave operators. Inequalities (III.19) and (III.7) give an $r$ independent absolutely integrable bound for the direct terms of $\Delta^+(f_\alpha,r)$.

Next, we have to show that also the cross term $(R_\beta^{(-)}\Psi(t),P(r)R_\gamma^{(-)}\Psi(t))$ has an $r$ independent absolutely integrable bound. Again by adding and subtracting $(R_\beta^{(-)}\Psi(t),P(r)\Phi'_\gamma(t))$ and using inequality (II.9), we see that it is sufficient to show that $(\Phi'_\beta(t),P(r)\Phi'_\gamma(t))$ is $L^1(t)$ uniformly in $r$. Since our proof of this statement depends on the properties of the projection operator $P(r)$, we postpone it till the Appendix. There we show that if the asymptotic wavepackets belong to $\mathscr{F}$ and if the internal cluster wavefunctions are bounded almost everywhere in momentum space, then the cross term $(\Phi'_\beta(t),P(r)\Phi'_\gamma(t))\in L^1\,(t\geqslant1)$ uniformly in $r$ (Lemma A1). The restriction $t\geqslant1$ is not essential for our argument since the norms appearing in Eq. (III.15) have values between 0 and 1.

So the Lebesgue dominated convergence theorem[17] allows us to interchange the $r$ limit and the $t$ integration. However the integrand is now zero since

$$\begin{aligned}&\lim_{r\to\infty}\left[\sum_\beta\|P(r)R_\beta^{(-)}\Psi(t)\|^2-\sum_\beta\|P(r)\Phi'_\beta(t)\|^2+\sum_\beta\left(R_\beta^{(-)}\Psi(t),P(r)\sum_{\gamma\neq\beta}R_\gamma^{(-)}\Psi(t)\right)\right]\\&=\|f_\alpha\|^2-\sum_\beta\|f'_\beta\|^2+\sum_\beta\left(R_\beta^{(-)}\Psi(t),\sum_{\gamma\neq\beta}R_\gamma^{(-)}\Psi(t)\right)=0.\end{aligned} \qquad \text{(III.20)}$$

The last equality in (III.20) is a consequence of the unitarity property (II.16) of the $S$ matrix and the orthogonality of the $R_\beta$. Thus (III.17) is established for $\Delta^+(f_\alpha;r)$. A similar but simpler argument applies to $\Delta^-(f_\alpha,r)$. In this case the cross terms are absent and one does not need the unitarity of $S_{\beta\alpha}$.

The eliminatation of $\Psi(t)$ from the problem is accomplished by employing Lemma 1. First define $T^\pm(f_\alpha;r)$ by

$$T^+(f_\alpha;r)=\tfrac{1}{2}\int_0^\infty dt\left[\sum_\beta\|P(r)\Phi'_\beta(t)\|^2-\|P(r)\Phi_\alpha(t)\|^2\right], \qquad \text{(III.21)}$$

$$T^-(f_\alpha;r)=\tfrac{1}{2}\int_{-\infty}^0 dt\left[\|P(r)\Phi_\alpha(t)\|^2-\sum_\beta\|P(r)\Phi'_\beta(t)\|^2\right], \qquad \text{(III.22)}$$

Then the symmetric definition of time delay Eq. (III.14) can be written

$$T_\alpha(f_\alpha;r)=T^+(f_\alpha;r)+T^-(f_\alpha;r)+\Delta^+(f_\alpha;r)+\Delta^-(f_\alpha;r). \qquad \text{(III.23)}$$

In this form $T^\pm(f_\alpha;r)$ contain only the asymptotic channel operators $J_\beta$, $\widetilde{H}_\beta$, and $S_{\beta\alpha}$. The exact waves are restricted to the terms $\Delta^\pm(f_\alpha;r)$ which vanish in the $r\to\infty$ limit.

Finally, let us display explicitly the $S$ matrix form of $T^\pm(f_\alpha;r)$. Define the channel "projection" operator

$$P_\beta(r)=J_\beta^\dagger P(r)J_\beta. \qquad \text{(III.24)}$$

This operator $P_\beta(r)$ converges strongly to $I_\beta$ as $r\to\infty$, but is not technically a projection since it is not idempotent. In terms of $P_\beta(r)$ we can write

$$\sum_\beta \|P(r)\Phi'_\beta(t)\|^2 = \sum_\beta (f_\alpha, \exp(i\widetilde{H}_\alpha t)S^\dagger_{\beta\alpha}P_\beta(r)S_{\beta\alpha}\exp(-i\widetilde{H}_\alpha t)f_\alpha). \tag{III.25}$$

In Eq. (III.25) we used Eq. (II.15) to intertwine the order of the operators. The second term in the integrand of $T^\pm(f_\alpha;r)$ may be expressed in a form similar to Eq. (III.25),

$$\|P(r)\Phi_\alpha(t)\|^2 = (f_\alpha, \exp(i\widetilde{H}_\alpha t)P_\alpha(r)\exp(-i\widetilde{H}_\alpha t)f_\alpha). \tag{III.26}$$

Combining Eqs. (III.21), (III.22), (III.25), and (III.26) gives the final result. Let

$$K_\alpha(r) = \sum_\beta S^\dagger_{\beta\alpha}P_\beta(r)S_{\beta\alpha} - P_\alpha(r), \tag{III.27}$$

then,

$$T^+(f_\alpha;r) = \tfrac{1}{2}\int_0^\infty dt\,(f_\alpha, \exp(i\widetilde{H}_\alpha t)K_\alpha(r)\exp(-i\widetilde{H}_\alpha t)f_\alpha), \tag{III.28}$$

$$T^-(f_\alpha;r) = -\tfrac{1}{2}\int_{-\infty}^0 dt\,(f_\alpha, \exp(i\widetilde{H}_\alpha t)K_\alpha(r)\exp(-i\widetilde{H}_\alpha t)f_\alpha). \tag{III.29}$$

It will be the task of the next section to obtain explicit values for the integrals (III.28) and (III.29).

## IV. THE TIME DELAY RELATION

In this section we complete the proof of the connection between time delay and the on-shell $S$ matrix and its energy derivatives by calculating the limit $r\to\infty$ of $T_\alpha(f_\alpha;r)$.

First, let us briefly describe the coordinate systems we employ. The initial channel is restricted to the collision of two fragments. The relative momentum of these two clusters is described by the three-dimensional vector $\mathbf{Q}_\alpha$. The kinetic energy associated with this motion is $\widetilde{Q}^2_\alpha = Q^2_\alpha/2m_\alpha$ where $m_\alpha$ is a reduced mass that can be written down as a function of the particle masses $m_i$. The outgoing channel is characterized by the $3(N_\beta - 1) = n$-dimensional vector $\mathbf{Q}'_\beta$ describing the relative motion of the $N_\beta$ clusters; $N_\beta = 2,3,\ldots,N$. Again, the kinetic energy associated with this motion is $\widetilde{Q}'^2_\beta = Q'^2_\beta/2m_\beta$ with $m_\beta$ the appropriate reduced mass. Observe that the $N_\beta = 2$ channel is the elastic ($\beta = \alpha$) or rearrangement ($\beta\neq\alpha$) channel; the $N_\beta = N$ channel represents total breakup.

Secondly, we state the conditions on the $S$ matrices necessary for our derivation. We know from Eq. (II.15) that the $S$ matrix conserves energy. We explicitly want to take out that energy conserving $\delta$ function by defining reduced $s$ matrices in the following way

$$\langle \mathbf{Q}_\beta | S_{\beta\alpha} | \mathbf{Q}_\alpha\rangle = \delta(E_\beta - E_\alpha)(m_\beta Q_\beta^{n-2} m_\alpha Q_\alpha)^{-1/2}\langle \widehat{Q}_\beta | s_{\beta\alpha}(E_\alpha)|\widehat{Q}_\alpha\rangle, \tag{IV.1}$$

where $E_\alpha = \widetilde{Q}^2_\alpha - \epsilon^\alpha$ and $E_\beta = \widetilde{Q}^2_\beta - \epsilon^\beta$. The kernel $\langle \widehat{Q}_\beta | s_{\beta\alpha}(E_\alpha | \widehat{Q}_\alpha\rangle$ represents an operator that will map square integrable functions with respect to the measure $d\widehat{Q}_\alpha$, i.e., $L^2(\widehat{Q}_\alpha)$ into $L^2(\widehat{Q}_\beta)$. The energy dependence indicated on the right of Eq. (IV.1) means that for such $S_{\beta\alpha}$ operator we have a one-parameter family of operators $s_{\beta\alpha}(E_\alpha)$. The momentum and reduced mass factors are chosen such that the operator relations $S$ obeys (unitarity, time-reversal invariance,$\cdots$) are also valid for $s(E)$ on the reduced angular space. We furthermore assume that the $S$ matrices preserve the space $\mathcal{F}$. Specifically, we suppose that any incoming wavepacket $f_\alpha$ in $\mathcal{F}$ will be mapped by $S_{\beta\alpha}$ into an infinitely differentiable function of compact support $f'_\beta$. We also restrict $f_\alpha$ to have no support in the intermediate neighborhood of the thresholds for the $\beta$ channel. To get an idea how physical this condition is, we will work out a typical realization, in terms of the reduced $s$-matrices, that will be useful in the course of the derivation. We know that if $f_\alpha\in\mathcal{F}$, then $f_\alpha\in L^1$. Employing Eq. (III.6) and Hölder's inequality, it is straightforward to verify that $f'_\beta\in L^1$ if

$$\|\tau_{\beta\alpha}(E)\|^2_{HS} = \int d^{n-1}\widehat{Q}_\beta d\widehat{Q}_\alpha |\langle \widehat{Q}_\beta|\tau_{\beta\alpha}(E)|\widehat{Q}_\alpha\rangle|^2 < C < \infty, \tag{IV.2}$$

for all $E$, in any finite interval of energy. Here $\tau_{\beta\alpha}(E)$ is the reduced $t$ operator defined by the following matrix elements

$$\langle \widehat{Q}_\beta|\tau_{\beta\alpha}(E)|\widehat{Q}_\alpha\rangle = \langle \widehat{Q}_\beta|s_{\beta\alpha}(E)|\widehat{Q}_\alpha\rangle - \delta_{\beta\alpha}\delta(\widehat{Q}_\beta - \widehat{Q}_\alpha). \tag{IV.3}$$

Except for simple kinematical factors, $\|\tau_{\beta\alpha}(E)\|^2_{HS}$ is proportional to the cross section for scattering from channel $\alpha$ to $\beta$. So restriction (IV.2) is nothing more than the requirement that the physical cross sections be finite. At this point, it is worthwhile to note that, since the incoming channel $\alpha$ is restricted to be a two-cluster channel, the scattering processes we consider are free of disconnected channels (e.g., the three-to-three scattering channel in the three-body problem). A third condition we then impose on the reduced $s$ matrices is that they are once differentiable with respect to $E_\alpha$ for fixed angles.

We now have all the necessary information to carry out the first step in our derivation of the time delay relation, namely the introduction of the Abel limit to do the $t$-integration.

*Lemma* 2: Let $f_\alpha\in\mathcal{F}$. Assume that the reduced $s$ matrices satisfy condition (IV.2), then

$$T_\alpha(f_\alpha;r) = i\int d^3\mathbf{Q}'_\alpha d^3\mathbf{Q}_\alpha f^*_\alpha(\mathbf{Q}'_\alpha)\frac{\langle \mathbf{Q}'_\alpha|K_\alpha(r)|\mathbf{Q}_\alpha\rangle}{\widetilde{Q}'^2_\alpha - \widetilde{Q}^2_\alpha}f_\alpha(\mathbf{Q}_\alpha) + \Delta^+(f_\alpha;r) + \Delta^-(f_\alpha;r), \tag{IV.4}$$

where the integral is defined as a principal-value integral.

*Proof*: Consider $T^{+}(f_\alpha;r)$ given by Eq. (III.28) or equivalently by Eq. (III.21). Using the estimate (III.12) for the $\alpha$- and $\beta$-channel term, and the additional fact that the total integrand is bounded by $2\|f_\alpha\|^2$ for all $t$, we have that $(f_\alpha,\exp(i\widetilde{H}_\alpha t)K_\alpha(r)\exp(-i\widetilde{H}_\alpha t)f_\alpha)\in L^1(t)$. the dominated convergence theorem then allows us to intorduce the Abel limit form of $T^{+}(f_\alpha;r)$, viz.,

$$T^{+}(f_\alpha;r)$$
$$=\tfrac{1}{2}\lim_{\epsilon\to 0+}\int_0^\infty dt\, e^{-\epsilon t}(f_\alpha,\exp(i\widetilde{H}_\alpha t)K_\alpha(r)\exp(-i\widetilde{H}_\alpha t)f_\alpha). \tag{IV.5}$$

We next want to prove that the momentum space integrals on the right side of Eq. (IV.5) exists absolutely. To do this we write out the scalar product in momentum space and consider the first term, viz.,

$$\lim_{\epsilon\to 0+}\int_0^\infty dt\sum_\beta\int d^n\mathbf{Q}'_\beta\,|f'^{*}_\beta(\mathbf{Q}'_\beta)$$
$$\times\exp[i(\widetilde{Q}'^2_\beta-\widetilde{Q}^2_\beta)t-\epsilon t]\langle\mathbf{Q}'_\beta|P_\beta(r)|\mathbf{Q}_\beta\rangle f'_\beta(\mathbf{Q}_\beta)|\,. \tag{IV.6}$$

Since $\langle\mathbf{Q}'_\beta|P_\beta(r)|\mathbf{Q}_\beta\rangle$ is uniformly bounded, as shown in Lemma A2 in the Appendix, and since $f'_\beta\in L^1(\mathbf{Q}'_\beta)$ by condition (IV.2) we immediately have that the first term is $L^1$. A parallel argument gives us that the second term in Eq. (IV.5) is $L^1$. Since also the $t$-integration exists absolutely, we can change the order of integration by Fubini's theorem[17] and do the integral over $t$ first. A completely analogous argument can be built up for $T^{-}(f_\alpha;r)$. Thus

$$T^{\pm}(f_\alpha;r)$$
$$=\frac{i}{2}\int d^3\mathbf{Q}'_\alpha\,d^3\mathbf{Q}_\alpha f^{*}_\alpha(\mathbf{Q}'_\alpha)\frac{\langle\mathbf{Q}'_\alpha|K_\alpha(r)|\mathbf{Q}_\alpha\rangle}{\widetilde{Q}'^2_\alpha-\widetilde{Q}^2_\alpha\pm i0}f_\alpha(\mathbf{Q}_\alpha). \tag{IV.7}$$

Taking the symmetric combination of $T^{+}$ and $T^{-}$ completes the proof of the lemma.

The second step in our proof of the time delay relation is to calculate the principal-value integral appearing in Eq. (IV.4) in the limit $r\to\infty$, using Fourier transform properties of the sphere.

*Lemma 3*: Let $f_\alpha(\mathbf{Q}_\alpha)\in\mathscr{F}$. Assume furthermore that $\langle\widehat{Q}_\beta|s_{\beta\alpha}(E)|\widehat{Q}_\alpha\rangle$ satisfies the three conditions stated in Eqs. (IV.1)–(IV.3). Then

$$T_\alpha(f_\alpha;\infty)$$
$$=-i\sum_\beta\int d^3\mathbf{Q}'_\alpha\,d^3\mathbf{Q}_\alpha\,d^{n-1}\widehat{Q}'_\beta\frac{\delta(E_\alpha-E'_\alpha)}{(m_\alpha Q_\alpha m_\alpha Q'_\alpha)^{1/2}}$$
$$\times f^{*}_\alpha(\mathbf{Q}'_\alpha)\Big[\langle\widehat{Q}'_\alpha|s_{\beta\alpha}(E'_\alpha|\widetilde{Q}'_\alpha\rangle^{*}$$
$$\times\frac{d}{dE_\alpha}\langle\widehat{Q}'_\beta|s_{\beta\alpha}(E_\alpha)|\widehat{Q}_\alpha\rangle\Big]f_\alpha(\mathbf{Q}_\alpha). \tag{IV.8}$$

*Proof*: Consider Eq. (IV.4) and write out the kernel $\langle\mathbf{Q}'_\alpha|K_\alpha(r)|\mathbf{Q}_\alpha\rangle$ using Eq. (III.27). This gives

$$T_\alpha(f_\alpha;\infty)$$
$$=\lim_{r\to\infty}i\int d^3\mathbf{Q}'_\alpha\,d^3\mathbf{Q}_\alpha f^{*}_\alpha(\mathbf{Q}'_\alpha)\frac{2m_\alpha}{Q'^2_\alpha-Q^2_\alpha}$$
$$\times\Big(\sum_\beta\int d^n\mathbf{Q}'_\beta d^n\mathbf{Q}_\beta\langle\mathbf{Q}'_\beta|S_{\beta\alpha}|\mathbf{Q}'_\alpha\rangle^{*}$$
$$\times\langle\mathbf{Q}'_\beta|P_\beta(r)|\mathbf{Q}_\beta\rangle\langle\mathbf{Q}_\alpha|S_{\beta\alpha}|\mathbf{Q}_\alpha\rangle$$
$$-\langle\mathbf{Q}'_\alpha|P_\alpha(r)|\mathbf{Q}_\alpha\rangle\Big)f_\alpha(\mathbf{Q}_\alpha). \tag{IV.9}$$

We now want to apply the properties of the projection operators $P(r)$. Lemma A4, proved in the Appendix, can immediately be used to calculate the limit $r\to\infty$ of the second term in Eq. (IV.9). The result is

$$2m_\alpha i\int d^3\mathbf{Q}'_\alpha f^{*}_\alpha(\mathbf{Q}'_\alpha)\frac{d}{dQ_\alpha}\left(\frac{Q_\alpha\,f_\alpha(Q_\alpha\widehat{Q}'_\alpha)}{Q'_\alpha\,Q_\alpha+Q'_\alpha}\right)_{Q_\alpha=Q'_\alpha}. \tag{IV.10}$$

In the first term of Eq. (IV.9) we first have to remove the energy conserving $\delta$ functions appearing in the $s$ matrices. We realize this by introducing the reduced $s$ matrices defined in Eq. (IV.1) and integrating over $Q_\alpha$ and $Q'_\alpha$. We get then

$$\lim_{r\to\infty}i\sum_\beta\int d\widehat{Q}'_\alpha d\widehat{Q}_\alpha d^n\mathbf{Q}'_\beta d^n\mathbf{Q}_\beta f^{*}_\alpha([2m_\alpha(\widetilde{Q}'^2_\beta-\epsilon^\beta+\epsilon^\alpha)]^{1/2}\widehat{Q}'_\alpha)$$
$$\times\frac{2m_\alpha[2m_\alpha(\widetilde{Q}'^2_\beta-\epsilon^\beta+\epsilon^\alpha)]^{1/4}}{Q'^{(n-2)/2}_\beta}\langle\widehat{Q}'_\beta|s_{\beta\alpha}(E'_\beta)|\widehat{Q}'_\alpha\rangle^{*}\frac{\langle\mathbf{Q}'_\beta|P_\beta(r)|\mathbf{Q}_\beta\rangle}{Q'^2_\beta-Q^2_\beta}$$
$$\times\langle\widehat{Q}_\beta|s_{\beta\alpha}(E_\beta|\widehat{Q}_\alpha\frac{[2m_\alpha(\widetilde{Q}^2_\beta-\epsilon^\beta+\epsilon^\alpha)]^{1/4}}{Q^{(n-2)/2}_\beta}f_\alpha([2m_\alpha(\widetilde{Q}^2_\beta-\epsilon^\beta+\epsilon^\alpha)]^{1/2}\widehat{Q}_\alpha).$$

Next, we use Lemma A4 to calculate this limit (IV.11). We split up the result in two parts. The first part, containing the derivative of the wavepacket and the kinematical factors, will be shown to cancel the contribution (IV.10). This is exactly what we expect physically, since otherwise time delay would depend on the shape of the incoming wavepacket. The second part, containing the derivative of the $s$ matrix only, will give us the final answer Eq. (IV.8).

Let us consider the first part of the limit (IV.11) which, after some algebra, can be written as

$$-i\sum_\beta\int d\hat{Q}'_\alpha d\hat{Q}_\alpha d^n\mathbf{Q}'_\beta\frac{2m_\alpha^2}{m_\beta}\frac{1}{Q_\beta'^{n-2}}f_\alpha^*([2m_\alpha(\widetilde{Q}_\beta'^2-\epsilon^\beta+\epsilon^\alpha)]^{1/2}\hat{Q}'_\alpha)$$
$$\times\langle\hat{Q}'_\beta|s_{\beta\alpha}(E'_\beta)|\hat{Q}'_\alpha\rangle^*\langle\hat{Q}'_\beta|s_{\beta\alpha}(E'_\beta)|\hat{Q}_\alpha\rangle\Big(\frac{1}{4}\frac{1}{[2m_\alpha(\widetilde{Q}_\beta'^2-\epsilon^\beta+\epsilon^\alpha)]^{1/2}}$$
$$\times f_\alpha([2m_\alpha(\widetilde{Q}_\beta'^2-\epsilon^\beta+\epsilon^\alpha)]^{1/2}\hat{Q}_\alpha)+\frac{1}{2}\frac{df_\alpha([2m_\alpha(\widetilde{Q}_\beta^2-\epsilon^\beta+\epsilon^\alpha)]^{1/2}\hat{Q}_\alpha)}{d[2m_\alpha(\widetilde{Q}_\beta^2-\epsilon^\beta+\epsilon^\alpha)]^{1/2}}\Big|_{Q_\beta=Q'_\beta}\Big).\qquad\text{(IV.12)}$$

In this expression, we change the variable of integration $Q'_\beta$ to $[2m_\alpha(\widetilde{Q}_\beta'^2-\epsilon^\beta+\epsilon^\alpha)]^{1/2}\equiv Q'_\alpha$. This leads to

$$-2m_\alpha i\sum_\beta\int d^3\mathbf{Q}'_\alpha d\hat{Q}_\alpha d^{n-1}\hat{Q}'_\beta f_\alpha^*(\mathbf{Q}'_\alpha)\langle\hat{Q}'_\beta|s_{\beta\alpha}(E'_\alpha)|\hat{Q}'_\alpha\rangle^*$$
$$\times\langle\hat{Q}'_\beta|s_{\beta\alpha}(E'_\alpha)|\hat{Q}_\alpha\rangle\Big(\frac{1}{4Q_\alpha'^2}f_\alpha(Q'_\alpha\hat{Q}_\alpha)+\frac{1}{2Q'_\alpha}\frac{d}{dQ'_\alpha}f_\alpha(Q'_\alpha\hat{Q}_\alpha)\Big).\qquad\text{(IV.13)}$$

Using unitarity for the reduced $s$ matrices, we see that this contribution (IV.13) cancels expression (IV.10).

Let us finally consider the second part of the limit (IV.11), viz.,

$$-i\sum_\beta\int d\hat{Q}'_\alpha d\hat{Q}_\alpha d^n\mathbf{Q}'_\beta\frac{m_\alpha[2m_\alpha(\widetilde{Q}_\beta'^2-\epsilon^\beta+\epsilon^\alpha)]^{1/2}}{Q_\beta'^{n-1}}f_\alpha^*([2m_\alpha(\widetilde{Q}_\beta'^2-\epsilon^\beta+\epsilon^\alpha)]^{1/2}\hat{Q}'_\alpha)$$
$$\times\langle\hat{Q}'_\beta|s_{\beta\alpha}(E'_\beta)|\hat{Q}'_\alpha\rangle^*\frac{Q'_\beta}{m_\beta}\frac{d}{dE_\beta}\langle\hat{Q}'_\beta|s_{\beta\alpha}(E_\beta)|\hat{Q}_\alpha\rangle\Big|_{E_\beta=E'_\beta}$$
$$\times f_\alpha([2m_\alpha(\widetilde{Q}_\beta'^2-\epsilon^\beta+\epsilon^\alpha)]^{1/2}\hat{Q}_\alpha).\qquad\text{(IV.14)}$$

Making again the change of variables $Q'_\beta\to[2m_\alpha(\widetilde{Q}_\beta'^2-\epsilon^\beta+\epsilon^\alpha)]^{1/2}\equiv Q'_\alpha$, we easily see, after some algebra, that (IV.14) becomes equal to the right-hand side of Eq. (IV.8). This completes the proof of Lemma 3.

To conclude, we want to discuss the results of Lemma 3 in more detail. From formula (IV.8), we know that $T_\alpha(f_\alpha;\infty)$ contains an energy conserving $\delta$ function. Furthermore, if we go back to the definition of $T_\alpha(f_\alpha;r)$ for finite $r$, given by Eqs. (III.12), (III.13), and (III.14), and write out both terms in the $t$-integrand, it is straightforward to show that the factor $\exp[i(\widetilde{Q}_\alpha'^2-\widetilde{Q}_\alpha^2t]$ also leads to this energy conserving $\delta$ function. Therefore, we are justified in defining a reduced time delay $q_\alpha(E_\alpha;r)$ by the following formula

$$T_\alpha(f_\alpha;r)=\int d^3\mathbf{Q}'_\alpha d^3\mathbf{Q}_\alpha\frac{\delta(E_\alpha-E'_\alpha)}{m_\alpha Q_\alpha}f_\alpha^*(\mathbf{Q}'_\alpha)\times\langle\hat{Q}'_\alpha|q_\alpha(E_\alpha;r)|\hat{Q}_\alpha\rangle f_\alpha(\mathbf{Q}_\alpha).\qquad\text{(IV.15)}$$

The results of Lemma 3 can then be written as

$$q_\alpha(E)\equiv\underset{r\to\infty}{\text{w-lim}}\,q_\alpha(E;r)=-i\sum_\beta s_{\beta\alpha}^\dagger(E)\frac{d}{dE}s_{\beta\alpha}(E),\qquad\text{(IV.16)}$$

where $q_\alpha(E)$ is an operator acting on the reduced space $L^2(\hat{Q}_\alpha)$.

## V. CONCLUSIONS

The analysis given in Sec. IV extends the theory of time delay to the $N$-body collision problem. Furthermore, the method of proof is substantially simpler than the approach we employed earlier in the three-body problem.[6] The reason for this rests with the fact that our first approach depended on evaluating matrix elements of the exact wave operator $\Omega_\alpha^{(\pm)}$. The singularity structure of $\Omega_\alpha^{(\pm)}$ is much more complicated then that of the $S$ matrix. In addition the results found here are somewhat different than we originally anticipated. In the two-body problem it is a simple exercise to show that the free reference transit time is independent of the choice of incoming or outgoing anymptotic channel, viz.,

$$\int_{-\infty}^{+\infty}dt\,\|P(r)\Phi_{\text{in}}(t)\|^2-\int_{-\infty}^{+\infty}dt\,\|P(r)\Phi'_{\text{out}}(t)\|^2=O(r^{-1}).\qquad\text{(V.1)}$$

In a general $N$-body scattering situation,the transit time for the $N_\beta$ freely evolving clusters is computed in lemma A5. Specialized to the single channel two body case, the result reads

$$\int_{-\infty}^{+\infty}dt\,\|P(r)\Phi_{\text{in}}(t)\|^2=2r\int d\mathbf{p}\,V^{-1}|f_{\text{in}}(\mathbf{p})|^2+O(r^{-1}),\qquad\text{(V.2)}$$

where $V$ is the radial velocity $|\mathbf{p}/m|$. This integral in turn may be written as the inner product $2r(f_{\text{in}},(2mH_0)^{-1/2}f_{\text{in}})$. Let $f'_{\text{out}}=Sf_{\text{in}}$ be the outgoing wavepacket, then intertwining and unitarity of the $S$ matrix give

$$(f'_{\text{out}},(2mH_0)^{-1/2}f'_{\text{out}})=(f_{\text{in}},(2mH_0)^{-1/2}f_{\text{in}}),\qquad\text{(V.3)}$$

which shows the validity of the statement (V.1). In the $N$-body case one might think that an appropriate generalization of property (V.1), viz.,

$$\int_{-\infty}^{+\infty} dt\, \|P(r)\Phi_\alpha(t)\|^2 - \int_{-\infty}^{+\infty} dt \sum_\beta \|P(r)\Phi'_\beta(t)\|^2 = O(r^{-1}), \tag{V.4}$$

still holds. The integral on the left is the transit time of the incident asymptotic wave $\Phi_\alpha(t)$ through the sphere $P(r)$. The integral on the right is the transit time of all the outgoing asymptotic states. The general formula for this difference (V.4) is given by the $\delta$ function part of Eq. (IV.7). Equation (V.4) would obviously imply that $T^{\pm}(f_\alpha;r)$ and $T_\alpha(f_\alpha;r)$ all lead to the same result in the $r\to\infty$ limit. Let us look at the results of Lemma A5 in this case, viz.,

$$\int_{-\infty}^{+\infty} dt\, \|P(r)\Phi_\alpha(t)\|^2 = 2r\left(\frac{m_0}{m_\alpha}\right)^{1/2} \int d\,\mathbf{Q}_\alpha V_\alpha^{-1} |f_\alpha(\mathbf{Q}_\alpha)|^2 + O(r^{-1}), \tag{V.5}$$

where $V_\alpha = |\mathbf{Q}_\alpha/m_\alpha|$. We remark that both formula (V.2) and (V.5) have an obvious classical interpretation. The transit time is proportional to the diameter of the sphere divided by the radial velocity. Using this result (V.5) in Eq. (V.4) we find that the left-hand side of the latter has a term which is linear in $r$. This introduces a linear divergence as $r\to\infty$ in both $T^{\pm}(f_\alpha;r)$. For this reason they are not acceptable definitions of time delay in $N$-body scattering except in the limited energy sector in which only elastic two-cluster scattering occurs (e.g., elastic collisions below the rearrangement threshold). Then obviously argument (V.3) applies to show (V.4) is valid. However, in the situation of several open channels, where $\epsilon^\alpha \neq \epsilon^\beta$, property (V.4) does not hold. This is the reason why the conclusions here differ from the ones made in our earlier three-particle work. Indeed, there we employed $T_\alpha^{\text{in}}(f_\alpha;r)$ as definition of time delay, which gives the same $r\to\infty$ limit as $T_\alpha^{+}(f_\alpha;r)$ because of Lemma 1. But that analysis also depended upon an ansatz [Eq. (A.30) of Ref. 6] that is equivalent to Eq. (V.4).

The derivation given in Secs. III and IV establishes that $T_\alpha(f_\alpha;r)$ is the meaningful definition of time delay in the $r\to\infty$ limit. It is interesting to note that this definition is essentially the same as the definition Smith originally gave in the time-independent statement of the problem [Eqs. (38) and (39) of Smith's paper in Ref. 1]. To see how this comes about write out the sum of the asymptotic transit times in a reduced operator form in analogy with Eq. (IV.15)

$$T_\alpha^{\text{asym}}(f_\alpha;r) \equiv \tfrac{1}{2}\int_{-\infty}^{+\infty} dt \left[\|P(r)\Phi_\alpha(t)\|^2 + \sum_\beta \|P(r)\Phi'_\beta(t)\|^2\right] = \int d^3\mathbf{Q}'_\alpha d^3\mathbf{Q}_\alpha \frac{\delta(E_\alpha - E'_\alpha)}{m_\alpha Q_\alpha} f^*_\alpha(\mathbf{Q}'_\alpha) \times \langle \hat{Q}'_\alpha | \sigma_\alpha(E_\alpha;r) | \hat{Q}_\alpha\rangle f_\alpha(\mathbf{Q}_\alpha). \tag{V.6}$$

If we use Eq. (V.5) to express $T_\alpha^{\text{asym}}(f_\alpha;r)$ in terms of the $S$ matrices and take out the energy $\delta$ functions, we are led to the reduced form

$$\sigma_\alpha(E;r) = m_0^{1/2} r\left(\frac{m_\alpha^{1/2}}{Q_\alpha} + \sum_\beta s^\dagger_{\beta\alpha}(E)\frac{m_\beta^{1/2}}{Q_\beta} s_{\beta\alpha}(E)\right) + O\left(\frac{1}{r}\right). \tag{V.7}$$

Employing these results and the definitions (III.3), (III.10), (III.13), and (III.14), we can easily derive the following time-independent representation for the kernel of the time-delay operator $q_\alpha(E;r)$ entering Eq. (IV.15)

$$\langle \hat{Q}'_\alpha | q_\alpha(E;r) | \hat{Q}_\alpha\rangle = 2\pi m_\alpha Q_\alpha \int_{|\mathbf{X}_0|\le r} d\,\mathbf{X}_0 \langle \mathbf{X}_0 | \Omega^{(+)} | Q_\alpha \hat{Q}'_\alpha\rangle^* \times \langle \mathbf{X}_0 | \Omega_\alpha^{(+)} | \mathbf{Q}_\alpha\rangle - \langle \hat{Q}'_\alpha | \sigma_\alpha(E;r) | \hat{Q}_\alpha\rangle. \tag{V.8}$$

Except for some mass factors and wavefunction normalizations, Eq. (V.8) is the starting point for Smith's study of time delay. One cannot expect to find exactly the same mathematical formulas as Smith gives in his Eq. (38) and (39) since they are based upon idealized and approximate forms for the $N$-body wavefunctions $\langle \mathbf{X}_0 | \Omega_\alpha^{(+)} | \mathbf{Q}_\alpha\rangle$.

To conclude this section, we indicate several reasons why the theory of time delay is of interest. It represents an observable that is a residual consequence of the detailed time-dependent dynamical evolution of the $N$-body scattering system. For two-particle systems considered in a specific partial wave the time delay is proportional to the energy derivative of the phase shift. Thus the general theory of time delay provides a method of defining a universal phase shift-like functional that is a characteristic of the scattering process. So far the applications of time delay have exploited this analogy with the phase shift. The theory has been used to derive an extended Levinson's theorem[18] in the two-body problem and to study the role of the collision process in statistical mechanics.[19-21] It has been particularly useful in the determination of the density expansion of the equation of state for a chemically reacting quantum gas.[21] The multispecies virial coefficients that enter this expansion all turn out to be Laplace transforms in the energy variable of the trace of the time delay operator $q_\alpha(E)$. Thus the equation of state is sensitive only to the time delay of the collision process.

## APPENDIX

This Appendix studies the nature of projection operators on a $n$-dimensional space. Particularly, we investigate the behavior of the projection operators when acting on the types of generalized functions encountered in this problem.

We assume that our operators act on the set of functions $\mathcal{F}$. We are interested in the $r\to\infty$ limit of the projection operators $P_\beta(r)$ for all $\beta$. For a limit of this kind we establish convergence of our results in the weak sense.

Since $P_\beta(r)$ converges to the identity in the strong sense, we immediately have

$$\lim_{r\to\infty} (f'_\beta, P_\beta f_\beta) = (f'_\beta, f_\beta), \tag{A1}$$

where $f'_\beta$ and $f_\beta$ are any functions in $\mathcal{F}$. We now want to express the left-hand side of Eq. (A.1) in momentum space. Therefore, we first introduce some further notation. Let the $3(N-1)=n_0$ dimensional vector $\mathbf{Q}_0$, describing the relative motion of $N$ free particles, be split up into a $3(N_\beta-1)=n$ dimensional vector $\mathbf{Q}_\beta$ describing the relative motion of $N_\beta$ clusters and a $(n_0-n)$-dimensional vector $\mathbf{Q}^\beta$ representing the internal coordinates of the clusters. We will call the conjugate coordinates related to these momenta, respectively $\mathbf{X}_0$, $\mathbf{X}_\beta$, and $\mathbf{Y}_\beta$. Then, the left-hand side of Eq. (A1) can be written as

$$(f'_\beta,P_\beta(r)f_\beta) = \int d^n\mathbf{Q}'_\beta\, d^n\mathbf{Q}_\beta f'^*_\beta(\mathbf{Q}_\beta)\langle\mathbf{Q}'_\beta|P_\beta(r)|\mathbf{Q}_\beta\rangle f_\beta(\mathbf{Q}_\beta), \tag{A2}$$

where according to Eqs. (II.6), (II.7), and (III.24)

$$\langle\mathbf{Q}'_\beta|P_\beta(r)|\mathbf{Q}_\beta\rangle = \int d^{n_0-n}\mathbf{Q}'_\beta d^{n_0-n}\mathbf{Q}^\beta\Big(\prod_{bl}\Psi^\beta_{bl}\Big)(\mathbf{Q}'^\beta)^* \times\langle\mathbf{Q}'_0|P(r)|\mathbf{Q}_0\rangle\Big(\prod_{bl}\Psi^\beta_{bl}\Big)(\mathbf{Q}^\beta), \tag{A3}$$

and

$$F(\mathbf{Q}_0)\equiv F_\beta(\{\mathbf{Q}^\beta,\mathbf{Q}_\beta\})=\Big(\prod_{bl}\Psi^\beta_{bl}\Big)(\mathbf{Q}^\beta)f_\beta(\mathbf{Q}_\beta). \tag{A4}$$

The kernel for $P(r)$ given in formula (A3) is

$$\langle\mathbf{Q}'_0|P(r)|\mathbf{Q}_0\rangle = \frac{1}{(2\pi)^{n_0}}\int_{|\mathbf{x}|<r} d^{n_0}\mathbf{x}\,\exp[i(\mathbf{Q}_0-\mathbf{Q}'_0)\cdot\mathbf{x}]. \tag{A5}$$

Introducing a $n_0$-dimensional spherical coordinate system, a straightforward computation gives

$$\langle\mathbf{Q}'_0|P(r)|\mathbf{Q}_0\rangle = \Big(\frac{r}{2\pi|\mathbf{Q}_0-\mathbf{Q}'_0|}\Big)^{n_0/2}J_{n_0/2}(r|\mathbf{Q}_0-\mathbf{Q}'_0|), \tag{A6}$$

wher $J$ is the Bessel function of the first kind.

We now want to prove a series of lemma's used in Secs. III and IV.

*Lemma A* 1: Let $f_\beta(\mathbf{Q}_\beta)\in\mathcal{F}(\mathbf{Q}_\beta)$ and $f_\gamma(\mathbf{Q}_\gamma)\in\mathcal{F}(\mathbf{Q}_\gamma)$. Let $(\Pi_{bl}\Psi^\beta_{bl})(\mathbf{Q}^\beta)$ and $(\Pi_{cl}\Psi^\gamma_{cl})(\mathbf{Q}^\gamma)$ be bounded almost everywhere. Set $I_{\beta\gamma}(r,t)$ to be

$$I_{\beta\gamma}(r,t)=(\Phi_\beta(t),P(r)\Phi_\gamma(t)),\quad \beta\neq\gamma, \tag{A7}$$

then

$$\int_1^\infty dt\,|I_{\beta\gamma}(r,t)|<C<\infty, \tag{A8}$$

where $C$ is a constant independent of $r$.

*Proof*: Using the expression (III.5) for the outgoing waves, the coordinate space representation of the time evolution operator (III.11) and the coordinate space equivalent of Eq. (A3), we can write Eq. (A7) in the following way,

$$\begin{aligned}I_{\beta\gamma}(r,t) &= M\int d\mathbf{X}'_\beta d\mathbf{Y}'_\beta d\mathbf{X}_\gamma d\mathbf{Y}_\gamma d\mathbf{X}''_0 f^*_\beta(\mathbf{X}'_\beta)\\ &\times\Big(\prod_{bl}\Psi^\beta_{bl}\Big)(\mathbf{Y}'_\beta)^*\delta(\mathbf{Y}''_\beta-\mathbf{Y}'_\beta)\\ &\times(2\pi it)^{-n_\beta/2}\exp[-i(\epsilon^\beta-\epsilon^\gamma)t]\\ &\times\exp[-i(\mathbf{X}'_\beta-\mathbf{X}''_\beta)^2/4t]\Theta(r^2-X''^2_0)(2\pi it)^{-n_\gamma/2}\\ &\times\exp[i(\mathbf{X}''_\gamma-\mathbf{X}_\gamma)^2/4t]\\ &\times\Big(\prod_{cl}\Psi^\gamma_{cl}\Big)(\mathbf{Y}_\gamma)\delta(\mathbf{Y}''_\gamma-\mathbf{Y}_\gamma)f_\gamma(\mathbf{X}_\gamma),\end{aligned} \tag{A9}$$

where all the reduced mass factors are put into the constant $M$ and where $n_\beta=3(N_\beta-1)$, $n_\gamma=3(N_\gamma-1)$.

We first assume that we are in a breakup scattering situation and take $N_\gamma>N_\beta$. We then introduce the notation $\mathbf{X}_\gamma=\{\mathbf{X}_\beta,\mathbf{Y}_{\beta-\gamma}\}$. Integrating out the $\mathbf{X}''_\beta$ part of Eq. (A9), using Eqs. (A5) and (A6), we get

$$\begin{aligned}I_{\beta\gamma}(r,t) &= M_1\int d\mathbf{X}'_\beta d\mathbf{X}_\beta d\mathbf{Y}_{\beta-\gamma}d\mathbf{Y}''_\beta f^*_\beta(\mathbf{X}'_\beta)\\ &\times\Big(\prod_{bl}\Psi^\beta_{bl}\Big)(\mathbf{Y}''_\beta)^*\Theta(r^2-\mathbf{Y}''_\beta)\\ &\times\exp[i(X^2_\beta-X'^2_\beta)/4t]\,t^{-n_\gamma/2}\Big(\frac{R}{|\mathbf{X}'_\beta-\mathbf{X}_\beta|}\Big)^{n_\beta/2}\\ &\times J_{n_\beta/2}\Big(\frac{R}{2t}|\mathbf{X}'_\beta-\mathbf{X}_\beta|\Big)\exp[i(\mathbf{Y}''_{\beta-\gamma}-\mathbf{Y}_{\beta-\gamma})^2/4t]\\ &\times\exp[-i(\epsilon^\beta-\epsilon^\gamma)t]\Big(\prod_l\Psi^\beta_l\Big)(\mathbf{Y}''_\gamma)f_\gamma(\{\mathbf{X}_\beta,\mathbf{Y}_{\beta-\gamma}\}),\end{aligned} \tag{A10}$$

where $R^2=r^2-\mathbf{Y}''^2_\beta$. Next, we consider the $\mathbf{X}_\beta$ and $\mathbf{X}'_\beta$ integral part in Eq. (A10). Introducing the variable

$$\mathbf{z}=\frac{R}{2t}(\mathbf{X}'_\beta-\mathbf{X}_\beta), \tag{A11}$$

that part can be written as

$$\begin{aligned}&t^{(n_\beta-n_\gamma)/2}\int d\mathbf{X}_\beta\,dz\,d\hat{z}\,f^*_\beta\Big(\mathbf{X}_\beta+2t\frac{\mathbf{z}}{R}\Big)\\ &\times\exp\{i[X^2_\beta-(\mathbf{X}_\beta+2t\mathbf{z}/R)^2]/4t\}\\ &\times z^{n_\beta/2-1}J_{n_\beta/2}(z)f_\gamma(\{\mathbf{X}_\gamma,\mathbf{Y}_{\beta-\gamma}\}).\end{aligned} \tag{A12}$$

Since $f_\beta$ is infinitely differentiable we may integrate (A12) $m$ times by parts employing the identity

$$\frac{1}{z}\frac{d}{dz}[z^{-n/2+1}J_{n/2-1}(z)] = -z^{-n/2}J_{n/2}(z). \quad (A13)$$

It is a straightforward exercise to show that all the surface terms vanish due to our conditions on the $f$'s. The result reads then

$$t^{(n_\beta - n_\gamma)/2}\int d\mathbf{X}_\beta\, dz\, d\hat{\mathbf{z}} z^{n_\beta/2-1-m}J_{n_\beta/2-m}(z)$$
$$\times \sum_{l=0}^{m} a_l z^l \frac{d^l}{dz^l}\left[f^*_\beta\left(\mathbf{X}_\beta + 2t\frac{\mathbf{z}}{R}\right)\right.$$
$$\left.\times \exp\{i[X_\beta^2 - (\mathbf{X}_\gamma + 2t\mathbf{z}/R)^2]/4t\}\right]f_\gamma(\{\mathbf{X}_\beta,\mathbf{Y}_{\beta-\gamma}\}), \quad (A14)$$

where the $a_l$ are constants obtained from sequence of partial integrations. Since it is not necessary to know what they are we do not bother to write them out.

We then have to consider the cases of $n_\beta$ even and $n_\beta$ odd separately. Start with $n_\beta$ odd. Here take $m = n_\beta/2 - \frac{1}{2}$. Then if we set $\mathbf{u} = 2t\mathbf{z}/R$ and if we work out the $l$ th derivative in Eq. (A14), we arrive at

$$|I_{\beta\gamma}(r,t)|$$
$$\leqslant M_2|t|^{(n_\beta - n_\gamma)/2}\int d\mathbf{X}_\beta\, d\mathbf{Y}_{\beta-\gamma}\, d\mathbf{u}\, d\mathbf{Y}''_\beta$$
$$\times\Theta(r^2 - Y''^2_\beta)\left|\left(\prod_{bl}\Psi^\beta_{bl}\right)(\mathbf{Y}''_\beta)\right|$$
$$\times\left|\left(\prod_{cl}\Psi^\gamma_{cl}\right)(\mathbf{Y}''_\gamma)\right|\left|\sin\frac{R}{2t}u\right|$$
$$\times\sum_{l=0}^{n_\beta/2-1/2}\sum_{m=0}^{l}\sum_{r=0}^{[m/2]}\left|u^{l-n_\beta}\frac{d^{l-m}f^*_\beta(\mathbf{X}_\beta+\mathbf{u})}{du^{l-m}}\right.$$
$$\left.\times\left(-\frac{i}{t}\right)^{m-r}\right|[(\mathbf{X}_\beta+\mathbf{u})\cdot\hat{u}]^{m-2r}$$
$$\times|f_\gamma(\{\mathbf{X}_\beta,\mathbf{Y}_{\beta-\gamma}\})|. \quad (A15)$$

The condition on the internal cluster wavefunction $\Psi^\beta_{bl}$ and $\Psi^\gamma_{cl}$ ensures that the $\mathbf{Y}''_\beta$ integral in Eq. (A15) exists. If we furthermore restrict $t\geqslant 1$ and introduce again $\mathbf{X}'_\beta = \mathbf{u} + \mathbf{X}_\beta$, we get

$$|I_{\beta\gamma}(r,t)|$$
$$\leqslant M_3 t^{(n_\beta - n_\gamma)/2}\int d\mathbf{X}'_\beta\, d\mathbf{X}_\beta\, d\mathbf{Y}_{\beta-\gamma}\sum_{l,m,r}$$
$$\times\frac{X'^{m-2r}_\beta}{|\mathbf{X}'_\beta - \mathbf{X}_\beta|^{n_\beta - l}}\left|\frac{d^{l-r}f^*_\beta(\mathbf{X}'_\beta)}{dX'^{l-r}_\beta}\right||f_\gamma(\{\mathbf{X}_\beta,\mathbf{Y}_{\beta-\gamma}\})|.$$
$$(A16)$$

Splitting up this integral in a $|\mathbf{X}'_\beta - \mathbf{X}_\beta|\geqslant 1$ part and a $|\mathbf{X}'_\beta - \mathbf{X}_\beta| < 1$ part and using the condition on the $f$'s, we can finally show that

$$|I_{\beta\gamma}(r,t)|\leqslant M_4 t^{(n_\beta - n_\gamma)/2}, \quad (A17)$$

where $M_4$ is finite and independent of $r$. A parallel argument works for $n_\beta$ even.

In the beginning of the proof we have assumed that $N_\gamma > N_\beta$. It is clear that the case $N_\gamma < N_\beta$ can be discussed in exactly the same way if we interchange the role of the $\gamma$ and $\beta$ indices. Only when $N_\gamma = N_\beta$, which describes rearrangement scattering, the analysis seems to break down. In this case, we still have to consider two different situations. First, if we have total rearrangement such that all of the initial internal particle coordinates are a complete set, then there are no intermediate integrations over $\mathbf{X}''_0$ in Eq. (A9) because of the $\delta$ functions. In that case we can derive directly from this expression (A9) that

$$|I_{\beta\gamma}(r,t)|$$
$$\leqslant M_1|t|^{-(n_\beta + n_\gamma)/2}\int d\mathbf{X}'_\beta\, d\mathbf{Y}'_\beta\, d\mathbf{X}_\gamma\, d\mathbf{Y}_\gamma|f_\beta(\mathbf{X}'_\beta)|$$
$$\times\left|\left(\prod_{bl}\Psi^\beta_{bl}\right)(\mathbf{Y}'_\beta)\right|\left|\left(\prod_{cl}\Psi^\gamma_{cl}\right)(\mathbf{Y}_\gamma)\right||f_\gamma(\mathbf{X}_\gamma)|, \quad (A18)$$

which immediately leads to

$$|I_{\beta\gamma}(r,t)|\leqslant M_2|t|^{-(n_\beta + n_\gamma)/2}. \quad (A19)$$

Secondly, if we are in a partial rearrangement situation, we still have to do the intermediate integrations over some of the $\mathbf{X}''_0$ coordinates, say $\mathbf{X}''_{\beta\gamma}$, in Eq. (A9). We can then repeat the breakup discussion, where $n_\beta$ is now replaced by the dimension of $\mathbf{X}''_{\beta\gamma}$.

So we have proved in general that

$$|I_{\beta\gamma}(r,t)|\leqslant M'|t|^{-p}, \quad (A20)$$

where $M'$ is a finite constant independent of $r$ and where $|p|\geqslant\frac{3}{2}$. Integration over $t$ of this result completes the proof of the lemma.

*Lemma A 2*: For $\Pi_{bl}\Psi^\beta_{bl}\in L^2(R^{n_0 - n})$ we have that

$$|\langle\mathbf{Q}'_\beta|P_\beta(r)|\mathbf{Q}_\beta\rangle| < C(r)\left\|\prod_{bl}\Psi^\beta_{bl}\right\|^2, \quad (A21)$$

where

$$C(r) = r^n\left(\frac{m_0}{m_\beta}\right)^{n/2}\frac{1}{2^{n-1}\pi^{n/2}\Gamma(n/2)}. \quad (A22)$$

*Proof*: it is convenient to carry out the proof in coordinate space. So we write

$$\langle\mathbf{Q}'_\beta|P_\beta(r)|\mathbf{Q}_\beta\rangle$$
$$= \frac{1}{(2\pi)^n}\int d^n\mathbf{X}'_\beta\, d^n\mathbf{X}_\beta$$
$$\times\exp[i(\mathbf{Q}_\beta\cdot\mathbf{X}_\beta - \mathbf{Q}'_\beta\cdot\mathbf{X}'_\beta)]\langle\mathbf{X}'_\beta|P_\beta(r)|\mathbf{X}_\beta\rangle.$$
$$(A23)$$

Using the coordinate space equivalent of Eq. (A3), we can write formula (A23) as

$$\langle \mathbf{Q}'_\beta | P_\beta(r) | \mathbf{Q}_\beta \rangle = \frac{1}{(2\pi)^n} \int d^n \mathbf{X}_\beta \Theta(2m_0 r^2 - \widetilde{X}^2_\beta) \exp[i(\mathbf{Q}_\beta - \mathbf{Q}'_\beta)\cdot \mathbf{X}_\beta] \int d^{n_0 - n} \mathbf{Y}_\beta \Theta(2m_0 r^2 - \widetilde{X}^2_\beta - \widetilde{Y}^2_\beta) \left| \left( \prod_{bl} \Psi^\beta_{bl})(\mathbf{Y}_\beta) \right) \right|^2, \quad \text{(A24)}$$

where $\widetilde{X}^2_\beta = 2m_\beta X^2_\beta$ and $\widetilde{Y}^2_\beta = 2\mu_\beta Y^2_\beta$. The second integral on the right is bounded by $\|\Pi_{bl}\Psi^\beta_{bl}\|^2$. So, taking the absolute value of Eq. (A24) completes the proof of the lemma.

*Lemma A 3*: For $F'$ and $F\in\mathscr{F}(\mathbf{Q}_0)$, the following property holds

$$\lim_{r\to\infty} \int d^{n_0}\mathbf{Q}'_0 d^{n_0}\mathbf{Q}_0 F'^*(\mathbf{Q}'_0) \frac{\langle \mathbf{Q}'_0 | P(r) | \mathbf{Q}_0\rangle}{Q'^2_\beta - Q^2_\beta} F(\mathbf{Q}_0) = - \int d^{n_0}\mathbf{Q}'_0 F'^*(\mathbf{Q}'_0) \frac{d}{dQ_\beta} \left[ \left( \frac{Q_\beta}{Q'_\beta} \right)^{(n-1)/2} \frac{F(\{\mathbf{Q}'^\beta, Q_\beta \hat{Q}'_\beta\})}{Q_\beta + Q'_\beta} \right]_{Q_\beta = Q'_\beta}, \quad \text{(A25)}$$

where the integral on the left is defined as a principal-value integral.

*Proof*: Introducing the variables

$$\mathbf{z} = r(\mathbf{Q}_0 - \mathbf{Q}'_0), \quad \mathbf{x} = r(\mathbf{Q}_\beta - \mathbf{Q}'_\beta), \quad \mathbf{y} = r(\mathbf{Q}^\beta - \mathbf{Q}'^\beta),$$

where $\mathbf{x}$ and $\mathbf{y}$ are respectively $n$- and $(n_0 - n)$-dimensional vectors, one has

$$z = (x^2 + y^2)^{1/2}, \quad Q'^2_\beta - Q^2_\beta = -\frac{x}{r}\left( \frac{x}{r} + 2Q'_\beta \cos\Theta_1 \right),$$

where $\Theta_1$ is the angle between $\mathbf{Q}'_\beta$ and $\mathbf{x}$. So we have, using Eq. (A6),

$$\int d^{n_0}\mathbf{Q}'_0 d^{n_0}\mathbf{Q}_0 F'^*(\mathbf{Q}'_0) \frac{\langle \mathbf{Q}'_0 | P(r) | \mathbf{Q}_0\rangle}{Q'^2_\beta - Q^2_\beta} F(\mathbf{Q}_0) = - \int d^{n_0}\mathbf{Q}'_0 \, d^{n_0}\mathbf{z} F'^*(\mathbf{Q}'_0) \frac{J_{n_0/2}(z)}{(2\pi z)^{n_0/2}} \frac{F(\{\mathbf{Q}'_\beta + \mathbf{x}/r, \mathbf{Q}'^\beta + \mathbf{y}/r\})}{x/r(x/r + 2Q'_\beta\cos\Theta_1)}. \quad \text{(A26)}$$

To evaluate the right-hand side of Eq. (A26) we introduce a $n$-dimensional spherical coordinate system $\{x/r, \Theta_1, \Theta_2, \ldots, \Theta_{n-1}\}$ to describe the vector $\mathbf{x}/r$. We note that if we perform the integration over $d\hat{x}$ first the denominator vanishes only when the $\Theta_1$ integration is carried out. The integration over $\mathbf{y}$ causes no special problems. If we furthermore note that $z/r\to 0$ if and only if $x/r \to 0$ and $y/r \to 0$, then we are motivated to write the right-hand side of Eq. (A26) as

$$- \int d^{n_0}\mathbf{Q}'_0 \, d^n\mathbf{x} \, d^{n_0 - n}\mathbf{y} F'^*(\mathbf{Q}'_0) \frac{J_{n_0/2}(z)}{(2\pi z)^{n_0/2}} G_{x/r}(\mathbf{Q}'_\beta, \mathbf{Q}'^\beta + \mathbf{y}/r), \quad \text{(A27)}$$

where

$$G_{x/r}(\mathbf{Q}'_\beta, \mathbf{Q}'^\beta + \mathbf{y}/r) = \tfrac{1}{2} \int_{-1}^{+1} d\cos\Theta_1 \frac{\sin^{n-3}\Theta_1}{x/r(x/r + 2Q'_\beta\cos\Theta_1)} \bar{F}_{x/r}(\mathbf{Q}'_\beta, \cos\Theta_1, \mathbf{Q}'^\beta + \mathbf{y}/r), \quad \text{(A28)}$$

$$\bar{F}_{x/r}(\mathbf{Q}'_\beta, \cos\Theta_1, \mathbf{Q}'^\beta + \mathbf{y}/r) = \frac{\Gamma(n/2)}{\pi^{n/2}} \int_0^\pi \sin^{n-3}\Theta_2 \, d\Theta_2 \cdots \int_0^{2\pi} d\Theta_{n-1} F\left( \left\{ \mathbf{Q}'_\beta + \frac{\mathbf{x}}{r}, \mathbf{Q}'^\beta + \frac{\mathbf{y}}{r} \right\} \right), \quad \text{(A29)}$$

and

$$\frac{x}{r} = \left( \frac{z^2}{r^2} - \frac{y^2}{r^2} \right)^{1/2}.$$

Examination of these formulas indicate that $G_{x/r}$ is the average value of the function

$$F\left( \left\{ \mathbf{Q}'_\beta + \frac{\mathbf{x}}{r}, \mathbf{Q}'^\beta + \frac{\mathbf{y}}{r} \right\} \right) (Q'_\beta - Q'^2_\beta)^{-1}$$

summed over the surface of a sphere centered at $\mathbf{Q}'_\beta$ with radius

$$\left( \frac{z^2}{r^2} - \frac{y^2}{r^2} \right)^{1/2}.$$

The function $\bar{F}_{x/r}(\mathbf{Q}'_\beta, \cos\Theta_1, \mathbf{Q}'^\beta + \mathbf{y}/r)$ is the nonsingular part of the average and integral (A28) is the integral over the singular part.

The next step in our proof is to show that $G_{x/r}(\mathbf{Q}'_\beta,\mathbf{Q}'^\beta+\mathbf{y}/r)$ is continuous in $z/r$. Since we know that $G_{x/r}$ is integrable with respect to $d^nz$ because of Eq. (A29) and the fact that $F(\{\mathbf{Q}'_\beta+\mathbf{x}/r,\mathbf{Q}'^\beta+\mathbf{y}/r\})$ is integrable, we shall then be justified in using the weak convergence of $P(r)$ to conclude that

$$\lim_{r\to\infty} -\int d^{n_0}\mathbf{Q}'_0\, d^n\mathbf{x}\, d^{n_0-n}\mathbf{y} F'^*(\mathbf{Q}'_0)\frac{J_{n_0/2}(z)}{(2\pi z)^{n_0/2}}G_{x/r}(\mathbf{Q}'_\beta,\mathbf{Q}'^\beta+\mathbf{y}/r)$$
$$= -\int d^{n_0}\mathbf{Q}'_0 F'^*(\mathbf{Q}'_0)G_0(\mathbf{Q}'_\beta,\mathbf{Q}'^\beta+0). \tag{A30}$$

So, let us investigate the behavior of

$$G_{x/r}\left(\mathbf{Q}'_\beta,\mathbf{Q}'^\beta+\frac{\mathbf{y}}{r}\right) \quad \text{as} \quad \frac{z}{r}=\frac{(x^2+y^2)^{1/2}}{r}\to 0.$$

If we define

$$w=\frac{x}{r}\cos\Theta_1+\frac{x^2}{2Q'_\beta r^2}, \tag{A31}$$

then Eq. (A28) becomes

$$G_{x/r}\left(\mathbf{Q}'_\beta,\mathbf{Q}'^\beta+\frac{\mathbf{y}}{r}\right)$$
$$=\frac{r}{4Q'_\beta x}\int_{-x/r+x^2/2Q'_\beta r^2}^{x/r+x^2/2Q'_\beta r^2} dw\left(1-\frac{x^2}{4Q'^2_\beta r^2}-\frac{r^2w^2}{x^2}+\frac{w}{Q'_\beta}\right)^{(n-3)/2}\overline{F}\left(\mathbf{Q}'_\beta,w,\mathbf{Q}'^\beta+\frac{\mathbf{y}}{r}\right)\frac{1}{w}. \tag{A32}$$

This is a principal-value integral of the form

$$\frac{1}{b+a}\int_{-a}^{b}\frac{dw}{w}h(w)=\frac{1}{b+a}\int_{-a}^{b}dw\frac{h(w)-h(0)}{w}+\frac{h(0)}{b+a}\int_{-a}^{b}\frac{dw}{w},\quad a>0,\ b>0, \tag{A33}$$

where $h$ is a differentiable function. We need to find the value of Eq. (A33) when $a\to 0$ and $b\to 0$. Since the integrand of the first integral on the left of Eq. (A33) is continuous, we may use the mean value theorem to write

$$\frac{1}{b+a}\int_{-a}^{b}\frac{dw}{w}h(w)=\left[\frac{h(w_1)-h(0)}{w_1}\right]+\frac{h(0)}{b+a}\ln\frac{b}{a}, \tag{A34}$$

where $w_1$ is some point in the interval $(b,-a)$. As the interval size goes to zero the factor in the square brackets becomes the derivative of $h$ at $w=0$. The second factor is just a constant times $h(0)$. If we apply formula (A34) to Eq. (A32), we obtain in the limit $z/r\to 0$.

$$G_0(\mathbf{Q}'_\beta,\mathbf{Q}'^\beta)=\frac{n-2}{4}\frac{1}{Q'^2_\beta}F(\{\mathbf{Q}'_\beta,\mathbf{Q}'^\beta\})+\frac{1}{2Q'_\beta}\frac{d}{dQ'_\beta}F(\{\mathbf{Q}'_\beta,\mathbf{Q}'^\beta\})$$

$$=\frac{d}{dQ_\beta}\left[\left(\frac{Q_\beta}{Q'_\beta}\right)^{(n-1)/2}\frac{F(\{\mathbf{Q}'_\beta,Q_\beta\widehat{Q}'_\beta\})}{Q_\beta+Q'_\beta}\right]_{Q_\beta=Q'_\beta} \tag{A35}$$

Substituting this result into Eq. (A30) completes the proof of Lemma A3.

*Lemma A 4*: For $f'_\beta$ and $f_\beta\in\mathscr{F}(\mathbf{Q}_\beta)$ we have

$$\lim_{r\to\infty}\int d^n\mathbf{Q}'^*_\beta\, d^n\mathbf{Q}_\beta f'^*_\beta(\mathbf{Q}'_\beta)$$
$$\times\frac{\langle\mathbf{Q}'_\beta|P_\beta(r)|\mathbf{Q}_\beta\rangle}{Q'^2_\beta-Q^2_\beta}f_\beta(\mathbf{Q}_\beta)=-\int d^n\mathbf{Q}'_\beta f'^*_\beta(\mathbf{Q}'_\beta)\frac{d}{dQ_\beta}\left[\left(\frac{Q_\beta}{Q'_\beta}\right)^{(n-1)/2}\frac{f_\beta(Q_\beta\widehat{Q}'_\beta)}{Q_\beta+Q'_\beta}\right]_{Q_\beta=Q'_\beta}. \tag{A36}$$

*Proof*: This result follows immediately from Lemma A3 and the Eqs. (A3) and (A4).

*Lemma A 5*: For $f'_\beta, f_\beta\in\mathscr{F}(\mathbf{Q}_\beta)$ we have

$$\int d^n\mathbf{Q}'_\beta\, d^n\mathbf{Q}_\beta f'^*_\beta(\mathbf{Q}'_\beta)\delta(\widetilde{Q}^2_\beta-\widetilde{Q}'^2_\beta)\langle\mathbf{Q}'_\beta|P_\beta(t)|\mathbf{Q}_\beta\rangle f_\beta(\mathbf{Q}_\beta)\underset{r\to\infty}{\to} r\int d^n\mathbf{Q}_\beta f'^*_\beta(\mathbf{Q}_\beta)\frac{(m_0m_\beta)^{1/2}}{\pi Q_\beta}f_\beta(\mathbf{Q}_\beta)+O\left(\frac{1}{r}\right), \tag{A37}$$

for all $\beta$.

*Proof*: The left-hand side of Eq. (A37) can be written as

$$\int d^{\,n}\mathbf{Q}'_\beta\, d^{\,n}\mathbf{Q}_\beta\, d^{\,n_0-n}\mathbf{Y}_\beta\, d^{\,n_0-n}\mathbf{Y}'_\beta f'^{*}_\beta(\mathbf{Q}'_\beta)\delta(\widetilde{Q}^2_\beta-\widetilde{Q}'^2_\beta)\Big(\prod_{bl}\Psi^\beta_{bl}\Big)(\mathbf{Y}'_\beta)^*\langle\mathbf{Q}'_\beta\mathbf{Y}'_\beta|P(r)|\mathbf{Q}_\beta\mathbf{Y}_\beta\rangle\Big(\prod_{bl}\Psi^\beta_{bl}\Big)(\mathbf{Y}_\beta)f_\beta(\mathbf{Q}_\beta). \tag{A38}$$

The mixed matrix element of $P(r)$ can be calculated in a straightforward way so that we get for the expression (A38)

$$\int d^{\,n_0-n}\mathbf{Y}_\beta\Big|\Big(\prod_{bl}\Psi^\beta_{bl}\Big)(\mathbf{Y}_\beta)\Big|^2\Theta\Big(\frac{m_0}{\mu_\beta}r^2-Y^2_\beta\Big)\int d^{\,n}\mathbf{Q}_\beta\, d^{\,n-1}\widehat{Q}'_\beta m_\beta Q^{\,n-2}_\beta$$
$$\times f'^{*}_\beta(Q_\beta\widehat{Q}'_\beta)\Big(\frac{R}{2\pi|Q_\beta\widehat{Q}'_\beta-\mathbf{Q}_\beta|}\Big)^{n/2}J_{n/2}(R\,|Q_\beta\widehat{Q}'_\beta-\mathbf{Q}_\beta|)f_\beta(\mathbf{Q}_\beta), \tag{A39}$$

where

$$R=\Big(\frac{m_0}{m_\beta}\Big)^{1/2}r\Big(1-\frac{\mu_\beta}{m_0}\frac{Y^2_\beta}{r^2}\Big)^{1/2}. \tag{A40}$$

Introducing the variable

$$z=R\,|Q_\beta\widehat{Q}'_\beta-\mathbf{Q}_\beta|=\sqrt{2}Q_\beta R\,(1-\cos\theta)^{1/2}, \tag{A41}$$

where $\theta$ is the angle between $\widehat{Q}_\beta$ and $\widehat{Q}'_\beta$, and introducing a $n$-dimensional spherical coordinate system

$$\Big\{Q_\beta,\arccos\Big(1-\frac{z^2}{2Q^2_\beta R^2}\Big),\theta'_2,\ldots,\theta'_{n-1}\Big\}$$

to describe the vector $\mathbf{Q}'_\beta$, we obtain for the momentum integrals in Eq. (A40)

$$\int d^{\,n}\mathbf{Q}_\beta m_\beta Q^{\,n-4}_\beta f_\beta(\mathbf{Q}_\beta)R^{\,n-2}\int_0^{2Q_\beta R}dz\Big[1-\Big(1-\frac{z^2}{2Q^2_\beta R^2}\Big)^2\Big]^{(n-3)/2}z^{-n/2+1}J_{n/2}(z)c_n\overline{f'^{*}_\beta}\Big(Q_\beta,\arccos\Big(1-\frac{z^2}{2Q^2_\beta R^2}\Big)\Big). \tag{A42}$$

In this expression $\overline{f'^{*}_\beta}$ is the average of $f'^{*}_\beta$ over the angles $\theta'_2,\ldots,\theta'_{n-1}$, and $c_n$ is a coefficient depending on the dimension. In the following we will take $n$ odd, then $c_n=(n-2)!!\,[2^{n/2}\Gamma(n/2)(n-3)!!]^{-1}$.

Next, we use the method of Lemma A1 and do a number of partial integrations $m$, employing the identity (A13), in order to lower the index of the Bessel function in Eq. (A42). It is straightforward to see that the surface terms do not contribute due to the vanishing of the factor $1-(1-z^2/2Q^2_\beta R^2)^2$ at the endpoints. We then arrive at the following expression

$$\int d^{\,n}\mathbf{Q}_\beta m_\beta Q^{\,n-4}_\beta f_\beta(\mathbf{Q}_\beta)R^{\,n-2}c_n\int_0^{2Q_\beta R}dz\,z^{-n/2+m}J_{n/2-m}(z)\Big(\frac{d}{dz}\frac{1}{z}\Big)^{m-1}$$
$$\times\frac{d}{dz}\Big\{\Big[1-\Big(1-\frac{z^2}{2Q^2_\beta R^2}\Big)^2\Big]^{(n-3)/2}\overline{f'^{*}_\beta}\Big(Q_\beta,\arccos\Big(1-\frac{z^2}{2Q^2_\beta R^2}\Big)\Big)\Big\}. \tag{A43}$$

Taking $m=(n-3)/2$, we get after some algebra

$$\int d^{\,n}\mathbf{Q}_\beta f_\beta(\mathbf{Q}_\beta)\frac{m_\beta}{(2\pi)^{1/2}}\frac{R}{Q_\beta}\int_0^{2Q_\beta R}dz\,z^{-1/2}J_{3/2}(z)\Big[\Big(1-\frac{z^2}{2Q^2_\beta R^2}\Big)^{(n-3)/2}+\cdots\Big]\overline{f'^{*}_\beta}\Big(Q_\beta,\arccos\Big(1-\frac{z^2}{2Q^2_\beta R^2}\Big)\Big), \tag{A44}$$

where we have not explicitly written out the following terms because they still contain the factor $1-(1-z^2/2Q^2_\beta R^2)^2$ to a certain power different from zero, such that they will not contribute to the final result, as we will see immediately.

Indeed, to obtain Eq. (A37) we have to do one more partial integration. If we then take $z=Rx$, it is easy to check that all the integral terms disappear in the limit $r\to\infty$ because of the Riemann–Lebesgue lemma. Also the upper limits $x=2Q_\beta$ of all surface terms vanish for the same reason. The lower limits $x=0$ of the surface terms disappear if they contain the factor $1-(1-x^2/2Q^2_\beta)^2=0$. So, only the lower limit of the surface terms derived from the first term in Eq. (A44) gives a contribution. As a result, we get for Eq. (A39)

$$\int d^{n_0-n}\mathbf{Y}_\beta \left| \left( \prod_{bl} \Psi^\beta_{bl} \right)(\mathbf{Y}_\beta) \right|^2 \Theta\left( \frac{m_0}{\mu_\beta} r^2 - Y_\beta^2 \right) \int d^n\mathbf{Q}_\beta f'^*_\beta(\mathbf{Q}_\beta) \frac{Rm_\beta}{\pi Q_\beta} f_\beta(\mathbf{Q}_\beta). \tag{A45}$$

Looking back at Eq. (A40) for $R$ and expanding the square root in powers of $1/r^2$, it is clear that only the first term $(m_0/m_\beta)^{1/2}r$ survives in the limit $r \to \infty$. In that limit, we can finally use orthonormality of the internal cluster wavefunctions to complete the proof of the lemma.


[1]E.P. Wigner, Phys. Rev. **98**, 145 (1955); F.T. Smith, Phys. Rev. **118**, 349 (1960); M.L. Goldberger and K.M. Watson, *Collision Theory* (Wiley, New York, 1964), p. 485; J.M. Jauch and J.P. Marchand, Helv. Phys. Acta **40**, 217 (1967); H.M. Nussenzveig, Phys. Rev. D **6**, 1534 (1972). A more extensive list of references and the comparative discussion of different concepts of time delay is found in Ref. 7.

[2]D. Bollé and T.A. Osborn, Phys. Rev. D **11**, 3417 (1975).

[3]J.M. Jauch, K.B. Sinha, and B.N. Misra, Helv. Phys. Acta **45**, 398 (1972).

[4]R. Lavine, in *Scattering in Mathematical Physics*, edited by J.A. LaVita and J.P. Marchand (Reidel, Dordrecht, 1974), p. 141.

[5]Ph. A. Martin, Commun. Math. Phys. **47**, 221 (1976).

[6]T.A. Osborn and D. Bollé, Phys. Lett. B **52**, 13 (1974); J. Math. Phys. **16**, 1533 (1975).

[7]D. Bollé and T.A. Osborn, Phys. Rev. D **13**, 299 (1976).

[8]J.M. Jauch, Helv. Phys. Acta **31**, 661 (1958).

[9]M.N. Hack, Nuovo Cimento **13**, 231 (1959).

[10]W. Hunziker, in *Lectures in Theoretical Physics*, edited by A.O. Barut and W.E. Brittin (Gordan and Breach, New York, 1968), Vol. X-A.

[11]C. Chandler and A.G. Gibson, J. Math. Phys. **14**, 1328 (1973); **18**, 2336 (1977).

[12]The basic results of two-Hilbert space theory are not changed by the addition of three-body or four-body potentials, etc., provided that these potentials are sufficiently well behaved. See Ref. 10.

[13]L.D. Faddeev, *Mathematical Aspects of the Three-Body Problem in Quantum Scattering* (Davey, New York, 1965).

[14]J. Ginibre and M. Moulin, Ann. Inst. Henri Poincaré **21**, 97 (1974); L.E. Thomas, Ann. Phys. (N.Y.) **90**, 127 (1975).

[15]V.N. Efimov, Phys. Lett. B **33**, 563 (1970); Sov. J. Nucl. Phys. **12**, 589 (1971); Nucl. Phys. A **210**, 157 (1973).

[16]J.D. Dollard, Commun. Math. Phys. **12**, 193 (1969); J. Math. Phys. **14**, 708 (1973).

[17]F. Riesz and B. Sz-Nagy, *Functional Analysis* (Ungar, New York, 1955).

[18]T.A. Osborn and D. Bollé, J. Math. Phys. **18**, 432 (1977).

[19]F.T. Smith, J. Chem. Phys. **38**, 1034 (1963); Phys. Rev. **131**, 2803 (1963); D. Bedeaux, Physica **45**, 469 (1970).

[20]D. Bollé and H. Smeesters, Phys. Lett. A **62**, 290 (1977).

[21]T.A. Osborn and T.Y. Tsang, Ann. Phys. (N.Y.) **101**, 119 (1976); T.A. Osborn, Phys. Rev. A **16**, 334 (1977).

# Constructive methods based on analytic characterizations and their application to nonlinear elliptic and parabolic differential equations

M. Barnsley and D. Bessis

*Cen Saclay, BP n° 2, 91190 Gif-Sur-Yvette, France*
(Received 1 August 1978)

We consider functional equations $F(\phi,\gamma)=0$, in general nonlinear, whose physically meaningful solution $\phi[\gamma]$ depends on a real or complex parameter $\gamma$. We suppose that positivity plays an important role in the equation; for example, the physical solution $\phi[\gamma]$ may be forcedly positive for all $\gamma$ in some range because it describes a density or a temperature. Furthermore, $F$ may involve operators which themselves possess positivity properties; in particular we think of energy and diffusion operators of the Laplace type. This paper describes and illustrates how such positivity properties can lead one to discover the analytic nature of $\phi[\gamma]$, and shows how in many cases knowledge of this analytic characterization provides a method for reconstructing $\phi[\gamma]$ starting from the functional equation itself. First, it is shown how positivity can generate analytic characterizations with the aid of Bernsteins' theorems. Equipped with this knowledge, we describe methods for attacking the formal perturbation series associated with $F(\phi,\gamma)=0$ to yield monotone, convergent sequences of bounds on $\phi[\gamma]$. Finally two classes of nonlinear problems are considered; one involves a general uniform elliptic operator while the other involves a parabolic operator; both involve a quadratic nonlinearity. In each case the key positivity properties of the operators involved are provided, and it is shown how the analyticity "character analysis" of the physically relevant solution $\phi[\gamma]$ is effected. In the case of the quadratically nonlinear elliptic problem a bifurcation phenomenon is associated with the solution: Here the positivity analysis enables us to describe completely the behavior of the solution when the bifurcation parameter passes through its critical value.

## 1. INTRODUCTION

Since the review[1] by Baker in 1965 of the theory of Padé approximants, in which it was shown how the classical theorems supply a key approach to the approximation of certain wide classes of functions occurring in mathematical physics, a variety of interesting developments have been made.[2] In particular we note that the Padé approximant method has led to good numerical results and a helpful point of view in the case of functions which can be represented by a series of Stieltjes, and that the recognition of new applications has been based to some extent on the following observation. A function which is representable by a series of Stieltjes can always be characterized in the form[3]

$$F(z)=\langle h,\phi_z\rangle,\quad z\in\mathbb{C}, \tag{1.1}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product in a Hilbert space $\mathscr{H}$, $h\in\mathscr{H}$ is given, and $\phi$ is the solution of an equation of the form

$$[H+zV]\phi_z=h \tag{1.2}$$

where $H$ and $V$ are, for example, positive self-adjoint linear mappings from $\mathscr{H}$ into itself. The point is that problems with the underlying structure (1.2) often occur in mathematical physics (Fredholm integral equation, first order Rayleigh–Schrödinger perturbation equations in quantum mechanics, and in fact in almost any situation where the model is basically linear), and it is not infrequent that one is concerned with a linear functional of the solution such as (1.1), which may be a parameter dependent diffusion constant,[4] a linear response function such as the dynamical polarizability,[5] the tangent of a phase shift,[6] etc.

In all such cases one can use Padé approximants to bound or simply to approximate the *expectation value*

$$F(z)=\langle h,\phi_z\rangle=\langle h,[H+zV]^{-1}h\rangle \tag{1.3}$$

of the "resolvant" operator $[H+zV]^{-1}$. In this sense the Stieltjes character of the approximated function has no liason with the detailed nature of the problem (1.2), resting instead on its Hilbert space realization together with such almost heuristic or formal properties as the self-adjointness of the operators involved.

It is among the aims of this paper to show that in many cases, it is possible to analyze the solution $\phi_z$ of the underlying problem itself with the aid of generalized Padé approximants, and to try and pinpoint the type of problem amenable to such analysis. Certain work has already been carried out in this direction, studying for example the $L^2$ convergence of the Padé approximant trial vector[7] to $\phi_z$, but again the results rely on and are concerned with the Hilbert space description of the quantities involved, and not with more explicit details such as the pointwise values of $\phi_z$ when it is treated, say, as a continuous function embedded in some Sobolev space.

To make the above description more concrete, consider the following simple illustration. Let $\phi_z$ be the unique solution in the Hilbert space $L^2[0,1]$ of the problem

$$-\frac{d^2}{dx^2}\phi_z(x)+zV(x)\phi_z(x)=h(x),$$
$$\phi_z(0)=\phi_z(1)=0, \tag{1.4}$$

with $h\in L^2[0,1]$, $V(x)$ a continuous positive function for $x\in[0,1]$, and $z$ a real positive parameter. Then the spectral theorem[8] applied to the self-adjoint linear operator which is naturally associated with the left-hand side of (1.4) and the boundary conditions provides the fact that

 0022-2488/79/061135-11$01.00  

$$F(z) = \int_0^1 h(x)\phi_z(x)dx \qquad (1.5)$$

can be reexpressed as the Stieltjes integral

$$F(z) = \int_0^\infty \frac{d\mu(s)}{(1+sz)}, \qquad (1.6)$$

where $\mu(s)$ is a bounded monotone nondecreasing function on $0 \leqslant s < \infty$. That is, $F(z)$ can be represented by a series of Stieltjes and bounds for it can be obtained with the aid of Padé approximants. However in this problem one can also ask about the analytic structure in $z$ of the function $\phi_z$ itself, because $\phi_z(x)$ actually belongs to the Sobolev space $W_2^{(2)}[0,1]$, consisting of functions on the interval $x \in [0,1]$ which are square integrable, as are their first and second derivatives; and therefore it has a unique continuous representation. In fact, if $h(x) \geqslant 0$ for almost all $x \in [0,1]$, it can be shown[25] that

$$\phi_z(x) = \int_0^\infty \exp(-tz)d\nu_x(t), \quad x \text{ fixed in } [0,1], \qquad (1.7)$$

where $\nu_x(t)$ is a bounded monotone nondecreasing function on $0 \leqslant t < \infty$. Thus $\phi_z(x)$ is the Laplace transform of a positive measure in the variable $z$ for each fixed $x$ and is amenable to approximation and bounding procedures based upon generalized Padé approximants.

The existence of the two representations (1.6) and (1.7) are consequences of distinct positivity properties of the differential operator in (1.4). (1.6) follows essentially from the fact that the operator in (1.4) is positive according to

$$\int_0^1 \psi(x)\left(-\frac{d^2}{dx^2} + zV(x)\right)\psi(x)dx \geqslant 0, \quad z \geqslant 0 \qquad (1.8)$$

for all $\psi(x)$ which have square integrable second derivative and for which $\psi(0) = \psi(1) = 0$. This is the positivity of the operator in the Hilbert space sense. On the other hand, as we shall see, (1.7) is a consequence of the positivity implication

$$\left(-\frac{d^2}{dx^2}\psi(x) + zV(x)\psi(x) \geqslant 0 \text{ for all } x \in [0,1], \text{ and } \psi(0) = \psi(1) = 0\right) \Rightarrow \psi(x) \geqslant 0 \text{ for all } x \in [0,1] \qquad (1.9)$$

when $\psi(x)$ possesses a continuous second derivative on $[0,1]$.

In fact, we will show in Secs. 4 and 5 that the representation (1.7) remains valid in far more general circumstances. We cite the following two instances. In Sec. 4 we show for example that the problem

$$-\frac{d^2}{dx^2}\phi_z(x) + zV(x)\phi_z(x)^2 = h(x), \qquad (1.10)$$
$$\phi_z(0) = \phi_z(1) = 0,$$

possesses precisely one solution $\phi_z(x)$ which both belongs to $W_2^{(2)}[0,1]$ and which is *positive* for $0 < x < 1$; and this solution admits a representation of the form (1.7). In Sec. 5 we show for example that the problem

$$-\frac{\partial^2}{\partial x^2}\phi_z(x,t) + \frac{\partial}{\partial t}\phi_z(x,t) + zV(x)\phi_z(x)^2 = h(x),$$
$$\phi_z(0,t) = \phi_z(1,t) = 0 \quad \text{for all } t > 0, \qquad (1.11)$$
$$\phi_z(x,0) = \phi_0(x) > 0,$$

has only one smooth solution $\phi_z(x,t)$; this solution can also be expressed in the form (1.7). The problems (1.10) and (1.11) are merely to illustrate the kinds of results described in Secs. 4 and 5—the actual results obtained there are far more general.

In Sec. 2 we recall how positivity properties are related to analyticity properties and in particular to the characterization of functions as transforms of positive measures.

In Sec. 3 we describe how various types of functions can be bounded and in principle often reconstructed on the basis of information in the form of coefficients occurring in either a Taylor series or some other purely formal expansion of the function. We concentrate on those classes of functions whose analytic nature may be discovered from positivity considerations.

In Sec. 4 we consider a particular class of nonlinear elliptic boundary value problems and lay out the fundamental theorems which can be used to explore the positivity properties of the solutions of such equations. We then illustrate, first with an algebraic analogy, and then by means of a sequence of theorems whose proofs are briefly sketched, how the positivity analysis can in practice be carried out. The result is a novel method for constructing monotonically converging upper and lower bounds upon the solution of the problem considered, together with a detailed understanding of the nature of this solution. However, we stress that the method is far more general than the particular application and can be applied in many other cases. The reader will see how he himself could readily derive the representation (1.7) associated with the problem (1.4), for example.

In Sec. 5 we indicate how the kind of analysis described above also applies to nonlinear uniformly parabolic problems.

## 2. POSITIVITY AND ANALYTICITY

We begin by recalling how positivity properties of functions of a real variable can generate analyticity properties, by giving the simplest results of Bernstein.[9]

### A. Totally convex functions

Suppose that for some $\epsilon > 0$ we have $f(x) \in C^\infty(0,\epsilon)$ [i.e., $f(x)$ is infinitely differentiable for $0 < x < \epsilon$] and moreover

$$(-1)^k f^{(2k)}(x) \geqslant 0 \quad \text{for all } x \in (0,\epsilon) \text{ and } k = 0,1,2,\cdots.$$

Then $f(x)$ can be analytically continued throughout the complex plane to produce an entire function of order at most 1 and of type at most $\pi/\epsilon$. That is, $f(z)$ is holomorphic at all finite points $z \in \mathbb{C}$ and there exists a constant $c > 0$ such that

$$|f(z)| < ce^{|z|\pi/\epsilon} \quad \text{for all } z \in \mathbb{C}.$$

## B. Completely monotone functions

Suppose that $f(x) \in C^{\infty}(0,\infty)$ and moreover

$$(-1)^k f^{(k)}(x) \geqslant 0 \quad \text{for all } x \in (0,\infty) \text{ and } k = 0,1,2,\cdots.$$

Then $f(x)$ can be analytically continued throughout the region $C^+ = \{z \in \mathbb{C} \mid \text{Re}z > 0\}$, being expressible as the convergent integral

$$f(z) = \int_0^{\infty} e^{-zt} d\alpha(t) \quad \text{for all } z \in C^+,$$

where $\alpha(t)$ is a monotone nondecreasing function for $0 < t < \infty$ [i.e., $d\alpha(t)$ is a positive measure]. Conversely, if $f(z)$ can be expressed in the latter form, then it obeys the former differential inequalities.

## C. Absolutely monotone functions

Suppose that for some $\epsilon > 0$ we have $f(x) \in C^{\infty}(0,\epsilon)$ and moreover

$$f^{(2k)}(x) \geqslant 0 \quad \text{for all } x \in (0,\epsilon) \text{ and } k = 0,1,2,\cdots.$$

Then $f(x)$ can be analytically continued to a function which is holomorphic in the disk $\{z \in \mathbb{C} \mid |z - \epsilon/2| < \epsilon/2\}$.

## D. Stieltjes functions

Suppose that $f(x) \in C^{\infty}(0,\infty)$, that $f(x) \geqslant 0$, and that

$$(-1)^{k-1}[x^k f(x)]^{(2k-1)} \geqslant 0 \quad \text{for all } x \in (0,\infty) \text{ and } k = 1,2,3,\cdots.$$

Then $f(x)$ can be analytically continued throughout the region $C' = \{z \in \mathbb{C} | z \notin (-\infty,0)\}$, being expressible as the convergent integral

$$f(z) = P + \int_0^{\infty} \frac{d\alpha(t)}{(z+t)} \quad \text{for all } z \in C',$$

where $\alpha(t)$ is a monotone nondecreasing function for $0 < t < \infty$, and $P$ is a nonnegative constant. Conversely, if $f(z)$ can be expressed in the latter form, then it obeys the former differential inequalities, and in fact more generally

$$(-1)^{n+k}[x^k f(x)]^{(n)} \geqslant 0 \quad \text{for all } x \in (0,+\infty) \text{ and } n \geqslant k$$

where $k = 0,1,2,\cdots$.

These relations can be recast in a more suitable form, by introducing the quantities

$$f_0^{(m)}(x) = \frac{d^m}{dx^m} f(x)$$

and

$$f_n^{(m)}(x) = \frac{x f_{n-1}^{(m+1)}(x)}{m+1} + f_{n-1}^{(m)}(x),$$

then the necessary and sufficient condition for $f(x)$ to be a Stieltjes function is

$$(-1)^m f_n^{(m)}(x) \geqslant 0 \quad \text{for all } x \in (0,+\infty) \text{ and all } m,n = 0,1,2,\cdots.$$

## 3. POSITIVITY AND BOUNDS

We consider any functional equation

$$F(\phi,\gamma) = 0 \tag{3.1}$$

from which it is possible to extract a formal series expansion in powers of $\gamma$ for the solution of interest

$$\phi(\gamma) = \phi_0 + \gamma\phi_1 + \cdots + \gamma^n\phi_n + \cdots. \tag{3.2}$$

For simplicity we suppose that $\phi$ is real valued, but one can extend the following to the case where $\phi$ is complex, or even an element of a noncommutative algebra; see for example, Refs. 10 and 11.

Five cases can occur:

(i) The series (3.2) is rapidly convergent for the value of $\gamma$ of interest. One can then consider that (3.2) resolves the problem.

(ii) The series (3.2) is convergent but the rate of convergence is too slow to be useful. Thus (3.2) as it stands fails to resolve the problem.

(iii) The series (3.2) is divergent for all $\gamma$ but possesses asymptotic properties [see for example, Ref. 12]. If $\gamma$ is in the asymptotic region, then (3.2) may again resolve the problem because the "effective convergence" can be extremely good.

(iv) The series (3.2) is divergent for all $\gamma$ and has no asymptotic convergence property: The series is useless as it stands.

(v) Some or all the coefficients $\phi_0, \phi_1, \phi_2, \cdots$ are infinite but they possess finite regularizations $\phi_0(\epsilon), \phi_1(\epsilon), \phi_2(\epsilon), \cdots$ in terms of a parameter $\epsilon$, see Ref. 6, such that

$$\phi_k(\epsilon) \to \phi_k \quad \text{as } \epsilon \to 0_+. \tag{3.3}$$

This case occurs frequently for singular problems.

In dealing with problems which occur in mathematical physics, one often finds oneself in one of the situations (ii), (iv), or (v). For cases (ii) and (iv) one can often apply a convergence acceleration technique based on the use of Padé approximants and their generalizations. Such an approach may be placed on a firm mathematical foundation provided that one has some knowledge of the analytic sturcture of $\phi(\gamma)$. Our thesis is that such knowledge can often itself be derived from positivity properties which may be discovered directly from the original equation (3.1). We give some examples of this process in Secs. 4 and 5. The point is that one can often expect positivity properties to be built into physical problems: probabilities, temperatures, pressures, densities, etc., are naturally positive and therefore provide possible candidates.

In case (v) one can sometimes apply a convergence acceleration technique to the regulated series $\{\phi_k(\epsilon) : k = 0,1,2,\cdots\}$ followed by the exploitation of variational properties of the accelerated result in the parameter $\epsilon$, see below.

For the moment we return to the case where the $\phi_i$ exist and pose the following problem: Given that $\phi(\gamma)$ belongs to a precise class of functions and given a finite number $\phi_0, \phi_1, \ldots, \phi_N$ of coefficients occurring in the development (3.2), find the "best" lower and upper bounds for $\phi(\gamma)$:

$$\underline{B}_N(\gamma;\phi_0,\phi_1,\ldots,\phi_N) \leqslant \phi(\gamma) \leqslant \bar{B}_N(\gamma;\phi_0,\phi_1,\ldots\phi_N). \tag{3.4}$$

Since $\underline{B}_N$ and $\bar{B}_N$ must lie on the boundary of the range of values of all functions belonging to the class and possessing the coefficients $\phi_0, \phi_1, \ldots, \phi_N$ it follows that they must form monotonic sequences:

$$\underline{B}_1 \leqslant \underline{B}_2 \leqslant \cdots \leqslant \underline{B}_N \leqslant \cdots \leqslant \underline{B}_\infty \leqslant \phi \leqslant \bar{B}_\infty \leqslant \cdots \leqslant \bar{B}_N \leqslant \cdots \leqslant \bar{B}_2 \leqslant \bar{B}_1, \tag{3.5}$$

where $B_\infty$ is the limit of $B_N$ as $N \to \infty$. When the two limits are equal the problem is determinate,[13] otherwise it is indeterminate. In the latter case the limits may be nonetheless so close that one can say (3.5) resolves the problem. In fact, because of the presence of noise, one always finds oneself in the indeterminate case—"optimization over a class with noise." We note that the existence of $B_N$ has no relation to the convergence or otherwise of the original series (3.2), which may very well diverge for all $\gamma$. The problem is better understood if it is posed in terms of information theory: Given the class to which $\phi(\gamma)$ belongs and given a finite number of derivatives of $\phi$ at the origin, find the best possible bounds for $\phi(\gamma)$ at $\gamma = \gamma_0$.

We now consider a number of cases for which one can explcitly construct the $B_N$'s by supposing that $\phi$ admits a representation

$$\phi[\gamma] = \int_0^\infty K(x\gamma)\, d\mu(x) \tag{3.6}$$

where $d\mu(x)$ is a positive measure and $K(x\gamma)$ is a kernel which is completely monotone in the variable $x\gamma$. For example,

$$\begin{aligned} &K(x\gamma) = e^{-x\gamma}, \quad \text{the Laplace kernel;} \\ &K(x\gamma) = (1 + x\gamma)^{-1}, \quad \text{the Cauchy kernel;} \\ &K(x\gamma) = (1 + x\gamma)^{-p}, p > 0, \quad \text{the Hilbert kernel.} \end{aligned} \tag{3.7}$$

In all cases we associate the problem with an auxiliary Stieltjes function,

$$S(z) = \int_0^\infty \frac{d\mu(x)}{(1 + xz)}, \tag{3.8}$$

which has the formal development

$$S(z) = \sum_{n=0}^\infty (-z)^n \mu_n \quad \text{with } \mu_n = \int_0^\infty x_n\, d\mu(x). \tag{3.9}$$

Knowing a finite number of the $\mu_n$'s one constructs the corresponding Padé approximants $[P - 1/P](z)$ and $[P/P](z)$. They are written

$$[P - 1/P](z) = \sum_{i=1}^P \frac{w_i}{(x_i + z)}$$
and
$$[P/P](z) = \sum_{i=1}^P \frac{\tilde{w}_i}{(\tilde{x}_i + \tilde{z})} + \tilde{w}_0. \tag{3.10}$$

Here the $x_i$'s, $\tilde{x}_i$'s, $w_i$'s, and $\tilde{w}_i$'s are uniquely fixed by the set $\{\mu_0, \mu_1, \ldots \mu_{2P}\}$ according to the equations[14]

$$\begin{aligned} &\sum_{i=1}^P w_i(x_i)^{-(n+1)} = \mu_n \quad \text{for } n = 0,1,\ldots,2P - 1, \\ &\tilde{w}_0 + \sum_{i=1}^P \tilde{w}_i(\tilde{x}_i)^{-1} = \mu_0, \\ &\sum_{i=1}^P \tilde{w}_i(\tilde{x}_i)^{-(n+1)} = \mu_n \quad \text{for } n = 1,2,\ldots,2P. \end{aligned} \tag{3.11}$$

The requisite $\mu_n$'s are obtained from the $\phi_n$'s by noting that

$$\begin{aligned} \phi(\gamma) &= \int_0^\infty K(x\gamma)\, d\mu(x) \\ &= \sum_{n=0}^\infty (-1)^n \gamma^n K_n \int_0^\infty x^n\, d\mu(x), \end{aligned} \tag{3.12}$$

where the completely monotone kernel $K(u)$ has been expanded,

$$K(u) = \sum_{n=0}^\infty (-1)^n K_n u_n, \tag{3.13}$$

so that

$$u_n = (-1)^n \phi_n / K_n, \quad n = 0,1,2,\cdots. \tag{3.14}$$

We now have the following result: For the completely monotone kernel $K(x\gamma)$,

$$\bar{B}_N(\gamma;\phi_0,\phi_1,\ldots,\phi_N) = \tilde{w}_0 K(0) + \sum_{i=1}^P \tilde{w}_i K(\gamma \tilde{x}_i) \tag{3.15}$$

and

$$\underline{B}_N(\gamma;\phi_0,\phi_1,\ldots,\phi_N) = \sum_{i=1}^P w_i K(\gamma x_i), \tag{3.16}$$

where $P$ is the integer part of $N/2$. We remark that the bounds are none other than Gaussian quadratures which use the set of zeros (the $x_i$'s and $\tilde{x}_i$'s) associated with the orthogonal polynomials constructed with respect to the measure $d\mu(x)$. Generalization of this problem is treated in Ref. 15. For the simpler case discussed here, see Ref. 16.

It is important to realize that these methods supply a mechanism for analytic continuation of exceptional quality. In some cases, on being given only a few coefficients occurring in the expansion, one is able to obtain precise bounds on the function out to five or ten times the radius of convergence of the expansion. Even more remarkable is the case where the series has zero radius of convergence.

We now come to case (v) and suppose that $\phi(\gamma)$ admits a regularization $\phi^\epsilon(\gamma)$ which can be expressed in the form (3.6), and which possesses a formal development

$$\phi^\epsilon(\gamma) = \sum_{p=0}^\infty \phi_p^\epsilon \gamma^P. \tag{3.17}$$

Then if we suppose that

$$\phi^\epsilon(\gamma) \leqslant \phi^0(\gamma) = \phi(\gamma), \tag{3.18}$$

we must have

$$\underline{B}_N^\epsilon(\gamma;\phi_0^\epsilon,\phi_1^\epsilon,\ldots,\phi_N^\epsilon) \leqslant \phi^\epsilon(\gamma) \leqslant \phi(\gamma), \tag{3.19}$$

whence

$$\underline{B}_N^{\sup} = \sup_\epsilon \underline{B}_N^\epsilon(\gamma;\phi_0^\epsilon,\phi_1^\epsilon,\ldots,\phi_N^\epsilon) \leqslant \phi(\gamma). \tag{3.20}$$

Albeit that the regularized Taylor series coefficients tend to infinity as $\epsilon$ tends to zero, the functional $B_N^\epsilon$ has a maximum for a certain $\epsilon = \epsilon(N,\gamma)$ which is the optimal choice on the basis of the given information at order $N$. In Fig. 1 we represent both $\phi^\epsilon(\gamma)$ and the $\underline{B}_N^\epsilon$ as functions of $\epsilon$ for fixed $\gamma$. One sees that $\underline{B}_1^{\sup}, \underline{B}_2^{\sup}, \ldots$form a monotone increasing sequence

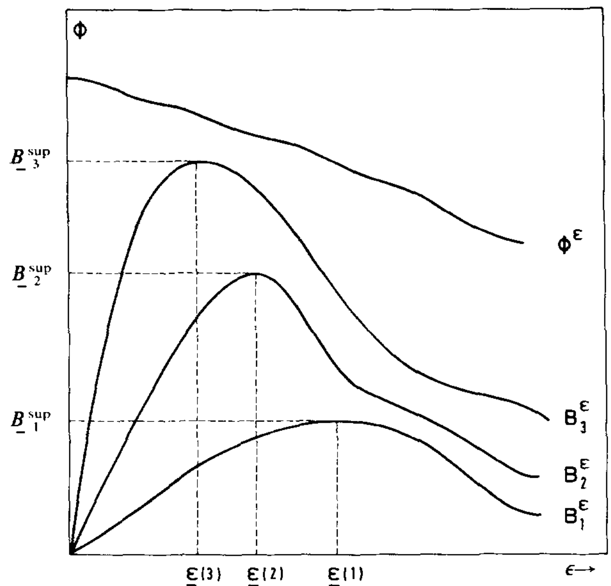


FIG. 1. The functions $\phi^{\epsilon}(\gamma)$ and $\underline{B}_N^{\epsilon}$ at fixed $\gamma$. $\underline{B}_1^{\sup}, \underline{B}_2^{\sup},\cdots$ forms a monotone increasing sequence of lower bounds for $\phi(\gamma)$.

of lower bounds to $\phi$, the sequence being necessarily convergent with limit $\underline{B}_{\infty}^{\sup} \leqslant \phi$ where the equality is attained if the moment problem associated with $\phi^{\epsilon}(\gamma)$ is determinate for each $\epsilon > 0$. The method described here is used extensively in Ref. 6.

A reason for particular interest in the methods described above is that $B_N(\gamma;\phi_0,\phi_1,\ldots,\phi_N)$ can have an analytic sturcture kindred to that of the exact solution, so that over and above its bounding properties $B_N(\gamma;\phi_0,\phi_1,\ldots,\phi_N)$ is a sensible *approximation* to $\phi(\gamma)$. For example, the analytic character of $\phi(\gamma)$ in (3.6) is to some extent carried through to the approximants (3.15) and (3.16) via the kernel. Furthermore, in the case where regularization is used as described above, even when the kernels belonging to the regularized problem possesses no singularities in $\gamma$, the optimized functions $\underline{B}_N^{\sup}(\gamma)$ may themselves have singularities which mimic those occurring in the true $\phi(\gamma)$. Lastly, in the case of differential equations where the boundary conditions are of Neumann–Dirichlet type, say,

$$\mathscr{F}(\phi,\gamma) = 0 \quad \text{over some region } D,$$
$$\mathscr{B}\phi = 0 \quad \text{on the boundary } \partial D \text{ of } D, \tag{3.21}$$

where

$$\mathscr{B}\phi \equiv \alpha(x)\phi(x,\gamma) + \sum_{i=1}^{n} \beta_i(x)\frac{\partial\phi(x,\gamma)}{\partial x_i}, \quad x\in\partial D, \tag{3.22}$$

one has the remarkable property that

$$\mathscr{B}B_N(\gamma;\phi_0,\ldots\phi_N) = 0. \tag{3.23}$$

That is to say that the approximations $B_N$ automatically satisfy the boundary conditions. The reason is that the $B_N$'s, while they may be complicated functions of the $\phi_n$'s, are nonetheless homogeneous of first degree in $(\phi_0,\phi_1,\ldots\phi_N)$. It is well known that the conditions $\mathscr{B}\phi_i = 0$ $(i = 0,1,2,\ldots)$ imply that $\mathscr{B}\mathscr{H}(\phi_0,\phi_1,\ldots\phi_N) = 0$ when $\mathscr{H}$ is homogeneous of first degree in the $\phi_i$.

## 4. CONSIDERATION OF A CLASS OF NONLINEAR ELLIPTIC EQUATIONS

As an illustration of how with the aid of positivity properties one can discover the analytic character of the physical solution of a parameter dependent functional equation, we consider the problem

$$L\phi - \lambda p\phi + \gamma q\phi^2 = f \text{ in } D,$$
$$B\phi = 0 \text{ on } \partial D, \quad \lambda\in\mathbb{R}, \quad \gamma\in\mathbb{R}_+. \tag{4.1}$$

$D$ is a bounded open region in $\mathbb{R}^n$ with boundary $\partial D$ and closure $\bar{D}$. The boundary is assumed to be $C^{2+\alpha}$ for some $0<\alpha<1$; that is, $\partial D$ can be mapped locally one-to-one onto an $(n-1)$-dimensional hyperplane by means of $C^{2+\alpha}$ mappings. By $C^{m+\alpha}(G)$ we mean the class of functions with continuous derivatives of all orders up to $m$ on the set $G$ and whose $m^{\text{th}}$ derivatives are uniformly Hölder continuous with exponent $\alpha$.

$L$ is a uniformly elliptic differential operator,

$$L\phi \equiv -\sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2\phi(x)}{\partial x_i\partial x_j} + \sum_{k=1}^{n} a_k(x)\frac{\partial\phi(x)}{\partial x_k} + a(x)\phi(x), \tag{4.2}$$

where $a_{ij}(x)\in C^{2+\alpha}(\bar{D})$, $a_k\in C^{1+\alpha}(\bar{D})$, and $a\in C^{\alpha}(\bar{D})$. The matrix $\|a_{ij}(x)\|$ is uniformly positive definite. Also $a(x)\geqslant 0$ for all $x\in\bar{D}$; $p$, $q$, and $f$ all belong to $C^{\alpha}(\bar{D})$; $p(x)>0$, $q(x)>0$, and $f(x)\geqslant 0$ for all $x\in\bar{D}$ with $f\not\equiv 0$.

$B$ is one or other of the boundary operators

$$B\phi \equiv \phi(x) \text{ on } \partial D, \tag{4.3}$$

$$B\phi \equiv \beta(x)\phi(x) + \frac{\partial\phi(x)}{\partial\nu} \text{ on } \partial D, \tag{4.4}$$

where $\partial/\partial\nu$ denotes the conormal derivative on $\partial D$. $\beta(x)\in C^{1+\alpha}(\partial D)$ and is nonnegative. $a(x)$ and $\beta(x)$ do not both vanish identically.

The problem is to discover the analyticity properties in $\gamma$ of the positive solution[17] of the problem (4.1) according to different values of $\lambda$, and thus to find out how the solution can be reconstructed for all $\gamma>0$ starting from the formal series expansion obtained for small $\gamma$. Before giving the essential general results for this problem we study a special case which sheds light on the question.

### A. The algebraic case

Suppose $\beta(x)\equiv 0$, $p = q \equiv 1$, and that both $a(x)$ and $f(x)$ are independent of $x$; then the solution to the problem is $\phi =$ const in $x$, where $\phi$ satisfies

$$(a-\lambda)\phi + \gamma\phi^2 = f. \tag{4.5}$$

The positive solution is thus

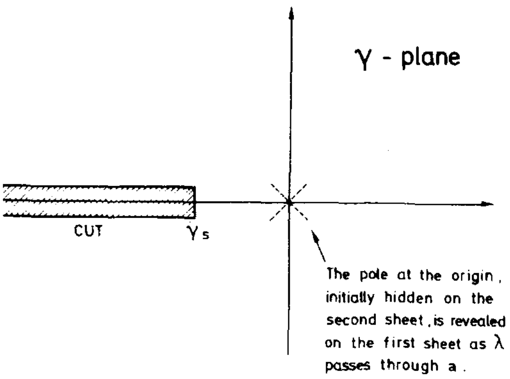


FIG. 2. The analytic structure of $\phi$ in (4.6). As $\lambda$ increases, $\gamma_s$ advances to the origin and then recedes, leaving a pole behind it.

$$\phi = \frac{-(a-\lambda) + [(a-\lambda)^2 + 4f\gamma]^{1/2}}{2\gamma}, \quad \gamma > 0. \quad (4.6)$$

*Analytic structure*: $\phi$ has a branch point of order two at $\gamma_s = -(a-\lambda)^2/4f$ and also one at infinity. We define the first sheet of the Riemann surface of $\phi(\gamma)$ by cutting the complex $\gamma$ plane from $-\infty$ to $\gamma_s$, see Fig. 2. $\phi$ has a pole at $\gamma = 0$ with residue $[|a-\lambda| - (a-\lambda)]/2$. When $\lambda < a$ this residue is zero while when $\lambda > a$ it is $|a-\lambda|$.

Noting that the second sheet of the Riemann surface is obtained simply by changing the sign of the square root in (4.6), we thus see that the pole remains hidden on the second sheet when $\lambda < a$ (subcritical regime) and that it appears on the first sheet only when $\lambda > a$ (supercritical regime). As $\lambda$ progresses from minus infinity towards $a$ the mobile sigularity $\gamma_s$ advances towards the origin, reaching it exactly when $\lambda = a$. The hidden pole is then uncovered and as $\lambda$ continues towards plus infinity the sigularity $\gamma_s$ now retreats back towards minus infinity, having achieved its task of revealing the pole.

*Positivity properties of the solution*: It is not hard to see that $\text{Im}\phi < 0$ when $\text{Im}\gamma > 0$ and that $\phi(\gamma)$ behaves like $\sqrt{f/\gamma}$ as $\gamma$ tends to infinity, from which it follows that $\phi(\gamma)$ is analytic in the cut complex plane, is a real Herglotz function, and thus can be represented

$$\phi(\gamma) = \int_0^\infty \frac{d\mu(\tau)}{1+\tau\gamma}, \quad (4.7)$$

where $d\mu(\tau)$ is a positive measure.

We have two cases:

(i) *Subcritical case* ($\lambda < a$): $\phi(\gamma)$ is holomorphic around $\gamma = 0$ and possesses an expansion

$$\phi(\gamma) = \phi_0 + \gamma\phi_1 + \cdots + \gamma^n\phi_n + \cdots. \quad (4.8)$$

In particular

$$\phi_0 = \int_0^\infty d\mu(\tau) < \infty, \quad (4.9)$$

so that the measure in (4.7) is bounded: Starting from the expansion (4.8) one can reconstruct the solution with the aid of Padé approximants as was explained in Sec. 3,

$$[P-1/P](\gamma) \leqslant \phi(\gamma) \leqslant [P/P](\gamma) \quad \text{for all } \gamma > 0. \quad (4.10)$$

Note that the radius of convergence of the series (4.8) tends to zero when $\lambda$ approaches $a$, while the bounds (4.10) remain valid for $\gamma$ as large as one likes.

(ii) *Supercritical case* ($\lambda > a$): It is clear that the solution which is holomorphic around the origin is now the negative one, as one sees on noting that

$$\phi_0 = \frac{f}{a-\lambda}, \quad \phi_1 = \frac{-f^2}{(a-\lambda)^3}, \ldots. \quad (4.11)$$

Conversely, the positive solution possesses a Laurent expansion

$$\phi = \frac{\tilde{\phi}_{-1}}{\gamma} + \tilde{\phi}_0 + \gamma\tilde{\phi}_1 + \cdots \quad (4.12)$$

with

$$(a-\lambda)\tilde{\phi}_{-1} + \tilde{\phi}^2_{-1} = 0, \quad (4.13)$$

$$(a-\lambda + 2\tilde{\phi}_{-1})\tilde{\phi}_0 = f, \quad (4.14)$$

$$(a-\lambda + 2\tilde{\phi}_{-1})\tilde{\phi}_1 = -\tilde{\phi}^2_0, \quad (4.15)$$

.................

It is seen that, when $\lambda > a$, (4.13) has a positive solution

$$\tilde{\phi}_{-1} = -(a-\lambda) > 0. \quad (4.16)$$

One has here a "critical phenomena" in the sense that the analytic continuation of the solution which is regular in $\gamma$ at the origin, starting from $\lambda < \lambda_c = a$, does not provide the correct answer. In the supercritical case it is necessary to start from the perturbation expansion, beginning with the positive solution of the associated equation given by putting $f = 0$ and $\gamma = 1$.

We remark however that, for $\lambda > a$,

$$\phi(\gamma) = \frac{\tilde{\phi}_{-1}}{\gamma} + \int_0^\infty \frac{d\tilde{\mu}(\tau)}{1+\gamma\tau},$$

$d\tilde{\mu}(\tau)$ is a positive measure, (4.17)

and that now once again $\int_0^\infty d\tilde{\mu}(\tau) < \infty$. Knowing $\tilde{\phi}_{-1}$ (which can be determined once and for all, independently of $f$!) one can now reconstruct from the perturbation series the positive solution by using Padé approximants to

$$\tilde{\phi}(\gamma) = \int_0^\infty \frac{d\tilde{\mu}(\tau)}{1+\gamma\tau}. \quad (4.18)$$

Moreover we notice that the supercritical positive solution is the analytic continuation of the subcritical negative solution, and that one may be interested in studying this solution also. The critical phenomena appears here as a simple permutation of Riemann sheets.

## B. The realistic case

In the realistic case we have not yet been able to obtain all of the analyticity and positivity properties of the algebraic case, but a major part of the preceding results have been extended to the general case. We are now going to summarize the essential results.

The two fundamental principles which allow one to approach the problem are:

(i) *Amann's theorem*[18]: Let $F(x,\phi)\in C^{\alpha}(\bar{D})$ for $x\in\bar{D}$ be such that $\partial F/\partial\phi$ exists and is continuous for $x\in\bar{D}$ and $\phi\in\mathbb{R}$. Suppose that there exists $\underline{\phi}$ and $\bar{\phi}\in C^{2+\alpha}(\bar{D})$ such that

$$\begin{matrix} L\underline{\phi}+F(x,\underline{\phi})\leqslant 0, & B\underline{\phi}\leqslant 0, \\ L\bar{\phi}+F(x,\bar{\phi})\geqslant 0, & B\bar{\phi}\geqslant 0, \end{matrix} \quad \text{and } \underline{\phi}\leqslant\bar{\phi}. \tag{4.19}$$

Then the problem

$$L\phi + F(x,\phi) = 0, \text{ in } D, \quad B\phi = 0 \text{ on } \partial D \tag{4.20}$$

possesses at least one solution $\phi\in C^{2+\alpha}(\bar{D})$ which satisfies

$$\underline{\phi}\leqslant\phi\leqslant\bar{\phi}. \tag{4.21}$$

(ii) *Positivity Lemma*[19]: We associate the operator $(L,B)$ with the eigenvalue problem: $L\psi = \lambda p\psi$ in $D$ and $B\psi = 0$ on $\partial D$. Then we recall that this operator, although it is not in general self-adjoint, has a purely discrete eigenvalue spectrum. The eigenvalue $\lambda_1$ having smallest real part is itself real and strictly positive, and the corresponding eigenfunction can always be chosen real and positive. Bearing this in mind, we have the following lemma.[19]

If, for some $\lambda\in\mathbb{R}$

$$\phi\in C^2(\bar{D}), \quad [L-\lambda p(x)]\phi\geqslant 0, \quad B\phi\geqslant 0, \quad \text{and } \phi>0, \tag{4.22}$$

then $\lambda\leqslant\lambda_1$, with $\lambda=\lambda_1$ if and only if the equality signs in (4.22) hold. Conversely, if

$$\phi\in C^2(\bar{D}), \quad [L-\lambda p(x)]\phi>0, \quad B\phi\geqslant 0, \quad \text{and } \lambda\leqslant\lambda_1, \tag{4.23}$$

then $\phi(x)>0$ for all $x\in D$.

Using the above theorem and lemma we obtain the following results, the detailed proofs of which can be found in Ref. 20.

*Proposition* 1: The given problem (4.1) possesses exactly one nonnegative solution $\phi(x)\in C^{2+\alpha}(\bar{D})$, for all $\lambda\in\mathbb{R}$ and $\gamma>0$. This solution is in fact strictly positive for all $x\in D$.

*Proof*: Amann's theorem yields the existence of at least one nonnegative solution: Zero is a lower solution and a suitably large positive constant is an upper solution. The strict positivity over $D$ of such a solution follows from the maximum principle. To establish uniqueness we suppose that $\phi_1$ and $\phi_2$ are both positive solutions: Then

$$[L+\gamma q(\phi_1+\phi_2)-\lambda p](\phi_1-\phi_2)=0 \text{ in } D$$
$$\text{and} \quad B(\phi_1-\phi_2)=0 \text{ on } \partial D. \tag{4.24}$$

But

$$[L+\gamma q(\phi_1+\phi_2)-\lambda p]\phi_1 = f+\gamma q\phi_1\phi_2>0 \text{ in } D$$
$$\text{while } B\phi_1=0 \text{ on } \partial D. \tag{4.25}$$

The positivity lemma tells us that $\lambda<\tilde{\lambda}_1$, where $\tilde{\lambda}_1$ is the smallest eigenvalue belonging to the problem

$$[L+\gamma q(\phi_1+\phi_2)]\psi=\lambda p\psi \text{ in } D \text{ with } B\psi=0 \text{ on } \partial D. \tag{4.26}$$

We conclude that $\phi_1-\phi_2\equiv 0$ in (4.24).

*Proposition* 2: About each point $\gamma_0>0$, $\phi(x,\gamma)$, the positive solution defined above, is holomorphic inside a circle of center $\gamma_0$ passing through the origin of the complex $\gamma$ plane, and

$$(-1)^n \frac{d^n}{d\gamma^n}\phi(x,\gamma)\bigg|_{\gamma=\gamma_0}>0,$$
$$(-1)^n \frac{d^n}{d\gamma^n}\{\gamma\phi(x,\gamma)\}\bigg|_{\gamma=\gamma_0}>0, \tag{4.27}$$

for all $x\in D$.

The full proof can be found in Ref. 20. Here we merely outline it: Let $\lambda\in\mathbb{R}$ and $\gamma_0>0$ be fixed, and let $\gamma\in\mathbb{C}$. Then we consider the formal series

$$\psi(\gamma_0,\rho)=\sum_{n=0}^{\infty}\frac{1}{n!}\psi_n(\gamma_0)\rho^n, \quad \rho=\gamma-\gamma_0. \tag{4.28}$$

The $\rho$-independent functions $\psi_n(\gamma_0)$ are supposed to satisfy the equations obtained by equating the coefficients of the different powers of $\rho$ when one substitutes $\psi$ for $\phi$ in (4.1), and where $\psi_0(\gamma_0)>0$. By making repeated use of the positivity lemma we show that the $\psi_n$'s have alternating signs and that the series converges absolutely and uniformly for $|\rho|<\gamma_0$. Using the compactness theorem[21] we then show that the function to which the series converges is indeed a solution of the problem. By uniqueness, $\psi(\gamma_0,\rho)$ is identified with $\phi(\gamma)$.

*Proposition* 3: By Bernstein's theorem one sees that there exist two measures $d\mu_{x,\lambda}(t)$ and $d\nu_{x,\lambda}(t)$ which are positive on $[0,\infty)$ and such that

$$\phi(\gamma)=\int_0^\infty e^{-\gamma t}d\mu_{x\lambda}(t), \quad \frac{\partial}{\partial\gamma}\{\gamma\phi(\gamma)\}=\int_0^\infty e^{-\gamma t}d\nu_{x\lambda}(t). \tag{4.29}$$

We next analyze the behavior of $\phi(\gamma)$ in the neighborhood of $\gamma=0$. To achieve this we need the following result: Let $\lambda_c$ be the eigenvalue (necessarily real and positive) having smallest real part corresponding to the problem

$$L\psi_c=\lambda_c p\psi_c \quad \text{in } D \text{ with } B\psi_c=0 \text{ on } \partial D. \tag{4.30}$$

Then one shows that $\lambda_c$ is a bifurcation point for the equation

$$L\theta-\lambda p\theta+q\theta^2=0 \quad \text{in } D, \quad B\theta=0 \text{ on } \partial D. \tag{4.31}$$

This equation possesses only one nonnegative solution when $\lambda<\lambda_c$, explicitly $\theta\equiv 0$, and two nonnegative solutions when $\lambda>\lambda_c$. The existence of a strictly positive solution $\theta_+(x)$ is established in this latter case by choosing $\epsilon\psi_c(x)>0$, with $\epsilon$ sufficiently small, as a lower solution. The uniqueness of $\theta_+(x)$ (strictly positive inside $D$) follows from the positivity lemma.

*Proposition* 4: For $\lambda\neq\lambda_c$, $\lambda\in\mathbb{R}$, $\phi(\gamma,x)$ is holomorphic inside a circle of finite radius, $|\gamma|<\Gamma$, about $\gamma=0$, with the exception of a simple pole at $\gamma=0$ of residue

$$\begin{matrix} 0, & \text{if } \lambda<\lambda_c, \\ \theta_+(x), & \text{if } \lambda>\lambda_c. \end{matrix} \tag{4.32}$$

*Proof*: One examines the formal development in increasing powers of $\gamma$ of the function $\gamma\phi(\gamma)$. By repeatedly using the positivity lemma, one shows that the series has coefficients of alternating signs and that it is absolutely and uniformly convergent for $|\gamma|$ sufficiently small and $\lambda\neq\lambda_c$, $\lambda\in\mathbb{R}$. The compactness theorem once again assures us that this series can be identified with $\gamma\phi(\gamma)$. When $\lambda<\lambda_c$ one finds that $\gamma\phi(\gamma)\to 0$ as $\gamma\to 0$. When $\lambda>\lambda_c$ one finds that $\gamma\phi(\gamma)\to\theta_*(x)$ as $\gamma\to 0$.

Combining the results in various of the above propositions, one can show that

*Proposition* 5: For all $\lambda\neq\lambda_c$, $\lambda\in\mathbb{R}$, one can write the positive solution of the problem in the form

$$\phi(\gamma,x)=\frac{\xi_0}{\gamma}+\int_0^\infty e^{-\gamma\tau}d\rho_{x,\lambda}(\tau) \tag{4.33}$$

where the positive measure $d\rho_{x,\lambda}(\tau)$ is bounded and

$$\xi_0=\begin{cases}0, & \text{if } \lambda<\lambda_c\\ \theta_*(x), & \text{if } \lambda>\lambda_c.\end{cases}$$

It is now possible to apply the results of Sec. 3 in order to obtain upper and lower bounds on the positive solution for all $\gamma>0$, with the aid of the generalized Padé approximants $B_N(\gamma)$.

## 5. APPLICATION TO NONLINEAR PARABOLIC EQUATIONS

### A. Definitions and basic principles

To show that similar results to those in Sec. 4 are true in the parabolic case we consider the problem

$$\begin{aligned}&\mathscr{L}\Phi-\lambda p\Phi+\gamma q\Phi^2=f(x,t)\quad\text{for }(x,t)\in Q=D\times(0,\infty),\\ &B\Phi=g(x,t)\quad\text{for }(x,t)\in\partial Q=D\times(0,\infty),\\ &\Phi(x,t=0)=\Phi_0(x)\quad\text{for }x\in D,\end{aligned}\tag{5.1}$$

where $\Phi$ is now a function of $x$ and $t$, and

$$\mathscr{L}\Phi=L\Phi+\frac{\partial\Phi}{\partial t}. \tag{5.2}$$

All quantities here are defined just as they were in Sec. 4 except for the prolongation of $L$ by the adjunction of $\partial/\partial t$, the conversion of the boundary conditions into those corresponding to an initial value problem, and the introduction of the functions $f(x,t)$ and $g(x,t)$ which are assumed to be smooth and nonnegative over their domains. The function $\Phi_0(x)$ is also supposed to be smooth and nonnegative. We have not introduced time dependences into the coefficients occurring in the first two equations on the left-hand side of (5.1), although much of what we say would withstand a certain amount of generalization in this direction.

Throughout this section $\lambda\in\mathbb{R}$ is fixed, and we are concerned only with the analyticity properties of solutions in the variable $\gamma$. One difference between (4.1) and (5.1) is that the boundary conditions in the latter are inhomogeneous; to make way for this additional flexibility we suppose that the following is true.

*Assumption* (*): (5.1) possesses a classical nonnegative solution, denoted briefly

$$\Phi(x,t;\gamma)=\Phi(\gamma)\quad\text{for each }\gamma>0. \tag{5.3}$$

For example, in Proposition 6 it is shown that this is certainly true when $g$ and $\Phi_0$ both vanish identically; and it is easy to envisage other cases.

Let $0<T<\infty$. We use the notation $Q_T$ for the region $D\times(0,T)$ and $\partial Q_T$ for the boundary $\partial D\times(0,T)$. $H^{\alpha+2,\alpha/2+1}(\bar{Q}_T)$ denotes the Hölder space consisting of functions $\theta$ defined on $(x,t)\in\bar{Q}_T$ possessing second derivatives $\partial^2\theta/\partial x_i\partial x_j$ $(i,j=1,2,\dots,N)$ which are uniformly Hölder continuous in $x$ throughout $\bar{Q}_T$ with exponent $\alpha$, and possessing the derivative $\partial\theta/\partial t$ which is uniformly Hölder continuous in $t$ throughout $\bar{Q}_T$ with exponent $\alpha/2$; see Ref. 22. $H^{\alpha+2,\alpha/2+1}(Q_T)$ is the Hölder space consisting of all $\theta\in H^{\alpha+2,\alpha/2+1}(\bar{R})$ for each domain $R$ with $\bar{R}\subset Q_T$. By a classical function we mean any function in $H^{\alpha+2,\alpha/2+1}(Q_T)$ which is defined and continuous on $\bar{Q}_T$.

In the following we will use repeatedly the following theorem which is a modified version[23] of one due to Sattinger.[21]

*Theorem*: Let $F(x,t;\Phi)$ be uniformly Hölder continuous in $x$, and in $t$, with exponents $\alpha$ and $\alpha/2$, respectively, throughout $\bar{Q}_T$. Let $\partial F/\partial\Phi$ exist and be continuous for $(x,t)\in\bar{Q}_T$ and $\Phi\in\mathbb{R}$. Suppose that there exist classical function $\underline{\Phi}$ and $\bar{\Phi}$ such that:

$$\begin{aligned}&\mathscr{L}\underline{\Phi}+F(x,t;\underline{\Phi})\leqslant 0\quad\text{for all }(x,t)\in Q_T,\\ &B\underline{\Phi}\leqslant g(x,t)\quad\text{for all }(x,t)\in\partial Q_T,\\ &\underline{\Phi}(x,t=0)\leqslant\Phi_0(x),\quad\text{for }x\in D;\end{aligned}\tag{5.4}$$

$$\begin{aligned}&\mathscr{L}\bar{\Phi}+F(x,t;\bar{\Phi})\geqslant 0\quad\text{for all }(x,t)\in Q_T,\\ &B\bar{\Phi}\geqslant g(x,t)\quad\text{for all }(x,t)\in\partial Q_T,\\ &\bar{\Phi}(x,t=0)\geqslant\Phi_0(x),\quad\text{for }x\in D;\end{aligned}\tag{5.5}$$

with $\underline{\Phi}\leqslant\bar{\Phi}$ throughout $Q_T$. Then the problem

$$\begin{aligned}&\mathscr{L}\Phi+F(x,t;\Phi)=0\quad\text{for }(x,t)\in Q_T,\\ &B\Phi=0\text{ on }\partial Q_T,\\ &\Phi(x,t=0)=\Phi_0,\end{aligned}\tag{5.6}$$

has a classical function $\Phi$ as solution, satisfying

$$\underline{\Phi}\leqslant\Phi\leqslant\bar{\Phi}\quad\text{for all }(x,t)\in\bar{Q}_T. \tag{5.7}$$

In the above the boundary conditions are satisfied in the limit as $(x,t)$ approaches the boundary, sometimes with restrictions on the possible paths by which the approach can be made; see for example Ref. 22, p. 404.

The proof of the above theorem makes important use of the maximum principle for parabolic operators. In its simplest form this consists in the fact that when $a\equiv 0$ the solutions to the problem $\mathscr{L}u=0$ assume their greatest and least values in $\bar{Q}_T$ on the boundary. More generally we have[22]

$$\begin{aligned}&u\geqslant 0\text{ in }Q_T,\quad Bu\geqslant 0\text{ on }\partial Q_T,\text{ and }u(t=0)\geqslant 0,\\ &\Rightarrow u\geqslant 0\text{ when }u\text{ is a classical function on }\bar{Q}_T.\end{aligned}\tag{5.8}$$

These results are closely related to the uniqueness theorems[22] which tell us for example that the problem

$$\begin{aligned}&\mathscr{L}\Phi+b(x,t)\Phi=f\text{ in }Q_T,\\ &B\Phi=0\text{ on }\partial Q_T,\quad\Phi(t=0)=0,\end{aligned}\tag{5.9}$$

has at most one classical solution. Here $b(x,t)$ is assumed to be continuous throughout $Q_T$.

*Proposition* 6: When either the function $\beta(x)$ in (4.4) is strictly positive for all $x\in\partial D$ or $g(x,t)\equiv 0$, the problem (5.1) possesses a unique classical solution for any $\lambda\in\mathbb{R}$ and all $\gamma>0$. This solution is nonnegative for all $(x,t)\in Q$. In the special case where $\Phi_0(x)\equiv 0, g(x,t)\equiv 0$, and $f(x,t)=f(x)$ is independent of $t$, this solution converges to the positive solution $\Phi[\gamma]$ of (4.1) as $t$ tends to infinity.

*Proof*: It is readily seen that an upper and lower solution pair for (5.1) is provided by a large enough positive constant, and zero, respectively. The existence of a classical solution of (5.1) is thus ensured by Sattinger's theorem.

Let $\Phi_1$ and $\Phi_2$ denote two classical solutions, then

$$[\mathscr{L}-\lambda p+\gamma q(\Phi_1+\Phi_2)](\Phi_1-\Phi_2)=0 \text{ in } Q_T \tag{5.10}$$

$$B(\Phi_1-\Phi_2)=0 \text{ on } \partial Q_T, \quad (\Phi_1-\Phi_2)=0 \text{ when } t=0.$$

Choosing $b(x,t)=\gamma q(\Phi_1+\Phi_2)-\lambda p$ we identify (5.10) with (5.9). It follows that the only classical solution of (5.10) is $\Phi_1-\Phi_2=0$ which establishes uniqueness.

The last statement in the proposition is a consequence of Ref. 24, Theorem 4.8. The upper and lower solution pair in Proposition 1, associated with the time independent problem, in fact provide, respectively, a decreasing and an increasing sequence of monotone interates, each sequence converging to the positive solution of the time independent problem.[24] Theorem 4.8 provides that in such a case the solution to the time dependent problem obtained by replacing $L$ by $\mathscr{L}$ and introducing initial data lying between the original upper and lower solution pair will always lie between the solutions found by taking the upper and lower solutions as initial data, and will converge as time goes to infinity to the positive solution of the time independent problem.

Q.E.D.

*Remark*: Whereas (4.1) may have many solutions, among which there is exactly one which is positive, the time dependent problem has only one solution. In the case $g(x,t)=0$ and $f(x,t)=f(x)$ we can think of the positive solution of (4.1) as being special: It corresponds to the solution picked out by the time dependent problem, for any smooth positive intitial data.

We will use the notation $\Phi[\gamma]$ for the classical solution of (5.1) provided by Proposition 6. In examining $\Phi[\gamma]$ we do not have available a result analogous to the positivity lemma of Sec. 4B (ii) and our central arguement is perforce more complicated than it was in Sec. 4. In place of the positivity lemma we repeatedly use Sattinger's theorem, given above.

Let both $\lambda\in\mathbb{R}$ and $\gamma_0>0$ be held fixed; let $\gamma\in\mathbb{C}$ be given and set $\rho=(\gamma-\gamma_0)$ as before. Then we investigate the formal series

$$\Psi=\Psi[\gamma_0,\rho]=\int_{n=0}^{\infty}\frac{1}{n!}\psi_n[\gamma_0]\rho^n, \tag{5.11}$$

where the $\rho$-independent functions $\psi_n[\gamma_0]$, $n=0,1,2,\ldots$ are supposed to satisfy the set of equations obtained by equating coefficients of the different powers of $\rho$ which occur in the formal expansion of

$$\mathscr{L}\Psi-\lambda p\Psi+(\gamma_0+\rho)q\Psi^2=f \text{ in } Q, \tag{5.12}$$

$$B\Psi=g \text{ on } \partial Q, \quad \Psi(x,t=0)=\Phi_0(x),\ x\in D.$$

The equations to be satisfied by the $\psi_n$'s are

$$\mathscr{L}\psi_0-\lambda p\psi_0+\gamma q\psi_0^2=f \text{ in } Q, \tag{5.13.0}$$

$$B\psi_0=g \text{ on } \partial Q, \quad \psi_0(x,t=0)=\Phi_0(x) \quad \text{for } x\in D,$$

$$\mathscr{L}\psi_n-\lambda p\psi_n+\gamma_0 q\sum_{m=0}^{n}\binom{n}{m}\psi_{n-m}\psi_m + nq\sum_{m=0}^{n-1}\binom{n-1}{m}\psi_{n-m-1}\psi_m=0 \text{ in } Q, \tag{5.13.n}$$

$$B\psi_n=0 \text{ on } \partial Q, \quad \psi_n(x,t=0)=0 \qquad \text{for } x\in D,\ n=1,2,3,\cdots.$$

*Proposition* 7: The $\psi_n$'s are uniquely defined classical functions for all $\gamma>0$. Moreover $\psi_0=\Phi[\gamma_0]$, and for each $n\in\{1,2,3,\cdots\}$ we have

$$(-1)^n\psi_n(x,t)\geqslant 0 \quad \text{and}$$
$$(-1)^{n-1}\{\gamma_0\psi_n(x,t)/n+\psi_{n-1}(x,t)\}\geqslant 0 \quad \text{for all } (x,t)\in Q. \tag{5.14.n}$$

*Proof*: By assumption (*), (5.13.0) possesses exactly one classical solution,

$$\psi_0=\Phi[\gamma_0]. \tag{5.15}$$

The remaining equations can be rewritten

$$[\mathscr{L}-\lambda p+2\gamma_0 q\psi_0]\psi_n+q\sum_{m=0}^{n-2}\frac{n!}{m!(n-m-1)!}\times\left(\frac{\gamma_0\psi_{m+1}}{m+1}+\psi_m\right)\psi_{n-m-1}+nq\psi_0\psi_{n-1}=0 \quad \text{in } Q, \tag{5.16.n}$$

$$B\psi_n=0 \text{ on } \partial Q, \quad \psi_n(x,0)=0 \text{ on } D, \quad n=1,2,3,\cdots.$$

The $n$th one of these equations is linear in $\psi_n$, involving it only in the term $[L-\lambda p+2\gamma_0 q\psi_0]\psi_n$, and the inhomogeneous part is a function of $(\psi_0,\psi_1,\ldots,\psi_{n-1})$. Thus the set (5.16.$n$) must be solved successively with $\psi_0=\Phi[\gamma_0]$.

Consider (5.16.1),

$$(\mathscr{L}-\lambda p+2\gamma_0 q\psi_0)\psi_1=-q\psi_0^2 \quad \text{in } Q, \tag{5.17}$$

$$\psi_1=0 \text{ on } \partial Q, \quad \psi_1=0 \text{ on } D.$$

We work in $Q_T$ and let $T$ tend to infinity, thereby providing results in $Q$. Noting that zero constitutes an upper solution while $-\psi_0/\gamma_0$ is a lower solution, we have from Sattinger's theorem that (5.17) has a classical solution $\psi_1(x)$ which, on comparing (5.17) with (5.9), is unique, and moreover

$$0\geqslant\psi_1(x,t)\geqslant-\psi_0(x,t)/\gamma_0 \quad \text{for all } (x,t)\in Q. \tag{5.18}$$

The latter is equivalent to the inequalities (5.14.$n$) when $n=1$.

An inductive arguement can now be set up by supposing that (5.13.n) has a unique classical solution $\psi_n$, and the inequalities (5.14.n) pertain for all $n\in\{1,2,\ldots,K\}$, for some integer $K \geqslant 1$. Then we consider the equations for $\psi_{K+1}$. The inductive hypothesis provides that the quantity

$$q \sum_{m=0}^{K-1} \frac{(K+1)!}{m!(K-m)!} \left( \frac{\gamma_0 \psi_{m+1}}{m+1} + \psi_m \right) \psi_{K-m}$$

$$+ (K+1) q \psi_0 \psi_K, \tag{5.19}$$

which appears in (5.16.$K+1$), has the same sign as $(-1)^K$. Hence *zero constitutes an upper or a lower solution associated with* (5.16.$K+1$), *according as K is even or odd respectively.* Now note that if (5.16.$K+1$) holds then we have the identities

$$[\mathscr{L} - \lambda p + 2\gamma_0 q\, \psi_0] \left( \frac{\gamma_0 \psi_{K+1}}{K+1} + \psi_K \right)$$

$$= -q \sum_{m=0}^{K-1} \frac{K!}{m!(K-m-1)!} \left( \frac{\gamma_0 \psi_{m+1}}{m+1} + \psi_m \right)$$

$$\times \left( \frac{\gamma_0 \psi_{K-m}}{K-m} + \psi_{K-m-1} \right), \tag{5.20}$$

$$B\left( \frac{\gamma_0 \psi_{K+1}}{K+1} + \psi_K \right) = 0 \text{ on } \partial D,$$

$$\left( \frac{\gamma_0 \psi_{K+1}}{K+1} + \psi_K \right) = 0 \quad \text{when } t = 0.$$

In particular, again using the inductive hypothesis, we see that the right-hand side of (5.20) has the same sign as $(-1)^K$ whence we deduce that $-(K+1)\psi_K/\gamma_0$ *constitutes a lower or an upper solution associated with* (5.16.$K+1$), *according K is even or odd respectively.*

Combining the italicized remarks above with the inductive assumption that $\psi_K$ has the same sign as $(-1)^K$, we see that zero and $-(K+1)\psi_K/\gamma_0$ provide an upper and lower solution for $\psi_{K+1}$, necessarily unique, and the inequalities (5.14.$n$) must be true for all $n \in \{1,2,\ldots,K+1\}$. This completes the induction. Q.E.D.

*Proposition* 8: For all $\rho \in C$ with $|\rho| < \gamma_0$ the series $\Psi[\gamma_0,\rho]$ is absolutely convergent, uniformly for $(x,t) \in \bar{Q}_T$, and we have $\Psi[\gamma_0,\rho] = \Phi[\gamma]$.

*Proof*: Throughout this proof $\rho \in C$ with $|\rho| < \gamma_0$ is fixed. Using the inequalities (5.14.$n$) we readily obtain

$$|\psi_n(x,t)| \leqslant n! |\psi_0(x,t)| / \gamma_0^n, \tag{5.21}$$

for $n = 0,1,2,\cdots$, for all $(x,t) \in \bar{Q}_T$. Hence

$$\sum_{n=0}^{\infty} \frac{1}{n!} |\psi_n(x,t)| \cdot |\rho|^n \leqslant |\psi_0(x,t)| / (1 - |\rho|/\gamma_0)$$

for all $(x,t) \in \bar{Q}_T$ (5.22)

and we see that the series $\Psi[\gamma_0,\rho]$ is absolutely convergent. Similarly,

$$|\Psi[\gamma_0,\rho](x,t) - \Psi_N[\gamma_0,\rho](x,t)|$$

$$\leqslant (|\rho|/\gamma_0)^{N+1} |\psi_0[\gamma_0](x,t)| / (1 - |\rho|/\gamma_0),$$

$$\text{for all } (x,t) \in \bar{Q}_T, \tag{5.23}$$

where $\Psi_N[\gamma_0,\rho]$ denotes the $N$th partial sum of $\Psi[\gamma_0,\rho]$, and since $\psi_0[\gamma_0](x,t)$ is continuous over $\bar{Q}_T$, the convergence is uniform.

Now let

$$\Theta_N = \mathscr{L} \Psi_N[\gamma_0,\rho] - \lambda p \Psi_N[\gamma_0,\rho] + \gamma q \Psi_N[\gamma_0,\rho]^2 - f. \tag{5.24}$$

Then using the definitive equations (5.13.$n$) we find

$$\Theta_N = \gamma_0 q \sum_{n=N+1}^{2N} \frac{1}{n!} \rho^n \sum_{m=n-N}^{N} \binom{n}{m} \psi_m \psi_n$$

$$+ \rho q \sum_{n=N}^{2N} \frac{1}{n!} \rho^n \sum_{m=n-N}^{N} \binom{n}{m} \psi_m \psi_n. \tag{5.25}$$

Hence from (5.21)

$$|\Theta(x,t)|$$

$$\leqslant \gamma_0 q(x) \sum_{n=N+1}^{2N} (|\rho|/\gamma_0)^n (2N+1-n) |\psi_0(x,t)|^2$$

$$+ |\rho| q(x) \sum_{n=N}^{2N} (|\rho|/\gamma_0)^n (2N+1-n) |\psi_0(x,t)|^2$$

$$\leqslant 2(N+1) q(x) |\psi_0(x,t)|^2 (|\rho|/\gamma_0)^{N+1} |\gamma_0| / (1 - |\rho| \gamma_0)$$

for all $(x,t) \in \bar{Q}_T$. (5.26)

Hence $\Theta_N$ converges uniformly over $\bar{Q}_T$ to zero as $N$ tends to infinity. This means we can write

$$[\mathscr{L} - a(x)] \Psi_N[\gamma_0,\rho] = h_N \text{ in } Q_T,$$

$$B\Psi_N[\gamma_0,\rho] = g \text{ on } \partial Q_T, \tag{5.27}$$

$$\Psi_N[\gamma_0,\rho](x,t=0) = \Phi_0(x) \text{ on } D,$$

where $h_N$ converges uniformly over $\bar{Q}_T$ to $f + (\lambda p - a)\Psi[\gamma_0,\rho] - \gamma q \Psi[\gamma_0,\rho]^2$ as $N$ tends to infinity. The argument now follows the same lines as those indicated in Ref. 24, p. 984: We conclude that the limiting function $\Psi[\gamma_0,\rho]$ is a classical solution of (5.1). Thus, using uniqueness, we must have $\Psi[\gamma_0,\rho] = \Phi[\gamma]$.

Having identified $\Phi[\gamma]$ with $\Psi[\gamma_0,\rho]$, our key result is immediate.

*Proposition* 9: For each $(x,t) \in Q$ and each $\lambda \in (-\infty, +\infty)$, there exists nondecreasing real cumulative distribution functions $\mu_{x,t}(s)$ and $\nu_{x,t}(s)$ on $0 \leqslant s < \infty$ such that

$$\Phi[\gamma](x,t) = \int_0^\infty \exp\{-\gamma s\} d\mu_{x,t}(s) \text{ for all } \gamma \text{ with}$$

$\text{Re}\gamma > 0$,

and

$$\frac{\partial}{\partial \gamma} \{\gamma \Phi[\gamma](x,t)\} = \int_0^\infty \exp\{-\gamma s\} d\nu_{x,t}(s) \text{ for all } \gamma$$

with $\text{Re}\gamma > 0$.

*Proof*: The above characterizations both follow from the Bernstein theorem described in Sec. 2.2. From Proposition 8 we have that $\Phi[\gamma]$, and hence also $\gamma\Phi[\gamma]$, are infinitely differentiable with respect to $\gamma$ for each $\gamma > 0$; and, using (5.14.$n$) together with the positivity of $\psi_0[\gamma] = \Phi[\gamma]$, we have

$$(-1)^n \frac{\partial^n}{\partial \gamma^n} \Phi[\gamma] = (-1)^n \psi_n[\gamma] \geqslant 0 \tag{5.28}$$

and

$$(-1)^n \frac{\partial^{n+1}}{\partial \gamma^{n+1}} \{\gamma \Phi[\gamma]\} = (-1)^n \{\gamma \psi_{n+1}[\gamma] + (n+1) \psi_n[\gamma]\} \geqslant 0 \tag{5.29}$$

for $n = 0,1,2,\cdots$ and all $\gamma > 0$.

Thus both $\Phi[\gamma]$ and $\partial\{\gamma\Phi[\gamma]\}/\partial\gamma$ are completely monotonic in $\gamma > 0$. Bernstein's theorem yields the representations (5.26) and (5.27) valid for all $\gamma > 0$, and analytic continuation ensures that they remain true for all $\gamma$ with $\mathrm{Re}\gamma > 0$. Q.E.D.


[1] G.A. Baker, Jr., Adv. Theor. Phys. **1**, 1 (1975).
[2] See for example, *The Padé Approximant in Theoretical Physics*, edited by G.A. Baker and J.L. Gammel (Academic, New York, 1970): *Padé Approximants and their Applications*, edited by P.R. Graves-Morris (Academic, London, 1972); *Essentials of Padé Approximants* by G.A. Baker, Jr. (Academic, New York, 1975).
[3] See for example: J. Nuttall, Phys. Rev. **157**, 1312 (1967); S.T. Epstein and M.F. Barnsley, J. Math. Phys. **14**, 314 (1973); M.F. Barnsley and P.D. Robinson, J. Inst. Math. Its Appls. **14**, 229 (1974); M.F. Barnsley, J. Math. Phys. **16**, 918 (1975).
[4] M.F. Barnsley and P.D. Robinson, J. Inst. Math. Its Appls. **14**, 251 (1974).
[5] P.W. Langhoff and M. Karplus, Phys. Rev. Lett. **19**, 1461 (1967).
[6] D. Bessis and M. Villani, J. Math. Phys. **16**, 462 (1975); D. Bessis, L. Epele, and M. Villani, J. Math. Phys. **15**, 2071 (1974).
[7] J.S.R. Chisholm, in *The Padé Approximant in Theoretical Physics*, cited in Ref. 2, pp. 171–82; see also Ref. 4.
[8] F. Riesz and B.Sz. -Nagy, *Functional Analysis* (Ungar, New York, 1965).
[9] D.V. Widder, *The Laplace Transform* (Princeton U. P., Princeton, N.J.,1941).
[10] D. Bessis, P. Méry, and G. Turchetti, Phys. Rev. D **15**, 2345 (1977).
[11] See the last reference in Ref. 2, Chap. 17.
[12] Hardy, *Divergent Series* (Oxford U.P., London, 1949).
[13] "Determinate" means here that the knowledge of all the derivatives at $\gamma = 0$ is sufficient to rebuild the function: This is always the case if $\gamma = 0$ is a regular point of $\Phi[\gamma]$.
[14] These are apparently unpleasant nonlinear algebraic equations. However one does not need to solve them explicitly in order to construct the Padé approximants $[(P-1)/P]$ and $[P/P]$ which are given by an unambiguous procedure which involves only the solution of a finite set of linear algebraic equations.
[15] S. Karlin and W.J. Studden, *Tchebycheff Systems* (Interscience, New York, 1966).
[16] J.C. Wheeler and R.G. Gordon, chapter 3 of *The Padé Approximant in Theoretical Physics*, edited by G.A. Baker and J.L. Gammel (Academic, New York, 1970).
[17] We see later that the problem (4.1) always has a positive solution.
[18] H. Amann, Indiana Univ. Math. J. **21**, 125 (1971).
[19] See Ref. 20. This lemma is a slight extension of one due to H.B. Keller and D.S. Cohen, J. Math. Mech. **16**, 1361 (1967).
[20] M.F. Barnsley and D. Bessis, Proc. R. Soc. Edin. "Padé approximant bounds on the positive solutions of some nonlinear elliptic equations," to appear.
[21] S. Agmon, A. Douglis, and L. Nirenberg, Commun. Pure Appl. Math. **12**, 623 (1959).
[22] O.A. Ladyzenskaja, V.A. Solonnikov, and N.N. Ural'ceva, Linear and Quasilinear Equations of Parabolic Type (Translations of the American Mathematical Society, Vol. 23, 1968).
[23] We have admitted a $t$ dependence in the $F(x,t;\ )$: this does not substantially effect the proof of the theorem given in Ref. 24.
[24] D.H. Sattinger Indiana Univ. Math. J. **21**, 979 (1972).
[25] C. Berg, J. Differential Eq. **24**, 323 (1977).

# Matrix products with application to classical statistical mechanics

Mark A. Novotny

*Stanford University, Stanford, California 94305* [a)]
(Received 28 July 1978)

A matrix product that generalizes the Kronecker matrix product is introduced and its properties are documented. This product is applied to classical statistical mechanics where its trace properties lead to both the transfer matrix and graphical expansion methods of evaluating the partition function.

## 1. INTRODUCTION

The purpose of this paper is to increase the usefulness of one of the most powerful mathematical tools of a physicist, the matrix formalism. This is accomplished by generalizing the Kronecker matrix product in order to study matrices that have a particular structure. This generalized matrix product has proved to have many interesting properties; those which we have found most useful are documented in this paper. The trace of a matrix assembled with the generalized Kronecker product is particularly interesting since it is a natural generalization of the trace of a matrix formed using Kronecker, Hadamard, or regular matrix multiplication. As a demonstration of the usefulness of this product, it is applied to the problem of putting the partition function of classical statistical mechanics into a tractable form.

It is shown that in the special case of a Markov process the partition function takes a transfer matrix formulation, while in general a graphical series expansion can be set forth. Hence this matrix product provides a connection between the transfer matrix and the high and low temperature series expansion techniques for the spin-$\frac{1}{2}$ Ising model.

## 2. DEFINITIONS OF MATRIX PRODUCTS

Let us review the standard matrix products[1,2] using the notation that for an $n\times p$ matrix $\mathbf{A}$ the elements are given by $\langle i|\mathbf{A}|j\rangle$ $1\leqslant i\leqslant n$, $1\leqslant j\leqslant p$. For $\mathbf{A}$ $n\times p$ and $\mathbf{B}$ $p\times m$ the regular matrix product, which is denoted by the juxtaposition of two matrices, yields a matrix $\mathbf{C}=\mathbf{AB}$ which is $n\times m$ and defined by $\langle i|\mathbf{C}|j\rangle=\Sigma_{l=1}^{p}\langle i|\mathbf{A}|l\rangle\langle l|\mathbf{B}|j\rangle$. The Hadamard product (element by element multiplication) is defined by $\mathbf{C}=\mathbf{A}\odot\mathbf{B}$, where all matrices are $n\times m$ and the elements are $\langle i|\mathbf{C}|j\rangle=\langle i|\mathbf{A}|j\rangle\langle i|\mathbf{B}|j\rangle$. For $\mathbf{A}$ $n\times p$ and $\mathbf{B}$ $m\times q$ the Kronecker product $\mathbf{C}=\mathbf{A}\otimes\mathbf{B}$ is of size $nm\times pq$ with matrix elements $\langle ij|\mathbf{C}|kl\rangle=\langle i|\mathbf{A}|k\rangle\langle j|\mathbf{B}|l\rangle$. Here the double index notation $\langle ij|$ stands for the lexicographical ordering of the elements from $\mathbf{A}$ and $\mathbf{B}$ to form the matrix elements of $\mathbf{C}$ in the standard way.[2,3,7]

We define a row product of two matrices where $\mathbf{A}$ is $n\times p$ and $\mathbf{B}$ is $n\times q$ to be a $n\times pq$ matrix $\mathbf{C}=\{\mathbf{A}\ \mathbf{B}\}$ with elements given by $\langle i|\mathbf{C}|jk\rangle=\langle i|\mathbf{A}|j\rangle\langle i|\mathbf{B}|k\rangle$. Similarly a column product of two matrices is defined when $\mathbf{A}$ is $p\times n$ and $\mathbf{B}$ is $q\times n$ to be a $pq\times n$ matrix

$$\mathbf{C}=\begin{Bmatrix}\mathbf{A}\\ \mathbf{B}\end{Bmatrix}$$

with elements given by $\langle jk|\mathbf{C}|i\rangle=\langle j|\mathbf{A}|i\rangle\langle k|\mathbf{B}|i\rangle$. The notation throughout this paper will be such that parentheses around an array will denote a matrix in the standard form, while curly brackets around an array of matrices will denote that the matrix is to be assembled using the above row and column products. Also the juxtaposition of matrices inside a curly bracket will not denote regular matrix multiplication unless the matrices are grouped together using parenthesis. The relation between the row and column products is immediately seen to be given by

$$\begin{Bmatrix}\mathbf{A}^{\mathrm{T}}\\ \mathbf{B}^{\mathrm{T}}\end{Bmatrix}^{\mathrm{T}}=\{\mathbf{A}\ \mathbf{B}\}. \tag{2.1}$$

This relationship allows us to study the properties of one of the products and obtain the corresponding properties of the other by the transpose relationship. After studying these products, it was found that they had been discussed to a limited extent elsewhere.[4–6]

## 3. PROPERTIES OF ROW AND COLUMN MATRIX PRODUCTS

Before documenting the properties of the matrix products, it would be desirable to understand the origin of most of these properties. Just as the Hadamard product of two square matrices is a principal submatrix of the Kronecker product of the two matrices,[2,7] it is obvious that the new product of two matrices forms a submatrix of the Kronecker product of the two matrices. This means that if $\mathbf{A}$ is $n\times p$ and $\mathbf{B}$ is $n\times q$, then $\{\mathbf{A}\ \mathbf{B}\}=\mathbf{P}_R(\mathbf{A}\otimes\mathbf{B})$, where $\mathbf{P}_R$ is an orthogonal projector from an $n^2$-dimensional vector space to a linear manifold of rank $n$. Using this fact makes the proofs of most of the properties of the new product trivial. Thus only one proof has been included (Appendix A); the others have been documented elsewhere.[8] To avoid explicit reference to the size of the matrices we shall hereafter assume that the sizes of the matrices in any equation is such that the matrix products are defined.

Some of the obvious but useful properties of the new matrix products include the distributive law,

$$\{(\mathbf{A}+\mathbf{B})\ \mathbf{C}\}=\{\mathbf{A}\ \mathbf{C}\}+\{\mathbf{B}\ \mathbf{C}\}, \tag{3.1}$$

and the associative law,

$$\{\mathbf{A}\ \{\mathbf{B}\ \mathbf{C}\}\}=\{\{\mathbf{A}\ \mathbf{B}\}\ \mathbf{C}\}. \tag{3.2}$$

[a)]Present address: Department of Physics and Astronomy, University of Georgia, Athens, Georgia 30602.

 0022-2488/79/061146-05$01.00  

The distributive law assures us that the product is well defined, while the associative law means that the set of all matrices of column degree $n$ under the column product forms a noncommutative semigroup with an identity (a monoid).[9] The associative law when both the row and column products are considered allows us to define

$$\begin{Bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{Bmatrix} \equiv \left\{\begin{Bmatrix}\mathbf{A}\\ \mathbf{C}\end{Bmatrix} \quad \begin{Bmatrix}\mathbf{B}\\ \mathbf{D}\end{Bmatrix}\right\} = \begin{Bmatrix}\{\mathbf{A} & \mathbf{B}\}\\ \{\mathbf{C} & \mathbf{D}\}\end{Bmatrix}. \tag{3.3}$$

Some simple rules for combining the conventional and new matrix products are

$$\{\mathbf{A}\ \ \mathbf{B}\}\odot\{\mathbf{C}\ \ \mathbf{D}\} = \{(\mathbf{A}\odot\mathbf{C})\ \ (\mathbf{B}\odot\mathbf{D})\}, \tag{3.4}$$

$$\{\mathbf{A}\ \ \mathbf{B}\}(\mathbf{C}\otimes\mathbf{D}) = \{(\mathbf{AC})(\mathbf{BD})\}, \tag{3.5}$$

and

$$\{\mathbf{A}\ \ \mathbf{B}\ \ \mathbf{C}\}\begin{Bmatrix}\mathbf{D}\\ \mathbf{E}\\ \mathbf{F}\end{Bmatrix} = (\mathbf{AD})\odot(\mathbf{BE})\odot(\mathbf{CF}). \tag{3.6}$$

Another property that will be useful in the next section is that if $\mathbf{Q}$ is a permutation matrix and $\mathbf{D}_1$, $\mathbf{D}_2$ are nonsingular diagonal matrices with $\mathbf{D} = \mathbf{D}_1\mathbf{D}_2 = \mathbf{D}_1\odot\mathbf{D}_2$, then

$$\mathbf{QD}\{\mathbf{A}\ \ \mathbf{B}\} = \{(\mathbf{QD}_1\mathbf{A})\ \ (\mathbf{QD}_2\mathbf{B})\}. \tag{3.7}$$

A more general version of this theorem will be stated and proved in Appendix A. It is easily shown[8] that the properties listed above can be extended to arrays with an arbitrary number of matrices.

The constant matrix $\mathbf{J}$ which has all elements equal to unity plays a unique role in a matrix formed using the generalized Kronecker products. Using subscripts to avoid confusion about the size of the matrices, we can list the interesting identity

$$\mathbf{C}_{l\times n}\{\mathbf{J}_{n\times m}\ \ \mathbf{A}_{n\times p}\ \ \mathbf{J}_{n\times q}\} = \{\mathbf{J}_{l\times m}\ \ (\mathbf{CA})_{l\times p}\ \ \mathbf{J}_{l\times q}\}, \tag{3.8}$$

where either the front or back constant matrices can be absent from both sides of the identity. It is of interest to note that since

$$\mathbf{A}\otimes\mathbf{B} = \begin{Bmatrix}\mathbf{A} & \mathbf{J}\\ \mathbf{J} & \mathbf{B}\end{Bmatrix} \tag{3.9}$$

the property $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD})$ can be derived by using Equations (3.3)–(3.9). In order not to obscure the structure of a matrix, we shall hereafter suppress the explicit listing of any constant matrices that enter the matrix product in an off-diagonal position. For example, the Kronecker product of three matrices will be denoted by

$$\begin{Bmatrix}\mathbf{A} & & \\ & \mathbf{B} & \\ & & \mathbf{C}\end{Bmatrix} \equiv \begin{Bmatrix}\mathbf{A} & \mathbf{J} & \mathbf{J}\\ \mathbf{J} & \mathbf{B} & \mathbf{J}\\ \mathbf{J} & \mathbf{J} & \mathbf{C}\end{Bmatrix} = \mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C}. \tag{3.9$'$}$$

The permutation matrix that interchanges vectors in a direct product representation also has an intuitive form using the new matrix products. For example, if $\mathbf{A}_i$, $1\leqslant i\leqslant 4$ are $n\times m$ matrices, then

$$\begin{Bmatrix}\mathbf{J} & \mathbf{I} & & \\ & \mathbf{J} & \mathbf{I} & \\ \mathbf{I} & & \mathbf{J} & \\ & & & \mathbf{I}\end{Bmatrix}\begin{Bmatrix}\mathbf{A}_1\\ \mathbf{A}_2\\ \mathbf{A}_3\\ \mathbf{A}_4\end{Bmatrix} = \begin{Bmatrix}\mathbf{A}_2\\ \mathbf{A}_3\\ \mathbf{A}_1\\ \mathbf{A}_4\end{Bmatrix}, \tag{3.10}$$

where $\mathbf{I}$ is the $n\times n$ identity matrix, $\mathbf{J}$ is the $n\times n$ constant matrix, and $\mathbf{J}$ matrices in the off-diagonal positions are indicated by blanks. It is important to realize that this intuitive form of regular matrix multiplication in which the $\mathbf{J}$ matrices play the role of a zero element is not possible when the number of nonconstant matrices in each row and column of the matrix products to be multiplied is different from one.

The trace properties of matrices under the new product is of special importance since it will be seen in the next section that the trace enters in classical statistical mechanics. It is interesting to note that the trace properties of the new product include as special cases the trace properties of Kronecker products

$$\mathrm{Tr}\left(\begin{Bmatrix}\mathbf{A}_1 & & \\ & \mathbf{A}_2 & \\ & & \mathbf{A}_3\end{Bmatrix}\right) = \mathrm{Tr}(\mathbf{A}_1\otimes\mathbf{A}_2\otimes\mathbf{A}_3) = \mathrm{Tr}(\mathbf{A}_1)\mathrm{Tr}(\mathbf{A}_2)\mathrm{Tr}(\mathbf{A}_3) \tag{3.11}$$

of Hadamard products

$$\mathrm{Tr}(\mathbf{A}_1\odot\mathbf{A}_2\odot\mathbf{A}_3) = \mathrm{Tr}\left(\begin{Bmatrix}\mathbf{A}_1 & \mathbf{I} & \\ & \mathbf{A}_2 & \mathbf{I}\\ & & \mathbf{A}_3\end{Bmatrix}\right), \tag{3.12}$$

and of regular matrix multiplication

$$\mathrm{Tr}(\mathbf{D}_1\mathbf{A}_1\mathbf{D}_2\mathbf{A}_2\mathbf{D}_3\mathbf{A}_3) = \mathrm{Tr}\left(\begin{Bmatrix}\mathbf{D}_1 & \mathbf{A}_1 & \\ & \mathbf{D}_2 & \mathbf{A}_2\\ \mathbf{A}_3 & & \mathbf{D}_3\end{Bmatrix}\right), \tag{3.13}$$

where $\mathbf{I}$ denotes the identity matrix and the $\mathbf{D}_i$ are diagonal matrices.

If all the matrices that enter the product in the diagonal position are square diagonal matrices, the matrix formed by the product is diagonal. This implies that

$$\mathrm{Tr}\left(\begin{Bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13}\\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23}\\ \mathbf{A}_{31} & \mathbf{A}_{32} & \mathbf{A}_{33}\end{Bmatrix}\right) = \mathrm{Tr}\left(\begin{Bmatrix}\mathbf{A}_{11} & \mathbf{B}_{12} & \mathbf{B}_{13}\\ & \mathbf{A}_{22} & \mathbf{B}_{23}\\ & & \mathbf{A}_{33}\end{Bmatrix}\right), \tag{3.14}$$

where $\mathbf{B}_{ij} = \mathbf{A}_{ij}\odot\mathbf{A}_{ji}^{\mathrm{T}}$. Equations (3.11)–(3.14) are easily generalized to include the trace of products of more than three matrices. In the general case it is possible to develop an expansion for the trace in terms of a vector expansion of the matrices that enter the new product. This is done in Sec. 4B, and in Appendix B it will be shown how this trace can be done graphically on the lattice of interactions.

## 4. APPLICATIONS

In this section the formalism is applied to classical statistical mechanics. In particular we shall limit ourselves to a system with only two-body and external field interactions, and will specialize to the case where each of the particles of the system has a countable number of energy levels. Section 4A will deal with the transfer matrix approach, while in Sec. 4B the method of obtaining high and low temperature expansions will be illustrated.

### A. Transfer matrix method

In classical statistical mechanics whenever the system has only short-range interactions the analysis of the system as a Markov process yields a transfer matrix method of evaluating the partition function.[10,11] The traditional method of obtaining the transfer matrix[12,13] is to partition the system into $M$ layers such that if $\mu_i$ denotes the configuration of layer $i$ the energy of any configuration is given by

$$\sum_{i=1}^{M} [E_i(\mu_i,\mu_{i+1}) + E_i(\mu_i)]$$

where usually $M+1=1$. Here $E_i(\mu_i,\mu_{i+1})$ is the interlayer interaction energy between layers $i$ and $i+1$ and $E_i(\mu_i)$ is the intralayer energy of layer $i$. The transfer matrix between layers $i$ and $i+1$ is then constructed by defining its elements to be $\langle\mu_i|\mathbf{P}_i|\mu_{i+1}\rangle \equiv \exp[-\beta E_i(\mu_i,\mu_{i+1}) - \beta E_i(\mu_i)]$, where $\beta=(k_B T)^{-1}$. The partition function is then given by $Z=\mathrm{Tr}(\mathbf{P}_1\mathbf{P}_2\cdots\mathbf{P}_M)$ and in order to perform the trace the structure of the transfer matrices $\mathbf{P}_i$ are studied.[3]

The approach taken in the present work is to deal directly with the two-particle transfer matrix $\mathbf{A}_{ij}$ between particles $i$ and $j$ and the matrix $\mathbf{D}_i$ that describes particles interacting with an external field. The matrices $\mathbf{D}_i$ are diagonal with elements given by $\exp[-\beta E_i(s)]$ where $E_i(s)$ is the energy of state $s$ when particle $i$ is in an external field. The elements of $\mathbf{A}_{ij}$ are $\exp[-\beta E_{ij}(s,t)]$ where $E_{ij}(s,t)$ is the energy associated with particle $i$ in state $s$ and particle $j$ in state $t$, while if the state described by this configuration is energetically forbidden $[E_{ij}(s,t)\to\infty]$ $\mathbf{A}_{ij}$ has a zero element in this position.

For the noninteracting $N$ particle problem the partition function is clearly given by the Kronecker product

$$\mathscr{Z} = \mathrm{Tr}(e^{-\beta H})$$

$$= \mathrm{Tr}\left(\begin{Bmatrix} \mathbf{D}_1 & & & \\ & \mathbf{D}_2 & & \\ & & \ddots & \\ & & & \mathbf{D}_N \end{Bmatrix}\right). \qquad (4.1)$$

If we now include two-body interactions some of the matrices in the off-diagonal positions of the product will no longer be constant matrices since for an interacting system some of the $E_{ij}(s,t)$ will no longer be zero. For a classical system the two-particle Hamiltonians satisfy $\exp[-\beta(H_{ij}+H_{kl})] = \exp(-\beta H_{ij})\exp(-\beta H_{kl})$, $1\leqslant i,j,k,l\leqslant N$, so the partition function is given by

$$\mathscr{Z} = \mathrm{Tr}(\mathfrak{A}) = \mathrm{Tr}\left(\begin{Bmatrix} \mathbf{D}_1 & \mathbf{A}_{12} & \mathbf{A}_{13} & \cdots & \mathbf{A}_{1N} \\ & \mathbf{D}_2 & \mathbf{A}_{23} & \cdots & \mathbf{A}_{2N} \\ & & \mathbf{D}_3 & & \vdots \\ & & & \ddots & \vdots \\ & & & & \mathbf{D}_N \end{Bmatrix}\right) \qquad (4.2)$$

where $\mathbf{A}_{ij}$, $1\leqslant i<j\leqslant N$ are the two-particle transfer matrices, some of which may be equal to $\mathbf{J}$. Provided the system studied is a Markov process, the transfer matrix can be derived from Eq. (4.2) by choosing the numbering of the particles in such a way that use of the generalizations of Eq. (3.3) and Eq. (3.14) gives a matrix in the form generalized from Eq. (3.13).

As an example consider a one-dimensional closed chain of identical spins that have $\mathbf{D}$ as an external field interaction matrix, and the two-particle transfer matrices $\mathbf{A}$ between nearest neighbors and $\mathbf{B}$ between next nearest neighbors. Assuming an even number of particles gives the partition function

$$\mathscr{Z} = \mathrm{Tr}\left(\begin{Bmatrix} \begin{Bmatrix} \mathbf{D} & \mathbf{A} \\ & \mathbf{D} \end{Bmatrix} & \begin{Bmatrix} \mathbf{B} & \\ \mathbf{A} & \mathbf{B} \end{Bmatrix} & & & \\ & \begin{Bmatrix} \mathbf{D} & \mathbf{A} \\ & \mathbf{D} \end{Bmatrix} & \begin{Bmatrix} \mathbf{B} & \\ \mathbf{A} & \mathbf{B} \end{Bmatrix} & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \ddots \\ \begin{Bmatrix} \mathbf{B} & \\ \mathbf{A} & \mathbf{B} \end{Bmatrix} & & & & \begin{Bmatrix} \mathbf{D} & \mathbf{A} \\ & \mathbf{D} \end{Bmatrix} \end{Bmatrix}\right)$$

$$= \mathrm{Tr}\left(\left(\begin{Bmatrix} \mathbf{D} & \mathbf{A} \\ & \mathbf{D} \end{Bmatrix}\begin{Bmatrix} \mathbf{B} & \\ \mathbf{A} & \mathbf{B} \end{Bmatrix}\right)^{N/2}\right). \qquad (4.3)$$

Equation (4.3) demonstrates that the transfer matrix obtained using this method takes a very compact form. It also illustrates that although the size and individual elements of the transfer matrix depend on the details of the system, the underlying structure of the transfer matrix depends only on which particles interact.

A partition function in the form of Eq. (4.2) also can facilitate the calculation of the expectation value of an observable $\mathscr{O}$. This is because it is possible to use the algebraic properties of the matrix products to simplify the calculation of the expectation value $\langle\mathscr{O}\rangle = \mathrm{Tr}(Oe^{-\beta H})/\mathscr{Z}$.

## B. Lattice expansions

This section developes an expansion for a matrix given by Eq. (4.2) and demonstrates that the traditional high and low temperature expansions of the spin-$\frac{1}{2}$ Ising model are special cases of this expansion. Appendix B will show how the general expansion can be done graphically on the lattice of interactions.

The trace of (4.2) can be written as

$$\mathrm{Tr}(\mathfrak{A}) = \sum_c \widehat{\mathbf{X}}_c^{\mathrm{T}}\mathfrak{A}\widehat{\mathbf{X}}_c, \qquad (4.4)$$

where $\widehat{\mathbf{X}}_c$ is the unit vector along the axis $c$ of the matrix $\mathfrak{A}$ and the sum is over the orthonormal complete set of vectors formed by the $\widehat{\mathbf{X}}_c$. The vector $\widehat{\mathbf{X}}_c^{\mathrm{T}}$ represents one configuration of the system and can be written as the product of the vectors $\hat{X}^{j\,\mathrm{T}}_{k}$ corresponding to particle $j$ being in state $k$. This allows us to define

$$\hat{X}^{\mathrm{T}}_{c} \equiv \hat{X}_{k_1k_2\cdots k_N}{}^{\mathrm{T}} = \{\hat{X}^{1\,\mathrm{T}}_{k_1}\hat{X}^{2\,\mathrm{T}}_{k_2}\cdots\hat{X}^{N\,\mathrm{T}}_{k_N}\}. \tag{4.5}$$

Substituting Eq. (4.5) into Eq. (4.4) and using the theorem of Appendix A gives

$$\mathrm{Tr}(\mathfrak{A}) = \sum_{k_1\cdots k_N} \hat{X}_{k_1\cdots k_N}{}^{\mathrm{T}}\mathfrak{A}\hat{X}_{k_1\cdots k_N}$$

$$= \sum_{k_1\cdots k_N} \prod_{ij=1}^{N}{}' \hat{X}^{i\,\mathrm{T}}_{k_i}\mathbf{A}_{ij}\hat{X}^{j}_{k_j} \tag{4.6}$$

where $\mathbf{A}_{ii} \equiv \mathbf{D}_i$ and the prime means that constant matrices do not have to be included in the product. Now for each $\mathbf{A}_{ij}$ we expand the unit vectors in Eq. (4.6) using a complete set of expansion vectors. This gives

$$\hat{X}^{j}_{k} = \sum_{m} \alpha^{ij}_{km}\vec{p}^{ij}_{m} \quad \text{and} \quad \hat{X}^{i\,\mathrm{T}}_{k} = \sum_{n} \gamma^{ij}_{kn}\vec{q}^{ij\,\mathrm{T}}_{n}$$

which puts Eq. (4.6) into the form

$$\mathrm{Tr}(\mathfrak{A}) = \sum_{k_1\cdots k_N} \prod_{ij=1}^{N}{}' \left[\sum_{mn} \alpha^{ij}_{k_jm}\gamma^{ij}_{k_in}\eta^{ij}_{nm}\right], \tag{4.7}$$

where $\eta^{ij}_{nm} \equiv \vec{q}^{ij\,\mathrm{T}}_{n}\mathbf{A}_{ij}\vec{p}^{ij}_{m}$. The freedom in picking the expansion vectors can in general be used to put the series into a convenient form, as will be demonstrated in Appendix B.

As an example let us consider the zero field spin-$\frac{1}{2}$ Ising model on any lattice, and let all interacting spins have the two-particle transfer matrix

$$\mathbf{A} = \begin{pmatrix} e^{J} & e^{-J} \\ e^{-J} & e^{J} \end{pmatrix}.$$

Choosing the expansion vectors $\vec{p}_1 = \vec{q}_1 = \binom{1}{1}$ and $\vec{p}_2 = \vec{q}_2 = \binom{1}{-1}$ gives

$$\eta_{nm} = \vec{q}^{\mathrm{T}}_{n}\mathbf{A}\vec{p}_{m} = 4\delta_{nm}\cosh(J)[\delta_{n1} + \tanh(J)\delta_{n2}]$$

and

$$\alpha_{kn} = \gamma_{kn} = \tfrac{1}{2} - \delta_{kn}\delta_{n2}.$$

Substituting these values into Eq. (4.7) yields

$$\mathscr{Z} = \cosh^{P}(J) \sum_{k_1\cdots k_N=1} \prod_{ij=1}^{N}{}' [1 + (-1)^{(k_i+k_j)}\tanh(J)], \tag{4.8}$$

where $P$ is the total number of interacting pairs. Equation (4.8) is the traditional high temperature expansion[14] as can be seen by writing out the product and doing the sum for each of the terms. The standard technique for deriving the high temperature expansion uses the identity $e^{J\kappa} = \cosh(J) \times [1 + \kappa\tanh(J)]$ which is valid for $\kappa = \pm 1$, and yields a simpler derivation of Eq. (4.8). However the present derivation has the advantage that it is easily generalized to other systems and other temperature regions. For example, low temperature expansions can be extracted from Eq. (4.7) by choosing the expansion vectors $\vec{p}^{ij}_{1}$ and $\vec{q}^{ij\,\mathrm{T}}_{1}$ to project out one of the ground states of the system, the other expansion vectors being chosen in a convenient manner.

## 5. SUMMARY

In this work a new type of matrix product has been introduced and some of its properties have been documented. We found that the trace properties were of particular interest since they are directly applicable to the evaluation of the partition function of a classical system with two-body interactions. This formulation led to an elegant procedure for deriving the transfer matrix for the partition function whenever the system had a Markov nature. In the general case the matrix trace was shown to yield a series expansion, and it was illustrated in Appendix B how the expansion could be carried out on the interaction lattice using bicolored bonds and vertex weight functions. This series expansion was seen to be quite general.

It was shown how the high and low temperature expansions could be extracted from the formalism. By deriving the traditional expansions and the transfer matrix from one starting point, a connection between these two techniques in classical statistical mechanics was demonstrated. A subsequent paper will use the formalism developed here to study the mapping between different systems in classical statistical mechanics.

## APPENDIX A

*Theorem*: Let $\mathbf{Q}$ be an $n \times m$ matrix with at most one nonzero element in each row, and let the size of $\mathbf{A}$ be $m \times l$ and $\mathbf{B}$ $m \times p$. Also let the two $n \times m$ matrices $\mathbf{Q}_1$, $\mathbf{Q}_2$ satisfy the conditions: (i) $\mathbf{Q} = \mathbf{Q}_1 \odot \mathbf{Q}_2$, (ii) if $\mathbf{Q}$ has a nonzero element somewhere in the row, then both $\mathbf{Q}_1$ and $\mathbf{Q}_2$ have zero elements in the row wherever $\mathbf{Q}$ does, and (iii) if $\mathbf{Q}$ has all zeros in a row, then either $\mathbf{Q}_1$ or $\mathbf{Q}_2$ has all zeros in that row. Then

$$\mathbf{Q}\{\mathbf{A}\ \ \mathbf{B}\} = \{(\mathbf{Q}_1\mathbf{A})\ \ (\mathbf{Q}_2\mathbf{B})\}.$$

*Proof*: Let $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant l$, $1 \leqslant q \leqslant p$, and let $t$ be the element of row $i$ of $\mathbf{Q}$ which is nonzero, assuming there is such an element. Then using elementary steps one has:

$$\langle i|\mathbf{Q}\{\mathbf{A}\ \ \mathbf{B}\}|jq\rangle$$

$$= \sum_{s=1}^{m} \langle i|\mathbf{Q}|s\rangle\langle s|\{\mathbf{A}\ \ \mathbf{B}\}|jq\rangle = (\langle i|\mathbf{Q}|t\rangle)(\langle t|\{\mathbf{A}\ \ \mathbf{B}\}|jq\rangle)$$

$$= (\langle i|\mathbf{Q}_1|t\rangle\langle i|\mathbf{Q}_2|t\rangle)(\langle t|\mathbf{A}|j\rangle\langle t|\mathbf{B}|q\rangle)$$

$$= \left(\sum_{s=1}^{m} \langle i|\mathbf{Q}_1|s\rangle\langle s|\mathbf{A}|j\rangle\right)\left(\sum_{u=1}^{m} \langle i|\mathbf{Q}_2|u\rangle\langle u|\mathbf{B}|q\rangle\right)$$

$$= \langle i|\mathbf{Q}_1\mathbf{A}|j\rangle\langle i|\mathbf{Q}_2\mathbf{B}|q\rangle = \langle i|\{(\mathbf{Q}_1\mathbf{A})\ \ (\mathbf{Q}_2\mathbf{B})\}|jq\rangle.$$

If all the elements in a row of $\mathbf{Q}$ are zero, then all the elements in that row of $\mathbf{Q}\{\mathbf{A}\ \ \mathbf{B}\}$ are zero. If one also notices that provided all the elements of one row of either $\mathbf{Q}_1$ or $\mathbf{Q}_2$ are zero the elements of that row of $\{(\mathbf{Q}_1\mathbf{A})\ \ (\mathbf{Q}_2\mathbf{B})\}$ are zero, then the theorem is proved.

## APPENDIX B

This Appendix documents a graphical technique of doing the sum in Equation (4.7). For simplicity we assume that all the matrices $\mathbf{A}_{ij}$ are $M \times M$. The trace is done graphically by following the steps below, where $P$ is the number of interacting pairs.

(1) Draw $M^{2P}$ identical directed graphs (interaction lattices). Each graph should have $N$ vertices and an arrow from point $i$ to $j$ if the $\mathbf{A}_{ij}$ matrix of Eq. (4.7) is not the constant matrix $\mathbf{J}$.

(2) Color the head and tail of each bond choosing one of $M$ colors, and make the coloring of each of the $M^{2P}$ graphs different. Each bond can be colored in $M^2$ ways so $M^{2P}$, the number of graphs, is the number of possible colorings.

(3) Assign a numerical factor to each bond of every graph. This numerical factor should be the $\eta^{ij}_{nm}$ of Eq. (4.7) for an arrow directed from point $i$ to point $j$ with tail of color $n$ and head of color $m$.

(4) Assign a vertex weight $\rho_l$ to every vertex of each graph. This vertex weight is found for vertex $l$ of any graph by: (i) multiplying together a $\gamma^{lj}_{kn}$ for each arrow directed from vertex $l$ to vertex $j$ with a tail of color $n$ and a $\alpha^{il}_{km}$ for each arrow entering vertex $l$ from vertex $i$ with a head of color $m$ (ii) summing this product over $k$, $1 \leqslant k \leqslant M$ to give the vertex weight $\rho_l$.

(5) Assign a numerical factor to each of the $M^{2P}$ graphs by multiplying together the vertex weights assigned to each of its $N$ vertices and the arrow weights assigned to each of its $P$ arrows.

(6) The trace expansion of Eq. (4.7) is then equal to the sum of the numerical factors assigned to each of the $M^{2P}$ graphs.

The procedure above is seen to be correct by writing out the product in Eq. (4.7) (which is the sum of $M^{2P}$ factors) and doing the summation over the $k_i$ $1 \leqslant i \leqslant N$ to give each factor the value assigned to one of the $M^{2P}$ graphs above.

This graphical technique can be simplified by choosing the expansion vectors of Eq. (4.7) in a convenient way. The choice $\vec{p}^{ij}_m = \vec{q}^{ij}_m$ allows the graphical technique to be done without distinguishing the head of the arrow from the tail. Furthermore only bonds of a single color need to be used if the expansion vectors are chosen to make $\eta^{ij}_{mn} \propto \delta_{mn}$. A further reduction in the number of graphs that are assigned nonzero weights can be obtained by using as expansion vectors any eigenvectors of a two-particle transfer matrix that have eigenvalue zero.

## ACKNOWLEDGMENTS

The author wishes to thank his dissertation advisor, Professor W.A. Little, for his support and his many useful discussions. Thanks are also due to Professor A.L. Fetter for his comments on the initial draft. Support from the National Aeronautical and Space Agency, Contract JPL 953752, and the National Science Foundation, Grant DMR 76-00726-A01 is acknowledged.

[1]C.R. Rao and S.K. Mitra, *Generalized Inverse of Matrices and Its Applications* (Wiley, New York, 1971).
[2]P.R. Halmos, *Finite-Dimensional Vector Spaces* (Van Nostrand, Princeton, N.J., 1958).
[3]K. Huang, *Statistical Mechanics* (Wiley, New York, 1973).
[4]C.G. Khatri and C.R. Rao, Sankhyā, Ser. A **30**, 167 (1968).
[5]C.R. Rao, J. Am. Stat. Assoc. **65**, 161 (1970).
[6]C.G. Khatri, J. Multivar. Anal. **1**, 70 (1971).
[7]M. Marcus and N.A. Khan, Can. Math. Bull. **2**, 81 (1959).
[8]M. Novotny, Ph.D. thesis, Stanford Univ. (1978) (unpublished).
[9]N. Jacobson, *Lectures In Abstract Algebra* (Van Nostrand, New York, 1951).
[10]H.A. Kramers and G.H. Wannier, Phys. Rev. **60**, 252 (1941).
[11]E.W. Montroll, Ann. Math. Stat. **18**, 18 (1947).
[12]E.H. Lieb, *Lectures In Theoretical Physics*, edited by K.T. Mahanthappa and W.E. Brittin (Gordon and Breach, New York, 1969), Vol. XI-D.
[13]W.J. Camp and M.E. Fisher, Phys. Rev. B **6**, 946 (1972).
[14]H.E. Stanley, *Introduction To Phase Transitions and Critical Phenomena* (Oxford U.P., New York, 1971).

# Jost solutions and Green's functions for the three-dimensional Schrödinger equation[a)]

Harry E. Moses

*Center for Atmospheric Research, College of Pure and Applied Science, University of Lowell, Lowell, Massachusetts 01854*

(Received 7 July 1978)

In an earlier paper, in which the minimal scattering data needed to reconstruct the scattering potential was found for the three-dimensional inverse problem, a Green's function appeared very naturally in this context. We use this Green's function to construct Jost solutions for the three-dimensional problem which are closely analogous to those for the one-dimensional problem. The Green's function and the Jost solutions differ from those given by Faddeev. The completeness relations of the new Jost solutions are given simply in terms of the scattering amplitude. They are the same as those given in an earlier attempt to obtain an algorithm of the Gelfand–Levitan type for the three-dimensional problem. However, it is not yet clear that the Jost solutions of the present paper are the same as those of the earlier paper, since different methods are used to define the two sets of Jost solutions. The one-dimensional problem is discussed in some detail to motivate our definition of the Jost solutions for the three-dimensional problem. The present paper is the first of three papers which report on research arising from the three-dimensional inverse problem.

## 1. INTRODUCTION

This is the first of a series of papers arising from our search for an algorithm of the Gel'fand–Levitan type for the inverse problem for the three-dimensional Schrödinger equation. Though algorithms of the Gel'fand–Levitan type have been given earlier (Refs. 1–3), we have been looking for others, which in a certain sense are more closely analogous to the algorithm for the one-dimensional problem. As yet we have not achieved our objective. However, we have obtained some interesting results for possible Jost solutions and corresponding Green's functions, which are direct generalizations for those for the one-dimensional problem and which fall squarely into the direct scattering problem as described in Ref. 4. The proposed Jost solutions differ from those used in Refs. 2 and 3 and have much simpler completeness relations in terms of the scattering amplitudes. These completeness relations are, in fact, identical to those for the Jost solutions of Ref. 1. In contrast to the method of obtaining the Jost solutions by means of appropriate Green's functions as in the present paper, a simple triangularity condition was assumed for the Gel'fand–Levitan kernel and the Jost function was constructed through the use of this kernel. Despite the agreement of the completeness relations between the Jost solutions of the present paper and those of Ref. 1, it is not yet clear to us that the two sets of Jost solutions are identical.

In the one-dimensional case, the completeness relations of the Jost solutions involve the minimal scattering and bound state data needed to recover the potential. In the three-dimensional case, the completeness relations involve redundant data. In the one-dimensional case, the completeness relations of the Jost solutions are essentially equivalent to the Gel'fand–Levitan equation or algorithm. Hence the present paper leads to the possibility that the three-dimensional algorithm will contain redundant data as its input and constraints will have to be added to remove the redundancy. (This difficulty also appears in the algorithms of Refs. 1–3.)

[a)]This research was supported by the Army Research Office under Grant Number DAAG29-78-G-0003.

The second paper of this series will rederive a nonlinear method of finding the scattering potential from the minimal data, this method being first given in Ref. 5. The rederivation will use the completeness relations of the Jost solutions, instead of those for the outgoing wavefunctions and will be carried out both for the one- and three-dimensional problems. The second paper is intended to add weight to our arguments that the three-dimensional Jost solutions of the present paper and the corresponding Green's function are particularly appropriate.

In the third paper we shall review the process of obtaining the Gel'fand–Levitan kernel in terms of the Green's function and the scattering potential for the one-dimensional case as a prototype of a similar calculation for the three-dimensional case. In particular, the "double Fourier transform" of the Green's function which gives the nature of the triangularity of the Gel'fand–Levitan kernel will be shown explicitly for three dimensions. Unlike the case for one dimension, the double Fourier transform (essentially the influence function of a homogeneous hyperbolic equation) is a distribution. Though formally this distribution has a simple appearance, it has a sufficiently formidable geometrical intepretation to make it difficult to write the integral equation for the Gel'fand–Levitan kernel in terms of the scattering potential. At this point we cannot yet express the scattering potential in terms of the Gel'fand–Levitan kernel. Hence, at this time, our kernel is at a distinct disadvantage with respect to that used in Refs. 2 and 3. Nevertheless, we believe the closeness of our approach to the one-dimensional analog makes the effort of obtaining the Gel'fand–Levitan kernel and its properties from our Green's function and the scattering potential a worthwhile objective.

These three papers, then, represent a status report of a new approach to a very difficult problem: the obtaining of useful algorithms for solving the inverse problem in three dimensions.

We shall refer very heavily to our earlier work (Refs. 1, 4, and 6) and, for the sake of brevity, will not repeat arguments which occur in these references. We shall also use a simpler notation where possible and relate it to the earlier

 0022-2488/79/061151-06$01.00  

notation. Also in the present paper we shall assume that there are no bound states.

## 2. THE GREEN'S FUNCTION FOR THE ONE-DIMENSIONAL PROBLEM

We define $\gamma_{\pm}(x)$ by

$$\gamma_{\pm}(x) = \lim_{\epsilon\to+0} \frac{1}{x\mp i\epsilon} = \pm \pi i\delta(x) + \frac{P}{x}, \qquad (1)$$

where $P$ stands for the principal part when $1/(x-x')$ is used as the kernel of an integral operator.

Let us consider the solutions $\psi(x|p)$ of the one-dimensional Schrödinger equation

$$\left[-\frac{d^2}{dx^2} + V(x)\right]\psi(x|p) = p^2\psi(x|p) \quad (-\infty < p < +\infty), \qquad (2)$$

which we shall denote by $\psi_{\pm}(x|p)$ which satisfy the integral equations

$$\psi_{\pm}(x|p) = \psi_0(x|p) + \int_{-\infty}^{\infty} G_{p\pm}(x|x')V(x')\psi_{\pm}(x'|p)\,dx', \qquad (3)$$

where $\psi_0(x|p) = (2\pi)^{-1/2}e^{ipx}$ is the eigenfunction of $H_0 = -d^2/dx^2$ and $G_{p\pm}(x|x') \equiv G_{p\pm}(x-x')$ are the outgoing $(-)$ and incoming $(+)$ Green's function given by

$$G_{p\pm}(x) = (1/2\pi)\int_{-\infty}^{\infty} \gamma_{\pm}(p^2-k^2)e^{ikx}\,dk\,. \qquad (4)$$

The eigenfunctions $\psi_{\pm}$ are called the outgoing and incoming wavefunctions.

The integral in Eq. (4) is readily carried out to give the well-known result

$$G_{p\pm}(x) = \pm\frac{i}{2|p|}\exp[\mp i|p|\,|x|]. \qquad (5)$$

The Jost solutions (from the left) are defined to solutions of Eq. (2) which satisfy the boundary condition

$$\lim_{x\to-\infty} f(x|p) = e^{ipx}. \qquad (6)$$

These functions are readily seen to obey the integral equation

$$f(x|p) = e^{ipx} + \int_{-\infty}^{\infty} G_{pJ}(x|x')V(x')f(x'|p)\,dx', \qquad (7)$$

where $G_{pJ}(x|x') \equiv G_{pJ}(x-x')$ is the Jost Green's function ($J$ stands for "Jost") given by

$$G_{pJ}(x) = \eta(x)\frac{\sin px}{p}. \qquad (8)$$

In Eq. (8) $\eta(x)$ is the Heaviside step function defined by

$$\eta(x) = 1 \quad \text{for } x>0, \quad \eta(x) = 0 \quad \text{for } x<0. \qquad (9)$$

We should like to write $G_{pJ}$ as a Fourier transform of a distribution $\gamma(p,k)$ in a manner analogous to Eq. (4). We shall show

$$G_{pJ}(x) = (1/2\pi)\int_{-\infty}^{\infty} \gamma(p,k)e^{ikx}\,dk\,, \qquad (10)$$

where

$$\gamma(p,k) = \eta(k)\gamma_{-}(p^2-k^2) + \eta(-k)\gamma_{+}(p^2-k^2). \qquad (11)$$

The relation given by Eqs. (10) and (11) seems not to be generally known. We first came across $\gamma(p,k)$ in carrying out research on the inverse problem of Ref. 5 in which a non-Gel'fand–Levitan technique was used. The three-dimensional analog of $\gamma(p,k)$ also appears in that paper. The three-dimensional analog will be used to define a Jost Green's function analogous to $G_{pJ}$ in the manner of Eq. (10). The three-dimensional Jost solutions will then be constructed as the solution of an integral equation which is the generalization of Eq. (7).

We note that $\gamma(p,k)$ is formed by projecting $\gamma_-$ and $\gamma_+$ on the positive and negative $k$ axes respectively. We reverse the projections to obtain $\hat{\gamma}(p,k)$:

$$\hat{\gamma}(p,k) = \eta(-k)\gamma_{-}(p^2-k^2) + \eta(k)\gamma_{+}(p^2-k^2). \qquad (12)$$

The corresponding Green's function $\widehat{G}_{pJ}(x)$ which is obtained from $\hat{\gamma}(p,k)$ as in Eq. (10) can readily be shown to be given by $\widehat{G}_{pJ}(x) = G_{pJ}(-x)$ and is used as in Eq. (7) to construct the Jost solutions $\hat{f}(x|p)$ from the right. These are solutions of Eq. (2) which satisfy the boundary condition

$$\lim_{x\to+\infty} \hat{f}(x|p) = e^{ipx}. \qquad (13)$$

We shall now give a simple proof that the use of Eq. (11) in Eq. (1) gives the Green's function $G_{pJ}(x)$ of Eq. (8). From Eqs. (1) and (11)

$$\gamma_{+}(p^2-k^2) = \gamma_{-}(p^2-k^2) + 2\pi i\delta(p^2-k^2) \qquad (14)$$

so that

$$\gamma(p,k) = \gamma_{-}(p^2-k^2) + \eta(-k)2\pi i\delta(p^2-k^2). \qquad (15)$$

But

$$\eta(-k)\delta(p^2-k^2) = \frac{1}{2|p|}\eta(-k)\delta(k+|p|). \qquad (16)$$

Thus on using Eq. (16) in Eq. (15), on substituting into Eq. (12), and finally on using Eqs. (4) and (5),

$$G_{pJ}(x) = -\frac{i}{2|p|}\exp(+i|p|\,|x|) + \frac{i}{2|p|}\exp(-i|p|x), \qquad (8')$$

which is identical to Eq. (8).

## 3. THE THREE-DIMENSIONAL GREEN'S FUNCTION

For three-dimensions, the outgoing and incoming wave functions are solutions of

$$[-\nabla^2 + V(\mathbf{x})]\psi(\mathbf{x}|\mathbf{p}) = p^2\psi(\mathbf{x}|\mathbf{p}) \quad (p = |\mathbf{p}|), \qquad (17)$$

which satisfy the integral equations

$$\psi_{\pm}(\mathbf{x}|\mathbf{p}) = \psi_0(\mathbf{x}|\mathbf{p}) + \int G_{p\pm}(\mathbf{x}|\mathbf{x}')V(\mathbf{x}')\psi_{\pm}(\mathbf{x}'|\mathbf{k})\,d\mathbf{x}' \qquad (18)$$

where

$$\psi_0(\mathbf{x}|\mathbf{p}) = (1/2\pi)^{3/2}\exp[i(\mathbf{p}\cdot\mathbf{x})] \qquad (18a)$$

and the Green's functions for the outgoing and incoming wave functions are given by

$$G_{p\pm}(\mathbf{x}|\mathbf{x}') = G_{p\pm}(\mathbf{x}-\mathbf{x}') = -(1/4\pi)\frac{\exp[\mp ip|\mathbf{x}-\mathbf{x}'|]}{|\mathbf{x}-\mathbf{x}'|}. \tag{18b}$$

As is well known, $G_{p\pm}(\mathbf{x})$ can be represented as the Fourier transform of the $\gamma_{\pm}$ functions as follows:

$$G_{p\pm}(\mathbf{x}) = (1/2\pi)^3\int \gamma_{\pm}(p^2-k^2)e^{i\mathbf{k}\cdot\mathbf{x}}d\,\mathbf{k}$$
$$(k=|\mathbf{k}|). \tag{19}$$

Equation (19) is the obvious analog of Eq. (9) for the one-dimensional case. It is our objective to give a $G_{pJ}(\mathbf{x})$ which is analogous to $G_{pJ}(x)$ of the one-dimensional problem and thereby define a three-dimensional Jost solution $f(\mathbf{x}|\mathbf{p})$ by

$$f(\mathbf{x}|\mathbf{p}) = e^{i\mathbf{p}\cdot\mathbf{x}} + \int G_{pJ}(\mathbf{x}-\mathbf{x}')V(\mathbf{x}')f(\mathbf{x}'|\mathbf{p})dx'. \tag{20}$$

Our method for finding this Green's function is to give a function analogous to the function $\gamma(p,k)$, Eq. (11), of the one-dimensional problem and take its Fourier transform. Our choice for this function will be denoted by $\gamma(p,k)$, where we now use optical coordinates. These coordinates are defined with respect to an axis, which we shall take to be the $z$ axis. For any vector $\mathbf{Y}$ we define the optical coordinates which consist of an "optical radius" $Y=|\mathbf{Y}|$ if the $z$ component of $\mathbf{Y}$ is positive and $Y=-|\mathbf{Y}|$, if $\mathbf{Y}$ has a negative $z$ component, and angular variables $\theta$, $\phi$, which give the direction of $\mathbf{Y}/Y$. This unit vector always points in the positive $z$ direction (i.e., has a positive or zero $z$ component). The range of $\theta$ and $\phi$ are $0<\theta<\pi/2$ and $0<\phi<2\pi$. That is, these angular variables are the usual ones for spherical polar coordinates when they specify the unit vector's direction. Henceforth, unless we specify otherwise, we shall mean optical coordinates when we use polar coordinates notation for vectors. Then in terms of optical coordinates we define $\gamma(p,k)$ precisely by Eq. (11) with, however, a three-dimensional intepretation. Thus $G_{pJ}$ is given by

$$G_{pJ}(\mathbf{x}) = (1/2\pi)^3\int e^{i\mathbf{k}\cdot\mathbf{x}}\gamma(p,k)\,d\,\mathbf{k}. \tag{21}$$

It is readily verified that

$$(p^2+\nabla^2)G_{pJ}(\mathbf{x}) = \delta(\mathbf{x}), \tag{22}$$

as is required of all Green's functions.

As is the case for the analogous Green's functions of Refs. 2 and 3, to evaluate the integral of Eq. (21) seems impossible. It may be that $G_{pJ}$ is a distribution.

We shall now compare this Green's function with those used in Refs. 2 and 3. Toward this end we note that our function $\gamma(p,k)$ is based on the choice of a particular axis, namely the $z$ axis. Actually one could pick any direction specified by the unit vector $\boldsymbol{\alpha}$. One would then obtain Green's functions for every $\boldsymbol{\alpha}$ and a corresponding set of Jost functions, which are equivalent to those obtained previously obtained by choosing $\boldsymbol{\alpha}$ to be a unit vector pointing in the positive $z$ direction. Let us denote the more general distribution by $\gamma_{\alpha}(p,\mathbf{k})$. Then

$$\gamma_{\alpha}(p,\mathbf{k}) = \eta(\boldsymbol{\alpha}\cdot\mathbf{k})\gamma_-(p^2-k^2) + \eta(-\boldsymbol{\alpha}\cdot\mathbf{k})\gamma_+(p^2-k^2). \tag{23}$$

When $\boldsymbol{\alpha}$ points in the positive $z$ direction, we recover $\gamma(p,k)$ when we note $\eta(k_z)=\eta(k)$ and $o(-k_z)=\eta(-k)$ in accordance with our definition of optical coordinates.

We also recover the one-dimensional case when we realize that in one dimension $\boldsymbol{\alpha}$ has only two directions, namely in the direction of the $k$ axis and opposite to that direction. The first case gives us $\gamma(p,k)$ and the second $\hat{\gamma}(p,k)$. We shall now show that the Green's functions of Refs. 2 and 3 differ from ours and do *not* reduce to the one-dimensional functions which are known to be correct. Thus, in this sense at least, our Jost functions are a closer analog of those for one dimension.

In Refs. 2 and 3, the distribution corresponding to $\gamma_{\alpha}$ is a function of the *vector* $\mathbf{p}$ instead of only the magnitude as in Eq. (23). In terms of our notation their distribution $\gamma_{\alpha}(\mathbf{p},\mathbf{k})$ is given by

$$\gamma_{\alpha}(\mathbf{p},\mathbf{k}) = \eta[\boldsymbol{\alpha}\cdot(\mathbf{p}-\mathbf{k})]\gamma_-(p^2-k^2) + \eta[-\boldsymbol{\alpha}\cdot(\mathbf{p}-\mathbf{k})]\gamma_+(p^2-k^2). \tag{24}$$

Clearly this distribution does not reduce to Eq. (11) in the one-dimensional case, as our distribution does. Indeed, the distribution corresponding to Eq. (24) in the one-dimensional case appears to be undefined, being of the form of a sum of undefined products of distributions.

## 4. NORMALIZATION AND COMPLETENESS RELATIONS FOR THE ONE-DIMENSIONAL JOST SOLUTIONS

We shall now develop the normaliztion and completeness relations for the one-dimensional Jost solutions. The method which we shall use will be more fundamental than that used in Ref. 6. Our intent is to use the one-dimensional treatment as a prototype of the three-dimensional one. The parallels are very close.

Instead of using $\psi_0(x|p)$ and $f(x|p)$ it will be convenient to define, using the notation of Refs. 4 and 6, $\langle x|H_0,A_0;E,a\rangle$ and $\langle x|H,A;E,a\rangle$ by

$$\langle x|H_0,A_0;E,a\rangle = [2|p|]^{-1/2}\psi_0(x|p), \tag{25}$$

$$\langle x|H,A;E,a\rangle = [4\pi|p|]^{-1/2}f(x|p). \tag{26}$$

In Eqs. (25) and (26) and later

$$E=p^2,\quad a=\mathrm{sgn}p=\pm 1. \tag{27}$$

We shall also find it useful to define $\langle x|H,A;E,a\rangle_{\pm}$ by

$$\langle x|H,A;E,a\rangle_{\pm} = [2|p|]^{-1/1}\psi_{\pm}(x|p). \tag{28}$$

Then $|H_0,A_0;E,a\rangle$ are the eigenkets of $H_0$ having the eigenvalue $E$. The operator $A_0$ is defined to be the operator (which commutes with $H_0$) whose eigenkets are chosen to be $|H_0,A_0;E,a\rangle$ with corresponding eigenvalue $a$. Similarly $|H,A;E,a\rangle_{\pm}$ and $|H,A;E,a\rangle$ are the eigenkets of $H$ and $A$ having the eigenvalues $E$ and $a$.

In Ref. 4 wave operators $U$, $U_{\pm}$ are introduced such that

$$\langle x|H,A;E,a\rangle = \langle x|U|H_0,A_0;E,a\rangle,$$
$$\langle x|H,A;E,a\rangle_\pm = \langle x|U_\pm|H_0,A_0;E,a\rangle. \tag{29}$$

Equations (3) and (7) are equivalent to the following integral equations for the wave operators:

$$U_\pm = I + \int_0^\infty \gamma_\pm(E-H_0)VU_\pm\delta(E-H_0)\,dE,$$
$$U = I + \int_0^\infty [\delta_{A_0,+1}\gamma_-(E-H_0) + \delta_{A_0,-1}\gamma_+(E-H_0)] \times VU\delta(E-H_0)dE, \tag{30}$$

where the operators $\delta(E-H_0)$ and $\delta_{A_0,a}$ are defined by

$$\delta(E-H_0) = \sum_a |H_0,A_0;E,a\rangle\langle H_0,A_0;E,a|, \tag{30a}$$
$$\delta_{A_0,a} = \int_0^\infty |H_0,A_0;E,a\rangle dE\,\langle H_0,A_0;E,a|. \tag{30b}$$

It should be noted that it follows from the definition of $\langle x|H_0,A_0;E,a\rangle$ that

$$\sum_a \int_0^\infty |H_0,A_0;E,a\rangle dE\,\langle H_0,A_0;E,a| = I \tag{31}$$

so that

$$\delta(E-H_0)|H_0,A_0;E',a\rangle = \delta(E-E')|H_0,A_0,E',a\rangle, \tag{32}$$
$$\delta_{A_0,a}|H_0,A_0;E,a'\rangle = \delta_{a',a}|H_0,A_0;E,a'\rangle.$$

In Ref. 4 it is shown that $U$ can be expressed in terms of $U_\pm$ as follows

$$U = U_-M_- = U_+M_+, \tag{33}$$

where $M_-$ and $M_+$ are operators which commute with $H_0$ and are generalizations of the Jost functions $f(k)$ and $f^*(k)$ respectively for the radial equation (see, e.g., Ref. 7). From the fact that there are no bound states

$$U_-U_-^\dagger = U_+U_+^\dagger = I, \tag{34}$$

where the dagger means Hermitian adjoint.

It is our intent to find the weight operator $W$

$$W = M_-^{-1}M_-^{-1\dagger} = M_+^{-1}M_+^{-1\dagger}. \tag{34'}$$

We see that

$$UWU^\dagger = I. \tag{35}$$

In Refs. 4 and 6 it is shown that Eq. (35) is equivalent to the Gel'fand–Levitan equation in one dimension. It is also equivalent to the Gel'fand–Levitan equation in three dimensions, if a triangular operator $K$ can be found such that $U = I + K$. The problem of finding such a $K$ or defining what triangularity means, for that matter, is the subject of Refs. 1–3 and of the present series of papers.

Returning to the one-dimensional problem, Eq. (35) also gives the completeness relations for the eigenfunctions $|H,A;E,a\rangle$, for Eq. (35) can be written

$$\sum_{a,a'}\int_0^\infty dE\,\langle x|H,A;E,a\rangle\langle a|\omega_c(E)|a'\rangle\langle H,A;E,a'|x'\rangle = \delta(x-x'),$$

where $\langle a|\omega_c(E)|a'\rangle$ is defined by

$$\langle H_0,A_0;E,a|W|H_0,A_0;E',a'\rangle = \delta(E-E')\langle a|\omega_c(E)|a'\rangle. \tag{36}$$

From Eq. (34)

$$\langle a|\omega_c(E)|a'\rangle = \sum_{a''}\langle a|\mu_\pm^{-1}(E)|a''\rangle\langle a''|\mu_\pm^{-1\dagger}(E)|a'\rangle, \tag{37}$$

where

$$\langle H_0,A_0;E,a|M_\pm^{-1}|H_0,A_0;E',a'\rangle = \delta(E-E')\langle a|\mu_\pm^{-1}(E)|a'\rangle,$$
$$\langle H_0,A_0;E,a|M_\pm^{-1\dagger}|H_0,A_0;E',a'\rangle = \delta(E-E')\langle a|\mu_\pm^{-1\dagger}(E)|a'\rangle. \tag{38}$$

From the definition of Hermitian adjoint it follows that

$$\langle a|\mu_\pm^{-1\dagger}(E)|a'\rangle = [\langle a'|\mu_\pm^{-1}(E)]|a\rangle^*. \tag{39}$$

From Eq. (1)

$$U = I + 2\pi i\delta_{A_0,-1}\int_0^\infty \delta(E-H_0)VU\delta(E-H_0)\,dE + \int_0^\infty \gamma_-(E-H_0)VU\delta(E-H_0)\,dE. \tag{40}$$

From Eq. (30a), it is readily seen that the solution of Eq. (40) is

$$U = U_-\left[I + 2\pi i\delta_{A_0,-1}\int_0^\infty \delta(E-H_0)VU\delta(E-H_0)dE\right]. \tag{41}$$

On comparing with Eq. (34), we have

$$M_- = [I + 2\pi i\delta_{A_0,-1}\int_0^\infty \delta(E-H_0)VU\delta(E-H_0)dE]$$
$$= I + 2\pi i\delta_{A_0,-1}\int_0^T \delta(E-H_0)VU_-\delta(E-H_0)dE\cdot M_- \tag{42}$$

Thus

$$\left[I - 2\pi i\delta_{A_0,-1}\int_0^\infty \delta(E-H_0)VU_-\delta(E-H_0)dE\right]M_- = 1 \tag{43}$$

and thus

$$M_-^{-1} = I - 2\pi i\delta_{A_0,-1}\int_0^\infty \delta(E-H_0)VU_-\delta(E-H_0)\,dE. \tag{44}$$

From Ref. 4 the scattering operator $S$ is given in terms of $U_-$ by

$$S = I - 2\pi i\int_0^T \delta(E-H_0)VU_-\delta(E-H_0)\,dE. \tag{45}$$

We now have a simple relationship between the operators $M_-^{-1}$ and $S$:

$$M_-^{-1} = \delta_{A_0,-1} S + \delta_{A_0,+1}, \tag{46}$$

from which, on using the fact that the scattering operator is unitary, we have

$$M_-^{-1\dagger} = S^{-1}\delta_{A_0,-1} + \delta_{A_0,+1}. \tag{47}$$

Finally, from Eq. (34)

$$W = I + \delta_{A_0,-1} S \delta_{A_0,+1} + \delta_{A_0,+1} S^\dagger \delta_{A_0,-1}, \tag{48}$$

which is equivalent to

$$\langle a|\omega_c(E)|a'\rangle = \delta_{a,a'} + \delta_{a,-1}\delta_{a',+1}\langle -1|S(E)|+1\rangle + \delta_{a,+1}\delta_{a',-1}[\langle -1|S(E)|+1\rangle]^*, \tag{49}$$

where $\langle a|S(E)|a'\rangle$ is defined by

$$\langle H_0,A_0;E,a|S|H_0,A_0;E',a'\rangle = \delta(E-E')\langle a|S(E)|a'\rangle. \tag{50}$$

Explicitly,

$$\langle a|S(E)|a'\rangle = \delta_{a,a'} - 2\pi i \int_{-\infty}^{+\infty} \langle H_0,A_0;E,a|x\rangle \times V(x)\langle x|H,A;E,a'\rangle_- \, dx \tag{51}$$

The usual reflection coefficient $b(p)$ is defined by

$$b(p) = \langle -1|S(E)|+1\rangle, \quad p = E^{1/2}. \tag{52}$$

Though $b(p)$ is defined only for positive $p$ in Eq. (52), one can extend the definition to negative $p$ by analytic continuation. One obtains the well-known result

$$b(-p) = [b(p)]^*. \tag{53}$$

On using Eqs. (49) and (53) into Eq. (36) and also using Eq. (26), we obtain the now familiar completeness relation

$$(2\pi)^{-1}\left[\int_{-\infty}^{+\infty} f(x|p)f^*(x'|p)dp + \int_{-\infty}^{+\infty} f^*(x|p)f^*(x'|p)b(p)dp\right] = \delta(x-x'). \tag{54}$$

As mentioned above, this completeness relation is equivalent to the Gel'fand–Levitan equation and contains the data needed to reconstruct the potential, namely the reflection coefficient $b(p)$.

Our procedure for obtaining the weight operator $W$ in this paper differs considerably from that which we used in our original work (Ref. 6). In Ref. 6, $W$ was obtained by first finding the functions $\langle a|\mu_-^{-1}(E)|a'\rangle$ by comparing the forms of the wavefunctions $f(x|p)$ and $\psi_-(x|p)$ for large negative $x$. By contrast, in the present paper, the relation of the Green's functions gives us the weight operator. This approach is preferable in three dimensions because one does not want to prejudge the asymptotic forms of the wavefunctions. As it turns out, however, in three dimensions the asymptotic assumptions of Ref. 1 give the same results as the Green's function approach. It is not yet clear to us whether the Jost solutions of the present paper are the same as those of Ref. 1. It seems unlikely, since the triangular properties of the Gel'fand–Levitan equation which come from the present series of papers are a generalization of those of Ref. 1, as will be seen in the third of the papers.

It should be mentioned in Refs. 2 and 3 the weight operator is also obtained by comparing the Green's function for the Jost function with that of the outgoing wave. However, because of the Green's function which they use, they obtain a more complicated result than we do, ours being the natural three-dimensional generalization of the weight operator of the present section.

## 5. NORMALIZATION AND COMPLETENESS RELATIONS FOR THE THREE-DIMENSIONAL JOST SOLUTIONS

It will be convenient to introduce the wavefunctions $\langle \mathbf{x}|H_0,A_0;E,a,\theta,\phi\rangle$ and $\langle \mathbf{x}|H,A;E,a,\theta,\phi\rangle$ defined by

$$\langle \mathbf{x}|H_0,A_0;E,a,\theta,\phi\rangle = E^{1/4}[(\sin\theta)/2]^{1/2}\psi_0(\mathbf{x}|\mathbf{p}), \tag{55}$$

$$\langle \mathbf{x}|H,A;E,a,\theta,\phi\rangle = (2\pi)^{-3/2}E^{1/4}[(\sin\theta)/2]^{1/2}f(\mathbf{x}|\mathbf{p}).$$

In Eq. (55), $a$ is the sign of the optical radius of $p$ ($=\pm 1$) and $\theta$, $\phi$ are the angular variables of the optical coordinates. Also $E = |\mathbf{p}|^2$ and $A_0,A$ denote collectively the operators whose eigenvalues are $a,\theta,\phi$. We note the completeness relationship

$$\sum_a \int_0^\infty dE \int_0^{\pi/2} d\theta \int_0^{2\pi} d\phi \times |H_0,A_0;E,a,\theta,\phi\rangle\langle H_0,A_0;E,a,\theta,\phi| = I. \tag{56}$$

As for the one-dimensional equation we can introduce an operator $U$ such that

$$\langle \mathbf{x}|H,A;a,\theta,\phi\rangle = \langle x|U|H_0,A_0;E,a,\theta,\phi\rangle \tag{57}$$

and the operators $U_\pm$ which can be used to construct the outgoing and incoming eigenfunctions of $H$. The procedure for finding the weight operator $W$ is identical to that for the one-dimensional problem of the previous section; all equations for the operators are identical. In particular, Eq. (48) is valid where the subscripts on the Kronecker $\delta$ refer only to the variable $a$. Thus writing

$$\langle H_0,A_0;E,a,\theta,\phi|W|H_0,A_0;E',\theta',\phi'\rangle = \delta(E-E')\langle a,\theta,\phi|\omega_c(E)|a',\theta',\phi'\rangle$$

and

$$\langle H_0,A_0;E,a,\theta,\phi|S|H_0,A_0;E',\theta',\phi'\rangle = \delta(E-E')\langle a,\theta,\phi|S(E)|a',\theta',\phi'\rangle, \tag{58}$$

using the fact that $S$ is unitary and the relation (proved in Ref. 1 with a slightly different notation)

$$\langle a,\theta,\phi|S(E)|a',\theta',\phi'\rangle = \langle -a',\theta',\phi'|S(E)|-a,\theta,\phi\rangle, \tag{59}$$

we have

$$\langle a,\theta,\phi|\omega_c(E)|a',\theta',\phi'\rangle$$

$$= \delta_{a,a'}(\theta - \theta')\delta(\phi - \phi') + \delta_{a,-1}\delta_{a',+1} \times \langle -1,\theta,\phi\,|S(E)|+1,\theta',\phi'\rangle + \delta_{a,+1}\delta_{a',-1}[\langle -1,\theta,\phi\,|S(E)|+1,\theta',\phi'\rangle]^*. \quad (60)$$

The completeness relation for the Jost solutions or equivalently the functions $\langle \mathbf{x}|H,A;a,\theta,\phi\rangle$ are obviously

$$\int_0^\infty dE \sum_{a,a'} \int_0^{\pi/2} d\theta \times \int_0^{2\pi} d\phi \langle \mathbf{x}|H,A;E,a,\theta,\phi\rangle \langle H,A;E,a,\theta,\phi\,|\mathbf{x}'\rangle$$
$$+ \int_0^\infty dE \int_0^{\pi/2} d\theta \int_0^{\pi/2} d\theta' \int_0^{2\pi} d\phi \times \int_0^{2\pi} d\phi' \langle \mathbf{x}|H,A;E,-1,\theta,\phi\rangle \times \langle -1,\theta,\phi\,|S(E)|+1,\theta',\phi'\rangle \langle H,A;E,+1,\theta',\phi'|\mathbf{x}'\rangle$$
$$+ \int_0^\infty dE \int_0^{\pi/2} d\theta \int_0^{\pi/2} d\theta' \int_0^{2\pi} d\phi \times \int_0^{2\pi} d\phi' \langle \mathbf{x}|H,A;E,+1,\theta,\phi\rangle \times [\langle -1,\theta,\phi\,|S(E)|+1,\theta',\phi'\rangle]^* \langle H,A;E,-1,\theta',\phi'|\mathbf{x}'\rangle = \delta(\mathbf{x}-\mathbf{x}'). \quad (61)$$

We can make this completeness relation take on a closer resemblance to the completeness relation for the one-dimensional Jost solutions [Eq. (54)] by introducing another form for the Jost solutions. Let us define $f(\mathbf{x}|p,\theta,\phi)$ by

$$f(\mathbf{x}|p,\theta,\phi) = p(\sin\theta)^{1/2} f(\mathbf{x}|\mathbf{p}) = (2\pi)^{3/2}(2)^{1/2}(E)^{1/4}\langle \mathbf{x}|H,A;E,a,\theta,\phi\rangle, \quad (62)$$

where $p,\theta,\phi$ are the optical coordinates of $\mathbf{p}$, $a = \operatorname{sgn} p$, $E = p^2$ as before. Also for positive $p$ let us define $b_p(\theta,\phi\,|\theta',\phi')$ by

$$b_p(\theta,\phi\,|\theta',\phi') = \langle -1,\theta,\phi\,|S(E)|+1,\theta',\phi'\rangle. \quad (63)$$

We can define $b_p(\theta,\phi\,|\theta',\phi')$ for negative $p$ by analytic continuation. In Ref. 1 it is shown that

$$b_{-p}(\theta,\phi\,|\theta',\phi') = [b_p(\theta,\phi\,|\theta',\phi')]^*. \quad (64)$$

Furthermore, from $f(\mathbf{x}|-\mathbf{p}) = [f(\mathbf{x}|\mathbf{p})]^*$,

$$f(\mathbf{x}|-p,\theta,\phi) = -[f(\mathbf{x}|p,\theta,\phi)]^*. \quad (65)$$

The completeness relation (61) takes the simpler form

$$\int_{-\infty}^{+\infty} dp \int_0^{\pi/2} d\theta \int_0^{2\pi} d\phi\, f(\mathbf{x}|p,\theta,\phi) f^*(\mathbf{x}'|p,\theta,\phi) + \int_{-\infty}^{+\infty} dp \int_0^{\pi/2} d\theta \int_0^{\pi/2} d\theta' \int_0^{2\pi} d\phi \int_0^{2\pi} d\phi' \times f^*(\mathbf{x}|p,\theta,\phi) b_p(\theta,\phi\,|\theta',\phi') f^*(\mathbf{x}'|,p,\theta',\phi') = \delta(\mathbf{x}-\mathbf{x}'). \quad (66)$$

Except for minor notational changes this completeness relation is identical to that for the corresponding eigenfunctions of $H$ in Ref. 1, where it was found without the use of a Green's function, but an assumption was made about the triangularity properties of the Gel'fand–Levitan kernel or, equivalently, about the asymptotic forms for the eigenfunctions.

To see the close analogy with the one-dimensional theory, one should compare Eq. (66) with its one-dimensional counterpart, Eq. (54).

The function $b_p(\theta,\phi\,|\theta',\phi')$ is simply related to the amplitude of the spherical wave which, for large $|\mathbf{x}|$ in $\psi_-(\mathbf{x}|\mathbf{p})$ of Eq. (18), describes the scattering. It is readily shown (see, e.g., Ref. 1)

$$\lim_{r\to-\infty} \psi_-(\mathbf{x}|\mathbf{p}) = \psi_0(\mathbf{x}|\mathbf{p}) - [2\pi \sin\lambda \sin\theta]^{-1/2} \frac{ie^{ip|r|}}{p|r|} b_p(\lambda,\sigma|\theta,\phi), \quad (67)$$

where $r,\lambda,\sigma$ are the optical coordinates of $\mathbf{x}$ and $p$ [taken $>0$ in Eq. (67)], $\theta$, $\phi$ are those of $\mathbf{p}$.

As mentioned earlier, the completeness relation (54) for the one-dimensional problem is equivalent to the Gel'fand–Levitan equation and contains the minimal data from which the scattering potential can be reconstructed, namely the reflection coefficient. It seems likely that the completeness relation (66) for three dimensions is equivalent to the Gel'fand–Levitan equation for an appropriate triangularity condition on the Gel'fand–Levitan kernel. However, the information that it contains is redundant in the following sense. It was shown in Ref. 5 that the potential can be reconstructed knowing $b_p(\theta,\phi\,|\theta,\phi)$, i.e., the reflection coefficient back along the same ray along which the incident wave was propagating, all directions of propagation being confined to a hemisphere. By contrast, the completeness relation Eq. (66) contains more data, namely $b_p(\theta,\phi\,|\theta',\phi')$. In Ref. 5 one can see that, in principle at least $b_p(\theta,\phi\,|\theta',\phi')$ can be found from $b_p(\theta,\phi\,|\theta,\phi)$ by first finding the potential $V(\mathbf{x})$, then finding the eigenfunction $\psi_-(\mathbf{x}|\mathbf{p})$, and, finally, using the asymptotic relation (67). The relation between $b_p(\theta,\phi\,|\theta',\phi')$ and $b_p(\theta,\phi\,|\theta,\phi)$ represents a constraint which may have to be used together with the Gel'fand–Levitan equation to provide an inverse method for the potential which depends on the minimal data. Since the method of Ref. 5 depends heavily on the fact that the potential $V(\mathbf{x})$ is diagonal in the $\mathbf{x}$ representation, that is, it is a multiplicative operator in that representation, the constraint is a consequence of this fact. It might be noted that we have already used the fact that the potential is a real operator in the derivation of the completeness relation (66) because Eq. (59) is a consequence of this fact. However, it is clear that we have not exhausted the relations that can be obtained from the fact that the potential is multiplicative. The constraint will be one of these relations.


[1] I. Kay and H.E. Moses, Nuovo Cimento **22**, 689 (1961). Also Commun. Pure Appl. Math. **14**, 435 (1961).
[2] R.G. Newton, "The Three-Dimensional Inverse Scattering Problem in Quantum Mechanics," Invited Lectures delivered at the 1974 Summer Seminar on Inverse Problems, Am. Math. Soc., August 5–16, 1974, UCLA.
[3] L.D. Faddeev, J. Sov. Math. **5**, 334 (1976).
[4] I. Kay and H.E. Moses, Nuovo Cimento **2**, 917 (1955).
[5] H.E. Moses, Phys. Rev. **102**, 559 (1956).
[6] I. Kay and H.E. Moses, Nuovo Cimento **3**, 276 (1956).
[7] R.G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), Chap. 12.

# The spectral properties of many-electron atomic Hamiltonians and the method of configuration interaction. III. Compactness proof associated with an infinite system of linear equations for $n$-electron atoms

M. H. Choudhury

*Department of Applied Mathematics, The University, Hull, England*
(Received 21 June 1977; revised manuscript received 28 December 1977)

An infinite system of linear equations is derived from the Schrödinger equation of an $n$-electron atomic system ($n \geq 3$). The linear operator defined by this system of equations is then shown to be compact in a region of the complex energy plane which excludes the various bound state and multiparticle scattering cuts (i.e., the essential spectrum of the Hamiltonian of the $n$-electron atomic system). It is further shown that the method can be used to deal with the case of the diatomic molecule. The above result then permits one, both in the case of the $n$-electron atomic system and the diatomic molecule, to truncate the infinite system of equations in question with the assurance that as the size of the truncated system is increased, the energy eigenvalues computed from the truncated system will uniformly converge to those of the original infinite system.

## 1. INTRODUCTION

In the final paper of this series, we derive from the Schrödinger equation of an $n$-electron atomic system an infinite system of linear equations which when truncated yields a system of $N \times N$ linear equations whose eigenvalues uniformly approximate the lowest $N$ eigenvalues of the original infinite system. We have already shown in the first of the present series of papers (hereafter referred to as I) that a naive generalization of the method adopted for a two-electron atomic system will not work. As mentioned in the previous paper (hereafter referred to as II), the idea is to obtain an infinite system of linear equations which define a compact linear operator in a suitable region of the complex energy plane. To accomplish this, we adopt a method whose underlying idea is that due to Weinberg,[1] and was applied by him to deal with multiparticle scattering processes. We shall not, however, use the diagrammatic language or techniques employed in that paper to derive our equations.

The proof of compactness that follows exhibits in a clear and detailed fashion the spectral properties of the Hamiltonian of the $n$-electron atomic system. The spectral structure of an $n$-electron Hamiltonian is a complex one. The reader will therefore appreciate that any analytical method (like the one we have used) which picks out and exhibits explicitly the various multiparticle cuts and their branch points in great detail will inevitably be an involved and lengthy study.

In Sec. 2 we derive from the Schrödinger equation of an $n$-electron atomic system ($n \geqslant 3$) an infinite system of linear equations. In Sec. 3 we show that the linear operator defined by this system of equations is compact in a region of the complex energy plane which excludes the various multiparticle cuts. For $n = 3$, we find that these multiparticle cuts which constitute the essential spectrum of the Hamiltonian are

(i) the bound state scattering cuts starting at the two-electron bound state energies (with the same nuclear charge) and extending to $+\infty$,

(ii) the bound state scattering cuts starting at the hydrogenic bound state energies and extending to $+\infty$,

(iii) the three-electron scattering cut starting at $E = 0$ and extending to $+\infty$.

Finally, in Sec. 4, we show how the method we have employed can be adopted to the situation of the diatomic molecule, that is, in studying the spectral properties of its Hamiltonian as well as to computing its energy eigenvalues and eigenfunctions. It is to be noted that, in this case, the Born–Oppenheimer separation will not be necessary. We end the section by summarizing our results and conclusions.

## 2. REDUCTION OF THE SCHRÖDINGER EQUATION FOR AN $n$-ELECTRON ATOMIC SYSTEM TO AN INFINITE SYSTEM OF LINEAR EQUATIONS

The Schrödinger equation for an $n$-electron atomic system can be written

$$H|\Psi\rangle = \Big[\sum_{i=1}^{n} H_{0i} + \sum_{\substack{i,j=1\\ i<j}}^{n} \frac{1}{r_{ij}}\Big]|\Psi\rangle = E\,|\Psi\rangle, \quad (2.1)$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$, $\mathbf{r}_i$ and $\mathbf{r}_j$ being the position operators of the $i$th and $j$th electrons respectively and

$$H_{0i} = -\frac{1}{2}\nabla_i^2 - Z/r_i, \quad i = 1,2,\ldots,n. \quad (2.2)$$

Let $\mathscr{H}_i$, $i = 1,2,\ldots,n$ be the space of states associated with the $i$th electron. The resolution of the identity in these spaces are given by

$$\sum_{\mathbf{n}_i} |\mathbf{n}_i\rangle\langle\mathbf{n}_i| + \Big(\int\!\!\sum_{\mathbf{k}_{Hi}}\Big)|\mathbf{k}_{Hi}\rangle\langle\mathbf{k}_{Hi}| = I_{Hi},$$
$$i = 1,2,\ldots,n, \quad (2.3)$$

 0022-2488/79/061157-13$01.00  

where $|\mathbf{n}_i\rangle$ and $|\mathbf{k}_{Hi}\rangle$ are the bound and continuum states associated with the hydrogenic Hamiltonian $H_{0i}$. The resolution of the identity in the $n$-electron product space $\mathscr{H} = \mathscr{H}_1 \otimes \mathscr{H}_2 \otimes \cdots \otimes \mathscr{H}_n$ is

$$I_H = I_{H1} \otimes I_{H2} \otimes \cdots \otimes I_{Hn} = \bigotimes_{i=1}^{n} I_{Hi}. \tag{2.4}$$

In particular, using (2.3), the resolution of the identity when $n = 3$ is

$$I_H = \sum_{\mathbf{n}_i} |\,\mathbf{n}_1,\mathbf{n}_2,\mathbf{n}_3\rangle\langle\mathbf{n}_1,\mathbf{n}_2,\mathbf{n}_3\,|$$

$$+ \sum_{\substack{\mathbf{n}i \\ i=1,2}} \int \sum_{\mathbf{k}_{H3}} |\mathbf{n}_1,\mathbf{n}_2,\mathbf{k}_{H3}\rangle\langle\mathbf{n}_1,\mathbf{n}_2,\mathbf{k}_{H3}|$$

+ (two similar terms)

$$+ \sum_{\mathbf{n}_1}\Big(\int \sum_{\substack{\mathbf{k}_{Hi} \\ i=2,3}}\Big) |\mathbf{n}_1,\mathbf{k}_{H2},\mathbf{k}_{H3}\rangle\langle\mathbf{n}_1,\mathbf{k}_{H2},\mathbf{k}_{H3}|$$

+ (two similar terms)

$$+ \Big(\int \sum_{\mathbf{k}_{Hi}}\Big) |\mathbf{k}_{H1},\mathbf{k}_{H2},\mathbf{k}_{H3}\rangle\langle\mathbf{k}_{H1},\mathbf{k}_{H2},\mathbf{k}_{H3}\,|, \tag{2.5}$$

which can be written in the abridged form

$$\Big(\int \sum_{\mathbf{v}_i} |\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\rangle\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3| = I_H\Big). \tag{2.5'}$$

If we use a complete discrete basis $\{|\boldsymbol{\alpha}_i\rangle\}$ in $\mathscr{H}_i$, then the resolution of the identity $I_H$ is given by

$$I_H = \sum_{\boldsymbol{\alpha}_i} |\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3\rangle\langle\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3|. \tag{2.6}$$

Equation (2.1) for $n = 3$ can be formally inverted to obtain the equations

$$|\Psi\rangle = G_i(E)V_i|\Psi\rangle, \quad i = 1,2,3, \tag{2.7}$$

where

$$G_i(E) = \Big(E - \sum_{s=1}^{3} H_{0s} - \frac{1}{r_{jk}}\Big)^{-1}, \quad i \neq j \neq k \tag{2.8}$$

and

$$V_i = \frac{1}{r_{ij}} + \frac{1}{r_{ik}}. \tag{2.9}$$

Beginning with the third of the equations (2.7) (i.e., when $i = 3$) and iterating with respect to the other two in succession, one obtains

$$|\Psi\rangle = G_1V_1G_2V_2G_3V_3|\Psi\rangle. \tag{2.10}$$

Taking the inner product with respect to $\langle\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3|$ and using (2.6), we obtain the infinite system of equations

$$\langle\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3|\Psi\rangle = \sum_{\boldsymbol{\alpha}_i'} \langle\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3|G_1V_1G_2V_2G_3V_3|\boldsymbol{\alpha}_1',\boldsymbol{\alpha}_2',\boldsymbol{\alpha}_3'\rangle \times\langle\boldsymbol{\alpha}_1',\boldsymbol{\alpha}_2',\boldsymbol{\alpha}_3'|\Psi\rangle. \tag{2.11}$$

It is easy to generalize equations (2.11) for the general case of $n$ electrons. Exactly the same procedure as that adopted in the three-electron case now yields

$$|\Psi\rangle = G_1V_1G_2V_2\cdots G_n V_n|\Psi\rangle, \tag{2.12}$$

where

$$G_i(E) = \Big(E - \sum_{s=1}^{n} H_{0s} - \sum_{\substack{j<k \\ j,k\neq i}}^{n} \frac{1}{r_{jk}}\Big)^{-1}, \quad i = 1,2,\ldots,n \tag{2.13}$$

and

$$V_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{1}{r_{ij}}, \quad i = 1,2,\ldots,n. \tag{2.14}$$

Taking the inner product with respect to $\langle\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_n\,|$ and using

$$\sum_{\boldsymbol{\alpha}_i} |\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n\rangle\langle\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n| = I_H, \tag{2.15}$$

we obtain from Eq. (2.12) the infinite system of equations

$$\langle\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n|\Psi\rangle = \sum_{\boldsymbol{\alpha}_i'} \langle\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n|G_1V_1G_2V_2\cdots G_nV_n|\boldsymbol{\alpha}_1',\ldots,\boldsymbol{\alpha}_n'\rangle \times\langle\boldsymbol{\alpha}_1',\ldots,\boldsymbol{\alpha}_n'|\Psi\rangle. \tag{2.16}$$

It will become evident that the proof of compactness for this general case is similar to that when $n = 3$. It will therefore suffice if we deal with the proof for compactness in the case of three electrons. Reverting back to the three-electron situation again, consider the equation

$$H_{jk}^{(2)}\,|\Psi\rangle = \Big(H_{0j} + H_{0k} + \frac{1}{r_{jk}}\,|\Psi\rangle = E\,|\Psi\rangle\Big),$$

$$j \neq k, \quad j,k = 1,2,3. \tag{2.17}$$

Let $\sigma(H_{jk}^{(2)})$ denote the spectrum of $H_{jk}^{(2)}$. This spectrum consists of the bound state poles and the various multiparticle scattering cuts associated with the resolvent $(E - H_{jk}^{(2)})^{-1}$. Let $\mathscr{H}_{jk}$ denote the space of states associated with the $(jk)$ subsystem. Denoting the bound states and the various continuum states of $\mathscr{H}_{jk}$ by $|\mathbf{v}_{jk}\rangle$, where the set of indices $\{\mathbf{v}_{jk}\}$ assume discrete or continuous values as appropriate, we have

$$H_{jk}^{(2)}\,|\mathbf{v}_{jk}\rangle = E_{\nu_{jk}}\,|\mathbf{v}_{jk}\rangle. \tag{2.18}$$

Here $E_{\nu_{jk}}$ are the energies corresponding to the states $|\mathbf{v}_{jk}\rangle$. We define

$$H_{jk} = \sum_{s=1}^{3} H_{0s} + \frac{1}{r_{jk}}. \tag{2.19}$$

Its resolvent is

$$G_i(E) = \Big(E - \sum_{s=1}^{3} H_{0s} - \frac{1}{r_{jk}}\Big)^{-1}. \tag{2.20}$$

Consider the product space $\mathscr{H} = \mathscr{H}_{jk} \otimes \mathscr{H}_i$. The resolution of the identity in $\mathscr{H}$ can also be written

$$\left(\int \sum_{\mathbf{v}_{jk}} |\mathbf{v}_{jk},\mathbf{v}_i\rangle\langle\mathbf{v}_{jk},\mathbf{v}_i| = I_H\right). \tag{2.21}$$

The spectrum of $H_{jk}$ consists of the set of points making up

(i) the poles at $E_{n_{jk}} + E_{n_i}$, where the $E_{n_{jk}}$ are the bound state energies of two-electron atoms as obtained from (2.18) and the $E_{n_i}$ are the hydrogenic bound state energies,

(ii) the bound state scattering cuts starting at $E_{n_{jk}}$ and extending to $+\infty$,

(iii) the bound state scattering cuts starting at $E_{n_i}$ and extending to $+\infty$,

(iv) the three-electron scattering cut starting at $E = 0$ and extending to $+\infty$.

Denoting this spectrum by $\sigma(H_{jk})$, we note that $G_i(E)$ is a bounded operator for $E \notin \sigma(H_{jk})$.

Finally, we mention that the basic idea behind the derivation of (2.10) or (2.12) is due to Weinberg.[1] It has been shown by Hunziger[2] that Weinberg's formulation works for a very large class of potentials which includes the Coulomb potential. It is important to note, however, that Weinberg's operator kernel $I(E)$,[3] while formally being similar to the operator $G_1V_1G_2V_2G_3V_3$, has important and significant differences. Our resolvent operators $G_i(E)$ are somewhat different. Our choice of $G_i(E)$ is basically motivated by the fact that the two-particle Schrödinger equation with a Coulomb type of potential is exactly solvable. This fact allows us to define our free resolvent as

$$G_0(E) = \left(E - \sum_{s=1}^{3} H_{0s}\right)^{-1},$$

while Weinberg's free resolvent is $G_0(E) = (E - H_0)^{-1}$, where $H_0$ is the total kinetic energy operator of the particles.

## 3. PROOF OF COMPACTNESS

In this section it will be shown that the operator

$$K(E) = G_1(E)V_1G_2(E)V_2G_3(E)V_3 \tag{3.1}$$

occurring in (2.10) is compact in a region $D_E$ (to be specified later) of the complex energy plane. Let us first note that the domain of

$$V_i = \frac{1}{r_{ij}} + \frac{1}{r_{ik}} \tag{3.2}$$

is a subset $D_V$ of $\mathbb{R}^3\times\mathbb{R}^3\times\mathbb{R}^3 = \mathbb{R}^9$ defined by

$$D_V = \{(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3) : (\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)\in\mathbb{R}^9, |\mathbf{r}_i - \mathbf{r}_j| \geqslant \epsilon, i,j,k = 1,2,3\}. \tag{3.3}$$

We have

$$V_{ij} = \frac{1}{r_{ij}} = \sum_{k=0}^{\infty} F_k(r_i,r_j)P_k(\cos\theta_{ij}), \tag{3.4}$$

where $\theta_{ij}$ is the angle between $\mathbf{r}_i$ and $\mathbf{r}_j$ and

$$F_k(r_i,r_j) = \frac{1}{r_i}\left(\frac{r_j}{r_i}\right)^k \theta(r_i - r_j) + \frac{1}{r_j}\left(\frac{r_i}{r_j}\right)^k \theta(r_j - r_i), \tag{3.5}$$

where $\theta(r)$ is the step function. We also require to define the sequence of potentials

$$V_{ij}^{(n)} = \sum_{k=0}^{n} F_k(r_i,r_j)P_k(\cos\theta_{ij}), \tag{3.6}$$

using which we define the potentials

$$V_i^{(n)} = V_{ij}^{(n)} + V_{ik}^{(n)}. \tag{3.7}$$

The compactness of $K(E)$ will be demonstrated by showing that

(i) the sequence of operators

$$K_n(E) = G_1(E)V_1^{(n)}G_2(E)V_2^{(n)}G_3(E)V_3^{(n)} \tag{3.8}$$

are compact whenever $E \in D_E$ and

$$\text{(ii)}\quad \|K - K_n\| \to 0 \quad \text{as } n\to\infty \tag{3.9}$$

in the uniform topology of the operator norm.

From the second resolvent equation

$$G_i(E) = G_{ij}^{(0)}(E) + G_{ij}^{(0)}(E)[H_{0k} + V_{jk}]G_i(E),$$

where

$$G_{ij}^{(0)}(E) = (E - H_{0i} - H_{0j})^{-1}, \tag{3.10}$$

one obtains

$$G_i(E) = [I - G_{ij}^{(0)}(E)(H_{0k} + V_{jk})]^{-1} G_{ij}^{(0)}(E). \tag{3.11}$$

We shall also require the second resolvent equation

$$G_k(E) = G_{ij}^{(0)}(E) + G_{ij}^{(0)}(E)[H_{0k} + V_{ij}]G_k(E). \tag{3.12}$$

Denote by $\sigma_0$ the set of points which make up

(i) the branch cut starting at $E_{n_i}$ and extending to $+\infty$,

(ii) the points $\{E_{n_i} + E_{n_j}\}$.

It is clear that $G_{ij}^{(0)}(E)$ is bounded for $E \notin \sigma_0$. Also $G_i(E)$ is bounded for $E \notin \sigma(H_{jk})$. Hence, from (3.11) we see that

$$[I - G_{ij}^{(0)}(E)(H_{0k} + V_{jk})]^{-1}$$

is bounded for $E \notin \sigma(H_{jk}) \cup \sigma_0$. We denote this region by $D_E$:

$$D_E = \{E : E \notin \sigma(H_{jk}) \cup \sigma_0\}. \tag{3.13}$$

Let us write (3.8) in the form

$$K_n(E) = [I - G_{12}^{(0)}(H_{03} + V_{23})]^{-1} G_{12}^{(0)} V_1^{(n)} G_2 V_2^{(n)} G_3 V_3^{(n)}.$$

Noting that $[I - G_{12}^{(0)}(H_{03} + V_{23})]^{-1}$ is bounded for $E \in D_E$ and using the fact that the product of a bounded operator and a compact operator is compact,[4] it is sufficient to show that the operator

$$G_{12}^{(0)}(E)V_1^{(n)}G_2(E)V_2^{(n)}G_3(E)V_3^{(n)} \tag{3.14}$$

is compact for $E \in D_E$. Substituting for $V_i^{(n)}$, $i = 1, 2, 3$, from (3.7) into (3.14), one obtains a sum of eight terms. There are two types of operators involved in this sum, namely

$$G_{12}^{(0)}V_{12}^{(n)}G_2V_{23}^{(n)}G_3V_{31}^{(n)} \tag{3.15}$$

and

$$G_{12}^{(0)}V_{12}^{(n)}G_2V_{23}^{(n)}G_3V_{23}^{(n)}. \tag{3.16}$$

It will therefore be sufficient if we prove that each of these operators are of the Hilberg–Schmidt type and therefore compact for $E \in D_E$, that is,

$$\left(\int\sum_{\mathbf{v}_i}\right)\left(\int\sum_{\mathbf{v}_i'}\right) |\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\, G_{12}^{(0)}V_{12}^{(n)}G_2V_{23}^{(n)}G_3V_{31}^{(n)}$$

$$|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|^2<\infty,\quad E\in D_E, \tag{3.17}$$

and

$$\Big(\int\sum_{\mathbf{v}_i}\Big)\Big(\int\sum_{\mathbf{v}_i'}\Big)|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_2V_{23}^{(n)}G_3V_{31}^{(n)}$$
$$|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|^2<\infty,\quad E\in D_E, \tag{3.18}$$

Considering (3.15) first and using (2.28), we find that

$$|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_2V_{23}^{(n)}G_3V_{31}^{(n)}\,|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|$$
$$=\Big|\Big(\int\sum_{\mathbf{v}_{12}''}\Big)\Big(\int\sum_{\mathbf{v}_3''}\Big)\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_2V_{23}^{(n)}\,|\mathbf{v}_{12}'',\mathbf{v}_3''\rangle$$
$$\times(E-E_{v_{12}''}-E_{v_3''})^{-1}\langle\mathbf{v}_{12}'',\mathbf{v}_3''|V_{31}^{(n)}|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|. \tag{3.19}$$

Using the result from Paper II, Appendix A,

$$\Big|\Big(\int\sum_{\mathbf{v}_{jk}}\Big)\Big(\int\sum_{\mathbf{v}_i}\Big)\langle\Phi_1|\mathbf{v}_{jk},\mathbf{v}_i\rangle\,(E-E_{v_{jk}}-E_{v_i})^{-1}\langle\mathbf{v}_{jk},\mathbf{v}_i|\Phi_2\rangle\Big|$$
$$\leqslant\text{const}\,|\langle\Phi_1|\Phi_2\rangle|,\quad E\notin\sigma(H_{jk}), \tag{3.20}$$

we see that (3.19) becomes

$$|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_2V_{23}^{(n)}G_3V_{31}^{(n)}\,|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|$$
$$\leqslant\text{const}\,|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_2V_{23}^{(n)}V_{31}^{(n)}\,|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|$$
$$\leqslant\text{const}\,[\,|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_{31}^{(0)}V_{23}^{(n)}V_{31}^{(n)}\,|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|$$
$$+|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_{31}^{(0)}H_{02}G_2V_{23}^{(n)}V_{31}^{(n)}\,|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|$$
$$+|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_{31}^{(0)}V_{31}G_2V_{23}^{(n)}V_{31}^{(n)}\,|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|\,],$$
$$E\in D_E, \tag{3.21}$$

where we have used (3.12) to obtain the last step. Taking the second term on the right-hand side of (3.21) and using (2.28) and (3.20) in succession, we obtain

$$|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_{31}^{(0)}H_{02}G_2V_{23}^{(n)}V_{31}^{(n)}\,|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|$$
$$=\Big|\Big(\int\sum_{\mathbf{v}_{31}''}\Big)\Big(\int\sum_{\mathbf{v}_2''}\Big)\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_{31}^{(0)}\,|\mathbf{v}_{31}'',\mathbf{v}_2''\rangle$$
$$\times E_{v_2''}(E-E_{v_{31}''}-E_{v_2''})^{-1}\langle\mathbf{v}_{31}'',\mathbf{v}_2''|V_{23}^{(n)}V_{31}^{(n)}|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|$$
$$\leqslant\text{const}\,|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_{31}^{(0)}V_{23}^{(n)}V_{31}^{(n)}\,|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|,$$
$$E\in D_E,$$

where, in the last step we have used the fact that $E_{v_2''}(E-E_{v_{31}''}-E_{v_2''})^{-1}$ is bounded for $E\in D_E$. Similarly, observing from (3.3), (3.4), and (3.5) that $V_{31}$ is a bounded function in its domain of definition $D_V$, one finds that the third term on the right-hand side of (3.21) yields the inequality

$$|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_{31}^{(0)}V_{31}G_2V_{23}^{(n)}V_{31}^{(n)}\,|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|$$
$$\leqslant\text{const}\,|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_{31}^{(0)}V_{23}^{(n)}V_{31}^{(n)}\,|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|.$$

Hence, using these inequalities [for the second and third term of (3.21)], we find that (3.21) becomes

$$|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_2V_{23}^{(n)}G_3V_{31}^{(n)}\,|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|$$
$$\leqslant\text{const}\,|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_{31}^{(0)}V_{23}^{(n)}V_{31}^{(n)}\,|\mathbf{v}_1'\mathbf{v}_2'\mathbf{v}_3'\rangle|,$$
$$E\in D_E. \tag{3.22}$$

Noting that $[1-E_{v_i}/(E-E_{v_i})]^{-1}$ is a bounded function for $E\in D_E$, we have the result

$$\Big|\Big(\int\sum_{\mathbf{v}_i}\Big)\langle\Phi_1|\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\rangle\,[1-E_{v_i}/(E-E_{v_i})]^{-1}$$
$$\times\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,\Phi_2\rangle\Big|$$
$$\leqslant\text{const}\,|\langle\Phi_1|\Phi_2\rangle|\quad\text{for } E\in D_E$$

Using this result, we have

$$|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,G_{12}^{(0)}V_{12}^{(n)}G_{31}^{(0)}V_{23}^{(n)}V_{31}^{(n)}\,|\mathbf{v}_1',\mathbf{v}_2',\mathbf{v}_3'\rangle|$$
$$=\Big|\Big(\int\sum_{\substack{\mathbf{v}_i''\\ i=1,2}}(E-E_{v_1})-E_{v_2})^{-1}$$
$$\times\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,V_{12}^{(n)}\,|\mathbf{v}_1'',\mathbf{v}_2'',\mathbf{v}_3\rangle$$
$$\times(E-E_{v_3})^{-1}\,[1-E_{v_i''}/(E-E_{v_3})]^{-1}$$
$$\times\langle\mathbf{v}_1'',\mathbf{v}_2'',\mathbf{v}_3|\,V_{23}^{(n)}V_{31}^{(n)}\,|\mathbf{v}_1',\mathbf{v}_2',\mathbf{v}_3'\rangle|$$
$$\leqslant\text{const}\,|(E-E_{v_1}-E_{v_2})^{-1}|\,|(E-E_{v_3})^{-1}|$$
$$\times|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,V_{12}^{(n)}V_{23}^{(n)}V_{31}^{(n)}\,|\mathbf{v}_1',\mathbf{v}_2',\mathbf{v}_3'\rangle|, \tag{3.23}$$

Hence, in order that (3.15) be compact and of the Hilbert-Schmidt type for $E\in D_E$, we see from (3.17) and (3.23) that it is sufficient to have

$$\Big(\int\sum_{\mathbf{v}_i}\Big)\Big(\int\sum_{\mathbf{v}_i'}\Big)|(E-E_{v_1}-E_{v_2})^{-1}|^2\,|(E-E_{v_3})^{-1}|^2$$
$$\times|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,V_{12}^{(n)}V_{23}^{(n)}V_{31}^{(n)}\,|\mathbf{v}_1',\mathbf{v}_2',\mathbf{v}_3'\rangle|^2<\infty,$$
$$E\in D_E.$$

Summing over the $v_i'$, we find that this compactness condition becomes

$$\Big(\int\sum_{\mathbf{v}_i}\Big)|(E-E_{v_1}-E_{v_2})^{-1}|^2\,|(E-E_{v_3})^{-1}|^2$$
$$\times|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,[V_{12}^{(n)}V_{23}^{(n)}V_{31}^{(n)}]^2\,|\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\rangle<\infty,$$
$$E\in D_E. \tag{3.24}$$

Similarly, one finds that the condition for the operator (3.16) to be compact and of Hilbert–Schmidt type can be written as

$$\Big(\int\sum_{\mathbf{v}_i}\Big)|(E-E_{v_1}-E_{v_2})^{-1}|^2\,|(E-E_{v_3})^{-1}|^2$$
$$\times|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,[V_{12}^{(n)}]^2\,[V_{23}^{(n)}]^4\,|\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\rangle|<\infty,$$
$$E\in D_E. \tag{3.25}$$

Defining

$$U_n^{(\alpha)}=\begin{cases}[V_{12}^{(n)}V_{23}^{(n)}V_{31}^{(n)}]^2, & \alpha=1,\\[2ex] [V_{12}^{(n)}]^2\,[V_{23}^{(n)}]^4, & \alpha=2,\end{cases} \tag{3.26}$$

we can write the compactness conditions (3.24), (3.25) concisely in the form

$$\Big(\int\sum_{\mathbf{v}_i}\Big)|(E-E_{v_1}-E_{v_2})^{-1}|^2\,|(E-E_{v_3})^{-1}|^2$$
$$\times|\langle\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3|\,U_n^{(\alpha)}\,|\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\rangle|<\infty,\quad E\in D_E. \tag{3.27}$$

Noting (2.8), we see that (3.27) can be written explicitly in the form

$$\sum_{\mathbf{n}_i} |E-E_{n_1}-E_{n_2}|^{-2}\,|E-E_{n_3}|^{-2}\,\langle \mathbf{n}_1,\mathbf{n}_2,\mathbf{n}_3|\;U_n^{(\alpha)}\;|\mathbf{n}_1,\mathbf{n}_2,\mathbf{n}_3\rangle$$

$$+\sum_{\substack{\mathbf{n}_i\\ i=1,2}}\Big(\int\!\!\sum_{\mathbf{k}_{H3}}\Big)|E-E_{n_1}-E_{n_2}|^{-2}\,|E-\tfrac{1}{2}k_3|^{-2}\,\langle \mathbf{n}_1,\mathbf{n}_2,\mathbf{k}_{H3}|\;U_n^{(\alpha)}\;|\mathbf{n}_1,\mathbf{n}_2,\mathbf{k}_{H3}\rangle$$

$$+\sum_{\substack{\mathbf{n}_i\\ i=2,3}}\Big(\int\!\!\sum_{\mathbf{k}_{H1}}\Big)|E-\tfrac{1}{2}k_1^2-E_{n_2}|^{-2}\,|E-E_{n_3}|^{-2}\,\langle \mathbf{k}_{H1},\mathbf{n}_2,\mathbf{n}_3|\;U_n^{(\alpha)}\;|\mathbf{k}_{H1},\mathbf{n}_2,\mathbf{n}_3\rangle+(\text{similar term})$$

$$+\sum_{\mathbf{n}_3}\Big(\int\sum_{\substack{\mathbf{k}_{Hi}\\ i=1,2}}\Big)|E-\tfrac{1}{2}k_1^2-\tfrac{1}{2}k_2^2|^{-2}\,|E-E_{n_3}|^{-2}\,\langle \mathbf{k}_{H1},\mathbf{k}_{H2},\mathbf{n}_3|\;U_n^{(\alpha)}\;|\mathbf{k}_{H1},\mathbf{k}_{H2},\mathbf{n}_3\rangle$$

$$+\sum_{\mathbf{n}_1}\Big(\int\sum_{\substack{\mathbf{k}_{Hi}\\ i=2,3}}\Big)|E-E_{n1}-\tfrac{1}{2}k_2^2|^{-2}\,|E-\tfrac{1}{2}k_3^2|^{-2}\,\langle \mathbf{n}_1,\mathbf{k}_{H2},\mathbf{k}_{H3}|\;U_n^{(\alpha)}\;|\mathbf{n}_1,\mathbf{k}_{H2},\mathbf{k}_{H3}\rangle+(\text{similar term})$$

$$+\Big(\int\!\!\sum_{\mathbf{k}_{Hi}}\Big)|E-\tfrac{1}{2}k_1^2-\tfrac{1}{2}k_2^2|^{-2}\,|E-\tfrac{1}{2}k_3^2|^{-2}\,\langle \mathbf{k}_{H1},\mathbf{k}_{H2},\mathbf{k}_{H3}|\;U_n^{(\alpha)}\;|\mathbf{k}_{H1},\mathbf{k}_{H2},\mathbf{k}_{H3}\rangle,\quad E\in D_E. \tag{3.28}$$

Hence, for compactness, it is sufficient to show that the individual terms of (3.28) are separately convergent

$$\sum_{\mathbf{n}_i}\langle \mathbf{n}_1,\mathbf{n}_2,\mathbf{n}_3|\;U_n^{(\alpha)}\;|\mathbf{n}_1,\mathbf{n}_2,\mathbf{n}_3\rangle<\infty,\quad E\in D_E, \tag{3.29}$$

$$\sum_{\substack{\mathbf{n}_i\\ i=1,2}}\Big(\int\!\!\sum_{\mathbf{k}_{H3}}\Big)|E-\tfrac{1}{2}k_3^2|^{-2}\,\langle \mathbf{n}_1,\mathbf{n}_2,\mathbf{k}_{H3}|\;U_n^{(\alpha)}\;|\mathbf{n}_1,\mathbf{n}_2,\mathbf{k}_{H3}\rangle<\infty,\quad E\in D_E, \tag{3.30}$$

$$\sum_{\substack{\mathbf{n}_i\\ i=2,3}}\Big(\int\!\!\sum_{\mathbf{k}_{H1}}\Big)|E-\tfrac{1}{2}k_1^2-E_{n_2}|^{-2}\,\langle \mathbf{k}_{H1},\mathbf{n}_2,\mathbf{n}_3|\;U_n^{(\alpha)}\;|\mathbf{k}_{H1},\mathbf{n}_2,\mathbf{n}_3\rangle<\infty\quad E\in D_E, \tag{3.31}$$

$$\sum_{\mathbf{n}_3}\Big(\int\sum_{\substack{\mathbf{k}_{Hi}\\ i=2,3}}\Big)|E-\tfrac{1}{2}k_1^2-\tfrac{1}{2}k_2^2|^{-2}\,\langle \mathbf{k}_{H1},\mathbf{k}_{H2},\mathbf{n}_3|\;U_n^{(\alpha)}\;|\mathbf{k}_{H1},\mathbf{k}_{H2},\mathbf{n}_3\rangle<\infty,\quad E\in D_E, \tag{3.32}$$

$$\sum_{\mathbf{n}_1}\Big(\int\sum_{\substack{\mathbf{k}_{Hi}\\ i=2,3}}\Big)|E-E_{n1}-\tfrac{1}{2}k_2^2|^{-2}\,|E-\tfrac{1}{2}k_3^2|^{-2}\,\langle \mathbf{n}_1,\mathbf{k}_{H2},\mathbf{k}_{H3}|\;U_n^{(\alpha)}\;|\mathbf{n}_1,\mathbf{k}_{H2},\mathbf{k}_{H3}\rangle<\infty,\quad E\in D_E, \tag{3.33}$$

$$\Big(\int\!\!\sum_{\mathbf{k}_{Hi}}\Big)|E-\tfrac{1}{2}k_1^2-\tfrac{1}{2}k_2^2|^{-2}\,\langle \mathbf{k}_{H1},\mathbf{k}_{H2},\mathbf{k}_{H3}|\;U_n^{(\alpha)}\;|\mathbf{k}_{H1},\mathbf{k}_{H2},\mathbf{k}_{H3}\rangle<\infty,\quad E\in D_E, \tag{3.34}$$

where we have used the fact that

$$|E-E_{n_1}-E_{n_2}|^{-2}\leqslant\text{const}\quad\text{and}\quad|E-E_{n_3}|^{-2}\leqslant\text{const},\quad E\in D_E.$$

We first consider (3.29). Using the addition theorem of spherical harmonics in the form

$$\sum_{m=-l}^{+l}\overline{Y}_{lm}(\theta,\phi)Y_{lm}(\theta,\phi)=(2l+1)/4\pi$$

and writing $d\Omega_i=\sin\theta_i\,d\theta_i\,d\phi_i$, $i=1,2,3$, we find that (3.29) becomes

$$\sum_{\mathbf{n}_i}\langle \mathbf{n}_1,\mathbf{n}_2,\mathbf{n}_3|\;U_n^{(\alpha)}\;|\mathbf{n}_1,\mathbf{n}_2,\mathbf{n}_3\rangle=(4\pi)^{-3}\sum_{n_i,l_i}(2l_1+1)(2l_2+1)(2l_3+1)\int_0^\infty r_1^2dr_1\int d\Omega_1\int_0^\infty r_2^2dr_2$$

$$\times\int d\Omega_2\int_0^\infty r_3^2dr_3\int d\Omega_3[R_{n_1l_1}(r_1)R_{n_2l_2}(r_2)R_{n_3l_3}(r_3)]^2\,U_n^{(\alpha)},\quad E\in D_E. \tag{3.35}$$

We have from (3.5) and (3.6)

$$[V_{ij}^{(n)}]^m\leqslant\text{const}\sum_{k=0}^{n}\left[\frac{1}{r_i^m}\left(\frac{r_j}{r_i}\right)^{mk}\theta(r_i-r_j)+\frac{1}{r_j^m}\left(\frac{r_i}{r_j}\right)^{mk}\theta(r_j-r_i)\right][P_k(\cos\theta_{ij})]^m, \tag{3.36}$$

where $m$ is a positive integer. Use of (3.36) yields

$$U_n^{(1)}=[V_{12}^{(n)}V_{23}^{(n)}V_{31}^{(n)}]^2$$

$$\leqslant\text{const}\sum_{\lambda_i=0}^{n}\left[\frac{1}{r_1^2}\left(\frac{r_2}{r_1}\right)^{2\lambda_1}\frac{1}{r_2^2}\left(\frac{r_3}{r_2}\right)^{2\lambda_2}\frac{1}{r_1^2}\left(\frac{r_3}{r_1}\right)^{2\lambda_3}\theta(r_1-r_2)\theta(r_2-r_3)\right.$$

$$+\frac{1}{r_1^2}\left(\frac{r_2}{r_1}\right)^{2\lambda_1}\frac{1}{r_3^2}\left(\frac{r_2}{r_3}\right)^{2\lambda_2}\frac{1}{r_1^2}\left(\frac{r_3}{r_1}\right)^{2\lambda_3}\theta(r_1-r_3)\theta(r_3-r_2)$$
$$+\frac{1}{r_2^2}\left(\frac{r_3}{r_2}\right)^{2\lambda_2}\frac{1}{r_3^2}\left(\frac{r_1}{r_3}\right)^{2\lambda_3}\frac{1}{r_2^2}\left(\frac{r_1}{r_2}\right)^{2\lambda_1}\theta(r_2-r_3)\theta(r_3-r_1)$$
$$+\frac{1}{r_2^2}\left(\frac{r_3}{r_2}\right)^{2\lambda_2}\frac{1}{r_1^2}\left(\frac{r_3}{r_1}\right)^{2\lambda_3}\frac{1}{r_2^2}\left(\frac{r_1}{r_2}\right)^{2\lambda_1}\theta(r_2-r_1)\theta(r_1-r_3)$$
$$+\frac{1}{r_3^2}\left(\frac{r_1}{r_3}\right)^{2\lambda_3}\frac{1}{r_1^2}\left(\frac{r_2}{r_1}\right)^{2\lambda_1}\frac{1}{r_3^2}\left(\frac{r_2}{r_3}\right)^{2\lambda_2}\theta(r_3-r_1)\theta(r_1-r_2)$$
$$+\frac{1}{r_3^2}\left(\frac{r_1}{r_3}\right)^{2\lambda_3}\frac{1}{r_2^2}\left(\frac{r_1}{r_2}\right)^{2\lambda_1}\frac{1}{r_3^2}\left(\frac{r_2}{r_3}\right)^{2\lambda_2}\theta(r_3-r_2)\theta(r_2-r_1)\Bigg]\,[P_{\lambda_1}(\cos\theta_{12})P_{\lambda_2}(\cos\theta_{23})P_{\lambda_3}(\cos\theta_{31})]^2, \tag{3.37}$$

$$U_n^{(2)}=[V_{12}^{(n)}]^2\,[V_{23}^{(n)}]^4\leqslant\text{const}\sum_{\lambda_i=0}^{n}\left[\frac{1}{r_1^2}\left(\frac{r_2}{r_1}\right)^{2\lambda_1}\frac{1}{r_2^4}\left(\frac{r_3}{r_2}\right)^{4\lambda_2}\theta(r_1-r_2)\theta(r_2-r_3)\right.$$
$$+\frac{1}{r_1^2}\left(\frac{r_2}{r_1}\right)^{2\lambda_1}\frac{1}{r_3^4}\left(\frac{r_2}{r_3}\right)^{4\lambda_2}\{\theta(r_3-r_1)\theta(r_1-r_2)+\theta(r_1-r_3)\theta(r_3-r_2)\}$$
$$+\frac{1}{r_2^2}\left(\frac{r_1}{r_2}\right)^{2\lambda_1}\frac{1}{r_2^4}\left(\frac{r_3}{r_2}\right)^{4\lambda_2}\{\theta(r_2-r_3)\theta(r_3-r_1)+\theta(r_2-r_1)\theta(r_1-r_3)\}$$
$$+\frac{1}{r_2^2}\left(\frac{r_1}{r_2}\right)^{2\lambda_1}\frac{1}{r_3^4}\left(\frac{r_2}{r_3}\right)^{4\lambda_2}\theta(r_3-r_2)\theta(r_2-r_1)[P_{\lambda_1}(\cos\theta_{12})]^2[P_{\lambda_2}(\cos\theta_{23})]^4, \tag{3.38}$$

where use has been made of the identities

$$\theta(r_3-r_2)\theta(r_1-r_2)=[\theta(r_3-r_1)+\theta(r_1-r_3)]\theta(r_3-r_2)\theta(r_1-r_2)=\theta(r_3-r_1)\theta(r_1-r_2)+\theta(r_1-r_3)\theta(r_3-r_2),$$
$$\theta(r_2-r_3)\theta(r_2-r_1)=\theta(r_2-r_3)\theta(r_3-r_1)+\theta(r_2-r_1)\theta(r_1-r_3)$$

to obtain the second and third terms on the right-hand side of (3.38). When the estimates (3.37) and (3.38) are substituted in (3.35), two inequalities are obtained (for $\alpha=1,2$). An examination of these inequalities reveal that there are two types of integrals involved whose angular integrations yield constants independent of $n_i$, $l_i$ and whose radial parts have the forms

$$\int_0^\infty dr_1\int_0^{r_1}dr_2\int_0^{r_2}dr_3\left[\frac{1}{r_1^2}\left(\frac{r_2}{r_1}\right)^{2\lambda_1}\frac{1}{r_2^2}\left(\frac{r_3}{r_2}\right)^{2\lambda_2}\frac{1}{r_1^2}\left(\frac{r_3}{r_1}\right)^{2\lambda_3}\right][R_{n_1l_1}(r_1)r_1R_{n_2l_2}(r_2)r_2R_{n_3l_3}(r_3)r_3]^2$$
$$=\int_0^1dt_1\int_0^1dt_2[t_1^{2(\lambda_1+\lambda_3)-1}\,t_2^{2(\lambda_2+\lambda_3)}]\int_0^\infty dr_1[R_{n_1l_1}(r_1)r_1R_{n_2l_2}(r_1t_1)r_1t_1R_{n_3l_3}(r_1t_1t_2)r_1t_1t_2]^2\,r_1^{-4} \tag{3.39}$$

and

$$\int_0^\infty dr_1\int_0^{r_1}dr_2\int_0^{r_2}dr_3\left[\frac{1}{r_1^2}\left(\frac{r_2}{r_1}\right)^{2\lambda_1}\frac{1}{r_2^4}\left(\frac{r_3}{r_2}\right)^{4\lambda_2}\right][R_{n_1l_1}(r_1)r_1R_{n_2l_2}(r_2)r_2R_{n_3l_3}(r_3)r_3]^2$$
$$=\int_0^1dt_1\int_0^1dt_2[t_1^{2\lambda_1-3}\,t_2^{4\lambda_2}]\int_0^\infty dr_1[R_{n_1l_1}(r_1)r_1R_{n_2l_2}(r_1t_1)r_1t_1R_{n_3l_3}(r_1t_1t_2)r_1t_1t_2]^2\,r_1^{-4}, \tag{3.40}$$

where the transformations $r_3/r_2=t_2$, $r_2/r_1=t_1$ have been made to obtain the right-hand sides of (3.39) and (3.40). We see from (3.35) that for compactness it is sufficient to show that

$$\sum_{n_i=1}^{\infty}\sum_{l_i=0}^{n_i-1}(2l_1+1)(2l_2+1)(2l_3+1)\sum_{\lambda_i=0}^{n}R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_2l_2,n_3l_3)<\infty,\quad \alpha=1,2,\quad E\in D_E, \tag{3.41}$$

where

$$R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_2l_2,n_3l_3)=\int_0^1dt_1\int_0^1dt_2\phi^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(t_1,t_2)\int_0^\infty dr[R_{n_1l_1}(r)rR_{n_2l_2}(rt_1)rt_1R_{n_3l_3}(rt_1t_2)rt_1t_2]^2\,r^{-4} \tag{3.42}$$

with

$$\phi^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(t_1,t_2)=\begin{cases}t_1^{2(\lambda_1+\lambda_3)-1}\,t_2^{2(\lambda_2+\lambda_3)} & \text{for } \alpha=1,\\[2ex] t_1^{2\lambda_1-3}\,t_2^{4\lambda_2} & \text{for } \alpha=2.\end{cases} \tag{3.43}$$

An examination of the remaining compactness conditions (3.30)–(3.34) reveal that it is sufficient to consider (3.30), (3.32), and (3.34) since the proof of (3.31) and (3.33) will be similar to those of (3.30) and (3.32) respectively. The reasoning led from (3.29) to (3.41) will, when applied to (3.30), (3.32), and (3.34) lead to

$$\sum_{\substack{n_i=1\\i=1,2}}^{\infty} \sum_{l_i=0}^{n_i-1} \sum_{l_3=0}^{\infty} (2l_1+1)(2l_2+1)(2l_3+1)\int_0^{\infty} dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2} \times\Big[\sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_2l_2,k_{H3}l_3)\Big] < \infty, \quad E\in D_E, \quad \alpha=1,2, \tag{3.44}$$

$$\sum_{n_3=1}^{\infty} \sum_{l_3=0}^{n_3-1} \sum_{\substack{l_i=0\\i=1,2}}^{\infty} (2l_1+1)(2l_2+1)(2l_3+1)\int_0^{\infty} dk_1\int_0^{\infty} dk_2 \frac{k_1^2k_2^2}{|E-\frac{1}{2}k_1^2-\frac{1}{2}k_2^2|^2} \times\Big[\sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(k_{H1}l_1,k_{H2}l_2,n_3l_3)\Big] < \infty, \quad \alpha=1,2, \quad E\in D_E, \tag{3.45}$$

$$\sum_{l_i=0}^{\infty} (2l_1+1)(2l_2+1)(2l_3+1)\int_0^{\infty} dk_1\int_0^{\infty} dk_2 \frac{k_1^2k_2^2}{|E-\frac{1}{2}k_1^2-\frac{1}{2}k_2^2|^2}\int_0^{\infty} dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2} \times\Big[\sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(k_{H1}l_1,k_{H2}l_2,k_{H3}l_3)\Big] < \infty, \quad \alpha=1,2, \quad E\in D_E, \tag{3.46}$$

where

$$R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_2l_2,k_{H3}l_3) = \int_0^1 dt_1\int_0^1 dt_2\, \phi^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(t_1,t_2)\int_0^{\infty} dr[R_{n_1l_1}(r)rR_{n_2l_2}(rt_1)rt_1R_{l_3}(k_3rt_1t_2)rt_1t_2]^2 r^{-4}, \quad \alpha=1,2, \tag{3.47}$$

$$R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(k_{H1}l_1,k_{H2}l_2,n_3l_3) = \int_0^1 dt_1\int_0^1 dt_2\, \phi^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(t_1,t_2)\int_0^{\infty} dr[R_{l_1}(k_1r)rR_{l_2}(k_2rt_1)rt_1R_{n_3l_3}(rt_1t_2)rt_1t_2]^2 r^{-4}, \quad \alpha=1,2, \tag{3.48}$$

$$R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(k_{H1}l_1,k_{H2}l_2,k_{H3}l_3) = \int_0^1 dt_1\int_0^1 dt_2\, \phi^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(t_1,t_2)\int_0^{\infty} dr[R_{l_1}(k_1r)rR_{l_2}(k_2rt_1)rt_1R_{l_3}(k_3rt_1t_2)rt_1t_2]^2 r^{-4}, \quad \alpha=1,2, \tag{3.49}$$

with $\phi^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(t_1,t_2)$ defined by (3.43). We first consider (3.41). For reasons exactly similar to those given in the previous paper [see comments following (3.39) of Paper II], we split up (3.41) and write it in the form

$$\sum_{n_i=2}^{\infty} \sum_{l_i=1}^{n_i-1} (2l_1+1)(2l_2+1)(2l_3+1) \sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_2l_2,n_3l_3) + \sum_{n_1=1}^{\infty} \sum_{\substack{n_i=2\\i=2,3}}^{\infty} \sum_{l_i=1}^{n_i-1} (2l_2+1)(2l_3+1) \sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_10,n_2l_2,n_3l_3)$$

$$+ \text{(two similar terms)} + \sum_{\substack{n_i=1\\i=2,3}}^{\infty} \sum_{n_i=2}^{\infty} (2l_1+1) \sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_20,n_30)$$

$$+ \text{(two similar terms)} + \sum_{n_i=1}^{\infty} \sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_10,n_20,n_30) < \infty, \quad \alpha=1,2, \quad E\in D_E.$$

It will therefore be sufficient to show that

$$\sum_{n_i=2}^{\infty} \sum_{l_i=1}^{n_i-1} (2l_1+1)(2l_2+1)(2l_3+1) \sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_2l_2,n_3l_3) < \infty, \quad E\in D_E, \tag{3.50}$$

$$\sum_{n_1=1}^{\infty} \sum_{\substack{n_i=2\\i=2,3}}^{\infty} \sum_{l_i=1}^{n_i-1} (2l_2+1)(2l_3+1) \sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_10,n_2l_2,n_3l_3) < \infty, \quad E\in D_E, \tag{3.51}$$

$$\sum_{\substack{n_i=1\\i=2,3}}^{\infty} \sum_{n_i=2}^{\infty} \sum_{l_i=1}^{n_i-1} (2l_1+1) \sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_20,n_30) < \infty, \quad E\in D_E, \tag{3.52}$$

$$\sum_{n_i=1}^{\infty} \sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_10,n_20,n_30) < \infty, \quad E\in D_E. \tag{3.53}$$

Denoting the infinite integral in (3.42) by $I(n_1l_1,n_2l_2,n_3l_3)$ and applying Holder's inequality in the form

$$\int_0^{\infty} |f(r)g(r)|dr \leqslant \Big(\int_0^{\infty} |f(r)|^p\Big)^{1/p} \Big(\int_0^{\infty} |g(r)|^{q/2}\Big)^{2/q} \frac{1}{p}+\frac{2}{q}=1 \tag{3.54}$$

to this integral, we have

$$I(n_1l_1,n_2l_2,n_3l_3)\leqslant\left(\int_0^\infty dr\frac{|R_{n_2l_2}(rt_1)rt_1|^{2p}}{r^4}\right)^{1/p}\left(\int_0^\infty dr\frac{|R_{n_1l_1}(r)rR_{n_3l_3}(rt_1t_2)rt_1t_2|^{q}}{r^4}\right)^{2/q}. \tag{3.55}$$

A further application of the Chauchy–Schwartz inequality on the second integral on the right-hand side yields

$$I(n_1l_1,n_2l_2,n_3l_3)$$
$$\leqslant\left(\int_0^\infty dr\frac{|R_{n_2l_2}(rt_1)rt_1|^{2p}}{r^4}\right)^{1/p}\left(\int_0^\infty dr\frac{|R_{n_3l_3}(rt_1t_2)rt_1t_2|^{2q}}{r^4}\right)^{1/q}\left(\int_0^\infty dr\frac{|R_{n_1l_1}(r)r|^{2q}}{r^4}\right)^{1/q}. \tag{3.56}$$

The substitution $\rho = rt_1$ in the first of the integrals on the right-hand side of (3.56) followed by the use of the inequality (II, Appendix B)

$$|R_{nl}(r)r|\leqslant\frac{\text{const}}{n^{1/2}} \tag{3.57}$$

in the form

$$|R_{n_2l_2}(\rho)\rho|^{2p-2}\leqslant\frac{\text{const}}{n^{(p-1)}}$$

gives

$$\int_0^\infty dr\,\frac{|R_{n_2l_2}(rt_1)rt_1|^{2p}}{r^4}\leqslant\text{const}\,\frac{t_1^3}{n_2^{(p-1)}}\int_0^\infty d\rho\,\frac{|R_{n_2l_2}(\rho)\rho|^2}{\rho^4}.$$

Hence, using the result[5]

$$\int_0^\infty dr\,\frac{|R_{nl}(r)r|^2}{r^4}=\frac{Z^4[3n^2-l(l+1)]}{2n^5(l+\frac{3}{2})(l+1)(l+\frac{1}{2})l(l-\frac{1}{2})}$$
$$\leqslant\text{const}\left[\left(1-\frac{l(l+1)}{3n^2}\right)/n^3l^5\right] \tag{3.58}$$

we find that

$$\left(\int_0^\infty dr\frac{|R_{n_2l_2}(rt_1)rt_1|^{2p}}{r^4}\right)^{1/p}\leqslant\text{const}\,\frac{t_1^{3/p}}{n_2^{(p+2)/p}\,l_2^{5/p}}. \tag{3.59}$$

Note that $l(l+1)/3n^2<1$ since $l\leqslant n-1$. Similarly

$$\left(\int_0^\infty dr\frac{|R_{n_3l_3}(rt_1t_2)rt_1t_2|^{2q}}{r^4}\right)^{1/q}\leqslant\text{const}\,\frac{(t_1t_2)^{3/q}}{n_3^{(q+2)/q}\,l_3^{5/q}}. \tag{3.60}$$

$$\left(\int_0^\infty dr\frac{|R_{n_1l_1}(r)r|^{2q}}{r^4}\right)^{1/q}\leqslant\text{const}\,\frac{1}{n_1^{(q+2)/q}\,l_1^{5/q}}. \tag{3.61}$$

These estimates when used in (3.56), yield an inequality which when used on the right-hand side of (3.42) gives

$$R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_2l_2,n_3l_3)$$
$$\leqslant\text{const}\int_0^1 dt_1\int_0^1 dt_2\,\phi^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(t_1t_2)$$
$$\times\left[\frac{t_1^{3(1/p+1/q)}\,t_2^{3/q}}{n_2^{(p+2)/p}\,l_2^{5/p}(n_1n_3)^{(q+2)/q}\,(l_1l_3)^{5/q}}\right]$$
$$\leqslant\text{const}\frac{1}{n_2^{(p+2)/p}\,l_2^{5/p}(n_1n_3)^{(q+2)/q}\,(l_1l_3)^{5/q}}, \tag{3.62}$$

provided that [note (3.43)]

$$3\left(\frac{1}{p}+\frac{1}{q}\right)>2. \tag{3.63}$$

The inequality (3.62) and the result

$$\sum_{l=1}^{n-1} l^{-\alpha}\leqslant\text{const}\,n^{-\alpha+1},\quad \alpha<1 \tag{3.64}$$

yields for the left-hand side of (3.50)

$$\sum_{n_i=2}^{\infty}\sum_{l_i=1}^{n_i-1}(2l_1+1)(2l_2+1)(2l_3+1)$$
$$\times\sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_2l_2,n_3l_3)$$
$$\leqslant\text{const}\sum_{n_i=2}^{\infty}\frac{1}{n_2^{(7-p)/p}(n_1n_3)^{(7-q)/q}}<\infty$$

provided that

$$(7-p)/p>1\quad\text{and}\quad(7-q)/q>1. \tag{3.65}$$

The conditions (3.63), (3.65), and [see (3.54)] $1/p+2/q=1$ imply that

$$\frac{2}{7}<\frac{1}{q}<\frac{1}{3}\quad,\quad\frac{1}{3}<\frac{1}{p}<\frac{3}{7}. \tag{3.66}$$

Hence there exist values of $1/p$ and $1/q$ which satisfy all three conditions (e.g., $(1/q)=13/42$ and $(1/p)=16/42$).

To demonstrate (3.51) we have [compare with (3.56)]

$$I(n_10,n_2l_2,n_3l_3)\leqslant\left(\int_0^\infty dr\frac{|R_{n_2l_2}(rt_1)rt_1|^{2p}}{r^4}\right)^{1/p}$$
$$\times\left(\int_0^\infty dr\frac{|R_{n_3l_3}(rt_1t_2)rt_1t_2|^{2q}}{r^4}\right)^{1/q}$$
$$\times\left(\int_0^\infty dr\,\frac{|R_{n_10}(r)r|^{2q}}{r^4}\right)^{1/q}$$
$$\leqslant\text{const}\,\frac{t_1^{3(1/p+1/q)}t_2^{3/q}}{n_2^{(p+2)/p}\,l_2^{5/p}\,n_3^{(q+2)/q}\,l_3^{5/q}}$$
$$\times\left(\int_0^\infty dr\frac{|R_{n_10}(r)r|^{2q}}{r^4}\right)^{1/q}, \tag{3.67}$$

where the last step is obtained by using (3.59) and (3.60). Noting (3.66), we set $Q=3+\epsilon$ ($\epsilon>0$, small), so that $p=3-\eta$, where $\eta$ is determined by $1/p+2/q=1$. Also using

(i) the estimate (3.57) in the form

$$|R_{n_10}(r)r|^{(1+2\epsilon)}\leqslant\text{const}\,n_1^{-(1+2\epsilon)/2},$$

(ii) the estimate (II, Appendix B)

$$|R_{n0}(r)|\leqslant\text{const}\,n^{-3/2}. \tag{3.68}$$

in the form $|R_{n_10}(r)|^3\leqslant\text{const}\,n_1^{-9/2}$,

(iii) the result[6]

$$\int_0^\infty dr \frac{|R_{n0}(r)r|^2}{r} = \frac{Z}{n^2}$$

we find that

$$\left(\int_0^\infty dr \frac{|R_{n_10}(r)r|^{2q}}{r^4}\right)^{1/q} \leqslant \text{const}\, \frac{1}{n_1^{(7+\epsilon)/q}}. \quad (3.69)$$

The substitution of the estimate (3.69) into (3.67) yields an inequality which when inserted into the right-hand side of (3.42) gives

$$R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_10,n_2l_2,n_3l_3) \leqslant \text{const} \frac{1}{n_1^{(7+\epsilon)/q}\, n_2^{(p+2)/p}\, l_2^{5/p}\, n_3^{(q+2)/q}\, l_3^{5/q}}. \quad (3.70)$$

Using (3.64), we find that the inequality (3.70) yields for the left-hand side of (3.51)

$$\sum_{n_1=1}^\infty \sum_{\substack{n_i=2\\ i=2,3}}^\infty \sum_{l_i=1}^{n_i-1} (2l_2+1)(2l_3+1) \sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_10,n_2l_2,n_3l_3)$$

$$\leqslant \text{const} \sum_{n_1=1}^\infty \sum_{n_i=2}^\infty \frac{1}{n_1^{(7+\epsilon)/q}\, n_2^{(7-p)/p}\, n_3^{(7-q)/q}} < 0.$$

Similarly, we have for (3.52)

$$I(n_1l_1,n_20,n_30) \leqslant \text{const} \frac{t_1^{3(1/p+1/q)} t_2^{3/q}}{n_1^{(q+2)/q}\, l_1^{5/q}\, n_2^{(7-\eta)/p}\, n_3^{(7+\epsilon)/q}}, \quad (3.71)$$

where, in obtaining (3.71), we have used the estimates (3.61),

$$\leqslant \left(\int_0^\infty dr \frac{|R_{n_20}(rt_1)rt_1|^{2p}}{r^4}\right)^{1/p} \leqslant \text{const} \frac{t_1^{3/p}}{n_2^{(7-\eta)/p}} \quad (3.72)$$

and

$$\left(\int_0^\infty dr \frac{|R_{n_30}(rt_1t_2)rt_1t_2|^{2q}}{r^4}\right)^{1/q} \leqslant \text{const} \frac{(t_1t_2)^{3/q}}{n_3^{(7+\epsilon)/q}}, \quad (3.73)$$

the last two estimates being obtained in a manner similar to that of (3.69). Hence [see (3.42)]

$$R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_20,n_30) \leqslant \text{const} \frac{1}{n_1^{(q+2)/q}\, l_1^{5/q}\, n_2^{(7-\eta)/p}\, n_3^{(7+\epsilon)/q}}$$

so that the left-hand side of (3.52), with the use of (3.64) becomes

$$\sum_{\substack{n_i=1\\ i=2,3}}^\infty \sum_{n_1=2}^\infty \sum_{l_1=1}^{n_1-1} (2l_1+1) \sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_20,n_30)$$

$$\leqslant \text{const} \sum_{n_1=2}^\infty \sum_{n_i=1}^\infty \frac{1}{n_1^{(7-q)/q}\, n_2^{(7-\eta)/p}\, n_3^{(7+\epsilon)/q}} < \infty.$$

Finally, in a similar fashion, the left-hand side of (3.53) becomes

$$\sum_{n_i=1}^\infty \sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_10,n_20,n_30)$$

$$\leqslant \text{const} \sum_{n_i=1}^\infty \frac{1}{(n_1n_3)^{(7+\epsilon)/q}\, n_2^{(7-\eta)/p}} < \infty.$$

This concludes the demonstration of (3.41).

We now consider (3.44). Proceeding as in the case of (3.41), we find that for compactness it is sufficient to show that

$$\sum_{\substack{n_i=2\\ i=1,2}}^\infty \sum_{l_i=1}^{n_i-1} \sum_{l_3=0}^\infty (2l_1+1)(2l_2+1)(2l_3+1) \times \sum_{\lambda_i=0}^{n} \int_0^\infty dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2} \times R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_2l_2,k_{H3}l_3) < \infty, \quad E\in D_E, \quad (3.74)$$

$$\sum_{n_1=1}^\infty \sum_{n_2=2}^\infty \sum_{l_2=1}^{n_2-1} \sum_{l_3=0}^\infty (2l_2+1)(2l_3+1) \times \left[\sum_{\lambda_i=0}^{n} \int_0^\infty dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2}\right] \times R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_10,n_2l_2,k_{H3}l_3) < \infty, \quad E\in D_E, \quad (3.75)$$

$$\sum_{\substack{n_i=1\\ i=1,2}}^\infty \sum_{l_3=1}^\infty (2l_3+1) \sum_{\lambda_i=0}^{n} \int_0^\infty dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2} \times R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_10,n_20,k_{H3}l_3) < 0, \quad E\in D_E. \quad (3.76)$$

Denoting the infinite integral in (3.47) by $I(n_1l_1,n_2l_2,k_{H3},l_3)$, we have [compared with (3.56)]

$$I(n_1l_1,n_2l_2,k_{H3},l_3) \leqslant \left(\int_0^\infty dr \frac{|R_{n_2l_2}(rt_1)rt_1|^{2p}}{r^4}\right)^{1/p} \left(\int_0^\infty dr \frac{|R_{n_1l_1}(r)r|^{2q}}{r^4}\right)^{1/q} \times \left(\int_0^\infty dr \frac{|R_{l_3}(k_3rt_1t_2)rt_1t_2|^{2q}}{r^4}\right)^{1/q}$$

$$\leqslant \text{const} \frac{t_1^{3(1/p+1/q)}\, t_2^{3/q}}{n_1^{(q+2)/q}\, l_1^{5/q}\, n_2^{(p+2)/p}\, l_2^{5/p}} \times \left(\int_0^\infty d\rho \frac{|R_{l_3}(k_3\rho)\rho|^{2q}}{\rho^4}\right)^{1/q}, \quad (3.77)$$

where the last step was obtained by use of the estimates (3.59), (3.61) and the transformation $\rho = rt_1t_2$. By transforming the infinite integral over in (3.77) to one over the infinite interval [0,1] and then using the mean value theorem of the integral calculus (see II, Appendix C where a similar trick is employed) one obtains

$$\left(\int_0^\infty d\rho \frac{|R_{l_3}(k_3\rho)\rho|^{2q}}{\rho^4}\right)^{1/q} = \left|R_{l_3}\left(\frac{k_3u_0}{1-u_0}\right)\right|^2 f(u_0,q), \quad u_0\in[0,1], \quad (3.78)$$

where

$$f(u_0,q) = \left(\frac{u_0}{1-u_0}\right)^2 (1-u_0)^{2/q} u_0^{-4/q}. \quad (3.79)$$

Substituting (3.78) into (3.77) and inserting the result into the right hand side of (3.47) yields

$$R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_2l_2,k_{H3}l_3) \leqslant \text{const} \frac{|R_{l_3}(k_3u_0/1-u_0)|^2 f(u_0,q)}{n_1^{(q+2)/q}\, l_1^{5/q}\, n_2^{(p+2)/p}\, l_2^{5/p}}.$$

We have therefore

$$\int_0^\infty dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_2l_2,k_{H3}l_3)$$
$$\leqslant \text{const}\,\frac{|R_{l_1}(k_{3m}u_0/1-u_0)|^2 f(u_0,q)}{n_1^{(q+2)/q}\, l_1^{5/q}\, n_2^{(p+2)/p}\, l_2^{5/p}}$$
$$\times\int_0^\infty dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2},$$

where $k_{3m}$ us the value at which $|R_{l_1}(\cdots)|^2$ attains its maximum value. Hence, the left hand side of (3.74) with the use of (3.64) becomes

$$\sum_{n_i=2}^{\infty}\sum_{\substack{l_i=1\\ i=1,2}}^{n_i-1}\sum_{l_3=0}^{\infty}(2l_1+1)(2l_2+1)(2l_3+1)$$
$$\times\sum_{\lambda_i=0}^{n}\int_0^\infty dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2}$$
$$R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_1l_1,n_2l_2,k_{H3}l_3)$$
$$\leqslant \text{const}\sum_{n_i=2}^{\infty}\sum_{l_3=0}^{\infty}\frac{(2l_3+1)\,|R_{l_3}(k_{3m}u_0/1-u_0)|^2 f(u_0,q)}{n_1^{(7-q)/q}\, n_2^{(7-p)/p}}$$
$$\times\int_0^\infty dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2}<\infty,\quad E\in D_E$$

where we have used the result [II, Appendix C]

$$\sum_{l=0}^{\infty}(2l+1)|R_l(kr)|^2<\infty. \tag{3.80}$$

Proceeding as in the case of (3.74), we have using (3.69), (3.59) and (3.78)

$$R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_10,n_2l_2,k_{H3}l_3)\leqslant \text{const}\,\frac{|R_{l_3}(k_3u_0/1-u_0)|^2 f(u_0,q)}{n_1^{(7+\epsilon)/q}\, n_2^{(p+2)/p}\, l_2^{5/p}}$$

so that the left hand side of (3.75) becomes

$$\sum_{n_1=1}^{\infty}\sum_{n_2=2}^{\infty}\sum_{l_2=1}^{n_2-1}\sum_{l_3=0}^{\infty}(2l_2+1)(2l_3+1)$$
$$\times\sum_{\lambda_i=0}^{n}\int_0^\infty dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2}R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_10,n_2l_2,k_{H3}l_3)$$
$$\leqslant \text{const}\sum_{n_1=1}^{\infty}\sum_{n_2=2}^{\infty}\sum_{l_3=0}^{\infty}(2l_3+1)\frac{|R_{l_3}(k_{3m}u_0/1-u_0)|^2 f(u_0,q)}{n_1^{(7+\epsilon)/q}\, n_2^{(7-p)/p}}$$
$$\times\int_0^\infty dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2}<\infty,\quad E\in D_E,$$

where in the last step we have used (3.80)

Similarly one shows that the left hand side of (3.76) satisfies

$$\sum_{n_i=1}^{\infty}\sum_{l_3=0}^{\infty}(2l_3+1)\sum_{\lambda_i=0}^{n}\int_0^\infty dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2}$$
$$\times R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(n_10,n_20k_{H3}\,l_3)$$
$$\leqslant \text{const}\sum_{n_i=1}^{\infty}\sum_{l_3=0}^{\infty}\frac{(2l_3+1)|R_{l_3}(k_{3m}u_0/1-u_0)|^2 f(u_0,q)}{n_1^{(7+\epsilon)/q}\, n_2^{(7-\eta)/p}}$$
$$\times\int_0^\infty dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2}<\infty,\quad E\in D_E.$$

This concludes the proof of (3.44). We next consider (3.45). As in the previous cases we split up the expression (as a sum of two terms) and find that for compactness it is sufficient to show that

$$\sum_{n_i=2}^{\infty}\sum_{l_i=1}^{n_i-1}\sum_{\substack{l_3=0\\ i=1,2}}^{\infty}(2l_1+1)(2l_2+1)(2l_3+1)$$
$$\times\int_0^\infty dk_1\int_0^\infty dk_2 \frac{k_1^2k_2^2}{|E-\frac{1}{2}k_1^2-\frac{1}{2}k_2^2|^2}$$
$$\times\left[\sum_{\lambda_i=0}^{n}R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(k_{H1}l_1,k_{H2}l_2n_3l_3)\right]<\infty,\quad E\in D_E, \tag{3.81}$$

$$\sum_{n_i=1}^{\infty}\sum_{\substack{l_i=0\\ i=1,2}}^{\infty}(2l_1+1)(2l_2+1)$$
$$\times\int_0^\infty dk_1\int_0^\infty dk_2 \frac{k_1^2k_2^2}{|E-\frac{1}{2}k_1^2-\frac{1}{2}k_2^2|^2}$$
$$\times\left[\sum_{\lambda_i=0}^{n}R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(k_{H1}l_1,k_{H2}l_2n_30)\right]<\infty,\quad E\in D_E, \tag{3.82}$$

Denoting the infinite integral in (3.48) by $I(k_{H1}l_1,k_{H2}l_2,n_3l_3)$, we have

$$I(k_{H1}l_1,k_{H2}l_2,n_3l_3)$$
$$\leqslant\left(\int_0^\infty dr\frac{|R_{l_2}(k_2rt_1)rt_1|^{2p}}{r^4}\right)^{1/p}\left(\int_0^\infty dr\frac{|R_{l_1}(k_1r)r|^{2q}}{r^4}\right)^{1/q}$$
$$\times\left(\int_0^\infty dr\frac{|R_{n_3l_3}(rt_1t_2)rt_1t_2|^{2q}}{r^4}\right)^{1/q},$$
$$\leqslant\text{const}\,\frac{(t_1t_2)^{3/q}}{n_3^{(q+2)/q}\,l_3^{5/q}}\left|R_{l_1}\left(\frac{k_1u_0}{1-u_0}\right)\right|^2 f(u_0,q)$$
$$\times\left(\int_0^\infty dr\frac{|R_{l_2}(k_2rt_1)rt_1|^{2p}}{r^4}\right)^{1/p}, \tag{3.83}$$

where in the last step we have used (3.60) and (3.78). Making the transformation $\rho=rt_1$ in the integral and using (3.78) one obtains the result

$$\left(\int_0^\infty dr\frac{|R_{l_2}(k_2rt_1)rt_1|^{2p}}{r^4}\right)^{1/p}=t_1^{3/p}\left|R_{l_2}\left(\frac{k_2u_0}{1-u_0}\right)\right|^2 f(u_0,p). \tag{3.84}$$

Substituting (3.84) into (3.83) and inserting the resulting inequality on the right hand side of (3.48) yields

$$R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(k_{H1}l_1,k_{H2}l_2,n_3l_3))$$
$$\leqslant\text{const}\,\frac{f(u_0,q)f(u_0,p)}{n_3^{(q+2)/q}\,l_3^{5/q}}\left|R_{l_1}\left(\frac{k_1u_0}{1-u_0}\right)\right|^2\left|R_{l_2}\left(\frac{k_2u_0}{1-u_0}\right)\right|^2.$$

Using polar coordinates $k_1=\rho\cos\theta$, $k_2=\rho\sin\theta$, we therefore have

$$\int_0^\infty k_1^2dk_1\int_0^\infty k_2^2dk_2\frac{R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(k_{H1}l_1,k_{H2}l_2n_3l_3)}{|E-\frac{1}{2}k_1^2-\frac{1}{2}k_2^2|^2}$$

$$\leqslant \frac{\text{const}}{n_3^{(q+2)/q}\, l_3^{5/q}} \int_0^{\pi/2} \sin^2\theta \cos^2\theta \, d\theta$$

$$\times \int_0^\infty \rho^5 d\rho \left| R_{l_1}\Big(\frac{\rho u_0 \cos\theta}{1-u_0}\Big)\right|^2 f(u_0,q) \left| R_{l_2}\Big(\frac{\rho u_0 \sin\theta}{1-u_0}\Big)\right|^2$$

$$\times f(u_0,p)/|E - \tfrac{1}{2}\rho^2|^2. \qquad (3.85)$$

Noting that $\rho^4/|E-\frac{1}{2}\rho^2|^2$ is bounded for $E \in D_E$ and using

(i) the mean value theorem of the integral calculus for the integration over $\theta$ in the interval $[0,\pi/2]$,

(ii) the result (II, Appendix C)

$$\int_0^\infty \rho \left| R_{l_1}\Big(\frac{\rho u_0 \cos\theta}{1-u_0}\Big)\right|^2 \left| R_{l_2}\Big(\frac{\rho u_0 \sin\theta}{1-u_0}\Big)\right|^2 d\rho$$

$$= \left| R_{l_1}\Big(\frac{u_0 v_0 \cos\theta_0}{(1-u_0)(1-v_0)}\Big)\right|^2 \left| R_{l_2}\Big(\frac{u_0 v_0 \sin\theta_0}{(1-u_0)(1-v_0)}\Big)\right|^2$$

$$\times \frac{v_0}{(1-v_0)^3}, \qquad v_0 \in [0,1],$$

we find that (3.85) reduces to

$$\int_0^\infty k_1^2 dk_1 \int_0^\infty k_2^2 dk_2 \frac{R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(k_{H1}l_1,k_{H2}l_2n_3l_3)}{|E-\frac{1}{2}k_1^2-\frac{1}{2}k_2^2|^2}$$

$$\leqslant \text{const}\, \frac{1}{n_3^{(q+2)/q}\, l_3^{5/q}}$$

$$\times \left| R_{l_1}\Big(\frac{u_0 v_0 \cos\theta_0}{(1-u_0)(1-v_0)}\Big)\right|^2 f(u_0,q)$$

$$\times \left| R_{l_2}\Big(\frac{u_0 v_0 \sin\theta_0}{(1-u_0)(1-v_0)}\Big)\right|^2 f(u_0,p)\, \frac{v_0}{(1-v_0)^3}.$$

Hence, using (3.64), we find that (3.81) becomes

$$\sum_{n_3=2}^\infty \sum_{l_3=1}^{n_3-1} \sum_{\substack{l_i=0\\ i=1,2}}^\infty (2l_1+1)(2l_2+1)(2l_3+1)$$

$$\times \int_0^\infty dk_1 \int_0^\infty dk_2 \frac{k_1^2 k_2^2}{|E-\frac{1}{2}k_1^2-\frac{1}{2}k_2^2|^2}$$

$$\times \Big[ \sum_{\lambda_i=0}^{n} R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(k_{H1}l_1,k_{H2}l_2n_3l_3)\Big],$$

$$\leqslant \text{const} \sum_{n_3=2}^\infty n_3^{-(7-q)/q}$$

$$\times \Big[ \sum_{l_1=0}^\infty (2l_1+1) \left| R_{l_1}\Big(\frac{u_0 v_0 \cos\theta_0}{(1-u_0)(1-v_0)}\Big)\right|^2 f(u_0,q)\Big]$$

$$\times \Big[ \sum_{l_2=0}^\infty (2l_2+1) \left| R_{l_2}\Big(\frac{u_0 v_0 \sin\theta_0}{(1-u_0)(1-v_0)}\Big)\right|^2 f(u_0,p)\Big]$$

$$\times \frac{v_0}{(1-v_0)^3} < \infty, \quad E \in D_E,$$

where in the last step we have used (3.80). Similarly, using (3.73), we find that (3.82) becomes

$$\sum_{n_3=1}^\infty \sum_{\substack{l_i=0\\ i=1,2}}^\infty (2l_1+1)(2l_2+1)$$

$$\times \int_0^\infty k_1^2 dk_1 \int_0^\infty k_2^2 dk_2 \frac{R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(k_{H1}l_1,k_{H2}l_2,n_3 0)}{|E-\frac{1}{2}k_1^2-\frac{1}{2}k_2^2|^2}$$

$$\leqslant \text{const} \sum_{n_3=2}^\infty n_3^{-(7+\epsilon)/q} \Big[ \sum_{l_1=0}^\infty (2l_1+1)$$

$$\times \left| R_{l_1}\Big(\frac{u_0 v_0 \cos\theta_0}{(1-u_0)(1-v_0)}\Big)\right|^2 f(u_0,q)\Big]$$

$$\times \Big[ \sum_{l_2=0}^\infty (2l_2+1) \left| R_{l_2}\Big(\frac{u_0 v_0 \sin\theta_0}{(1-u_0)(1-v_0)}\Big)\right|^2\Big]$$

$$\times f(u_0,p)\Big] \frac{v_0}{(1-v_0)^3} < \infty, \quad E \in D_E.$$

This concludes the proof of (3.45).

Finally, we consider (3.46). Denoting the infinite integral in (3.49) by $I(k_{H1}l_1,k_{H2}l_2,k_{H3}l_3)$. we have

$$I(k_{H1}l_1,k_{H2}l_2,k_{H3}l_3).$$

$$\leqslant \Big( \int_0^\infty dr \frac{|R_{l_2}(k_2 r t_1) r t_1|^{2p}}{r^4}\Big)^{1/p} \Big( \int_0^\infty dr \frac{|R_{l_1}(k_1 r) r|^{2q}}{r^4}\Big)^{1/q}$$

$$\times \Big( \int_0^\infty dr \frac{|R_{l_3}(k_3 r t_1 t_2) r t_1 t_2|^{2q}}{r^4}\Big)^{1/q}$$

$$\leqslant t_1^{3(\frac{1}{p}+\frac{1}{q})} t_2^{3/q} \left| R_{l_1}\Big(\frac{k_1 u_0}{1-u_0}\Big)\right|^2 f(u_0,q) \left| R_{l_2}\Big(\frac{k_2 u_0}{1-u_0}\Big)\right|^2$$

$$\times f(u_0,p) \left| R_{l_3}\Big(\frac{k_3 u_0}{1-u_0}\Big)\right|^2 f(u_0,q),$$

where we have used (3.78), (3.84), and the result

$$\Big( \int_0^\infty dr \frac{|R_{l_3}(k_3 r t_1 t_2) r t_1 t_2|^{2q}}{r^4}\Big)^{1/q}$$

$$= (t_1 t_2)^{3/q} \left| R_{l_3}\Big(\frac{k_3 u_0}{1-u_0}\Big)\right|^2 f(u_0,q)$$

obtained by the transformation $\rho = r t_1 t_2$ followed by the application of (3.78).

If we now proceed in a manner similar to that in the case of (3.81) we obtain

$$\sum_{l_i=0}^\infty (2l_1+1)(2l_2+1)(2l_3+1)$$

$$\times \sum_{\lambda_i=0}^{n} \int_0^\infty dk_1 \int_0^\infty dk_2 \frac{k_1^2 k_2^2}{|E-\frac{1}{2}k_1^2-\frac{1}{2}k_2^2|^2}$$

$$\times \int_0^\infty \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2} \leqslant R^{(\alpha)}_{\lambda_1\lambda_2\lambda_3}(k_{H1}l_1,k_{H2}l_2,k_{H3}l_3)$$

$$\leqslant \text{const} \Big[ \sum_{l_1=0}^\infty (2l_1+1) \left| R_{l_1}\Big(\frac{u_0 v_0 \cos\theta_0}{(1-u_0)(1-v_0)}\Big)\right|^2 f(u_0,q)\Big]$$

$$\times \Big[ \sum_{l_2=0}^\infty (2l_2+1) \left| R_{l_2}\Big(\frac{u_0 v_0 \sin\theta_0}{(1-u_0)(1-v_0)}\Big)\right|^2 f(u_0,p)\Big]$$

$$\times \frac{v_0}{(1-v_0)^3} \Big[ \sum_{l_3=0}^\infty (2l_3+1) \left| R_{l_3}\Big(\frac{k_{3m} u_0}{1-u_0}\Big)\right|^2 f(u_0,q)\Big]$$

$$\times \int_0^\infty dk_3 \frac{k_3^2}{|E-\frac{1}{2}k_3^2|^2} < \infty,$$

where, in the last step we have used (3.80) and $k_{3m}$ as the values of $k_3$ at which $|R_{l_i}(\cdots)|^2$ attains its maximum value. This completes the proof of (3.46).

We have therefore shown that the operator $K_n(E)$ defined by (3.8) is compact for $E\in D_E$. To show that the operator $K(E) = G_1V_1G_2V_2G_3V_3$ is compact for $E\in D_E$, all we need do is to show that

$$\|K - K_n\| \to 0, \quad \text{as } n\to\infty. \tag{3.9}$$

Noting that $V_i$ and $V_i^{(n)}$ have domain $D_V$ defined by (3.3) and are bounded, we observe that $(V_i - V_i^{(n)})\in L^\infty(D_V)$, where the norm on the Banach space $L^\infty(D_V)$ is defined by

$$\|f\| = \sup_{(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)\in D_V} |f(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)|.$$

Since $V_i - V_i^{(n)} \to 0$ as $n\to\infty$ and $G_i(E)\, i = 1,2,3$ are bounded operators for $E\in D_E$ we have

$$\begin{aligned}\|K - K_n\| &= \|G_1V_1G_2V_2G_3(V_3 - V_3^{(n)}) \\ &\quad + G_1V_1G_2(V_2 - V_{(2)}^{n})G_3V_3^{(n)} \\ &\quad + G_1(V_1 - V_1^{(n)})G_2V_2^{(n)}G_3V_3^{(n)}\| \\ &\leqslant \|G_1V_1G_2V_2G_3\|\,\|V_3 - V_3^{(n)}\| \\ &\quad + \|G_1V_1G_2\|\,\|V_2 - V_{(2)}^{n}\|\,\|G_3V_3^{(n)}\| \\ &\quad + \|G_1\|\,\|V_1 - V_1^{(n)}\|\,\|G_2V_2^{(n)}G_3V_3^{(n)}\| \to 0\end{aligned}$$

as $n\to\infty$, $E\in D_E$.

Hence the operator $K(E)$ is compact for $E\in D_E$.

## 4. ADOPTION TO THE CASE OF THE DIATOMIC MOLECULE

We now show that the method developed so far to deal with atomic systems can equally well be applied to the case of diatomic molecules. The Schrödinger equation for a diatomic molecule with respect to a reference system in which the center of mass of the nuclei is at the origin of coordinates is, to a very good approximation, given by

$$\Big(-\frac{\hbar^2}{2M}\nabla_R^2 - \frac{\hbar^2}{2m}\sum_{i=1}^{\infty}\nabla_i^2 + \frac{Z_AZ_B}{R} + \sum_{\substack{i,j=1\\ i<j}}^{n}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \times\sum_{i=1}^{n}\frac{Z_Ae^2}{|\mathbf{r}_i - \eta\mathbf{R}|} - \sum_{i=1}^{n}\frac{Z_Be^2}{|\mathbf{r}_i + (1-\eta)\mathbf{R}|}\Big]\Psi = E\Psi, \tag{4.1}$$

where $\mathbf{r}_i$ is the position of the $i$th electron (relative to the center of mass of the nuclei "$A$" and "$B$" with masses $M_A$ and $M_B$, respectively), $R = |\mathbf{R}|$ is the distance between the nuclei, $m$ is the mass of the electron,
$M = M_AM_B/(M_A + M_B)$ is the reduced mass,
$\eta = M_B/(M_A + M_B)$ and $Z_A$ and $Z_B$ are the charges on the nuclei "$A$" and "$B$" respectively. We have

$$\begin{aligned}\frac{1}{|\mathbf{r}_i - \eta\mathbf{R}|} &= \sum_{k=0}^{\infty}\Big[\frac{1}{r_i}\Big(\frac{\eta R}{r_i}\Big)^k\theta(r_i - \eta R) \\ &\quad + \frac{1}{\eta R}\Big(\frac{r_i}{\eta R}\Big)^k\theta(\eta R - r_i)\Big]P_k(\cos\theta_i) \\ &= \frac{\theta(r_i - \eta R)}{r_i} + \frac{\theta(\eta R - r_i)}{\eta R} \\ &\quad + \sum_{k=1}^{\infty} V_k(\mathbf{r}_i,\eta\mathbf{R}),\end{aligned} \tag{4.2}$$

$$V_k(\mathbf{r}_i,\eta\mathbf{R}) = \Big[\frac{1}{r_i}\Big(\frac{\eta R}{r_i}\Big)^k\theta(r_i - \eta R) + \frac{1}{\eta R}\Big(\frac{r_i}{\eta R}\Big)^k\theta(\eta R - r_i)\Big]P_k(\cos\theta_i) \tag{4.3}$$

and $\theta_i$ is the angle between $\mathbf{r}_i$ and $\mathbf{R}$. Similarly

$$\begin{aligned}\frac{1}{|\mathbf{r}_i + (1-\eta)\mathbf{R}|} &= \frac{\theta(r_i - (1-\eta)R)}{r_i} + \frac{\theta((1-\eta)R - r_i)}{(1-\eta)R} \\ &\quad + \sum_{k=1}^{\infty}(-1)^kV_k(\mathbf{r}_i,(1-\eta)\mathbf{R}).\end{aligned} \tag{4.4}$$

Also

$$\begin{aligned}\frac{\theta(r_i - \eta R)}{r_i} + \frac{\theta(\eta R - r_i)}{\eta R} &= \frac{\alpha}{r_i} + \frac{\alpha_0}{\eta R} + (1-\alpha)\frac{\theta(r_i - \eta R)}{r_i} - \alpha\frac{\theta(\eta R - r_i)}{r_i} \\ &\quad + (1-\alpha_0)\frac{\theta(\eta R - r_i)}{\eta R} - \alpha_0\frac{\theta(r_i - \eta R)}{\eta R},\end{aligned}$$

$\alpha, \alpha_0\in\mathbb{R}$

$$= \frac{\alpha}{r_i} + \frac{\alpha_0}{\eta R} + U_0(r_i,\eta R,\alpha,\alpha_0), \tag{4.5}$$

where

$$\begin{aligned}U_0(r_i,\eta R,\alpha,\alpha_0) &= (1-\alpha)\frac{\theta(r_i - \eta R)}{r_i} - \alpha\frac{\theta(\eta R - r_i)}{r_i} \\ &\quad + (1-\alpha_0)\frac{\theta(\eta R - r_i)}{\eta R} - \alpha_0\frac{\theta(r_i - \eta R)}{\eta R}.\end{aligned} \tag{4.6}$$

Similarly

$$\frac{\theta(r_i - (1-\eta)R)}{r_i} + \frac{\theta((1-\eta)R - r_i)}{(1-\eta)R} = \frac{\beta}{r_i} + \frac{\beta_0}{(1-\eta)R} + U_0(r_i,(1-\eta)R,\beta,\beta_0). \tag{4.7}$$

Using (4.2)–(4.7) we see that, for example, (4.1) can be written

$$\Big[-\frac{\hbar^2}{2M}\nabla_R^2 - \frac{[\{Z_A(\alpha_0/\eta) + Z_B[\beta_0/(1-\eta)]\}n - Z_AZ_B]e^2}{R} - \frac{\hbar^2}{2m}\sum_{i=1}^{n}\nabla_i^2 - \sum_{i=1}^{n}\frac{(Z_A\alpha + Z_B\beta)e^2}{r_i} + \sum_{\substack{i,j=1\\ i<j}}^{n}\frac{1}{r_{ij}}$$

$$+\sum_{i=1}^{n}\Big\{U_0(r_i,\eta R,\alpha,\alpha_0)+\sum_{k=1}^{\infty}V_k(\mathbf{r}_i,\eta R)$$
$$+U_0(r_i,(1-\eta)R,\beta,\beta_0)$$
$$+\sum_{k=1}^{\infty}(-1)^k V_k(\mathbf{r}_i,(1-\eta)R)\Big\}\Big]\Psi=E\Psi. \tag{4.8}$$

Equation (4.8) which we have just derived for the diatomic molecule is of the same form as the Schrödinger equation (2.1) for an $n$-electron atom. We can therefore derive from Eq (4.8) the usual conventional configuration interaction equations of the same form as (2.1) of $I$, truncate it, solve the eigenvalue problem associated with the truncated equations, thereby obtaining, for instance, the ground state energy level. Alternatively, one could use the system of equations (2.16) to obtain energy eigenvalues and eigenfunctions. Both these approaches have been made possible due to the derivation of Eq. (4.8) which, as we have noted, has exactly the same form as the Schrödinger equation (2.1) for an $n$-electron atom. The crucial step in the derivation of Eq. (4.8) is the identity (4.5). Note, that in our approach, the Born–Oppenheimer separation is not needed.

Finally, we summarize our results and conclusions. The Schrödinger equation for an $n$-electron atomic system ($n\geqslant 3$) is reduced to an infinite system of linear equations in such a way that the linear operator defined by this system of equations is compact in a region of the complex energy plane which excludes the various bound state and multiparticle scattering cuts, (i.e., the essential spectrum of the Hamiltonian of the $n$-electron atomic system). This allows one to truncate this infinite system of equations with the assurance that as the size of the truncated equations is increased, the energy eigenvalues obtained from the truncated equations will also increase and uniformly tend to the eigenvalues of the original infinite system. Further, we have shown that the method can be used in the case of the diatomic molecule without the use of a Born–Oppenheimer separation.

[1] S. Weinberg, Phys. Rev. **133**, 232–56 (1964).
[2] W. Hunziker, Helv. Phys. Acta **39**, 451–62 (1966).
[3] See Ref. 1, Sec. 3.
[4] R. Schatten, *Norm Ideals of Completely Continuous Operators* (Springer-Verlag, Berlin, 1960), p. 13.
[5] H.A. Bethe and E.E. Saltpeter, *Quantum Mechanics of One and Two-electron Atoms* (Springer-Verlag, Berlin, 1957), p. 17.
[6] See Ref. 5, p. 17.

# Modified modal theory of transient response in layered media

L. Tsang and J. A. Kong

*Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

(Received 28 April 1978)

A modified modal theory is developed for the transient response due to a line source placed in a two-layer slab medium. The double integral is evaluated by deformation on the complex wave number plane followed by a second deformation from the real frequency axis to the steepest descent path on the complex frequency plane. We first show that causality is preserved such that before the arrival time of the direct wave or the head wave, whichever takes the least time, the transient response vanishes. With the preservation of causality, we proceed to find the complete transient response at all times. The results obtained from the modal theory are then checked with those generated from direct numerical calculations with computers. Comparisons are also made with the explicit inversion technique which is applicable to the present case but appears to have severe restrictions that the modal theory does not have in generalizing to the solution of other transient problems.

## I. INTRODUCTION

In this paper we study transient fields due to an infinite line source located in a slab medium by using a modified modal approach via double deformation. The line source can be either acoustical or electrical in nature. Time harmonic fields in layered media have been extensively studied[1–6] with the use of the Sommerfeld integral representations by appropriate deformation in the complex wavenumber plane. The classical modal approach to time harmonic excitations yields normal modes pertaining to the structure represented by poles situated between the Sommerfeld integration path and the steepest descent path on the complex wavenumber plane.

Let us focus our attention on one particular mode $l$. As frequency changes, the location of the pole representing the $l$ th modes moves on the complex wavenumber plane. Depending on the frequency, the mode $l$ can exist as a guided mode[1,3–7] (surface wave mode) or a leaky mode. Over certain frequency ranges, the mode exists as an unexcited mode because it lies outside the Sommerfeld integration path and the steepest descent path. Over these frequency bands, the mode amplitude is zero. If we integrate each individual mode amplitude over real frequencies, the Paley–Wiener criterion is violated[8] and the mode in timedomain becomes noncausal. In this paper we develop time domain modes that are causal by employing the technique of double deformation. The procedure consists first of a deformation to the steepest descent path on the complex wavenumber plane. Then there is a second deformation from the real frequency axis to the steepest descent path on the complex frequency plane. The double deformation technique has been used to investigate the long time response of slab geometry.[9] In this paper, we show that the complete causal transient response at all times can be obtained with this approach. To check the results obtained with this approach, we compare with those generated by direct numerical integration and by the method of explicit inversion (Cagniard's method of integration[10–11]). It is noted that while the explicit inversion technique can be applied to evaluate the transient field of our present problem, its inherent severe restrictions prevent its generalization to treat other problems. Such restrictions do not appear in the modal theory developed in this paper.

## II. FORMULATION

Consider a line source situated at the center of a layer of fluid with density $\rho$ and characteristic velocity $v$. The layer is bounded on both sides by another fluid of density $\rho_1$ and characteristic velocity $v_1$. The thickness of the layer is $2a$. We choose the coordinate origin to coincide with the line source (Fig. 1). Let the source be initially at rest and have the excitation function

$$x(t) = e^{-\alpha t} \sin\omega_0 t \, u(t), \tag{2.1}$$

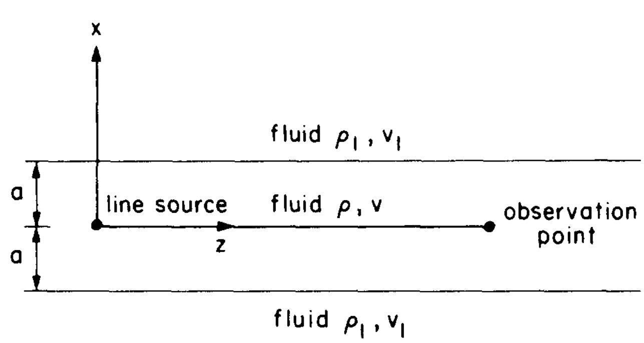


FIG. 1. Geometrical configuration of the problem.

a)This work was supported by the NSF Grant ENG76-01654, the Joint Services Electronics Program under Contract DAAG-29-78-C-0020, and by the Schlumberger-Doll Research Center.

  

where $u(t)$ is 1 for $t \geqslant 0$ and 0 for $t < 0$. The Fourier spectrum of the source excitation function is then

$$X(\omega) = \int_0^\infty dt\, x(t) e^{i\omega t} = \frac{\omega_0}{(\alpha - i\omega)^2 + \omega_0^2}. \tag{2.2}$$

For an observation point at $x = 0$ and a distance $z$ away from the source, the total transient response is

$$\phi(z,t) = -\frac{1}{2\pi^2} \mathrm{Re}\bigg( \int_0^\infty d\omega\, e^{-i\omega t} X(\omega) \times \int_{\mathrm{SIP}} dk_z \exp(ik_z z) \frac{g(\omega,k_z)}{f(\omega,k_z)} \bigg), \tag{2.3}$$

where the integration for $k_z$ follows the Sommerfeld integration path (SIP) which is slightly above the negative real $k_z'$ axis for $k_z' \leqslant 0$ and slightly below the positive real $k_z'$ axis for $k_z' > 0$. In (2.3):

$$g(\omega,k_z) = \frac{1}{ik_x} [1 + R_{01} \exp(i2k_x a)], \tag{2.4}$$

$$f(\omega,k_z) = 1 - R_{01} \exp(i2k_x a), \tag{2.5}$$

$$R_{01} = \frac{k_x - bk_{1x}}{k_x + bk_{1x}}, \tag{2.6}$$

$$b = \rho/\rho_1, \tag{2.7}$$

$$k_{1x} = (k_1^2 - k_z^2)^{1/2} = k_{1x}' + ik_{1x}'', \tag{2.8}$$

$$k_x = (k^2 - k_z^2)^{1/2} = k_x' + ik_x'', \tag{2.9}$$

$$k_1 = \omega/v_1, \tag{2.10}$$

$$k = \omega/v. \tag{2.11}$$

For electromagnetic wave exictation, we characterize the slab medium with permittivity $\epsilon$, permeability $\mu$, and the surrounding medium with $\epsilon_1$ and $\mu_1$. All the above formulas still hold except that $b = \mu/\mu_1$, $v = (\mu\epsilon)^{-1/2}$, and $v_1 = (\mu_1\epsilon_1)^{-1/2}$. We shall assume that $v < v_1$. Along the Sommerfeld integration path (SIP) all $k_x'$, $k_x''$, $k_{1x}'$, and $k_{1x}''$ are positive. Here we use a single prime to denote the real part of a variable and a double prime to denote its imaginary part.

We now outline the steps in the modified modal approach. In (2.3), we have a double integral $I$ over both frequencies and wavenumber $k_z$. The first step is the time-harmonic modal approach by deforming the Sommerfeld path of integration (SIP) to the steepest descent path (SDP) in the complex $k_z$ plane. In the process of deformation, residue contributions due to the poles of guided modes and the leaky modes are included, so that

$$I = \int_0^\infty d\omega\, (\text{poles}) + \int_0^\infty d\omega \int_{\mathrm{SDP}} dk_z. \tag{2.12}$$

The next step consists of interchanging order of integration in the second term in (2.12) and having a second deformation in the complex $\omega$ plane to the steepest descent path. Residues of poles that are encountered in deformation are also taken into account. We thus have

$$I = \int_0^\infty d\omega(\text{poles}) + \int_{\mathrm{SDP}} dk_z\, (\text{poles}) + \int_{\mathrm{SDP}} dk_z \int_{\mathrm{SDP}} d\omega. \tag{2.13}$$

Equation (2.13) is the final answer for the modified modal approach. Poles in the first term are on the complex $k_z$ plane and are functions of real $\omega$. Poles in the second term are on the complex $\omega$ plane and are a function of $k_z$ on the SDP. For both sets of poles, we label them by the same mode index $l$.

In Sec. IV, we examine the location of poles in the second term in (2.13) as a function of $k_z$ on SDP. It is shown that for time less than either the head wave arrival time or the direct wave arrival time, the first term and the second term in (2.13) exactly cancel each other and the third term is also zero identically. Causality is thus proved. In Sec. V, we present numerical results of the solution as given in (2.13). The third term is seemingly highly singular in that there is a double pole on the path of integration and also the integrand blows up to the fourth power at the lower limit. Such seemingly singular behavior is dealt with. In Secs. VI and VII, comparisons of results are made with brute force numerical integration and the technique of explicit inversion, respectively.

## III. MODAL APPROACH

In the modified modal approach to be described in this paper, we perform the double integration in (3) by deforming integration paths to the steepest descent paths. We first make the transformation

$$k_z = k_1 \sin\theta, \tag{3.1}$$

$$k_{1x} = k_1 \cos\theta. \tag{3.2}$$

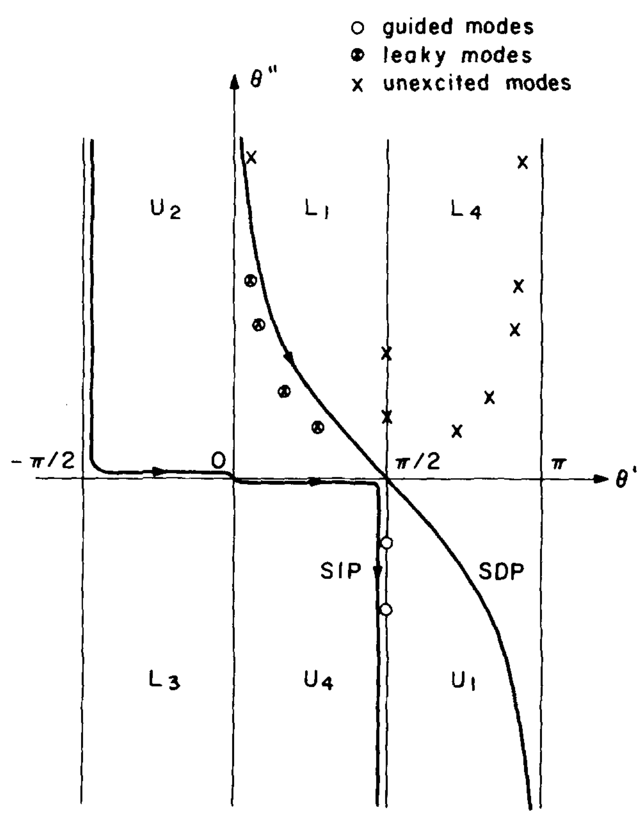


FIG. 2. Location of poles in the complex $\theta$ plane for a fixed frequency $\omega$.

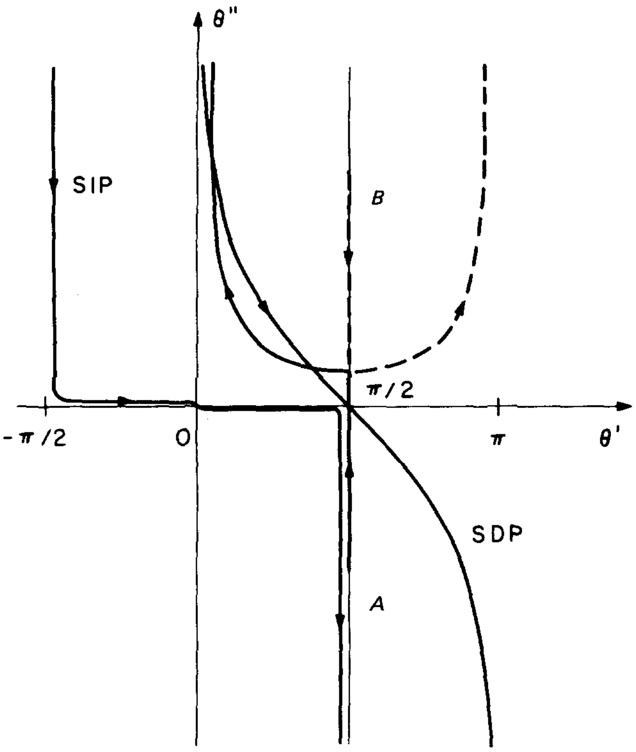


FIG. 3. Root locus $A$ represented by solid line and root locus $B$ represented by dotted line for mode $l = 1$ on the complex $\theta$ plane as a function of frequency. The parameters are $p = 1$, $b = 0.5$; $v = 1.4\times 10^5$ cm/sec, $v_1 = 3.5\times 10^5$ cm/sec, and $a = 2.5$ cm.

In the complex $\theta$ plane, for a fixed $\omega$ the Sommerfeld integration path (SIP) and the steepest descent path (SDP) are both illustrated in Fig. 2. The regions with $k''_{1x} > 0$ are marked $U$ and the regions with $k''_{1x} < 0$ are marked $L$. The subscripts on $U$ and $L$ denote the corresponding quadrants on the complex $k_z$ plane. Poles on the boundary of $U_1$ and $U_4$ represent guided wave modes and poles in region $L_1$ represent leaky wave modes.[1,3,4] The locations of the poles are determined by setting the determinant in the integrand of (2.3) equal to zero. We obtain the modal equation

$$1 - R_{01}\exp(i2k_x a) = 0. \quad (3.3)$$

We let $R_{01} = e^{i\phi}$ with $\phi$ complex and $-\pi < \phi < \pi$. The modal equation becomes

$$2k_x a + \phi = 2l\pi, \quad (3.4)$$

where $l = 0,1,2,\cdots$. Note that $k_x = (k^2 - k_z^2)^{1/2}$. Cutoff for a guided wave mode occurs at $k_z = k_1$ and $\phi = 0$. The cutoff frequency for the $l$th guided mode is seen to be

$$\omega_{l\,\mathrm{cut}} = \frac{l\pi}{a\sqrt{1/v^2 - 1/v_1^2}}. \quad (3.5)$$

We observe that for a fixed $\omega$ and $l\neq 0$ there are two solutions to the modal equation (3.4). As frequency varies, the locations of the two solutions on the complex $\theta$ plane also varies. We denote their loci $A$ and $B$ (Fig. 3). For $\omega > \omega_{l\,\mathrm{cut}}$, solution $A$ is a guided mode lying on the lower side of the vertical axis of $\theta' = \pi/2$ and $B$ lies on the upper side. As frequency $\omega$ decreases, $A$ and $B$ moves toward the $\theta = \pi/2$ and meets at $\omega = \omega_{l\,\mathrm{dou}}$. As frequency further decreases, $A$ moves into region $L_1$ and $B$ moves into region $L_4$. The locus $A$ crosses SDP at $\omega = \omega_{lf}$ and $\omega = \omega_{ls}$. For $\omega_{ls} < \omega < \omega_{lf}$, locus $A$ represents leaky wave modes. For $\omega < \omega_{ls}$ and $\omega_{lf} < \omega < \omega_{l\,\mathrm{cut}}$, locus $A$ represents modes that are not excited because they lie outside the region between SIP and SDP. For these two frequency bands, the poles have no residues and the amplitudes of the corresponding modes are zero. Thus if we restrict to real frequency, each individual mode is in violation of the Paley–Wiener criterion[8] and gives rise to noncausal results.

If we regard the contribution due to the $l$th mode as a summation of residues due to root $A$ over the frequencies when it is excited, we obtain a noncausal mode solution $\tilde{\phi}_l$ with

$$\tilde{\phi}_l(z,t) = \phi_{l1}(z,t) + \phi_{l2}(z,t), \quad l\geqslant 0. \quad (3.6)$$

Notice that there is no leaky mode for $l = 0$ so that $\phi_{02}(z,t) = 0$ where

$$\phi_{l1}(z,t) = -\frac{1}{2\pi^2}\mathrm{Re}\int_{\omega_{l\,\mathrm{cut}}}^{\infty} d\omega\, e^{-i\omega t}X(\omega) \times\left(2\pi i \exp[ik_{lz}(\omega)z]g_1(\omega,k_{lz}(\omega)) \Big/ \frac{\partial f_1}{\partial k_z}(\omega,k_{lz}(\omega))\right) \quad (3.7)$$

is the guided mode contribution to mode $l$ for $\omega_{l\,\mathrm{cut}} < \omega < \infty$, and

$$\phi_{l2}(z,t) = -\frac{1}{2\pi^2}\mathrm{Re}\int_{\omega_{ls}}^{\omega_{lf}} d\omega\, e^{-i\omega t}X(\omega) \times\left(2\pi i \exp[ik_{lz}(\omega)z]g_2(\omega,k_{lz}(\omega)) \Big/ \frac{\partial f_2}{\partial k_z}(\omega,k_{lz}(\omega))\right) \quad (3.8)$$

is the leaky mode contribution to mode $l$ over the frequency range $\omega_{ls} < \omega < \omega_{lf}$. In (3.7) and (3.8), $k_{lz}(\omega)$ is the pole in the $k_z$ plane as a function of $\omega$. The partial derivatives with respect to $k_z$ are taken by keeping $\omega$ constant. We find $k_{lz}(\omega)$ from

$$f(\omega,k_{lz}(\omega)) = 0 \quad (3.9)$$

for the $l$th mode. We used the subscripts 1 and 2 for the functions $f$ and $g$ to distinguish the regions $k''_{1x} > 0$ and $k''_{1x} < 0$. We let

$$g_1(\omega,k_z) = \frac{1}{ik_x}\left(1 - \frac{bk_{1x} - k_x}{bk_{1x} + k_x}\exp(i2k_x a)\right), \quad (3.10)$$

$$f_1(\omega,k_z) = 1 + \frac{bk_{1x} - k_x}{bk_{1x} + k_x}\exp(i2k_x a), \quad (3.11)$$

where $k''_{1x}\geqslant 0$. For $g_2$ and $f_2$ we replace $k_{1x}$ by $-k_{1x}$ in $g_1$ and $f_1$ such that

$$g_2(\omega,k_z) = \frac{1}{ik_x}\left(1 - \frac{bk_{1x} + k_x}{bk_{1x} - k_x}\exp(i2k_x a)\right), \quad (3.12)$$

$$f_2(\omega,k_z) = 1 + \frac{bk_{1x} + k_x}{bk_{1x} - k_x}\exp(i2k_x a), \quad (3.13)$$

where again we have $k''_{1x}\geqslant 0$. The reason for using $g_2$ and $f_2$ is to eliminate the necessity of remembering which function has $k''_{1x}\geqslant 0$ and which has $k''_{1x}\leqslant 0$. All four functions $g_1, f_1, g_2$, and $f_2$ have $k''_{1x}\geqslant 0$. Thus for guided modes $f_1(\omega,k_{lz}(\omega)) = 0$ and for leaky modes $f_2(\omega,k_{lz}(\omega)) = 0$.

## IV. CAUSALITY

In deformation from SIP to SDP, we noticed that on the $k_z$ plane, the SDP is a vertical line. We let

$$k_z = \frac{\omega}{v_1} + iq. \tag{4.1}$$

The complete transient solution $\phi(z,t)$ then consists of the modal fields plus the saddle point contributions. We obtain

$$\phi(z,t) = -\frac{1}{2\pi^2}\,\mathrm{Re}\int_0^\infty d\omega\, e^{-i\omega t} X(\omega)\, F(\omega), \tag{4.2}$$

where

$$F(\omega) = F_1(\omega) + F_2(\omega) + F_3(\omega), \tag{4.3}$$

with

$$F_1(\omega) = \int_0^\infty i dq \exp[i(\omega/v_1)z - qz]\left(\frac{g_1(\omega,k_z)}{f_1(\omega,k_z)} - \frac{g_2(\omega,k_z)}{f_2(\omega,k_z)}\right) \tag{4.4}$$

representing the contribution from the SDP,

$$F_2(\omega) = \sum_{l=1}^{\infty}\begin{cases} 2\pi i g_2(\omega,k_{lz}(\omega)) \Big/ \dfrac{\partial f_2}{\partial k_z}(\omega,k_{lz}(\omega)), & \omega_{ls} < \omega < \omega_{lf}, \\ 0, & \text{elsewhere,}\end{cases} \tag{4.5}$$

representing the leaky mode contributions, and

$$F_3(\omega) = \sum_{l=0}^{\infty}\begin{cases} 2\pi i g_1(\omega,k_{lz}(\omega)) \Big/ \dfrac{\partial f_1}{\partial k_z}(\omega,k_{lz}(\omega)), & \omega_{l\,\mathrm{cut}} < \omega, \\ 0, & \text{elsewhere,}\end{cases} \tag{4.6}$$

representing the guided mode contributions. We write

$$\phi(z,t) = \sum_{l=0}^{\infty} \tilde{\phi}_l(z,t) + \phi_s(z,t), \tag{4.7}$$

where $\tilde{\phi}_l(z,t)$ is the modal contribution as discussed in the last section and the SDP contribution is

$$\phi_s(z,t) = -\frac{1}{2\pi^2}\,\mathrm{Re}\int_0^\infty d\omega\, e^{-i\omega t} X(\omega)\, F_1(\omega) = -\frac{1}{2\pi^2}\,\mathrm{Re}\int_0^\infty i dq\, e^{-qz} H(q), \tag{4.8}$$

where

$$H(q) = \int_0^\infty d\omega\, X(\omega) \exp\left[-i\omega\left(t - \frac{z}{v_1}\right)\right]\left(\frac{g_1(\omega,k_z)}{f_1(\omega,k_z)} - \frac{g_2(\omega,k_z)}{f_2(\omega,k_z)}\right) \tag{4.9}$$

in view of (4.4).

We first examine the singularities of the integrand of $H(q)$ in the complex $\omega$ plane as a function of real $q$. For $0 < q < \infty$, we set

$$f_1(\omega,k_z = \omega/v_1 + iq) = 0, \tag{4.10}$$

$$f_2(\omega,k_z = \omega/v_1 + iq) = 0. \tag{4.11}$$

We label the solution corresponding to mode $l$, $\omega_l(q)$ and plot as a function of $q$ in Fig. 4. We note that the trajectory is in the lower-half $\omega$ plane in the pointed portion of the figure. We note that for the $l = 1$ mode, the locus of the solution for (4.11) crosses the real $\omega$ axis at $\omega = \omega_{l\,\mathrm{cut}}$, $\omega_{lf}$, and $\omega_{ls}$ corresponding to $q = 0$, $q_{lf}$, and $q_{ls}$, respectively. This is in accordance with the crossings of the locus $A$ with the SDP in the complex $\theta$ plane. For the $l = 0$ mode whose cutoff frequency is zero, we show the pole locus in the upper $\omega$ plane. Since the integration limits for $d\omega$ is from $\omega = 0$ to $\omega = \infty$, only poles on the right-half plane $\omega' \geqslant 0$ are of concern. The poles for $l = 0$ in the lower half-plane need separate attention.

For the bracket term in the integrand for $H(q)$ in (4.9), we find by using (3.10)–(3.13),

$$D(\omega,k_z) = \left(\frac{g_1(\omega,k_z)}{f_1(\omega,k_z)} - \frac{g_2(\omega,k_z)}{f_2(\omega,k_z)}\right)$$

$$= \frac{-i8bk_{1x}\exp(i2k_xa)}{[(bk_{1x}+k_x) + (bk_{1x}-k_x)\exp(i2k_xa)]\,[(bk_{1x}-k_x) + (bk_{1x}+k_x)\exp(i2k_xa)]}$$

$$= \frac{-i2bk_{1x}}{k_x^2 \sin^2 k_x a + b^2 k_{1x}^2 \cos^2 k_x a}. \tag{4.12}$$

We note that as $\omega \to \infty$ for a fixed $q$,

$$\lim_{\omega\to\infty} k_x^2 = \lim_{\omega\to\infty} \left[ \frac{\omega^2}{v^2} - \left( \frac{\omega}{v_1} + iq \right)^2 \right] = \frac{\omega^2}{v^2} \cos^2\theta_c, \tag{4.13}$$

where $\theta_c$ is the critical angle for which

$$\theta_c = \sin^{-1} \frac{v}{v_1}. \tag{4.14}$$

Thus at a fixed $q$ as $\omega \to \infty$,

$$\lim_{\omega\to\infty} |D(\omega,k_z)| \propto \exp(-2|k_x''|a) = \exp\left( -2 \frac{|\omega''|a}{v} \cos\theta_c \right). \tag{4.15}$$

By combining this decay factor with the other decay factors in the integrand of (4.9), we can determine whether we should deform the contour of integration upward or downward. For $\omega'' > 0$, we find the total decay factor to be $\omega''(t - z/v_1 - 2a \cos\theta_c/v)$ and for $\omega'' < 0$ it is $\omega''(t - z/v_1 + 2a \cos\theta_c/v)$. We conclude that

(1) For $t < z/v_1 - (2a/v) \cos\theta_c$, we can only deform upward.

(2) For $t > z/v_1 + (2a/v) \cos\theta_c$, we can only deform downward,

(3) For $z/v_1 - (2a/v) \cos\theta_c < t < z/v_1 + (2a/v) \cos\theta_c$, we can either deform upward or downward.

Notice in particular that

$$t_h = \frac{z}{v_1} + \frac{2a}{v} \cos\theta_c \tag{4.16}$$

is the arrivial time of the head wave.

We shall now show that our total transient response is causal. Notice that the arrival time of the direct wave is $t_d = z/v$ and the arrival time of the head wave is $t_h = z/v_1 + 2a \cos\theta_c/v$. By causal, we mean that $\phi(z,t) = 0$ for $t < t_h$ or $t < t_d$ depending on whether $t_h < t_d$ or $t_d < t_h$.

We first deform the original integration path from the positive real axis $\omega'$ to the positive imaginary axis $\omega''$. The residues due to poles on the upper $\omega$ plane in the first quadrant contribute to the integral. We set $\omega = ip$ along the imaginary $\omega''$ axis. Therefore, for $t < t_h$,

$$H(q) = \int_0^\infty idp \exp[p(t - z/v_1)] X(\omega = ip) D\left[\omega = ip, k_z = i\left(\frac{p}{v_1} + q\right)\right] + 2\pi i \sum_{l=0}^{\infty} S_l(q) + 2\pi i \sum_{l=1}^{\infty} T_l(q), \tag{4.17}$$

where

$$S_l(q) = \exp[-i\omega_l(q)(t - z/v_1)] X(\omega_l(q)) \frac{g_1(\omega_l(q), k_z = \omega_l(q)/v_1 + iq)}{(\partial f_1/\partial\omega)(\omega_l(q), k_z = \omega_l(q)/v_1 + iq)} \tag{4.18}$$

and

$$T_l(q) = \begin{cases} -\exp[-i\omega_l(q)(t - z/v_1)] X(\omega_l(q)) \dfrac{g_2(\omega_l(q), k_z = \omega_l(q)/v_1 + iq)}{(\partial f_2/\partial\omega)|(\omega_l(q), k_z = \omega_l(q)/v_1 + iq)}, & \text{for } q_{lf} < q < q_{ls}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.19}$$

The minus sign for $T_l(q)$ is due to the fact that $D = g_1/f_1 - g_2/f_2$. The partial derivatives with respect to $\omega$ are taken by keeping $q$ constant. The solutions for the $l$th mode are expressed as $\omega_l(q)$,

$$f_1\left(\omega_l(q), k_z = \frac{\omega_l(q)}{v_1} + iq\right) = 0, \qquad f_2\left(\omega_l(q), k_z = \frac{\omega_l(q)}{v_1} + iq\right) = 0$$

for all $q$. We note that $\omega_l(q)$ is an analytic function of $q$.

Substituting (4.19) in (4.8), we obtain for $t < t_h$,

$$\phi_s(z,t) = \frac{1}{2\pi^2} \mathrm{Re}\left\{ \int_0^\infty dq\, e^{-qz} \int_0^\infty dp \exp[p(t - z/v_1)] D(\omega,k_z) \right\} + \sum_{l=0}^{\infty} \phi_{l3}(z,t) + \sum_{l=1}^{\infty} \phi_{l4}(z,t), \tag{4.20}$$

where

$$\phi_{l3}(z,t) = -\frac{1}{2\pi^2} \mathrm{Re} \int_0^\infty idq\, e^{-qz} 2\pi i S_l(q), \tag{4.21}$$

$$\phi_{l4}(z,t) = -\frac{1}{2\pi^2}\,\mathrm{Re}\int_{q_{lf}}^{q_{ls}} i\,dq\,e^{-qz}2\pi i T_l(q). \tag{4.22}$$

We note that on the imaginary $\omega$ axis

$$\omega = ip, \tag{4.23}$$

$$k_z = i\left(\frac{p}{v_1}+q\right). \tag{4.24}$$

$p$ and $q$ are real and nonnegative.

$$k_{1x} = \sqrt{q\left(q+\frac{2p}{v_1}\right)} \quad \text{is pure real} \tag{4.25}$$

and

$$k_x = \left[-\frac{p^2}{v^2}+\left(\frac{p}{v_1}+q\right)^2\right]^{1/2} \quad \text{is pure real or pure imaginary.} \tag{4.26}$$

Thus from (4.12) we know that $D(\omega = ip, k_z = ip/v_1 + iq)$ is pure imaginary so that the first term in (4.20) is zero. Therefore, for $t < t_h$,

$$\phi_s(z,t) = \sum_{l=0}^{\infty}\phi_{l3}(z,t) + \sum_{l=1}^{\infty}\phi_{l4}(z,t). \tag{4.27}$$

We now change the integration variable from $q$ to $\omega_l = \omega_l(q)$ in (4.21). Since for all $q$

$$f_1\left(\omega_l(q), k_z = \frac{\omega_l(q)}{v_1}+iq\right) = 0 \tag{4.28}$$

we find

$$\frac{df_1}{dq} = 0 = \frac{\partial f_1}{\partial q}\bigg|_{\omega_l} + \frac{\partial f_1}{\partial \omega_l}\bigg|_q \frac{d\omega_l}{dq} = i\frac{\partial f_1}{\partial k_z}\bigg|_{\omega_l} + \frac{\partial f_1}{\partial \omega_l}\bigg|_q \frac{d\omega_l}{dq}$$

and therefore,

$$dq = \left(i\frac{\partial f_1}{\partial \omega_l}\bigg|_q \bigg/ \frac{\partial f_1}{\partial k_z}\bigg|_{\omega_l}\right) d\omega_l. \tag{4.29}$$

We obtain from (4.21)

$$\phi_{l3}(z,t) = \frac{1}{2\pi^2}\,\mathrm{Re}\int_{\omega_{l\,\mathrm{cut}}}^{\infty} d\omega_l\,2\pi i X(\omega_l)\exp(-i\omega_l t)\exp(ik_{lz}z)g_1(\omega_l,k_{lz}) \bigg/ \frac{\partial f_1}{\partial k_z}\bigg|_{\omega_l}, \tag{4.30}$$

where

$$k_{lz} = \frac{\omega_l}{v_1} + iq. \tag{4.31}$$

The integration contour runs in the complex $\omega$ plane from $\omega_{l\,\mathrm{cut}}$ to $\infty$ along $C_{l3}$ which is the mapping of the roots for mode $l$ for real $q$ satisfying (4.28) (Fig. 5).

Similarly we obtain from (4.22)

$$\phi_{l4}(z,t) = -\frac{1}{2\pi^2}\,\mathrm{Re}\int_{\omega_{lf}}^{\omega_{ls}} d\omega_l\,2\pi i X(\omega_l)\exp(-i\omega_l t)\exp(ik_{lz}z)g_2(\omega_l,k_{lz}) \bigg/ \left(\frac{\partial f_2}{\partial k_z}\bigg|_{\omega_l}\right). \tag{4.32}$$

The integration contour on the complex $\omega$ plane runs from $\omega_{lf}$ to $\omega_{ls}$ along $C_{l4}$ (Fig. 5).

Notice that $\phi_{l4}(z,t)$ cancels $\phi_{l2}(z,t)$ if we deform the integration path from $C_{l4}$ to the real $\omega$ axis because there is no singularity between the paths. We may also deform the integration along $C_{l3}$ for $\phi_{l3}(z,t)$ to the real axis if integration over the arc at $\infty$ vanishes. If the deformation is permitted, then $\phi_{l3}(z,t)$ will cancel $\phi_{l1}(z,t)$. We note that $f_1(\omega,k_z) = 0$ at $\omega_l$ and $k_{lz}$ implies that $g_1(\omega_l,k_{lz}) = 2/ik_{lx}$ does not decay exponentially. Similarly by letting

$$R_{01}\exp(i2k_x a) = 1$$

in the expression for $[\partial f/\partial k_z]_\omega$ we see that $[\partial f/\partial k_z]_\omega(\omega_l,k_{lz})$ does not decay exponentially for $|\omega_l| \to \infty$. According to the modal equation $f_1(\omega_l,k_{lz}) = 0$, $k_{lz} \to \omega_l/v$ + small terms as $|\omega_l| \to \infty$. Thus the exponential decay dependence of the integrand is

$$\exp[-i\omega_l(t - z/v)].$$

Deformation from $C_{l3}$ to real $\omega$ axis is permitted for $t < t_d = z/v$. We conclude that for $t < t_d$, $\phi_{l1}(z,t) + \phi_{l3}(z,t) = 0$.

As a consequence of deforming upward for $t < t_h$, and later by deforming $C_{l3}$ and $C_{l4}$ back to the real axis, we have

(1) $t < t_h < t_d$, $\phi(z,t) = 0.$

(2) $t < t_d < t_h$, $\phi(z,t) = 0.$

(3) $t_d < t < t_h$, $\phi(z,t) = \sum_{l=0}^{\infty} (\phi_{l1}(z,t) + \phi_{l3}(z,t)).$

Thus the earliest arrival is either the head wave (if $t_h < t_d$) or the direct wave ($t_d < t_h$). Causality is observed.

## V. COMPLETE TRANSIENT SOLUTION

For $t > z/v_1 - 2a\cos\theta_c/v$, we deform downward on the $\omega$ plane to the negative imaginary $\omega$ axis. Besides taking the residues of the poles due to $f_2(\omega,k_z) = 0$, we also take into account singularities due to the source function $X(\omega)$ which has a pole at

$$\omega_x = -i\alpha + \omega_0. \tag{5.1}$$

We find

$$H(q) = M(q) - \sum_{l=1}^{\infty} 2\pi i U_l(q) - \sum_{l=1}^{\infty} 2\pi i V_l(q) - 2\pi i \ \text{Residue}[X(\omega_x)] \exp[-i\omega_x(t - z/v_1)] D(\omega_x, k_z = \omega_x/v_1 + iq) \tag{5.2}$$

where

$$U_l(q) = \begin{cases} -X(\omega_l(q)) \exp[-i\omega_l(q)(t - z/v_1)] g_2(\omega_l(q), k_z = \omega_l(q)/v_1 + iq) \Big/ \left[ \dfrac{\partial f_2}{\partial \omega}\Big|_q \left(\omega_l(q), k_z = \dfrac{\omega_l(q)}{v_1} + iq\right)\right], & \text{for } 0 < q < q_{lf}, \\ 0. & \text{otherwise}, \end{cases} \tag{5.3}$$

$$V_l(q) = \begin{cases} -X(\omega_l(q)) \exp[-i\omega_l(q)(t - z/v_1)] g_2(\omega_l(q), k_z = \omega_l(q)/v_1 + iq) \Big/ \left[ \dfrac{\partial f_2}{\partial \omega}\Big|_q \left(\omega_l(q), k_z = \dfrac{\omega_l(q)}{v_1} + iq\right)\right], & \text{for } q_{ls} < q < \infty, \\ 0. & \text{otherwise}. \end{cases} \tag{5.4}$$

In (5.2), we left out the $l = 0$ term. Due to the strong interaction between the $l = 0$ mode and the negative imaginary $\omega''$ axis which is the steepest descent path on the $\omega$ plane, we denote the combined contributions as $M(q)$ which will be studied in detail later.

The complete transient solution is then found to be

$$\phi_s(z,t) = \Phi(z,t) + \sum_{l=1}^{\infty} \phi_{l5}(z,t) + \sum_{l=1}^{\infty} \phi_{l6}(z,t) + \phi_x(z,t), \tag{5.5}$$

where

$$\Phi(z,t) = -\frac{1}{2\pi^2} \text{Re} \int_0^{\infty} idq\, e^{-qz} M(q), \tag{5.6}$$

$$\phi_{l5}(z,t) = \frac{1}{2\pi^2} \text{Re} \int_0^{q_{lf}} idq\, e^{-qz}\, 2\pi i U_l(q), \tag{5.7}$$

$$\phi_{l6}(z,t) = \frac{1}{2\pi^2} \text{Re} \int_{q_{ls}}^{\infty} idq\, e^{-qz}\, 2\pi i V_l(q), \tag{5.8}$$

and

$$\phi_x(z,t) = \frac{1}{2\pi^2} \text{Re} \int_0^{\infty} idq\, e^{-qz}\, 2\pi i \ \text{Residue}[X(\omega_x)] \exp[-i\omega_x(t - z/v_1)] D(\omega_x, k_z = \omega_x/v_1 + iq). \tag{5.9}$$

Following a similar procedure in obtaining $\phi_{l3}(z,t)$ and $\phi_{l4}(z,t)$ in the last section, we find

$$\phi_{l5}(z,t) = \frac{1}{2\pi^2} \text{Re} \int_{C_{l5}} d\omega_l \exp(-i\omega_l t + ik_{lz} z)\, 2\pi i X(\omega_l) g_2(\omega_l, k_{lz}) \Big/ \left( \frac{\partial f_2}{\partial k_z}\Big|_{\omega_l} \right) \tag{5.10}$$

$$\phi_{l6}(z,t) = \frac{1}{2\pi^2} \text{Re} \int_{C_{l6}} d\omega_l \exp(-i\omega_l t + ik_{lz} z)\, 2\pi i X(\omega_l) g_2(\omega_l, k_z) \Big/ \left( \frac{\partial f_2}{\partial k_z}\Big|_{\omega_l} \right) \tag{5.11}$$

where $C_{l5}$ extends from $\omega_{l\,\text{cut}}$ to $\omega_{lf}$ corresponding to $q = 0$ and $q = q_{lf}$, and $C_{l6}$ extends from $\omega_{ls}$ to $\infty$ corresponding to $q = q_{ls}$ and $q \to \infty$ (Fig. 5).

Thus for $t > z/v_1 - 2a\cos\theta_c/v$, the total response is

$$\phi(z,t) = \Phi(z,t) + \phi_{01}(z,t) + \sum_{l=1}^{\infty} \phi_l(z,t) + \phi_x(z,t), \tag{5.12}$$

where

$$\phi_l(z,t) = \phi_{l1}(z,t) + \phi_{l2}(z,t) + \phi_{l5}(z,t) + \phi_{l6}(z,t). \tag{5.13}$$

The four components of the time-domain modal solution $\phi_l(z,t)$ are given by (4.26), (4.28), (5.10), and (5.11).

We make the following remarks:

(1) The solutions $\phi_{l2}$, $\phi_{l5}$, and $\phi_{l6}$ have the same integrand. The integrand for $\phi_{l1}$ is essentially the same except for the function $g_1$ and $f_1$ instead of $g_2$ and $f_2$. At the point of $q = 0$ or equivalently $\omega = \omega_{l\,\mathrm{cut}}$, $k_{1x} = 0$ and as a consequence $f_1 = f_2$ and $g_1 = g_2$.

(2) We cannot deform $C_{l5}$ to the real frequency axis between $\omega_{lf} < \omega < \omega_{l\,\mathrm{cut}}$ as we have done for $C_{l3}$ and $C_{l4}$ because there is a double root at $\omega = \omega_{l\,\mathrm{dou}}$ and $\partial f/\partial k_z$ vanishes at that point.

(3) Note that the expression (5.5) is also valid for $t > z/v_1 - 2a\cos\theta_c/v$ because we can deform either upward or downward for $z/v_1 - 2a\cos\theta_c/v < t < t_h$. Thus we have two alternative expressions for the total transient solution in this time interval.

From (5.6)

$$\Phi(z,t) = -\frac{1}{2\pi^2}\,\mathrm{Re}\int_0^{\infty} idq\, e^{-qz} M(q) = -\frac{1}{2\pi^2}\int_0^{\infty} dq\, e^{-qz}\,\mathrm{Im}[M(q)]. \tag{5.14}$$

We now examine the roots for $l = 0$, $q$ real of $f_2(\omega, k_z = \omega/v_1 + iq) = 0$. On the lower half-plane, as $q \to \infty$ there are two roots lying on the two sides of the imaginary axis. As $q$ decreases, these two roots approach each other and at $q = q_{0\,\mathrm{dou}}$, they merge to form a double root. As $q$ is further decreased, they both move up the $\omega''$ axis but with different speeds (Fig. 6).

On the negative imaginary axis we set

$$\omega = -ip \tag{5.15}$$

so that

$$k_{1x} = \left[q\left(q - \frac{2p}{v_1}\right)\right]^{1/2}, \tag{5.16}$$

$$k_x^2 = \left(q - \frac{p}{v_1} - \frac{p}{v}\right)\left(q - \frac{p}{v_1} + \frac{p}{v}\right). \tag{5.17}$$

In the region $p < v_1 q/2$, $k_{1x}$ is purely real and $D(\omega = -ip, k_z = -ip/v_1 + iq)$ is purely real. The roots of $f_2 = 0$ for the $l = 0$ mode when appearing on the imaginary axis, always fall on the region where $k_{1x}$ is purely imaginary.

The residues of the roots are proportional to $-g_2/(\partial f_2/\partial\omega)$ and in view of (5.15) are equal to $ig_2/(\partial f_2/\partial p)$. Since the roots are governed by $f_2 = 0$ where $f_1 \neq 0$, the residues are equal to that of $D(\omega, k_z = \omega/v + iq)$ with respect to $\omega$. In view of the fact that $\partial/\partial\omega = i\partial/\partial p$, the residues due to these roots are seen to be purely imaginary because $D$ assumes real values in these regions. The contributions of these poles to $M(q)$ are proportional to $2\pi i$ times the residues, which are purely real numbers. Thus they do not contribute to $\phi(z,t)$ as seen from (5.14). We conclude that when the steepest descent path is taken along the negative $\omega''$ axis, it makes no difference whether we circle above or below the poles or take the principal values of those poles that lie on the imaginary axis. Therefore, we find

$$\Phi(z,t) = -\frac{1}{2\pi^2}\int_0^{\infty} dq\, e^{-qz} I(q) \tag{5.18}$$

with

$$I(q) = \left[\mathrm{Re}\,2\pi V_0(q) + \mathrm{P.V.}\int_{qv_1/2}^{\infty} dp\,\exp[-p(t - z/v_1)]X(-ip)D\left(\omega = -ip, k_z = -\frac{ip}{v_1} + iq\right)\right], \tag{5.19}$$

where

$$V_0(q) = \begin{cases} -X(\omega_l(q))\exp[-i\omega_l(q)(t - z/v_1)]\,g_2\!\left(\omega_l(q), k_z = \dfrac{\omega_l(q)}{v_1} + iq\right) \Big/ \dfrac{\partial f_2}{\partial\omega}\Big|_q\!\left(\omega_l(q), k_z = \dfrac{\omega_l(q)}{v_1} + iq\right), & q > q_{0\mathrm{dou}}, \\ 0, & \text{otherwise}. \end{cases} \tag{5.20}$$

The lower limit in the principal value of the integral in (5.19) is due to the fact that for $0 < p < qv_1/2$, $D(\omega, k_z)$ is purely imaginary and that portion is not contributing to $\Phi(z,t)$. The value of $M(q)$ at $q = q_{0\,\mathrm{dou}}$ when there is a double pole on the imaginary axis presents no problem because $M(q)$ is a continuous function of $q$ for $0 < q < \infty$.

To evaluate the principal value in (5.19), we note that there are in general four poles close to the initial point $p = qv_1/2$. These are the two poles for $l = 0$, $f_2 = 0$ and lying in the lower $\omega$ plane. The other two poles are due to $l = 0$, $f_1 = 0$, and lying on two sides of the imaginary axis on the upper half $\omega$ plane. We shall carry out the modified asymptotic method by subtracting out these pole sigularities.[1]

We make the transformation of variables

$$q = Q/a, \tag{5.21}$$

$$p = v_1(P+Q)/2a, \tag{5.22}$$

to obtain

$$I(q) = \mathrm{Re}\left[2\pi V_0(q) + \mathrm{P.V.}\,\frac{v_1 b}{a^2} e^{-mQ}\sqrt{Q}\int_0^\infty dP\,\frac{e^{-mP}\sqrt{P}}{W(P,Q)}\,X\left(-i\frac{v_1}{2a}(P+Q)\right)\right], \tag{5.23}$$

where

$$m = \frac{v_1(t - z/v_1)}{2a}, \tag{5.24}$$

$$W(P,Q) = k_x^2 \sin^2 k_x a + b^2 k_{1x}^2 \cos^2 k_x a, \tag{5.25}$$

$$k_x^2 = -\frac{1}{4a^2}\left[P^2(R^2-1) + Q^2(R^2-1) + 2PQ(R^2+1)\right], \tag{5.26}$$

$$k_{1x}^2 = -PQ/a^2, \tag{5.27}$$

$$R = v_1/v. \tag{5.28}$$

As $Q \to 0$, the four poles locate at, respectively,

$$P = \left(\frac{Q}{S}\right)^{1/3},\quad \left(\frac{Q}{S}\right)e^{i2\pi/3},\quad \left(\frac{Q}{S}\right)^{1/3}e^{-i2\pi/3},\quad SQ^3, \tag{5.29}$$

where

$$S = \frac{(R^2-1)}{16b^2}. \tag{5.30}$$

The first two poles are due to $f_2 = 0$ and $l = 0$. The last two poles are due to $f_1 = 0$ and $l = 0$ lying on the upper half $\omega$ plane. In Fig. 7 we sketch the locations of the four poles on the complex $P$ plane for various values of $Q$. For $Q < Q_{0\mathrm{dou}} = aq_{0\mathrm{dou}}$, two poles lie on the real $P$ axis. For $Q > Q_{0\mathrm{dou}}$, all four poles lie on the complex $P$ plane

$$Q_{0\mathrm{dou}} = q_{0\mathrm{dou}} a. \tag{5.31}$$

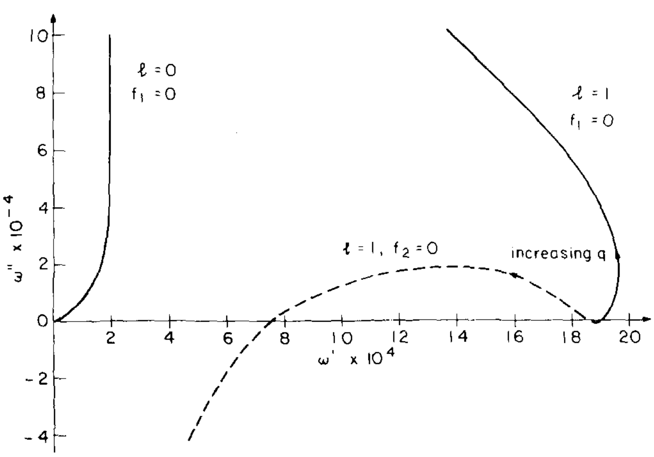


FIG. 4. Location of poles on the complex $\omega$ plane for $l = 0$ and $l = 1$ modes as a function of real $q$. The parameters are $\rho = 1$, $\rho_1 = 2$, $v = 1.4\times10^5$ cm/sec, $v_1 = 3.5\times10^5$ cm/sec, $a = 2.5$ cm, $\omega_0 = 8\pi\times10^4$ sec$^{-1}$, $\alpha = 0.79\omega_0/\pi$, and $z = 40$ cm. The intersections of the $l = 1$ mode trajectory with the real $\omega$ axis are at $q = 0$, $\omega_{1\mathrm{cut}} = 1.92\times10^5$ and $q_{1f} = 0.085$, $\omega_{1f} = 1.864\times10^5$.

Using the modified asymptotic technique, we find that

$$I(q) = \frac{v_1 b}{2} e^{-mQ}\sqrt{Q}\,\mathrm{Re}\Bigg[\frac{1}{2m}\left(\frac{\pi}{m}\right)^{1/2} Y_0(Q) + \sum_{n=1}^{4}\left[\frac{1}{2m}\left(\frac{\pi}{m}\right)^{1/2}\frac{R_n(Q)}{P_n(Q)} + \frac{R_n(Q)}{\sqrt{m}} \times \left[\sqrt{\pi} + i\pi l_n(Q) w(l_n(Q))\right]\right] - 2\pi i\sqrt{P_4(Q)} \times R_4(Q) e^{-mP_4(Q)} u(Q - Q_{0\mathrm{dou}})\Bigg], \tag{5.32}$$

where $u$ is the unit step function. $P_n(Q)$ are the complex location of the four poles with $n = 1,2,3,4$ as a function of $Q$ and $R_n(Q)$ is the residue defined by

$$R_n(Q) = X\left(-i\frac{v_1}{2a}\left[P_n(Q) + Q\right]\right)\Big/\frac{\partial W}{\partial P}\Big|(P_n(Q),Q). \tag{5.33}$$

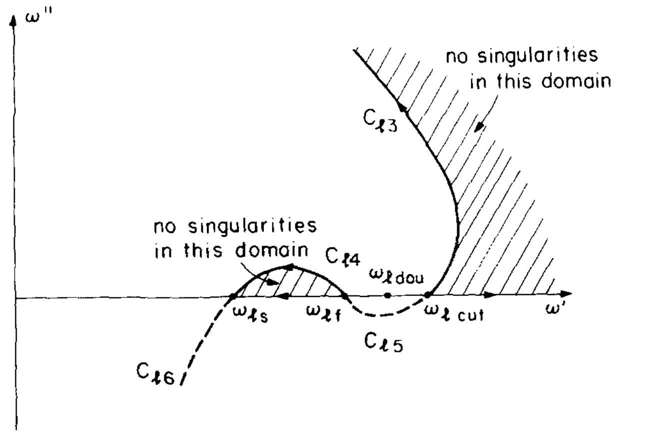


FIG. 5. Contours of $C_{l3}$, $C_{l4}$, $C_{l5}$, and $C_{l6}$ in the complex $\omega$ plane.

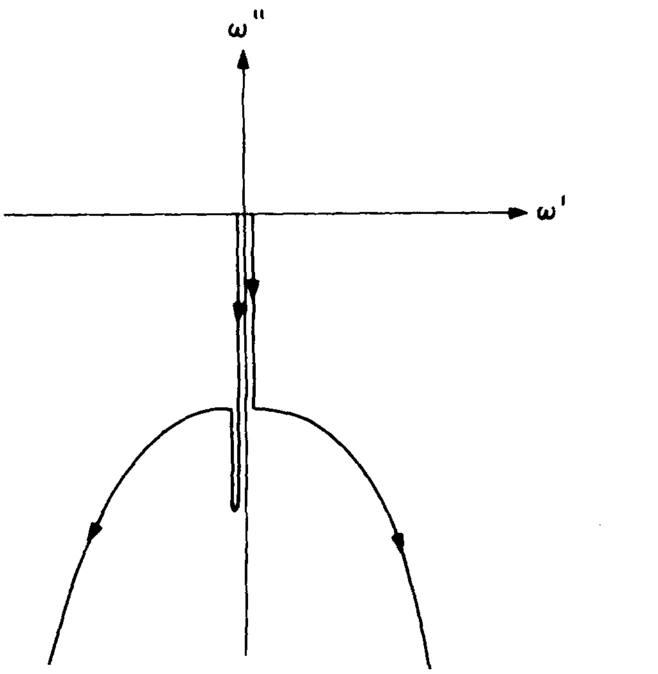


FIG. 6. Root locus for $f_2(\omega,k_z = (\omega/v_1) + iq) = 0$ for real $q$, $0 < q < \infty$ and $l = 0$ on the lower-half $\omega$ plane. For $q$ less than $q_{0dou}$ there are two roots on the negative $\omega''$ axis. For $q$ greater than $q_{0dou}$, they are situated symmetrically on two sides of the $\omega''$ axis.

$$l_n^2(Q) = mP_n(Q), \quad l_n''(Q) \geqslant 0. \tag{5.34}$$

the $w$ function in (5.32) is related to the incomplete error function[12]

$$\mathrm{Re}[\sqrt{P_n(Q)}] \geqslant 0, \tag{5.35}$$

$$Y_0(Q) = X\left(-i\frac{v_1}{2a}Q\right)\Big/ k_{x0}''^2 \sinh^2(k_{x0}''a) \tag{5.36}$$

and

$$k_{x0}'' = \frac{Q(R^2-1)^{1/2}}{2a}. \tag{5.37}$$

The continuity of the function $I(q)$ at $q = q_{0dou}$ can be shown in spite of the presence of a double root on the imaginary axis.

As $Q \to 0$, the locations of the four roots are as follows:

$$P_j(Q) = \left(\frac{Q}{S}\right)^{1/3} e^{i2\pi(j-1)/3}, \quad j = 1,2,3, \tag{5.38}$$

$$P_4(Q) = SQ^3, \tag{5.39}$$

where

$$S = \frac{(R^2-1)^2}{16b^2}, \tag{5.40}$$

$$\left.\frac{\partial W}{\partial P}\right|_{P_j(Q)} = \frac{3Qb^2}{a^2}, \quad j = 1,2,3, \tag{5.41}$$

$$\left.\frac{\partial W}{\partial P}\right|_{P_4(Q)} = -\frac{b^2Q}{a^2}. \tag{5.42}$$

Thus from (5.33), we find

$$R_j(Q) = O\left(\frac{1}{Q}\right), \quad j = 1,2,3. \tag{5.43}$$

From (5.36)

$$Y_0(Q) = O\left(\frac{1}{Q^4}\right) \tag{5.44}$$

which will mean a nonconvergent integral. It can be shown that the singular behavior of $Y_0(Q)$ cancels that of $R_j(Q)/P_j(Q)$ so that as $Q \to 0$,

$$Y_0(Q) + \sum_{j=1}^{4} \frac{R_j(Q)}{P_j(Q)} = O(1). \tag{5.45}$$

Also the $1/Q$ behavior of $R_j(Q)$ cancels each other in $\Sigma_{j=1}^4 R_j(Q)$, so that the dependence of $I(q)$ as $Q \to 0$ is

$$\lim_{Q\to 0} I(q) = -\frac{\pi v_1}{Q^{1/3}b} e^{-mQ} \frac{X(0)}{\sqrt{3}S^{1/6}}. \tag{5.46}$$

As a function of $Q$, $I(q)e^{mQ}$ behaves like $Q^{-1/3}$ as $Q \to 0$ and is continuous at $Q_{0dou}$.

Because of the $Q^{-1/3}$ behavior as $Q \to 0$, to evaluate $\Phi(z,t)$ as given by (5.18), we can use Laguerre's quadrature of $-\frac{1}{3}$ order. The zeros and weights of $L_n^{-1/3}(x)$ are calculated in ways similar to that in Ref. 13.

We can also use ordinary asymptotics to calculate $\Phi(z,t)$ by using (5.46) in (5.14). This gives a closed form solution for $\Phi(z,t)$,

$$\Phi(z,t) \approx \frac{1}{2\pi a}\frac{v_1}{b}\frac{X(0)}{\sqrt{3}S^{1/6}}\frac{\Gamma(\frac{2}{3})}{(z/a+m)^{2/3}}. \tag{5.47}$$

This is a good approximate solution to $\Phi(z,t)$ but not accurate enough to cancel the noncausal behavior of $\phi_{01}(z,t)$.

We remark that the location of the poles are a property of the medium parameters and is independent of $z$ and $t$. Once they are computed, they can be stored and used for all $z$ and $t$. Locations of the poles are calculated with the Newton–Rapson method with the initial approximation being the Taylor expanded value of the pole due to the neighboring $\omega$ or $q$. That is we use the location of the previous $\omega$ or $q$ and then (4.29) to find the initial approximation of the pole of this $\omega$ or $q$.

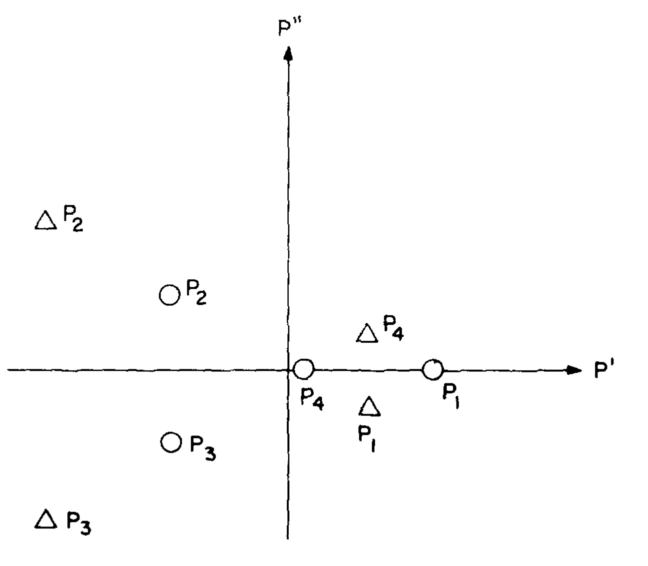


FIG. 7. Locations of our poles in the complex $P$ plane: (1) 0 for $Q < Q_{0dou}$; (2) $\Delta$ for $Q > Q_{0dou}$.

## VI. COMPARISON WITH NUMERICAL INTEGRATION RESULTS

The modified modal solutions as presented in the previous sections can be compared with results obtained by direct numerical calculations. The integration can be carried out along the real $k_z'$ axis and the Laplace contour in the complex frequency plane[10]

$$\phi(z,t) = -\frac{1}{\pi^2}\mathrm{Re}\Big(e^{\omega'' t}\int_0^\infty d\omega'\, e^{-i\omega' t}X(\omega)$$

$$\times \int_0^\infty dk_z \cos k_z z\, \frac{g(\omega,k_z)}{f(\omega,k_z)}\Big), \qquad (6.1)$$

where $\omega = \omega' + i\omega''$ with $\omega''$ positive and finite. We further separate the response from the direct arrival as follows,

$$\phi(z,t) = \phi_r(z,t) + \phi_d(z,t), \qquad (6.2)$$

where

$$\phi_r(z,t) = -\frac{1}{\pi^2}\mathrm{Re}\Big(e^{\omega'' t}\int_0^\infty d\omega'\, e^{-i\omega' t}X(\omega)$$

$$\times \int_0^\infty dk_z \frac{\cos k_z z}{ik_x}\,\frac{2R_{01}\exp(i2k_x a)}{1 - R_{01}\exp(i2k_x a)}\Big) \qquad (6.3)$$

is the response and

$$\phi_d(z,t) = -\frac{1}{\pi^2}\mathrm{Re}\Big(\int_0^\infty d\omega'\, e^{-i\omega t}X(\omega)$$

$$\times \int_0^\infty dk_z \frac{\cos k_z z}{ik_x}\Big) \qquad (6.4)$$

is the direct arrival. We evaluate the integrals in (6.3) by direct numerical integration. For the integrals in (6.4), we evaluate by convolving the source function with the Green's function for the line source

$$\phi_d(z,t) = -\frac{1}{2\pi}\int_0^\infty d\tau\, x(t-\tau)G(z,t),$$

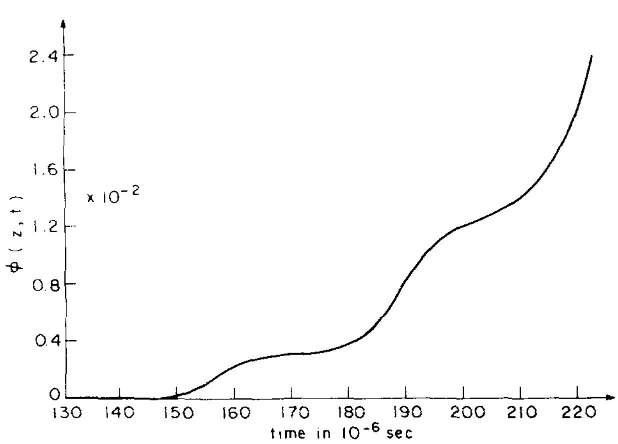


FIG. 8. Transient response up to 220 $\mu$sec after initial excitation for $\rho = 1$, $\rho_1 = 2$, $v = 1.4\times 10^5$ cm/sec, $v_1 = 3.5\times 10^5$ cm/sec, $a = 2.5$ cm, $\omega_0 = 8\pi\times 10^4$ sec$^{-1}$, $\alpha = 0.79\omega_0/\pi$, and $z = 40$ cm.

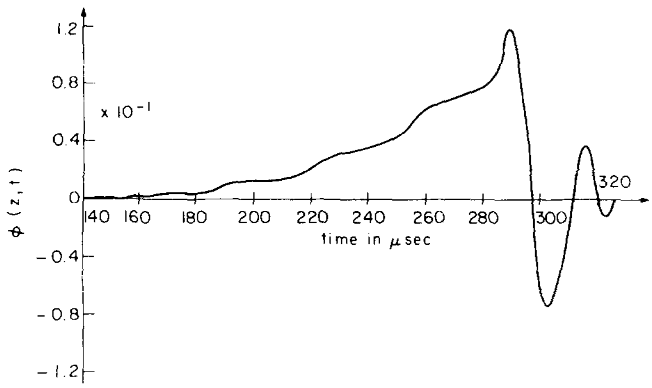


FIG. 9. Transient response up to 325 $\mu$sec after initial excitation.

where

$$G(z,t) = \begin{cases} 0, & \text{for } t < z/v, \\ -2v\; z\Big[\Big(\frac{vt}{z}\Big)^2 - 1\Big]^{1/2}, & \text{for } t > z/v. \end{cases} \qquad (6.5)$$

In Fig. 8, we plot the early arrival of the transient waveform. Notice that for the chosen parameters, the arrival time for the head wave is

$$t_h = \frac{z}{v_1} + \frac{2a\cos\theta_c}{v} = 146\,\mu\text{sec}$$

and the arrival time for the direct wave is

$$t_d = \frac{z}{v} = 286\,\mu\text{sec}.$$

The results obtained with the numerical integration method and the modified modal method are indistinguishable as plotted in Fig. 8. In Fig. 9, we show the complete transient solution to a time of 325 $\mu$sec after initial excitation.

## VII. EXPLICIT INVERSION METHOD

The response $\phi_r(z,t)$ in (6.3) can also be expressed in the form of a convolution of the source function with a Green's function of multiple reflection. The Green's function can be evaluated by using the technique of explicit inversion.[10,11] We write

$$\phi_r(z,t) = -\frac{1}{2\pi}\int_0^\infty d\tau\, x(t-\tau)G_r(z,\tau), \qquad (7.1)$$

where

$$G_r(z,t) = \frac{1}{\pi}\int_L d\omega\, e^{-i\omega t}$$

$$\times \int_{-\infty}^\infty \frac{dk_z}{ik_x}\,\frac{R_{01}\exp(i2k_x a)\exp(ik_z z)}{1 - R_{01}\exp(i2k_x a)} \qquad (7.2)$$

with $L$ denoting the Laplace contour for $-\infty < \omega' < \infty$ with $\omega''$ positive and finite.

We expand the denominator in (7.2) in a power series

$$\frac{1}{1 - R_{01}\exp(i2k_x a)} = \sum_{n=0}^\infty R_{01}^n \exp(i2nk_x a). \qquad (7.3)$$

Let

$$k_z = \frac{\omega}{v}\sin\theta, \tag{7.4}$$

$$k_x = \frac{\omega}{v}\cos\theta. \tag{7.5}$$

Then

$$R_{01}(\omega,\theta) = \frac{\cos\theta - b\,(\sin^2\theta_c - \sin^2\theta)^{1/2}}{\cos\theta + b\,(\sin^2\theta_c - \sin^2\theta)^{1/2}} \tag{7.6}$$

becomes independent of frequency and the $\omega$ dependence of the integrand lies entirely in the exponent. Explicit inversion applied to each term yields

$$G_r(z,t) = -2\sum_{n=1}^{\infty} F_n(t) \tag{7.7}$$

with

$$F_n(t) = \begin{cases} -\dfrac{2v}{R_n}\,\mathrm{Im}\left(\dfrac{R_{01}^n}{\sin\beta}\right), & \text{for } t > 2na/v, \\ 0, & \text{for } t < 2na/v, \end{cases} \tag{7.8}$$

where

$$R_n = [z^2 + (2na)^2]^{1/2} \tag{7.9}$$

and

$$\cos\beta = vt/R_n. \tag{7.10}$$

We notice that $R_n$ is the distance between the observation point and the $n$th image source. From (7.10) we see that $\beta = \beta' + i\beta''$ and for $t < R_n/v$,

$$-\pi/2 \leqslant \beta' \leqslant 0, \quad \beta'' = 0,$$

while for $t > R_n/v$,

$$\beta' = 0, \quad 0 \leqslant \beta'' < \infty.$$

In (7.8)

$$R_{01} = \frac{\gamma_n - b\Omega_n}{\gamma_n + b\Omega_n}, \tag{7.11}$$

$$\gamma_n = \cos(\beta + \theta_{in}), \tag{7.12}$$

$$\Omega_n = [\sin^2\theta_c - \sin^2(\beta + \theta_{in})]^{1/2}, \tag{7.13}$$

$$\sin\theta_{in} = z/R_n. \tag{7.14}$$

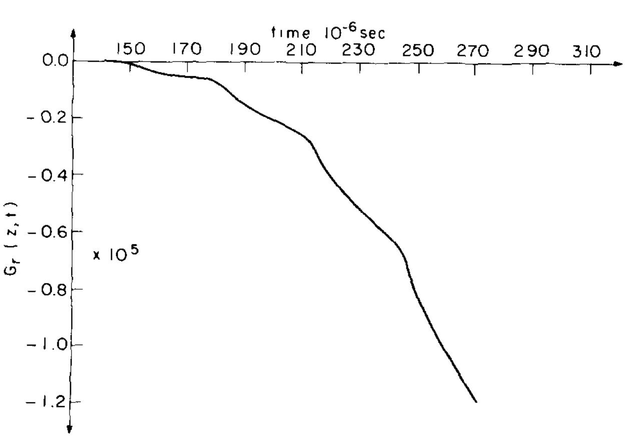


FIG. 10. The function $G_r(z,t)$ up to 280 $\mu$sec.

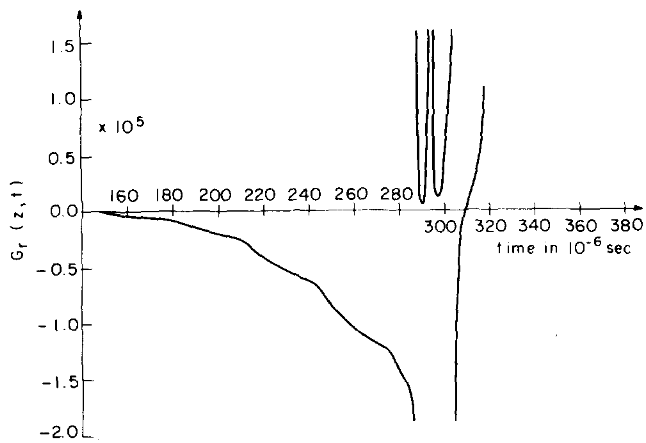


FIG. 11. The function $G_r(z,t)$ up to 320 $\mu$sec.

Thus $\theta_{in}$ corresponds to the observation angle for the $n$th image source. In (7.13) $\Omega_n$ lies in the fourth quadrant so that $0 \leqslant \Omega_n' < \infty$ and $-\infty < \Omega_n'' < 0$.

In Figs. 10 and 11 we plot $G_r(z,t)$ as a function of $t$. We note that head wave arrival times are

$$t_{hn} = \frac{z}{v_1} + \frac{2na\cos\theta_c}{v}, \quad n = 1,2,3,\cdots,$$

where $n$ indicates $n$th image. Thus the first head wave arrival time is at $t_{h1} = 146\,\mu$sec and the latter head waves arrive at time intervals $2a\cos\theta_c/v \approx 32\,\mu$sec apart.

The arrivals of the reflected waves are at

$$t_{rn} = \frac{[z^2 + (2na)^2]^{1/2}}{v}, \quad n = 1,2,3,\cdots.$$

For the chosen parameters in this study, we find $t_{r1} = 288\,\mu$sec and $t_{r2} = 295\,\mu$sec. Such reflected waves give rise to sharp spikes shown in Fig. 11.

Convolving $G_r(z,t)$ with $x(t)$ yields $\phi_r(z,t)$ which is then added to $\phi_d(z,t)$ to obtain the total response $\phi(z,t)$. When plotted the result is indistinguishable from the modified modal approach and the numerical integration approach. Thus confirming the correctness of all three methods.

The severe restriction on the method of explicit inversion is that at making the transformation (7.7), the function $R_{01}$ must become independent of frequency. All $\omega$ dependence can lie only in the exponent or exists as power series. Thus for example the explicit inversion technique will not be applicable to cylindrical geometry where we have the Hankel functions

$$H_n^{(1)}\left[\left(\frac{\omega^2}{v^2} - k_z^2\right)^{1/2} a\right] = H_n^{(1)}\left(\frac{\omega a}{v}\cos\theta\right),$$

$$H_n^{(1)}\left[\left(\frac{\omega^2}{v_1^2} - k_z^2\right)^{1/2} a\right] = H_n^{(1)}\left(\frac{\omega a}{v}(\sin^2\theta_c - \sin^2\theta)^{1/2}\right).$$

The explicit inversion technique is also not applicable to dispersive media, for instance, when

$$\epsilon = \epsilon' + i\sigma/\omega.$$

Then $R_{01}$ will be a function of $\omega$ even after the transformation (7.7)

The modified modal approach, on the other hand, does not seem to have such limitations as long as all singularities in the complex $\omega$ and $k_z$ plane are properly taken care of and may have a promising future for treating more complicated problems.


[1]L.B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves* (Prentice-Hall, Englewood Cliffs, N.J., 1973).
[2]L. Tsang, J.A. Kong, and G. Simmons, "Interference patterns of a horizontal electric dipole over layered dielectric media," J. Geophys. Res. **78**, 3287 (1973).
[3]J.R. Wait, *Electromagnetic Waves in Stratified Media* (MacMillan, New York, 1962).
[4]J.A. Kong, L. Tsang, and G. Simmons, "Geophysical subsurface probing with radio frequency interferometry," IEEE Trans. Antennas Propag. **22**, 616 (1974).
[5]L.M. Brekovskih, *Waves in Layered Media* (Academic, New York, 1960).
[6]M. Ewing, W. Jardetaky, and F. Press, *Elastic Waves in Layered Media* (McGraw-Hill, New York, 1957).
[7]A. Hessel, "General Characteristics of Travelling Wave Antennas," in *Antenna Theory*, edited by Collin and Zucker (McGraw-Hill, New York, 1969), Part 2, Chap. 19, pp. 151–258.
[8]B.P. Lathi, *Signals, Systems, and Communication* (Wiley, New York, 1965).
[9]J.H. Rosenbaum, "The long-time response of a layered elastic medium to explosive sound," J. Geophys. Res. **65**, 1557–614 (1960).
[10]L.B. Felsen, "Propagation and diffraction of transient fields in non-dispersive and dispersive media," edited by L.B. Felsen, Chapter 1 in *Transient Electromagnetic Fields* (Springer-Verlag, Berlin, 1976).
[11]L. Cagniard, *Reflection and Refraction of Progressive Waves*, translated and revised by E.A. Flinn and C.H. Dix (McGraw-Hill, New York, 1962).
[12]Abramowitz and Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
[13]T.S. Shao, T.C. Chen, and R.M. Frank, "Tables of zeros and Gaussian weights of certain associated Laguerre polynomials and the related generalized Hermite polynomials," in Math. Comput. **18**, 598 (1964).

# A method for determining a stochastic transition

John M. Greene

*Plasma Physics Laboratory, Princeton, New Jersey 08544*

(Received 6 November 1978)

A number of problems in physics can be reduced to the study of a measure-preserving mapping of a plane onto itself. One example is a Hamiltonian system with two degrees of freedom, i.e., two coupled nonlinear oscillators. These are among the simplest deterministic systems that can have chaotic solutions. According to a theorem of Kolmogorov, Arnol'd, and Moser, these systems may also have more ordered orbits lying on curves that divide the plane. The existence of each of these orbit types depends sensitively on both the parameters of the problem,, and on the initial conditions. The problem addressed in this paper is that of finding when given KAM orbits exist. The guiding hypothesis is that the disappearance of a KAM surface is associated with a sudden change from stability to instability of nearby periodic orbits. The relation between KAM surfaces and periodic orbits has been explored extensively here by the numerical computation of a particular mapping. An important part of this procedure is the introduction of two quantities, the residue and the mean residue, that permit the stability of many orbits to be estimated from the extrapolation of results obtained for a few orbits. The results are distilled into a series of assertions. These are consistent with all that is previously known, strongly supported by numerical results, and lead to a method for deciding the existence of any given KAM surface computationally.

## I. INTRODUCTION

Problems in many branches of physics can be reduced to the study of two-dimensional measure-preserving mappings. In one important application, these mappings are an abstract representation of the simplest nontrivial problem of classical mechanics, the motion of two coupled oscillators.[1] It is an intriguing problem because the corresponding equations are simple and deterministic, with solutions that are either ordered or chaotic. The type of solution depends sensitively on both the parameters of the system and on the initial conditions. The aim of this paper is to illustrate a point of view for the examination of the boundaries between these types of motion. The method adopted here is to first explore the problem empirically with the aid of a computer, and then use this insight to guide analytic calculations.

A number of authors[2-6] have recently written reviews of the subject covered in this paper, so that it need not be introduced in great detail. One physical example will be given here to provide a context for the remainder of this paper. Consider a particle constrained to the surface of a nonsymmetric bowl, i.e., moving in a potential $V(x,y)$ which has a minimum at $x = y = 0$. In general, the particle will go around and around the bowl on some irregular orbit. For ease in picturing and understanding this orbit, its dimensionality can be reduced by one by a method introduced by Poincaré. Consider a time at which the orbit crosses the ray $y = 0$, $x > 0$. This orbit is completely characterized by its position in the two-dimensional phase plane $(x,\dot{x})$ since the requirement $y = 0$ together with the conservation of energy can be used to complete the specification of the orbit in the full phase space. An orbit is then conveniently pictured through its successive intersections with this plane.

The orbits running around the bowl from intersection to intersection of the phase plane $(x,\dot{x})$ determine a mapping of the phase plane onto itself. By one of Poincaré's invariants, the area of a bundle of orbits is conserved in this mapping. Mappings with this area preserving property can be constructed analytically and these show the full range of orbit types as those arising from Hamiltonian differential equations. Thus, they represent a very convenient abstraction of dynamics, since they can be evaluated rapidly and accurately.

This paper is devoted to the study of a particular mapping that was introduced by Taylor,[7] and more recently treated extensively by Chirikov.[4] Termed "the standard mapping" by the latter author, it is

$$r_{n+1} = r_n - \frac{k}{2\pi}\sin 2\pi\theta_n, \quad \theta_{n+1} = \theta_n + r_{n+1}. \quad (1)$$

It transforms a point $(r_n,\theta_n)$ to the point $(r_{n+1},\theta_{n+1})$. In this space an orbit is a sequence of points generated by successive iterations of the mapping on an initial point $(r_0,\theta_0)$. One iteration of this mapping is thus analogous to one traversal of the particle around the bowl in the previous example. For this reason, the number of iterations that generate an orbit segment will be called the length of that segment.

The mapping of Eq. (1) is naturally periodic in both $\theta$ and $r$ with unit period. Thus, the domain $0 \leqslant r < 1$, $0 \leqslant \theta < 1$ will be treated as a torus.

Consider the standard mapping for the value $k = 0$. Then $r$ is a constant of the motion, and the mapping is integrable.[8] Orbits on invariant curves where $r$ is rational close on themselves after a finite number of iterations of the mapping, and thus are periodic. Surfaces with irrational $r$ are filled ergodically as the orbits are extended indefinitely.

 0022-2488/79/061183-19$01.00  

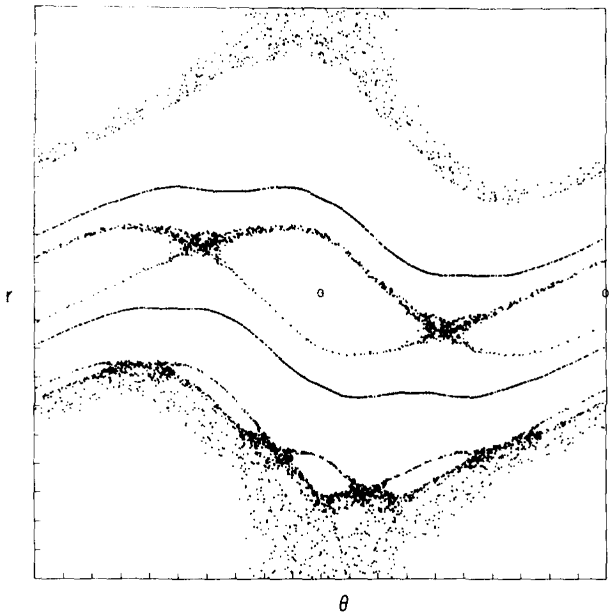


FIG. 1. Five orbits for the standard mapping with $k = 0.97$.

Five typical orbit segments for nonvanishing $k$ are shown in Fig. 1. Now a type of orbit appears that did not exist for the integrable case, one that apparently randomly fills out some area of the torus. These will be called stochastic, or two-dimensional orbits. Two such orbits are illustrated in Fig. 1. The two types of orbits that appeared in the integrable case are also found when $k$ is finite. According to a theorem of Kolmogorov, Arnol'd, and Moser,[9,10] for sufficiently small but finite $k$, there are orbits filling surfaces that in the limit as $k$ vanishes go continuously into surfaces with irrational $r$. These are one-dimensional orbits, or KAM surfaces. Two orbits of this type are illustrated in Fig. 1. Finally, according to a theorem of Poincaré and Birkhoff,[5] surfaces with rational $r$ are reduced to a finite number of periodic orbits when $k$ is nonvanishing. In fact, in Appendix A it is shown that for this mapping two such orbits survive from each rational surface. These orbits can be called zero-dimensional. One orbit of this type is denoted by the symbol 0 in Fig. 1. The resultant mapping, then, is a complex mixture of zero-, one-, and two-dimensional orbits.

Two KAM surfaces extending around the $r$, $\theta$ torus in the $\theta$ direction divide the torus in two. This divides the orbits into two classes, since by continuity and uniqueness, orbits in one region cannot cross the bounding KAM surfaces into the other region. Under this circumstance, two stochastic orbits such as shown in Fig. 1 are distinct and disconnected. Thus, these orbits cannot wander around the torus in the $r$ direction.

On the other hand, for sufficiently large values of $k$, orbits are seen to encircle the $r$, $\theta$ torus in the vertical, or $r$ direction. This behavior will be called connected stochasticity. The presence of connected stochasticity precludes the existence of horizontally encircling KAM surfaces.

We are thus led to the following picture. For small values of $k$, there are many KAM surfaces that encircle the $r$, $\theta$ torus horizontally. These divide the space into many compartments, each of which may contain stochastic orbits. For larger values of $k$, there are fewer such KAM surfaces, and individual stochastic orbits can occupy a greater area of the phase space. Finally, at some critical value of $k$, the last horizontally encircling KAM surface disappears. For larger values of $k$ there are vertically, encircling stochastic orbits. A major purpose of this paper is to calculate and describe the critical $k$.

Previous work on this problem for this mapping has been summarized by Chirikov.[4]

The method of approach used in this paper was first studied several years ago.[11] It is based on an examination of the stability of periodic orbits. These orbits are attractive points of departure since they are of finite length, and thus can be treated with arbitrary accuracy. It is shown here that there is necessarily a close relation between the stability of these orbits and the existence of nearby KAM surfaces.

This paper is an improvement over the previous paper in several respects. The particular mapping studied here is superior. There is now a well-defined problem for finding the critical $k$ for connected stochasticity that has no simple analog in the previous mapping. Also, this mapping is continuously connected to an integrable mapping through the parameter $k$, which is very useful conceptually. Among other developments that have been helpful is a new formulation of the problem of calculating the stability of periodic orbits given by Bountis and Helleman[12] that sheds considerable new insight. This is discussed in the next section, and described in more detail in Appendix B. Finally, present computers have permitted the calculation and sifting of much more data, allowing stronger statements of results to be given.

The quantities to be calculated in this paper are defined and discussed in Sec. II. The results of many numerical calculations are then distilled into a series of assertions given in Sec. III. Some of the evidence leading to these assertions is given in Sec. IV. Finally, the meaning of it all is discussed in Sec. V.

## II. DEFINITIONS

The material in this section has considerable overlap with similar material in the previous paper.[11] It is included here for completeness, and also to point out certain differences between these two papers.

Here we are interested particularly in the periodic orbits of the mapping given in Eq. (1). A periodic orbit is a finite set of points that transform among themselves under iteration of the mapping, and all of which are accessible from any one of the points. We will say that the orbit is of length $Q$ if the orbit closes after $Q$ iterations.

Not all of the periodic orbits are considered here. The class of interest can be defined succinctly as those periodic orbits that exist for all values of the parameter $k$, down to $k = 0$. Some of the other periodic orbits bifurcate out of shorter periodic orbits at a finite value $k$, and some just suddenly appear as $k$ is increased. One way of classifying all these orbits is through the bifurcation tree that produced them, as $k$ is varied from zero. This classification was called a hierarchy in the previous paper. Hopefully, it is somewhat clearer here, where the mapping can be connected to an integrable mapping by a continuous transformation, i.e., through variation of $k$.

A method for calculating all the periodic orbits of interest is given in Appendix A.

Similarly, attention in this paper is focussed only on those KAM surfaces that encircle the torus. Other KAM surfaces bifurcate out of periodic orbits, interspersed with the bifurcated periodic orbits discussed above. These, however, provide only limited impediment to the diffusion of many orbits, since they do not encircle the torus. The KAM surfaces of interest can also be defined as those that exist down to $k = 0$. The conclusion of this paper is that there is a close connection between the KAM surfaces and the periodic orbits that exist together down $k = 0$.

By extension, the KAM surfaces that bifurcate out of a periodic orbit are related to the interspersed periodic orbits that successively bifurcate out of the given orbit. All of the periodic orbits and KAM surfaces on a given branch of a bifurcation tree should be considered together as a system.

It is convenient to associate a winding number with the periodic orbits and KAM surfaces of interest. In the integrable limit, $k = 0$, this winding number is $q = 1/r$. For rational $r$, $r = P/Q$ with $P$ and $Q$ relatively prime, $Q$ is the length of the orbit before it closes, and

$$P \equiv \sum_{n=1}^{Q} r_n = \sum_{n=1}^{Q} (\theta_n - \theta_{n-1}) = \theta_Q - \theta_0, \qquad (2)$$

where $r_n$ and $\theta_n$ are the coordinates of the $n$th point of the periodic orbit. Then, from Eq. (1), $P$ and $Q$, and thus

$$q \equiv Q/P \qquad (3)$$

are well-defined and independent of $k$, and can be used to identify a given periodic orbit. Returning to the picture used in the Introduction where an iteration of the mapping was analogous to a traversal of a particle once around a bowl, $Q$ can be regarded as an angle, and it is reasonable to call $q$ a winding number. This winding number can be extended to KAM surfaces in the obvious way.

The nature, behavior, and characteristics of periodic orbits and KAM surfaces are not continuous functions of the winding number, $q$. It is observed that, in the neighborhood of a given periodic orbit, KAM surfaces and other longer periodic orbits are strongly perturbed. In perturbation theory, this effect appears to be a problem of small denominators, where the denominator is a measure of the distance between a perturbing periodic orbit and the region of interest. A good way to take account of this phenomena is to express winding numbers as continued fractions,[13]

$$q = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}}, \qquad (4)$$

where, since $q \geqslant 1$, the $a_n$'s are positive integers. This will be denoted

$$q = [a_0, a_1, a_2, \ldots, a_N]. \qquad (5)$$

Note that $[a_0,a_1,a_2]$ approaches $[a_0,a_1]$ when $a_2$ becomes large. Thus, the magnitude of the partial quotients, $a_n$, is a measure of the degree of isolation of the associated orbit.

The continued-fraction representation is unique up to an ambiguity in the last partial quotient

$$[a_0,a_1,\ldots,a_N] = [a_0,a_1,\ldots,a_N - 1,1] \qquad (6)$$

as can be seen from the definition. There is also an inversion symmetry around the midpoint of the standard mapping, with the result that orbits with winding numbers $q$ and $q/(q-1)$ are essentially identical. In the continued fraction representation, this means that winding numbers $[a_0,a_1,\ldots,a_N]$ and $[1,a_0 - 1,\ldots,a_N]$ are interchangeable. It is amusing that this is symmetric with the natural ambiguity of continued fractions.

It is sometimes useful to place a subscript on $q$ to indicate the number of partial quotients in its continued fraction representation.

Irrational numbers have unique representations as continued fractions, with an infinite number of partial quotients.[13] Thus, these numbers will be denoted $q_\infty$. Successive truncations of the infinite continued fraction yield rational approximations that are called the convergents[13] of $q$. These convergents yield, from among the periodic orbits of the given length or shorter, the one that most nearly approaches the surface of interest.

Other parameters can be calculated to further characterize the periodic orbits. Orbits in the immediate vicinity of the given orbit can be computed in the linear, differential approximation. The domain of this approximation is called the tangent space. It was well treated in the previous paper,[11] but the results will be summarized here.

The tangent space orbit $(\delta r_n, \delta\theta_n)$ at the point $(r_n,\theta_n)$ is given in terms of the initial conditions on the orbit $(\delta r_0, \delta\theta_0)$ at the point $(r_0,\theta_0)$, through a matrix $M$,

$$\begin{pmatrix}\delta\theta_n\\ \delta r_n\end{pmatrix} = M \begin{pmatrix}\delta\theta_0\\ \delta r_0\end{pmatrix}. \qquad (7)$$

This matrix $M$ can be computed as the product of matrices, one for each orbit section. Over the full cycle of a periodic orbit,

$$M = \prod_{i=1}^{Q} \begin{pmatrix}1 - k\cos 2\pi\theta_i & 1\\ -k\cos 2\pi\theta_i & 1\end{pmatrix}. \qquad (8)$$

The property of $M$ of greatest interest is its eigenvalues. These are the Floquet multipliers for the linear difference equation in the periodic tangent space. From the area-preserving property of the mapping,

$$\mathrm{Det} M = 1, \qquad (9)$$

the eigenvalues of $M$ depend only on its trace. As will be seen, to obtain the best analytic properties it is convenient to subtract two from the trace, and then it is convenient to scale it with a factor of $-4$. This leads to a definition of a quantity to be called the residue[11]

$$R = \tfrac{1}{4}(2 - \mathrm{Trace} M). \qquad (10)$$

The eigenvalues of $M$ are given in terms of the residue by

$$\lambda = 1 - 2R \pm 2[R(R-1)]^{1/2}. \quad (11)$$

When

$$0 < R < 1, \quad (12)$$

the eigenvalues are complex with magnitude unity. Under this condition, tangent space orbits, continued over many periods, rotate about the origin on ellipses. If we express

$$\lambda = \exp i\iota, \quad (13)$$

then $\iota$ is the average angle of rotation per period. It is given in terms of the residue,

$$R = \sin^2 \iota/2. \quad (14)$$

When $R < 0$ or $R > 1$, tangent space orbits lie on hyperbolas. Then the periodic orbit is said to be unstable since all the tangent space orbits march off to infinity, except those lying on the eigenvector of $M$ with an eigenvalue less than one.

In the previous paper[11] a theorem of Poincaré's was invoked to show that, for each rational $q$, there are as many periodic orbits with positive residue as there are with negative residue. It follows from the results of Appendix B that, for the mapping treated here, there is always one orbit of each kind when $k$ is small. These two periodic orbits with the same $q$ will be distinguished by $\pm$ superscripts.

For integrable mappings, all except a small, finite number of periodic orbits lie on surfaces composed of periodic orbits. Then there must be a line of periodic orbits in the tangent space also. A necessary condition for this is $R = 0$ ($\lambda = 1$). Thus, for the present mapping, all the residues for the periodic orbits of interest here vanish in limit as $k$ goes to zero.

The positive residue orbits are stable when $k$ is small, but the residues are seen to increase with $k$ and ultimately become larger than unity. At that point the corresponding orbits become unstable. This change in orbit character, from stable to unstable, as $k$ is increased, is the central concept of this paper that will be related to the disappearance of KAM surfaces.

Appendix B presents a result of Bountis and Helleman,[12] that the residue can also be written as the determinant of the $Q \times Q$ matrix,

$$R = -\tfrac{1}{4}\,\mathrm{Det}H \quad (15)$$

$$H = \begin{pmatrix} 2 - k\cos 2\pi\theta_1 & -1 & \cdots & -1 \\ -1 & 2 - k\cos 2\pi\theta_2 & & \\ 0 & & & \\ \cdots & & & \\ -1 & & & 2 - k\cos 2\pi\theta_Q \end{pmatrix}, \quad (16)$$

where $H$ is tridiagonal with additional $-1$'s in the corners.

It is apparent that when $k$ is large, the residue is proportional to $k^Q$. To prove this, one would need only to establish that periodic orbits do not approach either $\theta = \frac{1}{4}$ or $\theta = \frac{3}{4}$, and numerically the opposite tendency is observed.

In Appendix B it is shown that in the limit of small $k$ also, the residue is proportional to $k^Q$.

In the succeeding sections of this paper, the aim is to make sensible statements relating the residues of neighboring periodic orbits of different lengths. Since it appears that the magnitude of the residues is dominated by an exponential dependence on orbit length, it is natural to introduce a new quantity proportional to the $Q$th root of $R$. The new quantity will be called the mean residue, $f$,

$$f \equiv (|R|/\beta)^{1/Q}. \quad (17)$$

The quantity $\beta$ could be adjusted for convenience, as is discussed in Sec. IV C. From the considerations of Sec. IV E, the value $\beta = \frac{1}{4}$ is preferred for the cases of most interest, and that value has been used for all the numerical computations of this paper.

In the previous paper,[11] a slightly different definition of $f$ was used, which was the square of the value used here. For the present mapping, it is clear that it is preferable to have $f$ proportional to the perturbation $k$, rather than its square. In retrospect, this argument should have led to the present definition of $f$ in the previous paper also.

A quantity partially related to the mean residue has been used by other authors.[14,15]

$$h \equiv \frac{1}{Q}\ln|\lambda|, \quad (18)$$

where $\lambda$ is the eigenvalue of Eq. (11). When the residue is large, and therefore the eigenvalue is large,

$$\lambda \simeq -4R \quad (19)$$

so that

$$h \simeq \ln f. \quad (20)$$

However, in distinction to $h$, $f$ is a real analytic function of $k$ for both large and small values of $k$, and is thus considerably more useful in the same way that the residue is more useful than the eigenvalue.

The quantity $f$ can be evaluated for both positive and negative residue orbits. A superscript $\pm$ will be used to indicate the sign of the residue of the orbit.

One further property of the tangent space mapping is useful, that of the shapes of the conic section surfaces that are invariant when the mapping is extended over a full period. Since this quantity depends on more than the trace and determinant of $M$, we introduce the parameterization,[11]

$$M \equiv \begin{pmatrix} a+d & c+b \\ c-b & a-d \end{pmatrix}. \quad (21)$$

The condition on the determinant of $M$ can be written

$$a^2 + b^2 - c^2 - d^2 = 1. \quad (22)$$

First consider unstable periodic orbits so that the invariant surfaces are hyperbolas. It is straightforward[11] to establish that the angle $\gamma$ between the asymptotes of these hyperbolas is given by

$$\tan^2\gamma = \frac{a^2 - 1}{b^2} = \frac{4R(R-1)}{b^2}. \quad (23)$$

Thus, when the residue is small, the hyperbolas degenerate into straight lines, appropriately for integrable systems.

Further, it can be shown that in the stable case, where the invariant conics are ellipses, the same expression is related to the ratio of major, $\rho_+$, to minor, $\rho_-$, semiaxes,

$$\frac{4\rho_+^2\rho_-^2}{(\rho_+^2+\rho_-^2)^2}=\frac{1-a^2}{b^2}=\frac{4R(1-R)}{b^2}. \tag{24}$$

Again, small values of $R$ show the approach to the straight lines of an integrable system.

## III. ASSERTIONS

An outline of the numerical results that have been obtained for the standard mapping is given in this section. The results are presented as a series of assertions, or hypotheses. The evidence for the truth of these assertions will be given in the succeeding section. The emphasis here is on their significance, and on the relation between them. It will be seen that they are not independent, but since they have not been proven, it is undesirable to form a logical structure that is too rigid.

Assertion I:

$$k\leqslant f\leqslant 1+\tfrac{1}{2}k+\tfrac{1}{2}(k^2+4k)^{1/2}\leqslant k+2,$$

$$\frac{\partial f}{\partial k}\geqslant 1,\quad f^+\leqslant f^-.$$

This first assertion applies to each periodic orbit and provides some estimates and bounds on the magnitude of $f$. In fact, for the standard mapping, $f$ is close to linear in $k$. According to the third part of this assertion, the periodic orbit with negative residue has a slightly larger value of $f$ than the associated orbit with positive residue. Note that the upper bound on $f$ is consistent with the bound on the derivative, i.e. $\partial f_{\max}/\partial k\geqslant 1$.

Assertion II: Consider the mean residue, $f$, to be a function of the partial quotients of the continued-fraction representation of the winding number $q$ of a given periodic orbit. Then

$$f(a_0,\ldots,a_i,\ldots,a_N)>f(b_0,\ldots,b_i,\ldots,b_N)$$

if $a_i=b_i+1$ for some set of $b_i$'s that satisfy $b_i=b_{\max}=\max_j(b_j)$, and $a_i=b_i$ otherwise, with the further restriction

$$a_0=a_N=1.$$

Further,

$$\lim_{a_i\to\infty} f(a_0,a_1,\ldots,a_i,\ldots,a_N)>1.$$

This assertion compares orbits whose continued-fraction representations have the same number of partial quotients, and says that the mean residue is increased if any of the largest partial quotients are increased.

The first comment to be made about this assertion is that it is reasonable to consider $f$ to be a function of the partial quotients of $q$ rather than a function of $q$ directly. In fact, $f$ is not a well organized function of $q$ with continuity properties. Its dependence on the partial quotients is more orderly. The general thrust of this assertion is that periodic orbits with large partial quotients have large values of $f$, and thus tend more to instability. In other words, periodic orbits that are close to shorter periodic orbits tend to be more unstable than orbits that are further removed.

Unfortunately, the mean residue is not simply monotonic with respect to each partial quotient. Thus, the desired properties must be expressed in terms of some weaker statement. The most important use of a statement of this kind is the identification of the orbit with the minimum value of $f$, among all those with a given number of partial quotients in the continued-fraction expansion of its winding number. This provides a context, for example, for Assertion IV below. Thus Assertion II seems to be the cleanest statement that is both true and useful in this respect.

The symmetries associated with the partial fraction representation permit the restriction on $a_0$ and $a_N$ without eliminating any significant periodic orbit. Again, unfortunately, this restriction is necessary to avoid counterexamples.

The last part of this assertion, when combined with Assertion VI, leads to the conclusion that there is a stochastic region in the immediate vicinity of every chain of periodic orbits. Note that the bounds of Assertion I are consistent with the present inequality, and prevent the limit from diverging.

Assertion III: Consider an irrational winding number $q$, and its unique continued-fraction representation. Associated with this, consider the series of periodic orbits whose winding numbers are given by the successive truncations, or convergents, of this continued fraction, and calculate the mean residue for each. Then

$$f(q_\infty)\equiv\lim_{N\to\infty} f(q_N)$$

converges nontrivially, where

$$q_N\equiv[a_0,a_2,\ldots,a_N].$$

Further,

$$f^+(q_\infty)=f^-(q_\infty).$$

This assertion continues the definition of $f(q)$ to irrational values of the argument. According to the theory of continued fractions, the orbits that have been used in each approximation have the minimum separation from the chosen irrational orbit among all orbits of a given length. This has a clear meaning, at least, when there is a KAM surface with the chosen irrational $q$. Thus, the irrational is approached through a consistent sequence of rationals.

The statement that the convergence of $f$ is nontrivial means $f(q_\infty,k)$ is not identically unity. In that case, the interesting behavior associated with irrational winding numbers would be exhibited by some other function of the residue.

The second part of this assertion states that the same value of $f$ is achieved if the limit is taken using either the positive or negative residue orbits. While according to Assertion I, the negative residue orbit yields the larger value of $f$

for each finite approximation, the difference disappears in the limit. The significance is that near the irrational surface, associated positive and negative residue orbits will have the same character. See also the comments under Assertion VI.

Assertion IV: Define

$$q^*_\infty \equiv [1,1,1,...] = \tfrac{1}{2}(1 + \sqrt{5}).$$

Then

$$f(q^*_\infty) < f(q_\infty) \quad \text{for all } q_\infty \neq q^*_\infty.$$

The irrational number $q^*_\infty$ has been known for millenia as the golden mean. It has many interesting properties that are discussed in Niven,[13] and also by Gardner.[16] It is the number that is least easily approximable by rationals. It is thus the point at which the problems of small denominators are minimal, and the surface for which the conditions for the KAM theory are most easily satisfied.

Assertion IV follows if Assertion II is true for all sets of orbits with a finite number of partial quotients. Together with Assertion VI, it yields Assertion VII for the boundary of connected stochasticity for this mapping.

The details of this assertion are true only for the standard mapping considered in this paper. It should be borne in mind that the partial quotients determine the position of the orbit with respect to inhomogeneities over the full domain of the mapping, as well as its relation to nearby shorter periodic orbits. The former variations will always have a weak dependence on the partial quotients $a_i$ with large $i$. Thus, this assertion should be relevant, for general mappings, for all except the first few partial quotients. See also the comments on the next assertion.

Assertion V: Consider an irrational winding number $q^\dagger_\infty$ whose partial fraction representation has the property

$$a_i = 1 \quad \text{for all } i \geqslant N.$$

Choose that value of $k = k^\dagger_c$ such that the converged mean residue satisfies

$$f(q^\dagger_\infty, k^\dagger_c) = 1.$$

Then the associated residue converges with the limit,

$$R(q^\dagger_\infty, k^\dagger_c) = 0.25.$$

This assertion can be approached in the following manner. Consider an asymptotic representation of the residues and mean residues for the sequence of convergents of a particular $q_\infty$, in the limit of a large number of partial quotients. The length $Q$ of the associated periodic orbits can be taken to be the large parameter of the expansion. According to Assertion III, this asymptotic expansion for the residues can be written

$$R \simeq \gamma(Q, q_\infty, k) f^Q(q_\infty, k)[1 + \cdots], \tag{25}$$

where $\gamma$ need only satisfy

$$\gamma^{1/Q} \underset{Q\to\infty}{\longrightarrow} 1$$

Then Assertion V can be restated

$$\gamma(q^\dagger_\infty, k^\dagger_c) = 0.25. \tag{26}$$

The irrational numbers $q^\dagger_\infty$ are closely related to the golden mean that appeared in the previous assertion. They are an obvious generalization when attention is fixed on a subregion of a given mapping. It is remarkable that this assertion appears to be true for all such numbers, even though there are counterexamples to its generalization to the full set of irrational numbers. Since the surfaces corresponding to the winding numbers $q^\dagger_\infty$ have a considerable variation in their immediate environment, there is some hope that this assertion could be generalized to other mappings.

Picking up a loose end, the rather arbitrary value $\beta = \frac{1}{4}$ that appears in the definition of $f$, Eq. (17), has little or no influenece on the assertions of this section. Only in Assertions I and II will this number enter, and then, rather weakly.

The significance of this Assertion V is in determining a best value for $\beta$. A value of $\frac{1}{4}$ yields the most rapid convergence, in the sense of Assertion III, for those interesting winding numbers $q^\dagger_\infty$ in the vicinity of the critical limit, $k = k_c$, since the asymptotic expansion for the mean residue, from Eqs. (17) and (25), is

$$f = (\gamma/\beta)^{1/Q} f(q_\infty)[1 + \cdots], \tag{27}$$

and the leading term is independent of $Q$ only when $\beta = \gamma$.

It is interesting that in the preceding paper, similar considerations, if less coherently presented and less accurately evaluated, led also to the conclusion that $\beta$ should be given the value of $\frac{1}{4}$ for the most rapid convergence. It is very tempting to speculate that, for some hidden reason, $\beta = \frac{1}{4}$ is universally the desired value to best determine the KAM surfaces of most interest. The common factor between the corresponding orbits is not at all clear.

Assertion VI: The KAM surface with a given winding number $q_\infty$ exists if and only if

$$f(q_\infty) < 1.$$

This is perhaps the most striking of the various assertions of this section. An intuitive feeling for this assertion can be gleaned from a consideration of the definition of $f$, Eq. (17). When $f(q_\infty)$ is slightly smaller than unity, Assertion III yields the conclusion that the necessarily long, nearby periodic orbits corresponding to the convergents of $q_\infty$ will all have a small residue, $|R| \ll 1$. On the other hand, when $f(q_\infty)$ is slightly larger than unity, these residues will be large, $|R| \gg 1$. Thus, at the critical value of $k$, $R(q_\infty, k)$ will have an infinite discontinuity. It should not be surprising that this discontinuity is associated with other remarkable phenomena. Note that, from Assertion III, this discontinuity occurs simultaneously for positive and negative residue orbits.

Assertion VII: Connected stochasticity occurs for

$$k > k^*_c,$$

where

$$f(q^*_\infty, k^*_c) = 1, \quad k^*_c = 0.971635....$$

From Assertion IV.

Table I. Mean residue for $q = 3$ orbits.

| $k$ | $f^+$ | $f^-$ |
|---|---|---|
| 0.5 | 0.52068 | 0.52118 |
| 1.0 | 1.04378 | 1.05081 |
| 2.0 | 2.09494 | 2.16282 |
| 20.0 | 20.48158 | 21.44352 |
| 21.0 | 21.48851 | 22.46472 |

$$f(q_\infty, k_c^*) > 1, \quad \text{for all } q_\infty \neq q_\infty^*,$$

and thus, from Assertion VI, no other KAM surfaces exist that encircle the torus horizontally in the usual pictorialization. Therefore, for $k > k_c^*$, there are no impediments to orbits encircling the torus vertically. Evidence will be presented that this latter type of orbit does exist, then, at least for $k$'s that are slightly above the critical value. That is, it is shown that there is at most a very small range of the parameter $k$ for which there is neither a vertically encircling stochastic type orbit, nor a horizontally encircling KAM orbit. This is sort of reasonable in the following sense. If there are no vertically encircling orbits, then each orbit must have some upper and lower bound,

$$r_L(\theta) < r_i < r_U(\theta).$$

These bounds can be intuitively identified with KAM surfaces.

## IV. EVIDENCE

### A. Assertion I

The upper limit on $f$ has been obtained by considering the Jacobian matrix $M$ of Eq. (8). For any given length orbit, the trace of this matrix must always be less in absolute magnitude than the trace of the matrix obtained from an orbit restricted to $\theta = \frac{1}{2}$, i.e.,

$$|\text{Trace}M| < \text{Trace}\begin{pmatrix} 1+k & 1 \\ k & 1 \end{pmatrix}^Q = \lambda^Q + \lambda^{-Q}. \quad (28)$$

The inequality follows from the fact that the product of matrices on the right then involves the sum of positive terms, each of which has been maximized over conceivable orbits. The trace on the right has been evaluated by diagonalization, and thus represented in terms of the largest eigenvalue of each factor,

$$\lambda = 1 + \tfrac{1}{2}k + \tfrac{1}{2}(k^2 + 4k)^{1/2}. \quad (29)$$

When this estimate is used with the definition of $f$, an upper bound is obtained that decreases with orbit length. Since the general trend of Assertions II and III is that $f$ does not decrease with the orbit length, the leap has been made to minimize this bound over $Q$, with the result given in Assertion I.

A few periodic orbits are independent of $k$. Thus, they are easily found. For $q = 1$, the positive residue orbit is

$$r = 0, \quad \theta = 0 \quad (30)$$

and the negative residue orbit is

$$r = 0, \quad \theta = \tfrac{1}{2}. \quad (31)$$

Table II. Mean residue for various orbits with $k = 1$.

| $q$ | $Q/P$ | $f^+$ |
|---|---|---|
| [1,1,1,1,1] | $\frac{8}{5}$ | 1.0325 |
| [1,1,1,2,1] | $\frac{11}{7}$ | 1.0638 |
| [1,1,2,1,1] | $\frac{12}{7}$ | 1.0378 |
| [1,2,1,1,1] | $\frac{11}{8}$ | 1.1139 |
| [1,1,2,2,1] | $\frac{17}{10}$ | 1.0475 |
| [1,2,1,2,1] | $\frac{15}{11}$ | 1.1447 |
| [1,2,2,1,1] | $\frac{17}{12}$ | 1.0967 |
| [1,2,2,2,1] | $\frac{24}{17}$ | 1.1070 |
| [1,1,1,3,1] | $\frac{14}{9}$ | 1.1020 |
| [1,1,3,1,1] | $\frac{16}{9}$ | 1.0796 |
| [1,3,1,1,1] | $\frac{14}{11}$ | 1.2290 |

The corresponding mean residues are easily evaluated yielding

$$f^{\pm}(1,k) = k. \quad (32)$$

Similarly, the positive residue orbit for $q = 2$ is

$$r = \tfrac{1}{2}, \quad \theta = 0; \quad r = \tfrac{1}{2}, \quad \theta = \tfrac{1}{2} \quad (33)$$

and the corresponding mean residue is

$$f^+(2,k) = k. \quad (34)$$

These orbits thus test the lower bound on $f$ and its derivative.

The expansion for small $k$ given in Appendix B can yield as many asymptotic values of $f$ as one cares to evaluate. In this limit, the positive and negative residue orbits yield the same mean residues. A few such values are

$$f(3) = (\tfrac{9}{8})^{1/3}k,$$
$$f(4) = (\tfrac{5}{3})^{1/4}k,$$
$$f(5) = [(1675 + 375\sqrt{5})/768]^{1/5}k,$$
$$f(\tfrac{5}{2}) = [(1675 - 375\sqrt{5})/768]^{1/5}k. \quad (35)$$

Finally, in Table I, a few values of the mean residue are given for orbits with the winding number $q = 3$.

Table III. Residues and mean residues of golden mean convergents, $k = 0.971635$.

| $Q/P$ | $f^+$ | $R^+$ | $f^-$ | $R^-$ |
|---|---|---|---|---|
| 89/55 | 0.99998014 | 0.24956 | 1.000217 | − 0.25488 |
| 144/89 | 1.00001090 | 0.25039 | 1.000158 | − 0.25574 |
| 233/144 | 0.99999772 | 0.24987 | 1.000088 | − 0.25520 |
| 377/233 | 1.00000177 | 0.25017 | 1.000058 | − 0.25551 |
| 610/377 | 0.99999965 | 0.24995 | 1.000034 | − 0.25528 |
| 987/610 | 1.00000009 | 0.25002 | 1.000021 | − 0.25537 |

Table IV. Mean residues of golden mean convergents, $k = 0.9$.

| $Q/P$ | $f^+$ | $f^-$ |
|---|---|---|
| 55/34 | 0.92427 | 0.92428 |
| 89/55 | 0.92409 | 0.92409 |
| 144/89 | 0.92406 | 0.92406 |
| 233/144 | 0.92401 | 0.92401 |

Altogether, the results of this subsection should provide some feeling for the typical behavior of $f$ as a function of $k$. The most noteworthy result is that the positive residue increases monotonically as a function of $k$. It passes through unity, and thus the orbit becomes unstable without hesitation or indication of nonanalyticity in these parameters.

## B. Assertion II

In Table II, a series of values of mean residue is presented for a number of orbits whose winding number $q$ has five partial quotients. It is seen that minimizing the partial quotients yields the minimum mean residue. On the other hand, the fifth in the list is smaller than the second. This shows the difficulty of making useful true statements, without invalidating Assertion II.

The limit of a given partial quotient tending to infinity that is considered in the second part of this assertion is quite interesting. The winding number in this limit approaches the continued fraction that is truncated at the term $i-1$, as is clear from the continued fraction representation. Experimentally, the corresponding limiting orbits are closely associated with the negative residue orbit with the truncated widing number. This latter orbit will be referred to as the truncated orbit. It is observed that the limiting orbit approaches the homoclinic points of the truncated orbit[17] where each set of homoclinic points is defined as an orbit of infinite length that asymptotically, at either end of its trajectory, approaches points of the truncated orbit.[5] Thus, orbits that are close to the truncated orbit at several points will also be close to corresponding homoclinic points.

Now consider a set of limiting orbits with increasing partial quotients, as in the second part of Assertion II. These have an increasingly long residence near the truncated orbit with short bridges from one orbit section to the next. The beginning of this process can be seen in Fig. 3 of Ref. 11. The contributions to the Jacobian matrix $M$ from orbit sections neighboring the truncated orbit can be calculated as powers of the Jacobian matrix associated with the truncated orbit, and the contribution from the bridges is independent of $a_i$, when $a_i$ is large. As a result,

$$R(q=Q/P) \simeq (\lambda_t^{Q/Q_t} + \lambda_t^{-Q/Q_t})c \tag{36}$$

and

$$\lim_{a_i \to \infty} f(a_n) = (\lambda_t)^{1/Q_t}, \tag{37}$$

where $\lambda_t$ is the largest eigenvalue of the truncated orbit, $Q_t$ is the length of that orbit, and $c$ is a constant, for large $a_i$, associated with the bridges. Since the eigenvalue $\lambda_t$ is always larger than one, the limiting $f$ is also larger than one.

## C. Assertion III

Table III presents data relating to the convergence of the mean residue $f$ for a sequence of convergents to the golden mean. These have been chosen because the golden mean is the winding number of the greatest interest. The value of $k$ given here is the best approximation to the critical $k$ that has been evaluated. Attention for the moment should be focussed on the mean residue $f$, to the exclusion of $R$. It has been evaluated for both positive and negative residue points.

For the positive residue orbits, the convergence is oscillatory and the differences decrease approximately as $Q^{-2}$. Since $Q$ increases exponentially from convergent to convergent as powers of the golden mean, the convergence of $f$ is quite rapid and convincing.

The negative residue orbit exhibits convergence from above, the differences decrease more slowly, and they are more nearly proportional to $Q^{-1}$. The problem is that the value of $\beta$ in the definition of $f$ has been chosen to maximize the convergence of the positive residue orbit. For this set of orbits, the quantity $\gamma$ of Eq. (25) is slightly larger than 0.25, as can be seen from Table III. A slightly larger value of $\beta$ would provide faster convergence, without affecting the converged value, $f^-(q_\infty^*)$. Note also that, to within the limits of accuracy of the calculation, the converged $f$'s for the positive and negative residue orbits are identical.

As a further example, in Table IV the golden mean is again considered, but for a somewhat smaller value of $k$. Again monotonic convergence with differences proportional to $Q^{-1}$ is observed. It thus appears that the optimizing $\beta$ should be a function of $k$ as well as of the orbit.

Part of the reason for the excellent convergence of these cases lies in the regularity of the succeeding partial quotients. Orbits with random partial quotients are probably less interesting in light of Assertion II, and they are more difficult to calculate. The problem is that orbits become long very fast, there is a limited window

$$10^{-10} < |R| < 10^{9}$$

within which residues can be calculated accurately because of roundoff, and the residue depends exponentially on orbit length.

There are a couple of possibilities for generalizing the results that have been given here. Perhaps the convergence

Table V. Residue for convergents of $q = (143 + \sqrt{5})/38$, $k = 0.834365$.

| $Q/P$ | $R^+$ |
|---|---|
| 730/191 | 0.24766 |
| 1181/309 | 0.25166 |
| 1911/500 | 0.24924 |
| 3092/809 | 0.25079 |
| 5003/1309 | 0.24995 |

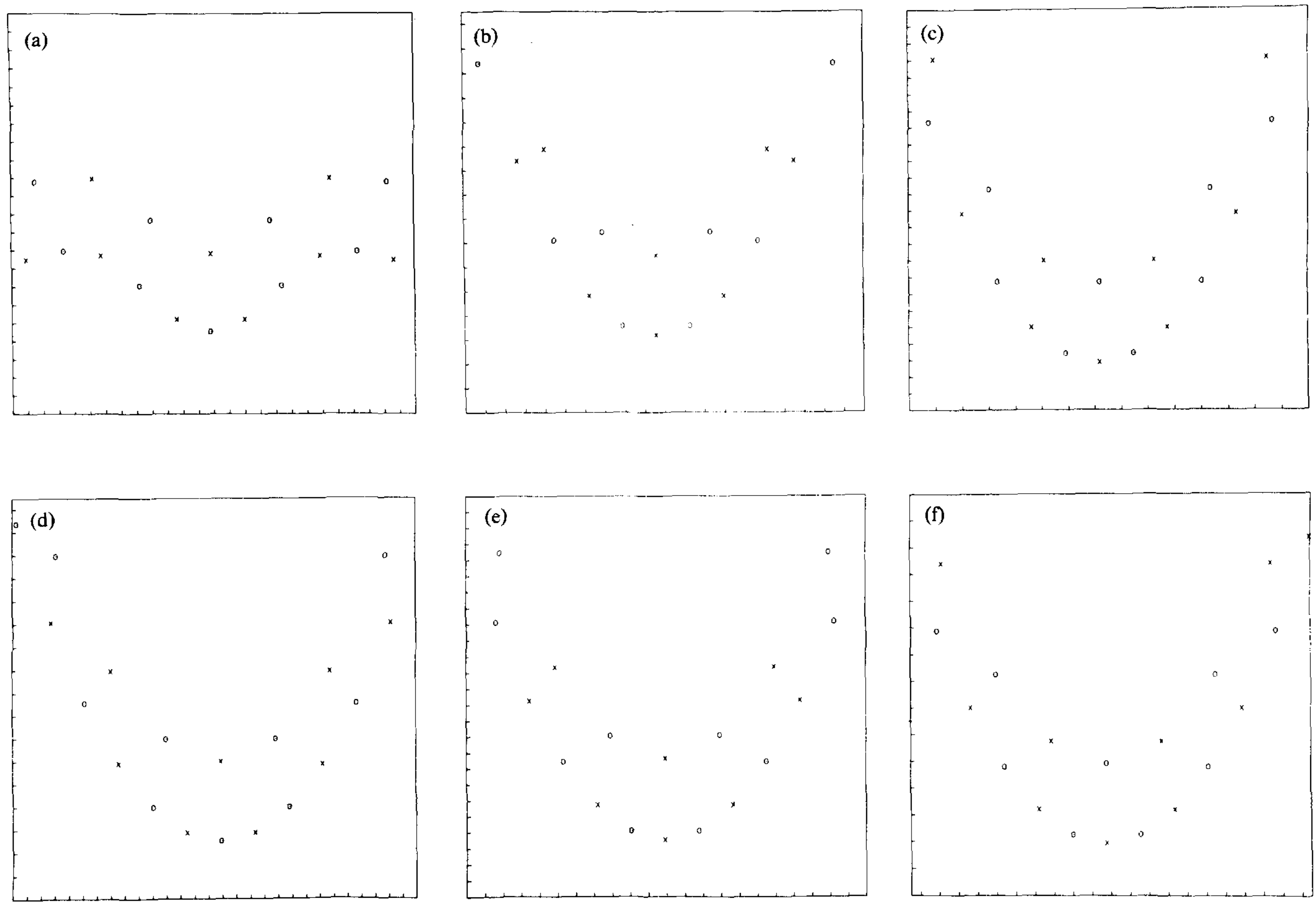


FIG. 2. Golden mean convergent periodic orbits with $k = 0.95$. In (a) positive and negative residue orbit segments are shown for $q = 55/34$ and $89/55$, that lie in a range of $\theta$, $\delta\theta = 5.2\cdot10^{-2}$ and a range of $r$, $\delta r = 1.1\cdot10^{-3}$. In (b) the orbits and ranges are $q = 144/89$ and $233/144$ with $\delta\theta = 2.0\cdot10^{-2}$ and $\delta r = 1.65\cdot10^{-4}$. In (c) they are $q = 377/233$ and $610/377$ with $\delta\theta = 7.5\cdot10^{-3}$ and $\delta r = 2.4\cdot10^{-5}$, in (d) $q = 987/610$ and $1597/987$ with $\delta\theta = 2.9\cdot10^{-3}$ and $\delta r = 3.5\cdot10^{-6}$, in (e) $q = 2584/1597$ and $4181/2584$ with $\delta\theta = 1.1\cdot10^{-3}$ and $\delta r = 5.0\cdot10^{-7}$, and in (f) $q = 6765/4181$ and $10946/6765$ with $\delta\theta = 4.2\cdot10^{-4}$ and $\delta r = 7.5\cdot10^{-8}$.

of $f(q^*_\infty)$ could be combined with the approximate monotonicity of $f$ with partial quotients discussed in Assertion II, and the upper bound established in Assertion I, to establish convergence in more general cases. Also, it might be possible to use the calculation of Appendix B to establish convergence in the limit of small $k$.

To conclude, the evidence for convergence seems quite strong for the interesting cases associated with the golden mean. It is at least credible that there should be convergence in a large generalization of this class.

## D. Assertion IV

No independent work has been done for this assertion. All the evidence assembled for Assertion II indicated that golden mean convergents yielded the smallest $f$, among all orbits whose winding number had a partial fraction representation with a given number of partial quotients. While this assertion follows from Assertion II, it is logically independent of the efforts to generalize it beyond the golden mean convergents. It has thus been given a separate number.

## E. Assertion V

Some evidence for this assertion has been given in Table III. The fact that the positive residue converges to $\frac{1}{4}$, accurate to at least four decimal places, is sufficiently remarkable to invite speculation that an integer is involved.

This inspired the calculation presented in Table V. Here an irrational winding number has been chosen whose first few partial quotients are arbitrary, but all of whose succeeding partial quotients are unity, $q = [3,1,4,1,\cdots] = (143 + \sqrt{5})/38$. Again, $k$ has been carefully selected to be close to the critical value for this winding number. The associated residue here is also approaching $\frac{1}{4}$ to a remarkable degree of accuracy.

## F. Assertion VI

The evidence for this assertion is given in Figs. 2, 3, and 4. In each of these figures, portions of periodic orbits have been plotted, with $\times$'s denoting negative residue orbits and $\circ$'s denoting positive residue orbits. In every case, the orbits chosen are golden mean convergents.

For Fig. 2, the value $k = 0.95$ has been chosen, and the corresponding mean residue had been evaluated,

$$f(q^*_\infty, 0.95) \simeq 0.977$$

so that the golden mean KAM surface is expected to exist. In

Fig. 2(a) the orbits shown have winding numbers $q = \frac{55}{34}$ and $\frac{89}{55}$, respectively. These two numbers are slightly larger and smaller, respectively, than the golden mean, and thus, the two orbits should enclose the golden mean KAM surface. This statement is true of the succeeding pairs of orbits also.

The next two golden mean convergents are shown in Fig. 2(b). Each successive golden mean convergent orbit has approximately $\phi$ times as many points as the preceeding, where the golden mean is denoted by $\phi$ for brevity. Thus, the horizontal scale has been expanded by approximately $\phi^2$ between Figs. 2(a) and 2(b), and also between succeeding frames. This accounts for the fact that each figure exhibits about the same number of points.

When examining the bottom of a parabola with increasing magnification, the vertical scale should be expanded as the square of the expansion of the horizontal scale to preserve the aspect. Thus, for each succeeding frame in this figure, the vertical scale has been expanded by $\phi^4$.

This figure then, is entirely consistent with the picture that these successive convergent orbits are closing down on a KAM surface that is represented very well by the first few terms of its Taylor expansion.

Expressions for the mapping in the tangent space of these periodic orbits are given in Eqs. (23) and (24). Since $R$ is very small and of the order $10^{-20}$ for the orbits of Fig. 2(f), the invariant ellipses associated with the positive residue orbit are extraordinarily long and thin, with an aspect ratio of the order $10^{10}$. Also, there is an extraordinarily small angle between asymptotes of the hyperbolas associated with the negative residue orbits. Even on the expanded scales used here, these figures are not resolvable from straight lines. Further, from Eq. (14), of the order of $10^{10}$ iterations are required to traverse these ellipses. Thus, about the same number of iterations would be required to distinguish the mapping in the portion of phase space delimited by Fig. 2(f) from the shear mapping of an integrable system.

Turn now to Fig. 3, which is very similar to Fig. 2, except that here $k = 0.971635$, which is the critical $k$ to the accuracy of this figure. Remarkably unlike the previous picture, here a new structure appears with each successive magnification. From Table III, it can be seen that each of these orbits

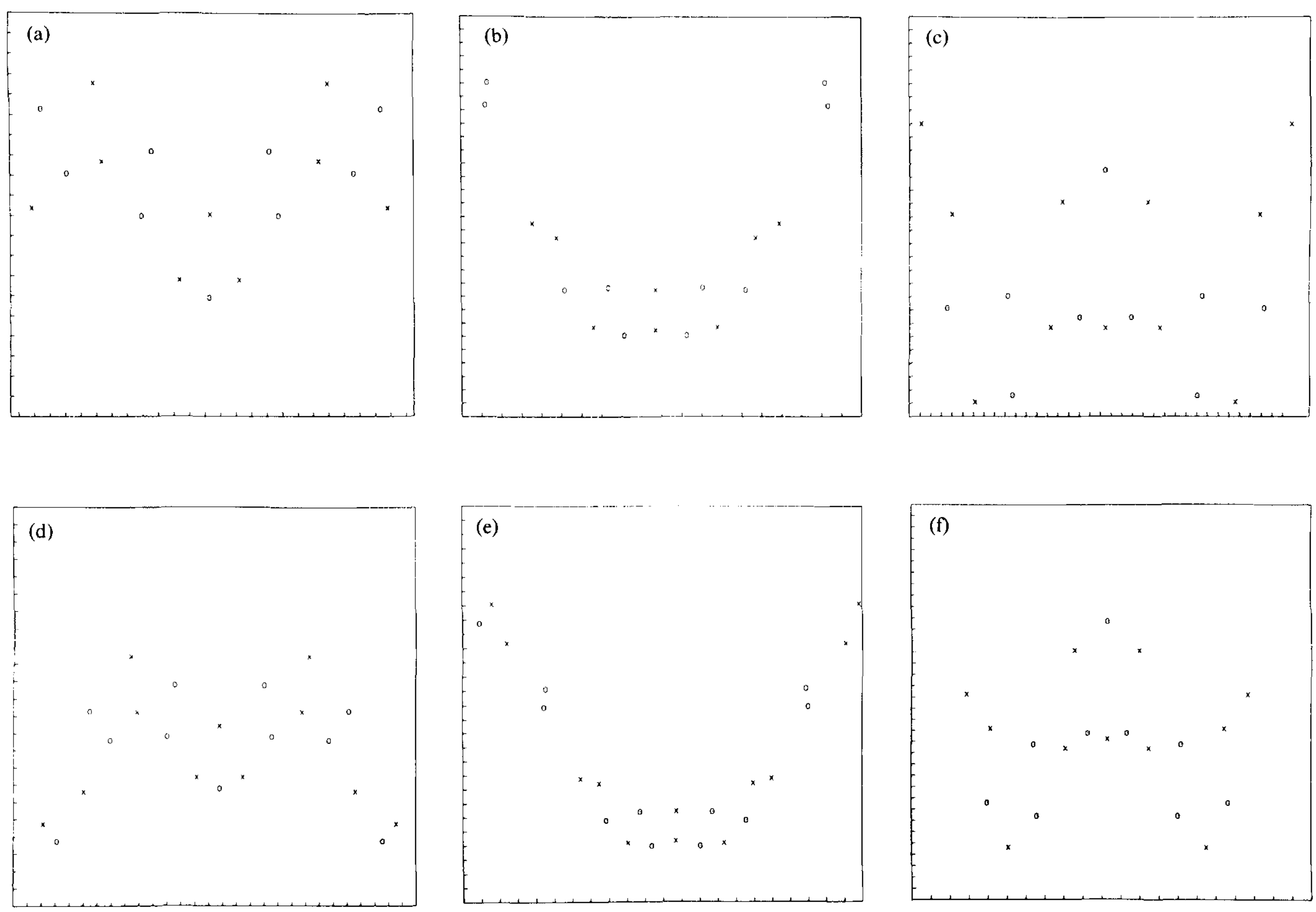


FIG. 3. Golden mean covergent periodic orbits with $k = 0.971635$. In (a) the orbits and ranges are $q = 55/34$ and $89/55$ with $\delta\theta = 5.2\cdot10^{-2}$ and $\delta r = 1.0\cdot10^{-3}$, in (b) $q = 144/89$ and $233/144$ with $\delta\theta = 2.0\cdot10^{-2}$ and $3.0\cdot10^{-4}$, in (c) $q = 377/233$ and $610/377$ with $\delta\theta = 7.5\cdot10^{-3}$ and $\delta r = 1.5\cdot10^{-5}$, in (d) $q = 987/610$ and $1597/987$ with $\delta\theta = 2.9\cdot10^{-3}$ and $\delta r = 4.6\cdot10^{-6}$, in (e) $q = 2584/1597$ and $4181/2584$ with $\delta\theta = 1.1\cdot10^{-3}$ and $\delta r = 1.4\cdot10^{-6}$, and in (f) $q = 6765/4181$ and $10946/6765$ with $\delta\theta = 4.2\cdot10^{-4}$ and $\delta r = 7.0\cdot10^{-8}$.

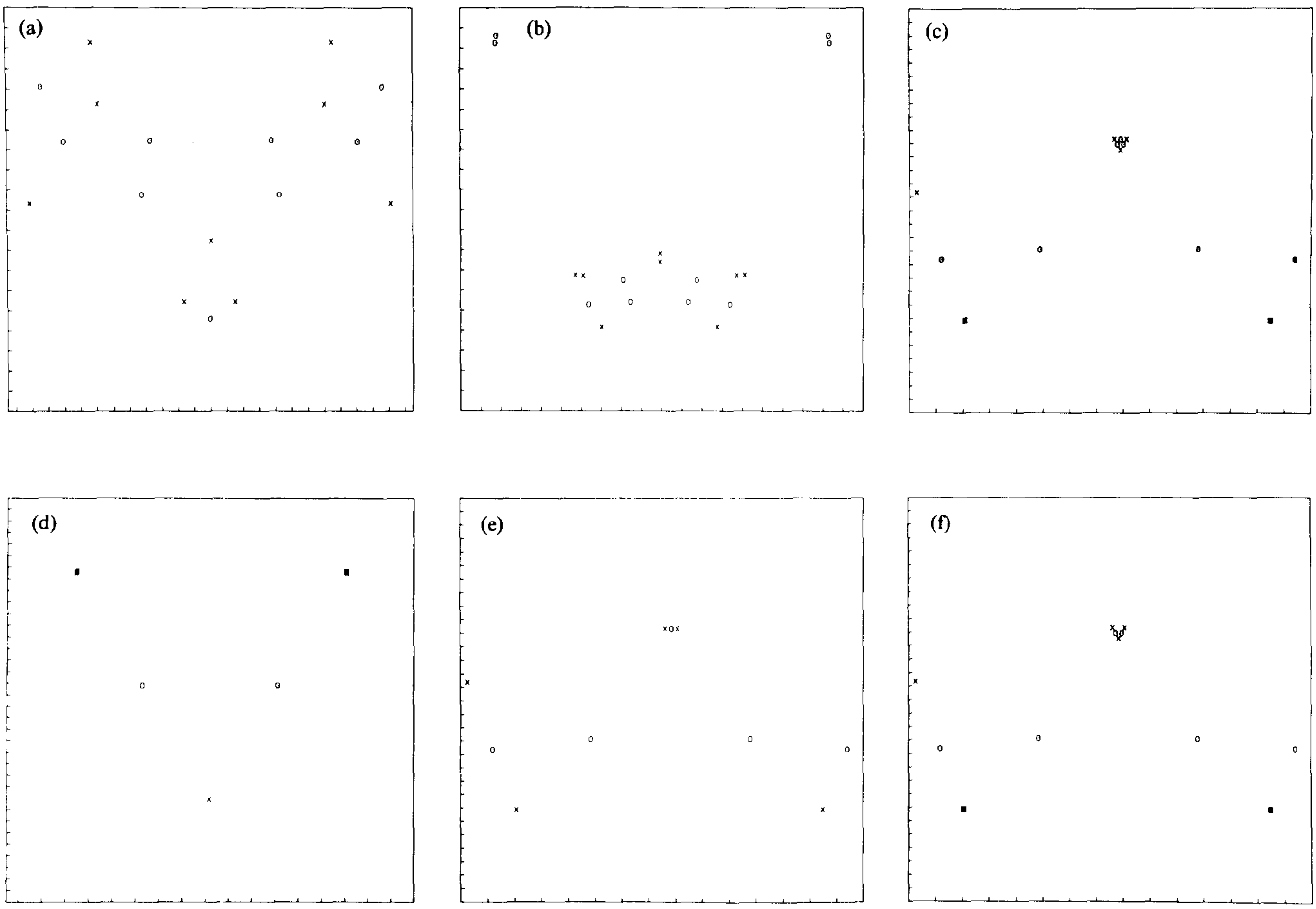


FIG. 4. Golden mean convergent orbits with $k = 0.99$. In (a) the orbits and ranges are $q = 55/34$ and $89/55$ with $\delta\theta = 5.0\cdot10^{-2}$ and $\delta r = 1.0\cdot10^{-3}$, in (b) $q = 144/89$ and $233/144$ with $\delta\theta = 2.0\cdot10^{-2}$ and $\delta r = 4.0\cdot10^{-4}$, in (c) $q = 377/233$ and $610/377$ with $\delta\theta = 7.5\cdot10^{-3}$ and $\delta r = 1.5\cdot10^{-4}$, and in (d) $q = 987/610$ with $\delta\theta = 3.5\cdot10^{-4}$ and $\delta r = 7.0\cdot10^{-6}$. In (e) and (f) orbits of (c) with the two values of $q$ have been plotted separately.

has very nearly the same residue, $R$. Thus, from Eqs. (23) and (24), the shape of the tangent space conic sections varies only with the parameter $b$. Evaluating this, it was found that these shapes would be similar from frame to frame, if the mean vertical magnification were 6.01/frame when averaged over a three-frame period. This was used here, rather than the magnification of $\phi^4 = 6.85$/frame used in the previous figure. It is seen that similarity is indeed achieved by this scaling. It seems natural to associate the structure on successive scales with a necessity to accommodate the invariant ellipses surrounding each stable orbit.

As further evidence in this direction, note that the threefold period in the structure is accompanied by a threefold period in the positions of the $\times$ and $\circ$ points, relative to the center of each frame.

After carefully observing that Figs. 3(c) and 3(f) each exhibit two $w$ shaped curves, one lying above the other, the picture emerges that the convergents for this value of $k$ are squeezing down on some nonanalytic curve that has a structure on every scale.

Finally, in preparation for the next figure, note that, while there is similarity, the configurations in Figs. 3(d), 3(e), and 3(f) are somewhat shrunken compared with Figs. 3(a), 3(b), and 3(c), respectively, indicating that there is a mild tendency toward the clustering of orbits.

Now consider Fig. 4 for which $k$ has been increased to 0.99. In each of these frames, the vertical exaggeration has been fixed at 50. This scaling yields a similarity of the tangent space configurations for this value of $k$, and indeed, there is a tendency for similarity of the wedges that appear here.

Aside from that, there is such a strong tendency toward the clustering of orbits, i.e., some of the black marks on Figs. 4(c) and 4(d) represent three or four points, that these orbits do not seem to be squeezing down on any kind of curve, even a highly singular one. Something strange is going on at this value of $k$.

To conclude then, the value of the mean residue, $f$, determines the value of the residue, $R$. This , in turn, has a strong influence on the shape of the tangent space figures through Eqs. (23) and (24). The latter fit smoothly against KAM surfaces when $k$ is below the critical value and rumple it nonanalytically at the critical value. Beyond the critical

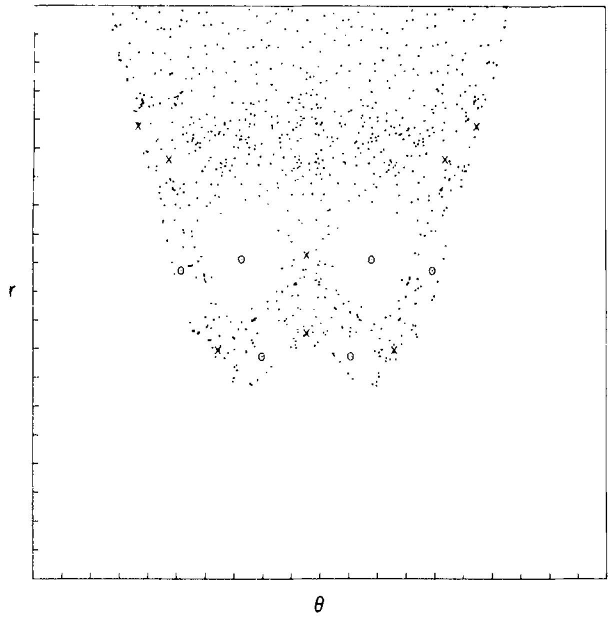


FIG. 5. Orbits near the golden mean with $d = 0.975$. Periodic orbits with positive and negative residues are shown for $q = 144/89$ and $233/144$, together with a segment of a stochastic orbit. The ranges of $r$ and $\theta$ in this figure are $\delta\theta = 2.0\cdot10^{-2}$ and $\delta r = 2.0\cdot10^{-4}$.

value there is no indication for the existence of KAM surfaces or any other organizing principle.

## G. Assertion VII

In the previous section, evidence was presented that KAM surfaces existed for values of $k$ up to the critical value. Here we concentrate on showing that stochastic orbits vertically encircling the torus exist for values of $k$ slightly exceeding the critical value. It is clear that horizontally encircling KAM orbits, and vertically encircling stochastic orbits cannot coexist. It is not so clear that there is no range of $k$ for which neither type of orbit exists. Nevertheless, it appears that this statement is true.

Five orbits are shown in Fig. 5 where $k$ has been chosen to be 0.975. Four of these are periodic, the positive and negative residue points with winding number $\frac{144}{89}$ and $\frac{233}{144}$ respectively. These are the last golden mean convergents for which there are stable orbits. Golden mean convergent orbits with length 377 or greater have residues greater than unity, and are therefore unstable.

A stochastic orbit is also shown on that figure. The initial condition for this orbit was near one of the $\times$ points on the figure. The interesting point is that it wanders back and forth across the region between the periodic orbits, the region where the KAM surface might exist. There is not the slightest indication of an invariant surface in this vicinity that divides phase space.

It was not possible to carry this orbit to the length required for it to encircle the torus vertically. It wandered slowly upward out of the picture, and beyond, but the diffusion of other orbits back into this region is really extraordinarily slow. Similar behavior has been found by Karney[18] in an inhomogeneous random walk problem, showing that it is consistent with a Markovian process.

The details of the stochastic orbit diffusion are governed by a numerical roundoff. One can rapidly lose thousands of digits of formal accuracy in such calculations. To show that the effects of Fig. 5 are not strictly roundoff, we present Fig. 6.

For this figure, $k$ is 0.97. From the criteria of Assertion VI, KAM surfaces exist on both sides of the stochastic orbit shown. The length of the orbit shown in this figure is a few hundred thousand. In another calculation, this orbit was taken to a length of $5\cdot10^7$, and it was entirely contained within the stochastic orbit region exhibited here, in spite of the fact that crude estimates show that the calculation suffered from truly fantastic numerical error.

How can this be? A more careful consideration of the numerical error is called for. A detailed calculation of this is given in Appendix C. It turns out that, at a given point, the errors from all the preceding parts of the orbit are, to great accuracy, spread out in only one direction. When this direction is calculated, it is found to be parallel to the apparent edge of the stochastic orbit in Fig. 6. Thus, numerical error can only lead to diffusion parallel to KAM surfaces, not across them.

As for numerical error in diffusing orbits, other calculations, similar in spirit,[11,19] have shown that there is almost always one exact orbit in the vicinity of any orbit determined numerically with a roundoff error that is small for each iteration.

These considerations lead to the conclusion that roundoff does not affect the essential features to be learned from Figs. 5 and 6.

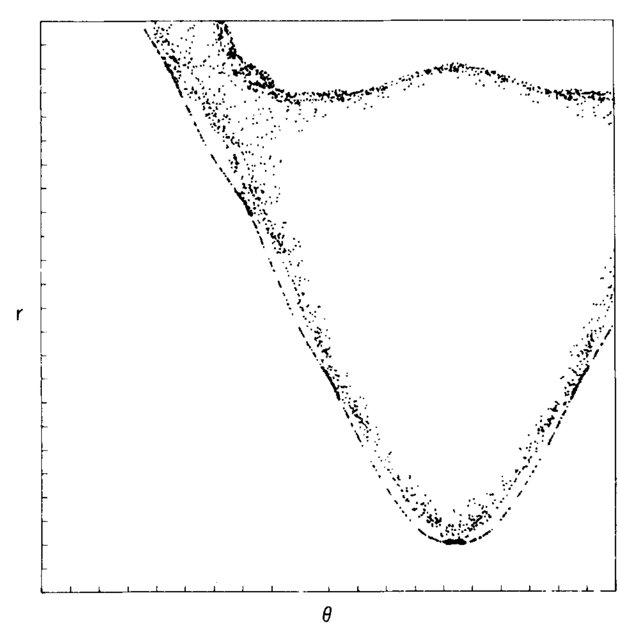


FIG. 6. Stochastic orbit for $k = 0.97$. The range of $r$ and $\theta$ in this figure is $\delta\theta = 1.0\cdot10^{-2}$ and $\delta r = 1.2\cdot10^{-4}$.

## V. DISCUSSION

This paper has explored the concept that there is a close relation between the existence of KAM surfaces and the stability of nearby periodic orbits. Since Moser has shown[20] that KAM surfaces are in the closure of the set of periodic orbits, it is reasonable that such a relation might exist. The relationship has been found to be very close when the problem is put in the proper perspective. In this section we first review that perspective.

Since not all orbits can be calculated and evaluated for stability, a credible method must be found for estimating this trait by extrapolation. Eigenvalues evaluated in the tangent space of the periodic orbit, as in Eqs. (10), (11), and (15), quantitatively characterize the stability of the orbit. These eigenvalues are unsuited for extrapolation, however, because they are not analytic functions of the parameters of the system. They have branch points where they turn from real to complex as in Eq. (11).

Information equivalent to these eigenvalues is contained in the residue, $R$, defined in Eqs. (10) or (15). This residue is a real continuous function of the parameters of the system. It is, thus, in this respect quite suitable for interpolation and extrapolation.

As a vehicle for extrapolation, the residue suffers from another problem. Orbits near a KAM surface have varying lengths and are longer, the closer the orbit is to the surface. It has been noted by several authors in the past[11,21,22] that the residue has an exponential dependence on orbit length. The methods of Appendix B of this paper allow this result to be established firmly for large and small $k$. As a result, nearby orbits may have vastly different residues. This exponential dependence on orbit length can be suppressed by defining the mean residue in Eq. (17).

Even this mean residue, or residue per unit length, is not a continuous function of relative location of the periodic orbit. More precisely, it does not depend continuously on the winding number $q$ defined in Eq. (3). The problem is that periodic orbits perturb nearby longer periodic orbits. An orbit of given length is more likely to be unstable if it is close to a shorter periodic orbit and more likely to be stable if it is relatively isolated. Since every rational number has its zone of perturbation, the dependence of $f$ on the winding number $q$ is very peculiar as seen for example in Ref. 11 or 21.

One way to deal with this problem is to express the winding number as a continued fraction as was done in Eq. (4). The magnitudes of the partial quotients of the continued fraction provide an estimate of all the perturbations on the given orbit. It is not yet possible to estimate the mean residue, $f$, quantitatively as a function of these partial quotients, but they do yield a very useful qualitative understanding. This understanding can be used to estimate the stability of orbits by extrapolation of known results for nearby orbits.

In order to relate the existence of a KAM surface to the stability of nearby periodic orbits, it is desirable to make a careful choice of this set of orbits. It should be a sequence of successively longer orbits that asymptotically, in some sense, approach the postulated KAM surface. Further, each member should be as close to the desired surface as feasible and as far from the perturbing effects of other, shorter orbits. A choice that meets these criteria is the set whose winding numbers are the convergents to the KAM winding number. These were defined in Sec. II.

These considerations determined the calculations made in this paper.

The results of a large number of these numerical calculations have been distilled into a series of assertions in Sec. III.

According to Assertion III, the sequence of mean residues for the convergents of a KAM surface converges to a limit. For the sequences calculated in Sec. IV C, adding a partial quotient had a smaller effect when the continued fraction representation of the winding number had many partial quotients.

When the limiting mean residue is less than one, beyond some orbit length all the residues will be less than one according to Eq. (17). In fact, the residues rapidly approach zero in the limit. Therefore, the positive residue orbits are stable. On the other hand, when the limiting mean residue is greater than one, the converse is true. In this case both the positive and negative residue orbits of the chosen set are unstable for orbits longer than some length.

Figures 2, 3, and 4 show graphically that there is a close relation between KAM surfaces and the stability of nearby periodic orbits. Repeating the discussion from Sec. IV F, the environment of an irrational winding number is entirely different depending on whether the nearby periodic orbits are stable or unstable. This is particularly true because, according to Eq. (17), the stability of these orbits is extreme, whether stable or unstable. It is thus reasonable that the value of the converged mean residue for an irrational winding number determines the existence of the corresponding KAM surface. This idea has been discussed in Ref. 22.

Convergence of the mean residue can be optimized by judicious choice of $\beta$ in Eq. (17). It appears that $\beta = \frac{1}{4}$ is the preferred value for all cases that have been considered to date. The underlying reason is given in Assertion V. This choice of $\beta$ yields remarkably good results. For example, consider the crudest approximation to the converged mean residue for the golden mean surface discussed in Sec. III. The leading convergents to this surface are the periodic orbits with $q$ of one and two. The mean residue was evaluated analytically for these orbits in Sec. IV A and is equal to $k$ in either case. Thus, the leading estimate for the converged mean residue of the golden mean surface is

$$f(q^*_\infty) \approx k.$$

This estimate yields a critical $k$ of unity, only a few percent different from the more exact value of Assertion VII! So, the method described in this paper seems well suited for giving rough estimates for stochastic behavior.

The accuracy of this estimate depends critically on choosing an optimum value of $\beta$ in Eq. (17). That is, it depends on Assertion V, that near the critical $k$ many orbits have residues near $\frac{1}{4}$. It is interesting to examine this criterion geometrically. From Eq. (14), the corresponding value of $\iota$ for these orbits is 60°. When a periodic orbit bifurcates out an orbit with period six times longer than itself, some nearby related KAM surface is on the edge of disappearance. This relationship has been noted by Lichtenberg[23] also.

Sufficient numerical work has been done on this problem to identify reasonable hypotheses, but considerable effort is now needed to provide proofs. Probably the most crucial of these is Assertion III, concerning the convergence of the mean $k$ residue for irrational winding numbers. It might be possible to use the methods of Appendix B to establish this hypothesis for small $k$ as a first step. Such a proof would show that the mean residue is indeed a fruitful concept.

Recently, Percival[24] has derived a variational principle for KAM surfaces that might also provide a foundation for further progress.

## ACKNOWLEDGMENTS

The author is particularly indebted to Dr. Bountis and Dr. Helleman for sharing their insights. Their help in pointing the way toward the calculation of Appendix B is especially appreciated. This paper has also benefited greatly from discussions with many people among whom are Dr. Kruskal, Dr. Moser, Dr. Kaufman, Dr. Lichtenberg, Dr. Meiss, Dr. Treve, Dr. Karney, Dr. Dewar, Dr. Oberman, Dr. Krommes, Dr. Rechester, and Dr. Finn. The hospitality of the Lawrence Livermore Laboratory, where part of this paper was written, was very helpful. Finally, the support and encouragement of Dr. Ford has been greatly appreciated. This work was supported by the United States Department of Energy Contract No. EY-76-C-02-3073.

## APPENDIX A: CALCULATION OF PERIODIC ORBITS

It is possible to use the symmetry of the mapping to reduce the problem of finding a given periodic orbit to one of finding the root of a function of one variable. This reduction from a two-dimensional problem to a one-dimensional problem vastly increases the speed and accuracy with which these orbits can be determined. This method has been described thoroughly by deVogeleare[25] but is included here for completeness.

The nature of the symmetry can be stated succinctly: The mapping is the product of two involutions. In other words, if the standard mapping is denoted by $T$, then

$$T = I_2 I_1, \tag{A1}$$

where $I_1$ is given by

$$\theta_n = -\theta_{n-1}, \quad r_n = r_{n-1} - \frac{k}{2\pi}\sin 2\pi\theta_{n-1}, \tag{A2}$$

and $I_2$ by

$$\theta_{n+1} = -\theta_n + r_n, \quad r_{n+1} = r_n. \tag{A3}$$

It is straightforward to show that

$$I_1^2 = 1, \quad I_2^2 = 1, \tag{A4}$$

so that these transformations are involutions.

Each of these involutions has lines of fixed points.[26] Namely,

$$I_1(r,\theta) = (r,\theta)$$

is satisfied by $\theta = 0$ or $\theta = \frac{1}{2}$ for any $r$, and

$$I_2(r,\theta) = (r,\theta)$$

is satisfied by $\theta = \frac{1}{2}r$ or $\theta = \frac{1}{2}(r+1)$.

It is now easy to show that if the initial value of an orbit is a fixed point of $I_1$,

$$I_1(r_0,\theta_0) = (r_0,\theta_0), \tag{A5}$$

and the $N$th iterate is also a fixed point of $I_1$,

$$I_1 T^N(r_0,\theta_0) = T^N(r_0,\theta_0), \tag{A6}$$

then the full orbit is periodic with length $2N$. Consider with the aid of Eqs. (A1), (A5), and (A6),

$$\begin{aligned} T^{2N}(r_0,\theta_0) &= T^{N-1} I_2 I_1 T^{N-1} I_2 I_1(r_0,\theta_0) \\ &= T^{N-1} I_2 T^{N-1} I_2(r_0,\theta_0). \end{aligned} \tag{A7}$$

From Eqs. (A1) and (A4),

$$I_2 T = I_1, \quad T I_1 = I_2, \tag{A8}$$

which, used alternately in Eq. (A7) yields

$$T^{2N}(r_0,\theta_0) = (r_0,\theta_0) \tag{A9}$$

as stated.

Thus, the problem of finding periodic orbits has been reduced to that of finding the root of any function that vanishes on the fixed lines of $I_1$, and with the independent variable taken to be the one-parameter family of fixed points of $I_1$. For example, fixed points can be found from the solutions of

$$\sin 2\pi\theta_N(r_0) = 0$$

with $\theta_0 = 0$.

The procedure can be readily generalized to include the fixed points of $I_2$ as either the initial or the final point of the computed orbit. All the periodic orbits that exist down to $k = 0$ can be determined this way.

## APPENDIX B: EXPANSION FOR SMALL $k$

In this Appendix an algorithm is derived for calculating the residue, $R$, in the limit of small $k$. It is shown that $R$ is proportional to $k^Q$, where $Q$ is the length of the periodic orbit. In the first few paragraphs, the relation between the two forms for calculating $R$, Eqs. (10) and (15), is derived, since the less familiar Bountis and Helleman[12] form is used in the argument of this Appendix.

The equations governing the orbits in the tangent space are found by differencing Eq. (1) and evaluating the coefficients on the periodic orbit considered, yielding

$$\begin{aligned} \delta r_{n+1} &= \delta r_n - k\cos 2\pi\theta_n \delta\theta_n, \\ \delta\theta_{n+1} &= \delta\theta_n + \delta r_{n+1}, \end{aligned} \tag{B1}$$

where $\theta_n$ is the coordinate of the $n$th point on the periodic orbit. A set of equations for the $2Q$ variables $(\delta\theta_1,\delta r_1,\ldots,\delta\theta_Q,\delta r_Q)$ is closed through the Floquet condition,

$$\delta r_{n+Q} = \lambda\delta r_n, \quad \delta\theta_{n+Q} = \lambda\delta\theta_n. \tag{B2}$$

Then $(\delta r_0,\delta\theta_0)$ is an eigenvector of $M$ of Eqs. (7) and (8), with eigenvalue $\lambda$.

These equations can be written in matrix form

$$\begin{pmatrix} 1 & 0 & -1 & 1 & 0 & 0 & \cdots \\ -k\cos 2\pi\theta_1 & 1 & 0 & -1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & -1 & 1 & \cdots \\ 0 & 0 & -k\cos 2\pi\theta_2 & 1 & 0 & -1 & \cdots \\ \vdots & \vdots & & & & & \\ -\lambda & \lambda & 0 & 0 & \cdots & 1 & 0 \\ 0 & -\lambda & 0 & 0 & \cdots & -k\cos 2\pi\theta_Q & 1 \end{pmatrix} \begin{pmatrix} \delta\theta_1 \\ \delta r_1 \\ \delta\theta_2 \\ \delta r_2 \\ \\ \delta\theta_Q \\ \delta r_Q \end{pmatrix} = \mathbf{J}\cdot\delta\mathbf{x} = 0. \tag{B3}$$

The condition that the determinant of $\mathbf{J}$ vanishes yields an equation for $\lambda$.

Adding each even numbered row to the row above yields

$$\mathrm{Det}\mathbf{J} = \mathrm{Det}\begin{pmatrix} M_1 & -1 & 0 & 0 & \cdots & \\ 0 & M_2 & -1 & 0 & \cdots & \\ \vdots & \vdots & & & & -1 \\ -\lambda & 0 & \cdots & & & M_Q \end{pmatrix}, \tag{B4}$$

where each element is now a $2\times 2$ matrix, and

$$M_i = \begin{pmatrix} 1 - k\cos 2\pi\theta_i & 1 \\ -k\cos 2\pi\theta_i & 1 \end{pmatrix}. \tag{B5}$$

Multiplying the first row by $M_2$, adding the second row, and continuing on to eliminate all but the diagonal term in the first row, we find

$$\mathrm{Det}\mathbf{J} = \mathrm{Det}\left(\prod_{i=1}^{Q} M_i - \lambda I\right), \tag{B6}$$

where $\mathrm{Det}M_i = 1$ has been used. This is the usual equation for the eigenvalue $\lambda$.

Alternatively, the fourth element in the first row of Eq. (B3) can be used to eliminate every other element in its row and column. The other 1's down that diagonal are treated similarly, as is the $\lambda$ in the second column next to last row. These elements can then be factored out of the determinant,[12] leaving

$$\mathrm{Det}H(\lambda) = 0 \tag{B7}$$

where

$$H(\lambda) = \begin{pmatrix} 2 - k\cos 2\pi\theta_1 & -1 & \cdots & & -\lambda^{-1} \\ -1 & 2 - k\cos 2\pi\theta_2 & -1 & \cdots & \\ \vdots & \vdots & & & \\ -\lambda & \cdots & & -1 & 2 - k\cos 2\pi\theta_Q \end{pmatrix} \tag{B8}$$

is a tridiagonal matrix with additional elements in the corners. Considering the cofactor of the element, $\lambda$, yields the result

$$\mathrm{Det}H(\lambda) = \mathrm{Det}H(1) - \lambda - \lambda^{-1} + 2. \tag{B9}$$

Since the residue $R$ is related to the eigenvalue $\lambda$, we obtain directly

$$R = -\tfrac{1}{4}\,\mathrm{Det}H(1). \tag{B10}$$

In some respects, a determinant is easier to calculate than the trace of a product.

The relation between these two forms for calculating $R$ is analogous to the relation between Hill's method and the shooting method for calculating the Floquet parameter for the Mathieu equation.

We now proceed to evaluate $\mathrm{Det}H(1)$. The argument $\lambda = 1$ will be understood for the rest of this section, and thus dropped from the notation. At $k = 0$,

$$H(k=0) = \begin{pmatrix} 2 & -1 & 0 & \cdots & & -1 \\ -1 & 2 & -1 & \cdots & & \\ 0 & -1 & 2 & \cdots & & \\ \vdots & & & & & \\ -1 & \cdots & & & -1 & 2 \end{pmatrix}. \tag{B11}$$

It can be seen immediately that the eigenvalues of this matrix are

$$\eta_j = 2(1 - \cos 2\pi j/Q) = 4\sin^2 \pi j/Q \tag{B12}$$

with eigenvectors,

$$\xi_{n,j} = \cos 2\pi nj/Q. \tag{B13}$$

The lowest eigenvalue vanishes, so

$$\mathrm{Det}H\,(k=0) = 0. \tag{B14}$$

The product of the others can be evaluated with the result[27]

$$\prod_{j=1}^{Q-1} \eta_j = Q^2. \tag{B15}$$

It follows from $\mathrm{Det}H = \Pi_j \eta_j$ that the lowest order novanishing approximation to $\mathrm{Det}H$ is $Q^2\eta_0$, where $\eta_0$ is the lowest order nonvanishing approximation to the lowest eigenvalue of the symmetric matrix $H$.

Since $H$ depends on the orbits, it is necessary to evaluate the orbits to some degree of accuracy. It will be shown by the result that the requisite order is $k^{Q-1}$.

The coordinate $r_n$ can be eliminated from the standard mapping, Eq. (1), yielding

$$-\theta_{n+1} + 2\theta_n - \theta_{n-1} = \frac{k}{2\pi}\sin 2\pi\theta_n. \tag{B16}$$

This has been written in recursion form. That is, if the periodic orbit is known to some degree of precision, and that estimate is used to evaluate the right-hand side of Eq. (B16), inverting the operator on the left yields an estimate improved by one order in $k$. The inversion is done with the periodicity condition

$$\theta_{n+Q} = \theta_n. \tag{B17}$$

We already know something about this operator, since it is $H\,(k=0)$. For one thing it is singular, and each element of the eigenvector of the vanishing eigenvalue is unity. Multiplying by this eigenvector yields the solvability condition

$$\sum_{n=1}^{Q} \sin 2\pi\theta_n = 0. \tag{B18}$$

A formal solution of these equations can be written in the form

$$2\pi\theta_n = 2\pi\theta_n^{(0)} + \sum_{j=0}^{\infty}\sum_{l=1}^{Q} \frac{a_{lj}(r^{(0)})k^{l+2j}}{2(1-\cos 2\pi l r^{(0)})}\sin 2\pi l\theta_n^{(0)}, \tag{B19}$$

where

$$r^{(0)} \equiv P/Q, \quad \theta_{n+1}^{(0)} = \theta_n^{(0)} + r^{(0)}, \tag{B20}$$

$P$ and $Q$ are relatively prime, $\theta_0^{(0)}$ will be determined later, and the $a_{lj}$ coefficients are evaluated as follows. Inserting this form into the right- and left-hand sides of Eq. (B16) yields

$$\sum_{j=0}^{\infty}\sum_{l=1}^{Q} a_{lj}k^{l+2j}\sin 2\pi l\theta_n^{(0)} = k\sin 2\pi\theta_n = k\sin\left(2\pi\theta_n^{(0)} + \sum_{j=0}^{\infty}\sum_{l=1}^{Q}\frac{a_{lj}k^{l+2j}}{2(1-\cos 2\pi l r^{(0)})}\sin 2\pi l\theta_n^{(0)}\right). \tag{B21}$$

The sine on the right can be expanded since the second term of its argument is small in powers of $k$. The factors of resulting products are combined to yield Fourier coefficients of $\sin 2\pi l\theta_n^{(0)}$ as powers of $k$. These determine the coefficients $a_{lj}$ on the left-hand side, order by order in $k$. Thus, the formal solution is consistent.

Since from Eq. (B20), the $\theta_n^{(0)}$ are equally spaced in the interval (0,1), the solvability condition is satisfied trivially for $l \neq Q$ and thus, up to the order of $k^Q$. At the order $k^Q$, the value of $\theta_0^{(0)}$ must be chosen so that $\sin 2\pi Q\theta_n^{(0)} = 0$. There are two such solutions, with $\cos 2\pi Q\theta_n^{(0)} = \pm 1$. With this solvability condition, the problem of a vanishing denominator, $1 - \cos 2\pi l r^{(0)}$, disappears. Further, it can be used to eliminate terms with $l > Q$ in favor of terms with $l < Q$.

Since the relation between the $\theta_n^{(0)}$ and $r^{(0)}$ is identical to the integral mapping for $k = 0$, and also from the fact that the solvability condition is satisfied in the lower orders by every point on the surface $r^{(0)} = \text{const}$, it is apparent that the formal solution above is equivalent to the transformation from an integrable mapping to one that is integrable to all orders. This is useful below so it will be pursued further here.

Consider the transformation

$$r = r(r^{(0)},\theta^{(0)}), \quad \theta = \theta(r^{(0)},\theta^{(0)}). \tag{B22}$$

If these functions satisfy

$$r(r^{(0)},\theta^{(0)}) = \theta(r^{(0)},\theta^{(0)}) - \theta(r^{(0)},\theta^{(0)} - r^{(0)}),$$

$$-\theta(r^{(0)},\theta^{(0)} + r^{(0)}) + 2\theta(r^{(0)},\theta^{(0)}) - \theta(r^{(0)},\theta^{(0)} - r^{(0)})$$

$$= \frac{k}{2\pi}\sin 2\pi\theta(r^{(0)},\theta^{(0)}) \tag{B23}$$

to all orders in $k$, then $r^{(0)}$ is a constant of the standard mapping to all orders. Eliminating $\theta^{(0)}$ between $r$ and $\theta$ would yield an implicit expression for the invariant $r^{(0)}$.

Indeed, the formal solution of these equations is very similar to that given above,

$$2\pi\theta = 2\pi\theta^{(0)} + \sum_{j=0}^{\infty}\sum_{l=1}^{\infty} \frac{a_{lj}(r^{(0)})k^{l+2j}}{2(1-\cos 2\pi l r^{(0)})} \sin 2\pi l\theta^{(0)}, \tag{B24}$$

where the $a_{lj}$ are determined by

$$\sum_{j=0}^{\infty}\sum_{l=1}^{\infty} a_{lj}k^{l+2j}\sin 2\pi l\theta^{(0)}$$
$$= k\sin\left(2\pi\theta^{(0)} + \sum_{j=0}^{\infty}\sum_{l=1}^{\infty} \frac{a_{lj}k^{l+2j}}{2(1-\cos 2\pi l r^{(0)})}\sin 2\pi l\theta^{(0)}\right). \tag{B25}$$

The subscripts on $\theta^{(0)}$ have been dropped since here $\theta$ depends continuously on $\theta^{(0)}$, the sum over $l$ has been extended to infinity, and the condition that $r^{(0)}$ be rational has been dropped. Indeed, it is better if it is not rational!

Now return to the problem of calculating the lowest eigenvalue of $H$,

$$H(\delta\theta_n) = \eta_0(\delta\theta_n),$$

where $\delta\theta_n$ is the eigenvector. This also can be expressed in recursion form,

$$-\delta\theta_{n+1} + 2\delta\theta_n - \delta\theta_{n-1} = (\eta_0 + k\cos 2\pi\theta_n)\delta\theta_n \tag{B26}$$

again, yielding the operator $H(k=0)$. The solvability condition is used to determine the eigenvalue,

$$\eta_0 = -k\frac{\Sigma_{n=1}^{Q}\delta\theta_n\cos 2\pi\theta_n}{\Sigma_{n=1}^{Q}\delta\theta_n}. \tag{B27}$$

As will be seen, the lowest order of the eigenvalue is the order of $k^Q$. Thus, we need to evaluate the eigenvector accurate to the order $k^{Q-1}$, with $\eta_0$ being a higher order term in Eq. (B26).

This can be done directly. Take the derivative of Eq. (B23) with respect to $\theta^{(0)}$, and evaluate it at the point $\theta^{(0)} = \theta_n^{(0)}$ which are the values of Eq. (B20). It can be verified immediately that these derivatives satisfy the same equation as $\delta\theta_n$, and thus,

$$\delta\theta_n = \left.\frac{\partial\theta}{\partial\theta^{(0)}}\right|_{\theta^{(0)}=\theta_n^{(0)}}. \tag{B28}$$

Since the mapping is effectively integrable to the desired order, the eigenvector is merely deformed by the transformation of Eq. (B22).

It then follows immediately that

$$\delta\theta_n k\cos 2\pi\theta_n = \left.\frac{k}{2\pi}\frac{\partial}{\partial\theta^{(0)}}\sin 2\pi\theta_n\right|_{\theta^{(0)}=\theta_n^{(0)}}$$
$$= \sum_{j=0}^{\infty}\sum_{l=1}^{\infty} l a_{lj}k^{l+2j}\cos 2\pi l\theta_n^{(0)}. \tag{B29}$$

In evaluating the eigenvalue from Eq. (B27), the sum over $n$ vanishes trivially for $l < Q$. Thus, the first nonvanishing term is proportional to $k^Q$. For $l = Q$, the cosines are either plus or minus one depending on the orbit as discussed above. Therefore, the eigenvalue has as its lowest nonvanishing estimate

$$\eta_0 = \mp a_{Q0}Qk^Q, \tag{B30}$$

yielding for the residue

$$R = \pm\tfrac{1}{4}a_{Q0}Q^3k^Q. \tag{B31}$$

The quantities $a_{Q0}$ can be calculated one by one using Eq. (B25). It would be exceedingly interesting to know some general properties of these coefficients.

## APPENDIX C: STRETCHING AND NUMERICAL ERRORS

In numerical work, it has been noticed that computed orbits are stochastic only in regions where stochasticity is expected and lie on surfaces when the mapping is integrable. This is surprising in the sense that there are huge numerical errors in calculating these orbits, yet these errors do not seem to fundamentally change the character of the orbit. A part of that question is examined in this Appendix.

These numerical errors are highly anisotropic. Here, we show that errors are generally parallel to KAM surfaces and do not lead to diffusion across them.

Consider a long segment of an orbit that is not periodic. There will be a certain small numerical error in calculating the first iteration of the mapping from $(r_1,\theta_1)$ to $(r_2,\theta_2)$. This error will have a probability distribution that can be crudely represented by a small circle. At the end of the considered segment, at the proint $(r_n,\theta_n)$, this circle will have become an ellipse, in the tangent space approximation, with a large aspect ratio. Thus, a small error will have become a large error in one direction.

First, let us calculate this ellipse. If the tangent space orbits are represented by

$$\begin{pmatrix}\delta\theta_n\\ \delta r_n\end{pmatrix} = N_{n,2}\begin{pmatrix}\delta\theta_2\\ \delta r_2\end{pmatrix}, \tag{C1}$$

where the notation distinguishes the matrix $N$ from the similar matrix for periodic orbits, $M$, then the initial values, and the adjoints, are given in terms of the endpoints by

$$\delta\mathbf{x}_0 = N_{n,2}^{-1}\delta\mathbf{x}_n, \quad \delta\mathbf{x}_0^\dagger = \delta\mathbf{x}_n^\dagger(N_{n,2}^{-1})^\dagger. \tag{C2}$$

The condition that the initial point lie on a circle of radius one,

$$\delta\mathbf{x}_0^\dagger\cdot\delta\mathbf{x}_0 = \delta\mathbf{x}_n^\dagger(N_{n,2}^{-1})^\dagger N_{n,2}^{-1}\delta\mathbf{x}_n$$
$$= \delta r_0^2 + \delta\theta_0^2 = 1 \tag{C3}$$

yields an expression for the ellipse at $(r_n,\theta_n)$. Thus, the axes of the ellipse are given in terms of the eigenvalues and eigenvectors of $(N_{n,2}^{-1})^\dagger N_{n,2}^{-1}$, or its inverse, $NN^\dagger$. If $N$ is parameterized

$$N_{n,2}=\begin{pmatrix}a_2+d_2 & c_2+b_2\\ c_2-b_2 & a_2-d_2\end{pmatrix} \tag{C4}$$

with

$$a_2^2+b_2^2-c_2^2-d_2^2=1, \tag{C5}$$

then

$$N_{n,2}N_{n,2}^{\dagger}=\begin{pmatrix}A_2+D_2 & C_2\\ C_2 & A_2-D_2\end{pmatrix}, \tag{C6}$$

where

$$\begin{aligned}A_2&=a_2^2+b_2^2+c_2^2+d_2^2=1,\\ C_2&=2(a_2c_2-b_2d_2),\\ D_2&=2(a_2d_2+b_2c_2).\end{aligned} \tag{C7}$$

The eigenvalues of this matrix are the squares of the major and minor semiaxes,

$$\begin{aligned}\rho_{\pm}^2&=A_2\pm(A_2^2-1)^{1/2}\\ &=[(a_2^2+b_2^2)^{1/2}\pm(c_2^2+d_2^2)^{1/2}]^2,\end{aligned} \tag{C8}$$

and the eigenvectors yield the angle, $\theta_e$, that this ellipse makes with the line $\delta r=0$.

$$\tan\theta_e=C_2/[(A_2^2-1)^{1/2}+D_2]. \tag{C9}$$

Note that the eigenvalues of the matrix $N_{n,2}$ do not enter, and in fact this result depends on both $a_2$ and $b_2$ whereas the eigenvalues of $N_{n,2}$ depend only on $a_2$. Nor is there any reason for eigenvalues to be important. They are appropriate for determining the properties of powers of $N_{n,2}$ and the nonperiodic orbit will never retrace this orbit segment.

Next consider the effect of the roundoff error that arises in computing $(r_1,\theta_1)$ from $(r_0,\theta_0)$. This error is propagated to the point $(r_n,\theta_n)$ over a slightly longer path, so the corresponding orbits in the tangent space are given by

$$N_{n,1}\equiv\begin{pmatrix}a_1+d_1 & c_1+b_1\\ c_1-b_1 & a_1-d_1\end{pmatrix}=N_{n,2}\begin{pmatrix}a+d & c+b\\ c-b & a-d\end{pmatrix},$$

where the second factor on the right propagates the tangent space orbits from $(r_0,\theta_0)$ to $(r_1,\theta_1)$.

Straightforward multiplication yields

$$\begin{aligned}a_1&=aa_2-bb_2+cc_2+dd_2, & b_1&=ab_2+ba_2+cd_2-dc_2,\\ c_1&=ac_2+bd_2+ca_2-db_2, & d_1&=ad_2-bc_2+cb_2+da_2,\end{aligned} \tag{C10}$$

and

$$\begin{aligned}a_1c_1-b_1d_1&=(ac-bd)(a_2^2-b_2^2+c_2^2-d_2^2)-2(ad+bc)(a_2b_2-c_2d_2)\\ &\quad+(a^2+b^2+c^2+d^2)(a_2c_2-b_2d_2),\\ a_1d_1+b_1c_1&=2(ac-bd)(a_2b_2+c_2d_2)+(ad+bc)\\ &\quad\times(a_2^2-b_2^2-c_2^2+d_2^2)\\ &\quad+(a^2+b^2+c^2+d^2)(a_2d_2+b_2c_2).\end{aligned} \tag{C11}$$

We next need some identities that follow from the determinant condition, $a^2+b^2=c^2+d^2+1$,

$$\begin{aligned}a^2-b^2+c^2-d^2&=2(ac-bd)\frac{ac+bd}{a^2+b^2}+\frac{a^2-b^2}{a^2+b^2},\\ ab-cd&=-(ac-bd)\frac{ad-bc}{a^2+b^2}+\frac{ab}{a^2+b^2},\\ ab+cd&=(ad+bc)\frac{ac+bd}{a^2+b^2}+\frac{ab}{a^2+b^2},\\ a^2-b^2-c^2+d^2&=2(ad+bc)\frac{ad-bc}{a^2+b^2}+\frac{a^2-b^2}{a^2+b^2}.\end{aligned} \tag{C12}$$

Using these results in Eq. (C11) yields

$$a_1c_1-b_1d_1=(a_2c_2-b_2d_2)S+(ac-bd)\frac{a_2^2-b_2^2}{a_2^2+b_2^2}-(ad+bc)\frac{2a_2b_2}{a_2^2+b_2^2}, \tag{C13}$$

$$a_1d_1+b_1c_1=(a_2d_2+b_2c_2)S+(ac-bd)\frac{2a_2b_2}{a_2^2+b_2^2}+(ad+bc)\frac{a_2^2-b_2^2}{a_2^2+b_2^2}, \tag{C14}$$

$$a_1^2+b_1^2=(a_2^2+b_2^2)S-(c^2+d^2), \tag{C15}$$

where

$$S\equiv a^2+b^2+c^2+d^2+2(ac-bd)\frac{a_2c_2+b_2d_2}{a_2^2+b_2^2}+2(ad+bc)\frac{a_2d_2-b_2c_2}{a_2^2+b_2^2}. \tag{C16}$$

We are interested in the cases $a_1^2+b_1^2$ and $a_2^2+b_2^2$ are both large, so that the ellipses for both roundoff errors are very long, but the coefficients, $a$, $b$, $c$, and $d$ are not large since the latter only carry the tangent space orbits over one iteration. It follows from Eq. (C15) that $S$ cannot be very small. It also follows from $\mathrm{Det}NN^{\dagger}=1$ that at least one of the factors $a_2c_2-b_2d_2$ and $a_2d_2+b_2c_2$ must be large. The two terms on the right of Eqs. (C13) and (C14) cannot be large. Hence, the two error ellipses must be very nearly parallel in the limit of interest.

It follows that the large accumulated errors at a given point are all in a direction that is characteristic of the given point. Errors perpendicular to this direction are only of the order of the roundoff, and thus, very small. Since the quantity $(a^2+b^2)$ generally grows exponentially with the orbit length, numerical convergence to the characteristic direction is quite rapid.

These quantities can be calculated on a KAM surface. Since this surface can be transformed to $r^{(0)}=\text{const}$ as in Appendix B, the tangent space orbit can be written in the space $(dr^{(0)},\delta\theta^{(0)})$,

$$N_{n,1}=\begin{pmatrix}1 & 0\\ s & 1\end{pmatrix}^n=\begin{pmatrix}1 & 0\\ ns & 1\end{pmatrix}. \tag{C17}$$

Thus, here $(a^2 + b^2)^{1/2}$ grows only linearly with $n$. From Eq. (C9) $\theta_e$ goes to zero in the limit of large $n$, so that the roundoff error is parallel to the KAM surface. Numerical calculations confirm this result for orbits that come close to KAM surfaces.


[1]M. Hénon and C. Heiles, Astron. J. **69**, 73 (1964).
[2]J. Ford, in *Fundamental Problems in Statistical Mechanics III*, edited by E.D.G. Cohen (North Holland, Amsterdam, 1975), p. 215.
[3]K.J. Whiteman, Rep. Prog. Phys. **40**, 1033 (1977).
[4]B.V. Chirikov, "A Universal Instability of Many-Dimensional Oscillator Systems," Phys. Rep. (to be published).
[5]M.V. Berry, in *Topics in Nonlinear Dynamics*, edited by S. Jorna, Am. Inst. Phys. Conf. Proc. (A.I.P., New York, 1978), Vol. 46, p. 16..
[6]Y.M. Treve, in *Topics in Nonlinear Dynamics*, edited by S. Jorna, Am. Inst. Phys. Conf. Proc. (A.I.P., New York, 1978), Vol. 46, p. 147.
[7]J.B. Taylor, unpublished (1968).
[8]V.I. Arnol'd and A. Avez, *Erogodic Problems of Classical Mechanics* (Benjamin, New York, 1968), p. 94.
[9]V.I. Arnol'd, Usp. Mat. Nauk. **18**, 13 (1963) [Russian Mathematical surveys **18** (5), 9 (1963)].
[10]J. Moser, Nachr. Akad. Wiss. Göttingen II Math.-Physik Kl. No. 1 (1962).
[11]J.M. Greene, J. Math. Phys. **9**, 760 (1968).
[12]T. Bountis, Ph. D. Thesis, Physics Dept., University of Rochester, N.Y. (1978); T. Bountis and R.H.G. Helleman, to be submitted to J. Math. Phys.; R.H.G. Helleman, in *Statistical Mechanics and Statistical Methods*, edited by U. Landman (Plenum, New York, 1977), p. 343, Eqs. (2.10)–(2.12).
[13]I. Niven, *Irrational Numbers* (Mathematical Association of America, Menasha, Wisconsin, 1956).
[14]C. Froeschlé, Astron. Astrophys. **9**, 15 (1970).
[15]G. Bennetin, L. Galgani, and J.-M. Strelcyn, Phys. Rev. A **14**, 2338 (1976).
[16]M. Gardner, *Second Scientific American Book of Mathematical Puzzles and Diversions* (Simon and Shuster, New York, 1961).
[17]J.M.A. Danby, Celest. Mech. **8**, 273 (1973).
[18]C.F.F. Karney (private communication).
[19]G. Benettin, M. Casartelli, L. Galgani, A. Giorgilli, and J.-M. Strelcyn, Nuovo Cimento B **44**, 183 (1978).
[20]J. Moser, Bull. Astron. **3**(3), 53 (1968).
[21]L.J. Laslett, E.M. McMillan, and J. Moser, Courant Institute Rep. NYO-1480-101, unpublished (1968).
[22]G.H. Lundsford and J. Ford, J. Math. Phys. **13**, 700 (1972).
[23]A.J. Lichtenberg (private communication).
[24]I.C. Percival, J. Phys. A **12**, L57 (1969).
[25]R. deVogelaere, *Contributions to the Theory of Nonlinear Oscillations*, edited by S. Lefschetz (Princeton U.P., Princeton, New Jersey, 1958), Vol. IV. p.53.
[26]J. Finn, Ph.D. Thesis, Physics Dept. University of Maryland (1974).
[27]I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965), p. 34.

# Adiabatic expansions of solutions of coupled second-order linear differential equations. II[a)]

S. A. Fulling

*Department of Mathematics, Texas A & M University, College Station, Texas 77843*
(Received 28 August 1978)

The "phase-integral" approach to higher-order WKB approximations, associated with the names of Fröman and Chakraborty, is generalized to systems of equations, written in vector notation as $\mathbf{h}''(t) + u^2\mathbf{M}(t)\mathbf{h}(t) = 0$, where $\mathbf{M}$ is positive definite and $u \to \infty$. Expansions are constructed in the form $\mathbf{h} \sim p^{-1/4}\mathbf{e} \exp(-iu\int^t p^{-1/2}\,dt')$, where $p$ and $\mathbf{e}$ are power series in $u^{-1}$ and $\mathbf{e}$ is (asymptotically) a unit vector. Expanding out the higher-order terms in the exponential yields the approximations studied in the first paper of this series, which are less uniform in $t$. (A search for still greater uniformity leads to nonlinear differential equations for the leading term in $\mathbf{e}$, which can be explicitly solved only in special cases, notably that in which all eigenvalues of $\mathbf{M}$ are distinct.) The expansions are proved to be valid asymptotic approximations on finite intervals where the eigenvalues of $\mathbf{M}$ are strictly positive and do not cross (i.e., the multiplicity does not depend on $t$).

## 1. INTRODUCTION

The first paper[1] of this series presented expansions as $u \to \infty$ for the solutions of

$$\mathbf{h}''(t) + u^2\mathbf{M}(t)\mathbf{h}(t) = 0, \qquad (1)$$

where $\mathbf{M}(t)$ is, for each $t$, a positive definite, self-adjoint operator in a finite-dimensional complex Hilbert space. The case of permanently degenerate (i.e., coinciding at all $t$) eigenvalues of $\mathbf{M}$ was handled neatly by employing projection operators onto the eigensubspaces, instead of bases of eigenvectors, whenever possible. If $\mathbf{M}(t)$ is sufficiently smooth, the expansions are defined and asymptotically valid away from points where two or more otherwise distinct eigenvalues cross (or merge). A sequel was promised which would offer a partial resolution of the crossing problem by the method of matched asymptotic expansions. This paper is not that one.[2]

In this paper a slightly different type of expansion is constructed, still for the case of no crossing but arbitrary permanent degeneracy of the eigenvalues. These expansions are better approximations (more uniform in $t$) than those of Ref. 1, and they are formally more appropriate for the application in quantum field theory which motivated this work.

To motivate the new type of expansion we consider the much-studied case of a single ordinary differential equation of the form (1), $M(t)$ now being a positive number. Three kinds of expansions appear in the literature:

(A)[3] $$h(t) \sim M(t)^{-1/4} \exp\left[\mp iu \int_c^t M(t')^{1/2}\,dt'\right] \times \sum_{s=0}^{\infty} (\mp iu)^{-s} a_s(t).$$

(B)[4–7] $$h(t) \sim W(t)^{-1/2} \exp\left[\mp iu \int_c^t W(t')\,dt'\right],$$

$$W(t)^{\nu} \sim M(t)^{\nu/2} + \sum_{s=1}^{\infty} u^{-2s} b_{2s}(t).$$

[Here $\nu$ may be 2, 1, $\frac{1}{2}$, etc.; the results for any choice of $\nu$ are asymptotically equivalent, since $M(t) \neq 0$ by assumption.]

(C)[8] $$h(t) \sim \exp\left\{\mp iu \int_c^t \left[M(t)^{1/2} + \sum_{s=1}^{\infty} u^{-s} c_s(t)\right]\right\}.$$

In Ref. 4 it is shown that the $c_s$ with $s$ odd (which are imaginary) are related to those with $s$ even in such a way that the integrals of the odd terms and their exponentials can be evaluated, to yield the real factor $[uW(t)]^{-1/2}$, where $b_{2s} = c_{2s}$ if $\nu = 1$. Thus the expansion of type C is merely an awkward way of writing the expansion of type B, obscuring the fact that the amplitude (modulus) of the asymptotic approximation is a local functional of $M(t)$ and its derivatives.[9]

To convert an expansion of type B to one of type A, one simply expands

$$\exp\left[\mp iu \int_c^t W(t')\,dt'\right] \Big/ \exp\left[\mp iu \int_c^t M(t')^{1/2}\,dt'\right]$$

as a power series in $u^{-1}$. Clearly this worsens the approximation except in the neighborhood of the starting point, $t = c$, since one is replacing oscillatory functions like $\exp(iu^{1-2s}\int b_{2s}\,dt)$ by truncated power series in possibly monotonic quantities like $\int b_{2s}\,dt$. The type-B expansion involves the *instantaneous effective frequency of oscillation*, $uW(t)$, of the solution; the type-A expansion, by insisting on the naive frequency, $uM(t)^{1/2}$, forces the appearance of undesirable "secular terms" which are nonuniform in $t$ (although of negative order in $u$).[10]

The expansion B (which is often called the "phase-integral approximation") suggests a natural decomposition of

a)Research supported in part by National Science Foundation Grant No. PHY 77-01432.

  

the solutions into a subspace of "positive frequency" and one of "negative frequency." Namely, a solution has positive (negative) frequency if its initial data at the point $c$ coincide with the values at that point of such an expansion and its derivative, with a minus (plus) sign in the exponent. (The unnatural sign convention is traditional in quantum field theory.) This direct sum decomposition is independent of $c$, up to the order in $u$ to which one has carried out the expansion [provided that $M(t)$ is sufficiently smooth for the expansion to be valid]. This fact has important implications for quantum field theory in time-dependent external gravitational (or other) potentials.[11] The present work will, it is hoped, allow the conclusions of Refs. 11 to be extended to space–time models in which the field equation does not decouple into individual normal modes.[12]

The expansions for coupled equations treated in this paper alleviate, but in general do not completely eliminate, the problem of secular terms. Section 5 outlines an attempt to solve the problem completely, which did not yield explicit formulas for the approximation to **h**. Apparently the approach of Sec. 5 can be used to define the "true" approximate normal modes of a coupled system, and the material of the rest of the paper and of Ref. 1 provides approximate representations of these functions which should be adequate for many practical purposes.

## 2. POLAR DECOMPOSITION OF A VECTOR FUNCTION

The characteristic feature of the expansion B discussed in Sec. 1 is the factorization of the solution into an amplitude (a positive real-valued function) and a phase (a complex-valued function of modulus unity): $h(t) = A(t)e^{-iS(t)}$. That $A$ is proportional to $|S'|^{-1/2}$ (for the exact solution; in this case for the approximation as well) is an immediate consequence of the fact that the Wronskian $h^{*\prime} h - h^* h'$ is an imaginary constant; this comes in turn from the differential equation (1) and the reality of $M$.

Our goal is to treat similarly the case where $\mathbf{h}(t)$ takes values in an $N$-dimensional Hilbert space, which we might as well identify with $\mathbb{C}^N$, so that the scalar product is the standard one,

$$(\mathbf{h},\mathbf{g}) = \sum_{j=1}^{N} h_j^* g_j, \tag{2}$$

and Eq. (1) stands for a system of $N$ coupled equations. The Hermiticity of **M** implies the conservation of a generalized Wronskian:

$$(\mathbf{h}',\mathbf{h}) - (\mathbf{h},\mathbf{h}') = iC, \tag{3}$$

where $C$ is a real number independent of $t$. A basis for all the solutions can be chosen to consist of $N$ solutions with $C = 1$ and their complex conjugates, which have $C = -1$. Henceforth we shall consider a solution with $C = 1$.

Any vector-valued function can be written as the product of a scalar amplitude, a scalar phase, and a unit vector:

$$\mathbf{h}(t) = A(t)\mathbf{e}(t)e^{-iS(t)}, \tag{4}$$

$$(\mathbf{e},\mathbf{e}) = 1. \tag{5}$$

We cannot require that **e** be real, since its components may have a relative phase. However, we can restrict the arbitrariness in the phase of **e** by requiring

$$(\mathbf{e},\mathbf{e}') = 0. \tag{6}$$

The real part of Eq. (6) already follows from Eq. (5); the geometrical significance of the imaginary part is that the phase variation of $\mathbf{h}(t)$ has been entirely concentrated into $S(t)$. To see that this is always possible, let $\hat{\mathbf{e}}(t)$ be an arbitrary differentiable function whose value at each $t$ is a unit vector; then

$$S(t) \equiv i\int_c^t (\hat{\mathbf{e}},\hat{\mathbf{e}}')\,dt' \tag{7}$$

is real, and $\mathbf{e} \equiv \hat{\mathbf{e}}e^{iS}$ satisfies Eq. (6). Readers of Ref. 1 will recognize Eq. (6) as the key property of a Kato eigenvector, but in the present context **e** will not generally be an eigenvector of **M**.

From Eqs. (4) and (6) one obtains

$$(\mathbf{h}',\mathbf{h}) - (\mathbf{h},\mathbf{h}') = 2iS'A^2.$$

Thus for a solution satisfying Eq. (3) with $C = 1$, one must have

$$A = (2S')^{-1/2} \quad \text{and} \quad S' > 0. \tag{8}$$

We introduce a quantity $\tilde{p}(t)$ corresponding to $W(t)^2$ in expansion B:

$$S'(t) = u\tilde{p}(t)^{1/2},$$

$$\mathbf{h} = (2u)^{-1/2}\tilde{p}^{-1/4}\mathbf{e}\exp\left[-iu\int_c^t \tilde{p}(t')^{1/2}\,dt'\right]. \tag{9}$$

Then Eq. (1) is equivalent to

$$0 = (\mathbf{M} - \tilde{p})\mathbf{e} - 2iu^{-1}\tilde{p}^{1/2}\mathbf{e}' - u^{-2}\tilde{p}\tilde{f}\mathbf{e} - \tfrac{1}{2}u^{-2}\tilde{p}^{-1}\tilde{p}'\mathbf{e}' + u^{-2}\mathbf{e}'', \tag{10}$$

where

$$\tilde{f}(t) = \tfrac{1}{4}\tilde{p}^{-2}\tilde{p}'' - \tfrac{5}{16}\tilde{p}^{-3}(\tilde{p}')^2. \tag{11}$$

## 3. CALCULATION OF THE TERMS IN THE EXPANSION

Equation (9) with Eqs. (5), (6), and (10) is an *exact* representation of *any* solution of Eq. (1) satisfying Eq. (3) with $C = 1$. The approximation begins when one assumes that $\tilde{p}$ and **e** can be approximated by polynomials in $u^{-1}$:

$$\tilde{p}(t) = \sum_{s=0}^{m+1} u^{-s}p_s(t), \tag{12}$$

$$\mathbf{e}(t) = \sum_{s=0}^{m} u^{-s}\mathbf{e}_s(t). \tag{13}$$

Anticipating that $p_0$ will be an eigenvalue of **M**, we shall write $p$ for $p_0$ and let

$$f \equiv \tfrac{1}{4}p^{-2}p'' - \tfrac{5}{16}p^{-3}(p')^2, \tag{14}$$

so that the notation is consistent with that of Ref. 1. We substitute Eqs. (12) and (13) into Eqs. (10), (6), and (5) and examine the results order by order in $u$. The general term is rather complicated to write out, so attention will be concentrated on the equations of the first few orders and their solvability. (See Appendix A for higher orders.)

The term of order $u^0$ in Eq. (10) is

$$(\mathbf{M}-p)\mathbf{e}_0=0, \tag{15}$$

confirming that $p$ must be an eigenvalue with eigenvector $\mathbf{e}_0$. (It follows that $\tilde{p}$ will be strictly positive for sufficiently large $u$.) Let $\mathbf{P}(t)$ be the orthogonal projection onto the space of all eigenvectors of $\mathbf{M}(t)$ with eigenvalue $p(t)$. According to Eqs. (5) and (6) in lowest order, $\mathbf{e}_0$ is a unit vector satisfying

$$(\mathbf{e}_0,\mathbf{e}_0')=0. \tag{16}$$

If there are no other linearly independent eigenvectors in $\mathscr{H}\equiv\mathbf{P}\mathbb{C}^N$, this is enough to establish that $\mathbf{e}_0(t)$ evolves under the Kato transformation (see Ref. 1, Sec. 2 and references therein):

$$\mathbf{e}_0'(t)=\mathbf{P}'(t)\mathbf{e}_0(t),\quad \mathbf{P}(t)\mathbf{e}_0'(t)=0, \tag{17}$$

or, equivalently,

$$\mathbf{e}_0(t)=\mathbf{U}(t)\mathbf{e}_0(c), \tag{18}$$

$$\mathbf{U}'(t)=\sum_{k=1}^{K}\mathbf{P}_k'(t)\mathbf{P}_k(t)\mathbf{U}(t),\quad \mathbf{U}(c)=\mathbf{1}, \tag{19}$$

where the sum is over the $K$ distinct eigenspaces of $\mathbf{M}(t)$.

Equations (17) and (18) should also be enforced if $p$ is a degenerate eigenvalue, but to see why we must look at the next-order term in Eq. (10), which is

$$(\mathbf{M}-p)\mathbf{e}_1=p_1\mathbf{e}_0+2ip^{1/2}\mathbf{e}_0'. \tag{20}$$

For consistency the projection of the right-hand side of this equation onto $\mathscr{H}$ must vanish. In view of Eq. (16), it follows that Eqs. (17) hold and that

$$p_1=0. \tag{21}$$

Now $\mathbf{e}_0(t)$ is completely determined by its (normalized) initial value, $\mathbf{e}_0(c)$.

The projection of Eq. (20) onto the orthogonal complement, $\mathscr{H}^{\perp}$, of the eigenspace can now be solved:

$$\mathbf{e}_1=\mathbf{e}_1^{\perp}+\mathbf{P}\mathbf{e}_1,\quad \mathbf{e}_1^{\perp}\equiv(\mathbf{1}-\mathbf{P})\mathbf{e}_1, \tag{22}$$

$$\mathbf{e}_1^{\perp}=2ip^{1/2}(\mathbf{M}-p)^{-1}\mathbf{e}_0'. \tag{23}$$

In Eq. (23) the inverse matrix $(\mathbf{M}-p)^{-1}$ is defined as an operator on and onto $\mathscr{H}^{\perp}$. It exists at all $t$ because of our initial "no crossing" assumption: The eigenvalues of $\mathbf{M}(t)$ acting on $\mathscr{H}^{\perp}(t)$ remain permanently distinct from $p(t)$.

The term of order $u^{-1}$ in Eq. (6) is

$$(\mathbf{e}_1,\mathbf{e}_0')+(\mathbf{e}_0,\mathbf{e}_1')=0. \tag{24}$$

A short manipulation converts this equation to

$$(\mathbf{e}_0,(\mathbf{P}\mathbf{e}_1)')=2i\,\mathrm{Im}(\mathbf{e}_0',\mathbf{e}_1^{\perp})$$

$$=4ip^{1/2}(\mathbf{e}_0',(\mathbf{M}-p)^{-1}\mathbf{e}_0'). \tag{25}$$

Before we can use Eq. (25) to find $\mathbf{P}\mathbf{e}_1$, we must look at the projection onto $\mathscr{H}$ of the $u^{-2}$ term of Eq. (10):

$$p_2\mathbf{e}_0=-2ip^{1/2}\mathbf{P}(\mathbf{e}_1')-pf\mathbf{e}_0+\mathbf{P}\mathbf{e}_0''. \tag{26}$$

The right-hand side must be a multiple of $\mathbf{e}_0$, so, projecting onto the other $p$-eigenvectors, we get

$$(\mathbf{P}-\mathbf{e}_0\otimes\mathbf{e}_0^*)(\mathbf{e}_1')=-\tfrac{1}{2}ip^{-1/2}(\mathbf{P}-\mathbf{e}_0\otimes\mathbf{e}_0^*)(\mathbf{e}_0''). \tag{27}$$

Therefore, combining Eqs. (25), (27), and (23), we have

$$\begin{aligned}\mathbf{P}[(\mathbf{P}\mathbf{e}_1)'] &=\mathbf{e}_0(\mathbf{e}_0,(\mathbf{P}\mathbf{e}_1)')+(\mathbf{P}-\mathbf{e}_0\otimes\mathbf{e}_0^*)(\mathbf{e}_1'-\mathbf{e}_1^{\perp\prime})\\ &=4ip^{1/2}(\mathbf{e}_0',(\mathbf{M}-p)^{-1}\mathbf{e}_0')\mathbf{e}_0\\ &\quad-\tfrac{1}{2}ip^{-1/2}(\mathbf{p}-\mathbf{e}_0\otimes\mathbf{e}_0^*)(\mathbf{e}_0'')\\ &\quad-2ip(\mathbf{P}-\mathbf{e}_0\otimes\mathbf{e}_0^*)\frac{d}{dt}[p^{1/2}(\mathbf{M}-p)^{-1}\mathbf{e}_0'],\end{aligned} \tag{28}$$

where $\mathbf{e}_0\otimes\mathbf{e}_0^*$ (or $|e_0\rangle\langle e_0|$) is the orthogonal projection onto $\mathbf{e}_0$. To solve Eq. (28) for $\mathbf{P}\mathbf{e}_1$ is a problem of a type solved in Ref. 1, Sec. 3:

$$(\mathbf{P}\mathbf{e}_1)(t)=\mathbf{U}(t)\hat{\mathbf{a}}_1(t), \tag{29}$$

$$\hat{\mathbf{a}}_1(t)=\int_c^t dt'\,\mathbf{U}(t')^{-1}\{\mathbf{P}[(\mathbf{P}\mathbf{e}_1)']\}(t')+(\mathbf{P}\mathbf{e}_1)(c). \tag{30}$$

Here $\hat{\mathbf{a}}_1(t)$ is in $\mathscr{H}(c)$, not $\mathscr{H}(t)$. If an explicit basis of Kato eigenvectors is introduced for $\mathscr{H}(t)$, the effect of the operator $\mathbf{U}(t)\mathbf{U}(t')^{-1}$ is to move the basis vectors outside of the integral, which then becomes a set of scalar indefinite integrals of the coefficient functions of the expansion of $\mathbf{P}[(\mathbf{P}\mathbf{e}_1)']$ in terms of that basis. (See the examples in Appendix B.)

The term of order $u^{-1}$ in Eq. (5) is

$$2\mathrm{Re}(\mathbf{e}_0,\mathbf{e}_1)=0; \tag{31}$$

that is, $(\mathbf{e}_0,\mathbf{P}\mathbf{e}_1)$ is imaginary. This is a constraint on the constant of integration, $\mathbf{P}\mathbf{e}_1(c)$, in Eq. (30). The determination of $\mathbf{e}_1$ is now complete. Also, Eq. (26) now yields

$$\begin{aligned}p_2&=-2ip^{1/2}(\mathbf{e}_0\mathbf{e}_1')+(\mathbf{e}_0,\mathbf{e}_0'')-pf\\ &=-(\mathbf{e}_0,\mathbf{P}'(\mathbf{M}-5p)(\mathbf{M}-p)^{-1}\mathbf{P}'\mathbf{e}_0)-pf,\end{aligned} \tag{32}$$

which is real, as it ought to be. [The second equality is derived with aid of Eqs. (17), (23), (24), the derivative of (16), and the Hermiticity of $\mathbf{P}'$.]

The continuation to higher orders is clear in principle. The projection onto $\mathscr{H}^{\perp}$ of the $s$th-order terms in Eq. (10) determines $\mathbf{e}_s^{\perp}$, while the projection onto $\mathbf{e}_0$ determines $p_s$. The $s$th-order part of Eq. (6) gives the component of $(\mathbf{P}\mathbf{e}_s)'$ proportional to $\mathbf{e}_0$, and the rest of $\mathbf{P}[(\mathbf{P}\mathbf{e}_s)']$ must be obtained by acting with $\mathbf{P}-\mathbf{e}_0\otimes\mathbf{e}_0^*$ on the term of order $s+1$ in Eq. (10). One can then find $\mathbf{P}\mathbf{e}_s$ by integration, imposing the $s$th-order term of Eq. (5) on the integration constant, $\mathbf{P}\mathbf{e}_s(c)$.

The recursion relations for the first few orders are recorded in Appendix A, and an example is worked out in Appendix B.

## 4. COMMENTS

It is not obvious from the algorithm that the $p_s$ are guaranteed to turn out *real* for $s>2$, but this can be established by a tedious *reductio ad absurdum*, which we omit.

Unlike the approximations constructed in Ref. 1, this approximation is not linear in its dependence on the starting values, $\mathbf{e}_0(c)$ and $\mathbf{Pe}_s(c)$. In particular, if $\mathbf{b}_1$ and $\mathbf{b}_2$ are orthogonal eigenvectors with the same eigenvalue $p$, then the expansion built on $\mathbf{e}_0 = A_1\mathbf{b}_1 + A_2\mathbf{b}_2$ ($A_j \in \mathbb{C}$) is not the corresponding linear combination of the expansions with $\mathbf{e}_0 = \mathbf{b}_1$ and $\mathbf{e}_0 = \mathbf{b}_2$. Nevertheless, the asymptotic validity of both series (Sec. 6) implies tht they must agree up through the order to which the expansion is carried out, both with each other and with the corresponding expansion of the type of Ref. 1. This can verified to the lower nontrivial order from the formulas (9), (15)–(16), (23), (28)–(30), and (32), and the corresponding formulas (16), (17), (22), (23), and (25)–(27) of Ref. 1. (See the example in Appendix B.)

A closely related effect is that $p_s$ ($s \geqslant 2$) will not be the same for all vectors $\mathbf{e}_0$ in a given eigenspace, since such quantities as $(\mathbf{e}_0,\mathbf{e}_0'')$ will differ.

A more positive observation is that $p_2$ does not depend upon the constant of integration, $\mathbf{Pe}_1(c)$, in Eqs. (29)–(30). Consequently, $p_2(t)$ can be computed as if $\mathbf{Pe}_1(t) = 0$, and hence does not involve an integral over $t'$. That is, the frequency correction depends only on the *local* behavior of $\mathbf{M}$ and its derivatives, not on "past history." One would expect this to hold true in all orders in the nondegenerate case, but it is not evident from the recursion relations for $p \geqslant 3$. [In the degenerate case, the component of the integral in $\mathbf{Pe}_s(t)$ perpendicular to $\mathbf{e}_0(t)$ in effect modifies the direction in the $p$-eigenspace in which $\mathbf{e}(t)$ points, and hence it *should* affect the frequency in higher orders, by the mechanism mentioned in the preceding paragraph.]

## 5. SKETCH OF A POSSIBLE REFINEMENT

The integrals in Eq. (30) and the analogous higher-order formulas for $\mathbf{Pe}_s$ are, alas, secular terms of the sort that were eliminated in the one-dimensional problem by abandoning type-A expansions for type-B expansions. Thus, although our results improve on Ref. 1,[13] they do not completely achieve the intended goal. Satisfying Eq. (6) order-by-order by a series (13) is evidently a sin of the same nature as satisfying $|e^{-iS}| = 1$ by

$$e^{-iS} = \exp\Big(-iu\int^t p^{1/2}\,dt'\Big)[1 - iu^{-1}S_1 - u^{-2}S_2 + \cdots]$$

(which would lead to an expansion of type A), and it suffers a similar punishment. A completely satisfactory resolution of this problem does not seem to be possible, but some relevant observations can be reported here.

It is clear that the offensive integral terms, which lie in the eigenspace $\mathscr{H}$, ought to be somehow absorbed into $\mathbf{e}_0(t)$. Let us continue to hypothesize the expansions (12) and (13), but allow the $\mathbf{e}_s$ to depend on $u$. Suppose that

$$\mathbf{e}_0' = \sum_{s=0}^{m} u^{-s}(\mathbf{e}_0')_s. \tag{33}$$

{This is analogous to writing

$$(e^{-iS})' = -iu\Big(\sum_{s=0}^{m+1} u^{-s}p_s\Big)^{1/2} e^{-iS},$$

which has $e^{-iS} = \exp[-iu \int (\Sigma\, u^{-s}p_s)^{1/2}\,dt']$ as its solution.} In deriving recursion relations, $\mathbf{e}_0, \mathbf{e}_1, \cdots$ and $(\mathbf{e}_0')_0, (\mathbf{e}_0')_1, \cdots$ are all to be formally treated as independent of $u$. The recursion relations will determine $\mathbf{e}_s$ ($s \geqslant 1$) in terms of $\mathbf{e}_0$ and $\mathbf{e}_0'$, and hence induce expansions like Eq. (33) for $\mathbf{e}_s'$.

Under these rules, the term of lowest order in Eq. (1) is still Eq. (15), so $\mathbf{e}_0$ is still an eigenvector with eigenvalue $p$. The normalization condition, $(\mathbf{e}_0,\mathbf{e}_0) = 1$, from Eq. (5) likewise survives. The leading part of Eq. (6), however, is now

$$(\mathbf{e}_0,(\mathbf{e}_0')_0) = 0. \tag{34}$$

Note next that $\mathbf{e}_0 = \mathbf{Pe}_0$ implies $\mathbf{e}_0' = \mathbf{Pe}_0' + \mathbf{P}'\mathbf{e}_0$, or

$$\sum_{s=0}^{m} u^{-s}(\mathbf{e}_0')_s = \sum_{s=0}^{m} u^{-s}\mathbf{P}[(\mathbf{e}_0')_s] + \mathbf{P}'\mathbf{e}_0. \tag{35}$$

It follows that

$$(\mathbf{e}_0')_s \in \mathscr{H} \equiv \mathbf{P}\mathbb{C}^N \quad \text{for } s > 0 \tag{36}$$

and that

$$(\mathbf{e}_0')_0^{\perp} = (\mathbf{e}_0')^{\perp} = \mathbf{P}'\mathbf{e}_0. \tag{37}$$

Proceeding to the next order, we obtain from Eq. (10)

$$(\mathbf{M} - p)\mathbf{e}_1 - p_1\mathbf{e}_0 - 2ip^{1/2}(\mathbf{e}_0')_0 = 0. \tag{38}$$

Projecting this equation with $\mathbf{P} - \mathbf{e}_0 \otimes \mathbf{e}_0^*$, we find that $(\mathbf{e}_0')_0$ has no component in $\mathscr{H}$ orthogonal to $\mathbf{e}_0$. This fact combines with Eqs. (34) and (37) to establish that

$$(\mathbf{e}_0')_0 = \mathbf{P}'\mathbf{e}_0; \tag{39}$$

to lowest order, $\mathbf{e}_0$ behaves as a Kato eigenvector. The projection of Eq. (38) onto $\mathbf{e}_0$ yields, as before, $p_1 = 0$, and the projection onto $\mathscr{H}^{\perp}$ yields

$$\mathbf{e}_1^{\perp} = 2ip^{1/2}(\mathbf{M} - p)^{-1}(\mathbf{P}'\mathbf{e}_0) \tag{40}$$

as the correct version of Eq. (23) in this context. Also, Eq. (5) in this order again leads to Eq. (31).

In place of Eqs. (24)–(25), the $u^{-1}$ part of Eq. (6) is now

$$(\mathbf{e}_1,(\mathbf{e}_0')_0) + (\mathbf{e}_0,(\mathbf{e}_0')_1) + (\mathbf{e}_0,(\mathbf{e}_1')_0) = 0. \tag{41}$$

Recall that our objective is to avoid recursion relations whose solutions involve indefinite integrals (except in contexts like the exponent of $e^{-iS}$, where they are irrelevant to the magnitude of the error of the approximation). This can be done by avoiding inhomogeneous differential equations like Eq. (28) for $\mathbf{Pe}_s$. Instead, we shall in each order simply choose $\mathbf{Pe}_s$ to be zero, except insofar as Eq. (5) decrees otherwise. The latter will always yield a condition of form

$$2\,\mathrm{Re}(\mathbf{e}_0,\mathbf{e}_s) = \text{prescribed function of } t, \tag{42}$$

but we are free to set

$$\mathrm{Im}(\mathbf{e}_0,\mathbf{e}_s) = 0, \quad (\mathbf{P} - \mathbf{e}_0 \otimes \mathbf{e}_0^*)(\mathbf{e}_s) = 0. \tag{43}$$

The inconsistencies which would thereby have arisen in the procedure of Sec. 3 are avoided this time because of the freedom to choose $(\mathbf{e}_0')_s$. In the case of $\mathbf{e}_1$, Eq. (42) is Eq. (31) and we have

$$\mathbf{Pe}_1 = 0. \tag{44}$$

Hence Eq. (41) becomes

$$(\mathbf{e}_0,(\mathbf{e}_0')_1) = -(\mathbf{e}_1,\mathbf{P}'\mathbf{e}_0) - (\mathbf{e}_0,(\mathbf{e}_1^{\perp\prime})_0)$$

$$= -(\mathbf{e}_1^\perp,\mathbf{P}'\mathbf{e}_0) + ((\mathbf{e}_0')_0,\mathbf{e}_1^\perp) - \frac{d}{dt}(\mathbf{e}_0,\mathbf{e}_1^\perp)$$

$$= 2i\,\mathrm{Im}(\mathbf{P}'\mathbf{e}_0,\mathbf{e}_1^\perp)$$

or

$$(\mathbf{e}_0,(\mathbf{e}_0')_1) = 4ip^{1/2}(\mathbf{e}_0,\mathbf{P}'(\mathbf{M}-p)^{-1}\mathbf{P}'\mathbf{e}_0). \qquad (45)$$

To finish the determination of $(\mathbf{e}_0')_1$, one must act with $\mathbf{P} - \mathbf{e}_0 \otimes \mathbf{e}_0^*$ on the second-order term of Eq. (10). The result combines with Eq. (45) to give

$$(\mathbf{e}_0')_1 = 4ip^{1/2}(\mathbf{e}_0,\mathbf{P}'(\mathbf{M}-p)^{-1}\mathbf{P}'\mathbf{e}_0)\mathbf{e}_0$$

$$-(\mathbf{P}-\mathbf{e}_0\otimes\mathbf{e}_0^*)[(\mathbf{e}_1')_0 + \tfrac{1}{2}ip^{-1/2}((\mathbf{e}_0')_0')_0]. \qquad (46)$$

From Eq. (39) one finds that

$$((\mathbf{e}_0')_0')_0 = \mathbf{P}''\mathbf{e}_0 + (\mathbf{P}')^2\mathbf{e}_0, \qquad (47)$$

and $(\mathbf{e}_1')_0 = (\mathbf{e}_1^{\perp\prime})_0$ can be calculated from Eq. (40).

The other equations of order $u^{-2}$ determine $p_2$ [for which the final version of Eq. (32) remains valid], $\mathbf{e}_2^\perp$,

$$2\mathrm{Re}(\mathbf{e}_0,\mathbf{e}_2) = -(\mathbf{e}_1^\perp,\mathbf{e}_1^\perp), \qquad (48)$$

and

$$(\mathbf{e}_0,(\mathbf{e}_0')_2) = -(\mathbf{e}_0,(\mathbf{e}_2')_0) - (\mathbf{e}_2,(\mathbf{e}_0')_0) - (\mathbf{e}_1,(\mathbf{e}_1')_0). \quad (49)$$

The $\mathbf{e}_2$ terms in Eq. (49) can be found from Eqs. (43) and (48) and the formula for $\mathbf{e}_2^\perp$.

In this way one can determine to any desired order the expression (33) for $\mathbf{e}_0'$. Then $\mathbf{e}_0$ is the solution of this differential equation with a given initial value, $\mathbf{e}_0(c)$, and $\mathbf{e}_s$ and $p_s$ are determined from $\mathbf{e}_0$ by the formulas derived in the foregoing discussion. The only problem is that the terms on the right-hand side of Eq. (33) depend on $\mathbf{e}_0(t)$ nonlinearly, so that it is not possible to write down an explicit general solution. For this reason the expansion constructed in this section is unlikely to be of much practical use. Its existence should be of some theoretical interest, however.

One exception to this gloomy conclusion is the case when the eigenvalue $p$ is not degenerate. There $\mathbf{P} - \mathbf{e}_0 \otimes \mathbf{e}_0^*$ is zero, and $\mathbf{e}_0$ is already known up to a phase factor, which cancels out of expressions like (46) and (48), which are always matrix elements of operators with respect to $\mathbf{e}_0$. Thus Eq. (33) takes the form

$$(\mathbf{e}_0,\mathbf{e}_0') = \text{prescribed function of } t \text{ and } u, \qquad (50)$$

and such a problem can always be solved by the method which led to Eq. (7). See the examples in Appendix B.

An alternative approach to this case is worth mentioning because of its relative conceptual simplicity. Returning to Sec. 2, let us relax Eq. (6), although keeping Eq. (5). Then Eq. (8) must be replaced by

$$A = (2u)^{-1/2}[\tilde{p}^{1/2} - u^{-1}\,\mathrm{Im}(\mathbf{e},\mathbf{e}')]^{-1/2}. \qquad (51)$$

This will complicate Eq. (10) and add many terms to the resulting recursion relations in the method of Sec. 3. But now we no longer have the constraints of the type of Eq. (25), so we may set $\mathrm{Im}(\mathbf{e}_0, \mathbf{e}_s) = 0$ by fiat as in the method discussed earlier in this section, but proceed otherwise as in Sec. 3. Thus if $\mathscr{H}$ (the eigensubspace) is one-dimensional, $\mathbf{e}_s(t)$ will be, except for phase, a purely local functional of $\mathbf{M}(t)$, without the polynomially growing secular terms which can arise from iterated indefinite integrals in the unmodified approach of Sec. 3. However, if $\mathbf{e}_0$ is not the only eigenvector in $\mathscr{H}$, this approach has little advantage over Sec. 3, since the equations of the type of Eq. (27) will force integral terms to be included in the approximation.

In the nondegenerate case the expansion based on Eq. (51) clearly must be equivalent to the one based on Eq. (33) [or (50)], with some changes in the meaning of $\mathbf{e}$ and $\tilde{p}$. Eq. (51) is slightly misleading, since it suggests that the amplitude of the approximation (9) changes when the condition $\mathrm{Im}(\mathbf{e},\mathbf{e}') = 0$ is dropped. Actually, it is the frequency, $p^{1/2}$, which changes, as a phase factor is transferred from $\mathbf{e}$ to $e^{iS}$.

## 6. ASYMPTOTIC VALIDITY

*Theorem*: Let $p(t)$ be one of the eigenvalues of a positive definite Hermitian matrix function $\mathbf{M}(t)$ which is continuously differentiable $m + 3$ times. Define $\mathbf{h}_m(t)$ by Eqs. (9), (12), and (13), with $p_s$ and $\mathbf{e}_s$ calculated by the recursive process described in Sec. 3 and Appendix A. Let $\mathbf{h}(t)$ be the solution of Eq. (1) with initial data $\mathbf{h}(c) = \mathbf{h}_m(c)$, $\mathbf{h}'(c) = \mathbf{h}_m'(c)$. Let

$$\mathbf{Z}_m = u^{1/2}(\mathbf{h} - \mathbf{h}_m). \qquad (52)$$

Then as $u \to \infty$,

$$\|\mathbf{Z}_m(t)\| = O(u^{-(m+1)}) \quad \text{and} \quad \|\mathbf{Z}_m'(t)\| = O(u^{-m}) \qquad (53)$$

throughout any bounded closed interval of $t$ containing $c$ and containing no point of crossing of the eigenvalues.

The proof is almost identical to that in Sec. 5 of Ref. 1, which should be consulted for details. By construction [in particular, the fulfillment of Eq. (10) through order $u^{-n}$] we have

$$\mathbf{Z}_n'' + u^2\mathbf{M}\mathbf{Z}_n = O(u^{-(n-1)}), \qquad (54)$$

and this conclusion is unaffected if $p_{n+1}$ and $\mathbf{P}\mathbf{e}_n$ are omitted from the sums (12)–(13). Since the Green function for the inhomogeneous equation is $O(u^{-1})$ and its derivative is bounded, Eq. (54) implies that $\mathbf{Z}_n = O(u^{-n})$ and $\mathbf{Z}_n' = O(u^{-(n-1)})$. If we let $n = m + 1$, Eq. (53) follows, since $\mathbf{Z}_m$ differs from the modified $\mathbf{Z}_{m+1}$ only by the term involving $u^{-(m+1)}\mathbf{e}_{m+1}^\perp$.

To construct $\mathbf{Z}_{m+1}''$ we need derivatives of order $m + 3$ of the eigenvalues and eigenprojections of $\mathbf{M}$. These exist, since we are avoiding crossing points.[14]

*Corollary*: Let $\{\mathbf{e}_{0(j)}(t)\}$ $(j = 1,\dots,N)$ be an orthonormal basis for $\mathbb{C}^N$ consisting of eigenvectors of $\mathbf{M}(t)$. Define corresponding approximate solutions $\mathbf{h}_{m(j)}(t)$. If $u$ is sufficiently large, the vectors $\mathbf{e}_{(j)}(t)$ [Eq. (13)] are linearly independent. Therefore, every solution of Eq. (1) is approximated in the sense of Eq. (53) by some linear combination of the $\mathbf{h}_{m(j)}$ and their complex conjugates.

One would expect the conclusions of Ref. 1 and this paper to apply in an infinite-dimensional Hilbert space if $\mathbf{M}(t)$ is a positive definite self-adjoint operator with entirely discrete spectrum—for example, an elliptic differential operator on a compact manifold with no vanishing eigenvalues, for which Eq. (1) is a classical hyperbolic partial differential equation. The main technical point to be established is that the operator $\mathbf{G}(t,t')$ describing the general solution of the Cauchy problem exists[15] and has the nice behavior at $u\rightarrow\infty$ described above for the finite-dimensional Green function. However, a more interesting extension of this work would be to obtain a better understanding of what happens when eigenvalues cross or merge; this might suggest a new, more powerful approach to these problems which would be applicable to very general operators, including those with continuous spectrum.

## ACKNOWLEDGMENTS

I am grateful to B.L. Hu for motivation and encouragement during the early stages of this work, especially during a brief period of hospitality extended to me in 1974 by the University of Maryland relativity group. I was able to return to the project during the 1978 summer program of the Aspen Center for Physics.

## APPENDIX A: RECURSION RELATIONS

The equations which come from Eq. (5), after $(\mathbf{e}_0,\mathbf{e}_0)=1$, are all of the form

$$2\mathrm{Re}(\mathbf{e}_0,\mathbf{e}_s) = -\sum_{n=1}^{s-1}(\mathbf{e}_n,\mathbf{e}_{s-n}). \tag{A1}$$

The equations from Eq. (6), after $(\mathbf{e}_0,\mathbf{e}_0')=0$, are of the form

$$(\mathbf{e}_0,(\mathbf{P}\mathbf{e}_s)') = 2i\,\mathrm{Im}(\mathbf{e}_0',\mathbf{e}_s^\perp) - \sum_{n=1}^{s-1}(\mathbf{e}_n,\mathbf{e}_{s-n}'). \tag{A2}$$

The first seven orders of Eq. (10) are

$$(\mathbf{M}-p)\mathbf{e}_0 = 0, \tag{A3a}$$

$$(\mathbf{M}-p)\mathbf{e}_1 = p_1\mathbf{e}_0 + 2ip^{1/2}\mathbf{e}_0', \tag{A3b}$$

$$(\mathbf{M}-p)\mathbf{e}_2 = 2ip^{1/2}\mathbf{e}_1' + [p_2 + \tfrac{1}{4}p''p^{-1} - \tfrac{5}{16}(p')^2p^{-2}]\mathbf{e}_0 + \tfrac{1}{2}p'p^{-1}\mathbf{e}_0' - \mathbf{e}_0'', \tag{A3c}$$

$$(\mathbf{M}-p)\mathbf{e}_3 = 2ip^{1/2}\mathbf{e}_2' + [p_2 + \tfrac{1}{4}p''p^{-1} - \tfrac{5}{16}(p')^2p^{-2}]\mathbf{e}_1 + \tfrac{1}{2}p'p^{-1}\mathbf{e}_1' - \mathbf{e}_1'' + p_3\mathbf{e}_0 + ip_2p^{-1/2}\mathbf{e}_0', \tag{A3d}$$

$$\begin{aligned}(\mathbf{M}-p)\mathbf{e}_4 = {}& [(\mathbf{M}-p)\mathbf{e}_3 \text{ with } \mathbf{e}_s \text{ replaced by } \mathbf{e}_{s+1}] \\ & + [p_4 + \tfrac{1}{4}p_2''p^{-1} - \tfrac{5}{8}p_2'p'p^{-2} - \tfrac{1}{4}p_2p''p^{-2} \\ & + \tfrac{5}{8}p_2(p')^2p^{-3}]\mathbf{e}_0 \\ & + [ip_3p^{-1/2} + \tfrac{1}{2}p_2'p^{-1} - \tfrac{1}{2}p_2p'p^{-2}]\mathbf{e}_0',\end{aligned} \tag{A3e}$$

$$\begin{aligned}(\mathbf{M}-p)\mathbf{e}_5 = {}& [(\mathbf{M}-p)\mathbf{e}_4 \text{ with } \mathbf{e}_s \text{ replaced by } \mathbf{e}_{s+1}] \\ & + [p_5 + \tfrac{1}{4}p_3''p'p^{-1} - \tfrac{5}{8}p_3'p'p^{-2} \\ & - \tfrac{1}{4}p_3p''p^{-2} + \tfrac{5}{8}p_3(p')^2p^{-3}]\mathbf{e}_0 \\ & + [ip_4p^{-1/2} + \tfrac{1}{2}p_3'p^{-1} \\ & - \tfrac{1}{2}p_3p'p^{-2} - \tfrac{1}{4}ip_2^2p^{-3/2}]\mathbf{e}_0',\end{aligned} \tag{A3f}$$

$$\begin{aligned}(\mathbf{M}-p)\mathbf{e}_6 = {}& [(\mathbf{M}-p)\mathbf{e}_5 \text{ with } \mathbf{e}_s \text{ replaced by } \mathbf{e}_{s+1}] \\ & + [p_6 + \tfrac{1}{4}p_4''p^{-1} - \tfrac{5}{8}p_4'p'p^{-2} \\ & - \tfrac{1}{4}p_4p''p^{-2} + \tfrac{5}{8}p_4(p')^2p^{-3} \\ & - \tfrac{1}{4}p_2p_2''p^{-2} - \tfrac{5}{16}(p_2')^2p^{-2} \\ & + \tfrac{5}{4}p_2p_2'p'p^{-3} + \tfrac{1}{4}p_2^2p''p^{-3} \\ & - \tfrac{15}{16}p_2^2(p')^2p^{-4}]\mathbf{e}_0 + [ip_5p^{-1/2} \\ & + \tfrac{1}{2}p_4'p^{-1} - \tfrac{1}{2}p_4p'p^{-2} - \tfrac{1}{2}ip_3p_2p^{-3/2} \\ & - \tfrac{1}{2}p_2p_2'p^{-2} + \tfrac{1}{2}p_2^2p'p^{-3}]\mathbf{e}_0'.\end{aligned} \tag{A3g}$$

[All terms involving $p_1$ have been omitted from Eqs. (A3c)–(A3g), since Eq. (A3b) implies that $p_1=0$.] The formulas for $\mathbf{e}_s^\perp$, $p_s$, and $(\mathbf{P}-\mathbf{e}_0\otimes\mathbf{e}_0^*)\mathbf{e}_{s-1}'$ are obtained by projecting Eqs. (A3) onto the appropriate subspaces.

Equations (A3) were generated by computer with a symbolic algebra program.[16] The first four were verified by hand calculation, including the $p_1$ terms prototypical of the expansion to higher orders.

## APPENDIX B: EXAMPLE

Let us consider the example treated in Ref. 1,

$$\mathbf{M}(t) = \begin{pmatrix} t & 0 & 0 \\ 0 & tC^2+S^2 & (t-1)CS \\ 0 & (t-1)CS & tS^2+C^2 \end{pmatrix}, \tag{B1}$$

$$C\equiv\cos t,\quad S\equiv\sin t,\quad 1<t<\infty,\quad c=\pi.$$

This has a double eigenvalue, $p=t$, with a Kato basis of eigenvectors

$$\mathbf{b}_1^{(1)} = \begin{pmatrix}1\\0\\0\end{pmatrix},\quad \mathbf{b}_1^{(2)} = \begin{pmatrix}0\\C\\S\end{pmatrix}, \tag{B2}$$

and an eigenvalue $p=1$ with Kato eigenvector

$$\mathbf{b}_2=\begin{pmatrix}0\\-S\\C\end{pmatrix}. \tag{B3}$$

The derivatives of these vectors are

$$\mathbf{b}_1^{(1)\prime}=0,\quad \mathbf{b}_1^{(2)\prime}=\mathbf{b}_2,\quad \mathbf{b}_2'=-\mathbf{b}_1^{(2)}. \tag{B4}$$

Applying the formulas in Secs. 2 and 3 of the present paper, one easily finds the adiabatic expansions corresponding to $\mathbf{e}_0=\mathbf{b}_1^{(1)},\mathbf{b}_1^{(2)}$, and $\mathbf{b}_2$, respectively:

$$\mathbf{h}_1^{(1)}=(2u)^{-1/2}[t+O(u^{-2})]^{-1/4}\mathbf{b}_1^{(1)}$$

$$\times\exp\Big\{-iu\int_\pi^t[t'+u^{-2}\tfrac{5}{16}(t')^{-2}+O(u^{-3})]^{1/2}dt'\Big\}, \tag{B5a}$$

$$\mathbf{h}_1^{(2)}=(2u)^{-1/2}[t+O(u^{-2})]^{-1/4}\Big(\mathbf{b}_1^{(2)}+2iu^{-1}t^{1/2}$$

$$\times(1-t)^{-1}\mathbf{b}_2+4iu^{-1}\int_\pi^t\frac{(t')^{1/2}}{1-t'}dt'\mathbf{b}_1^{(2)}+O(u^{-2})\Big)$$

$$\times\exp\Big\{-iu\int_\pi^t\Big[t'+u^{-2}\Big(\frac{5t'-1}{1-t'}+\frac{5}{16}(t')^{-2}\Big)$$

$$\times+O(u^{-3})\Big\}^{1/2}dt', \tag{B5b}$$

$$\mathbf{h}_2=(2u)^{-1/2}\Big(\mathbf{e}_2-2iu^{-1}(t-1)^{-1}\mathbf{b}_1^{(2)}+4iu^{-1}$$

$$\times\int_\pi^t\frac{dt'}{t'-1}\mathbf{b}_2+O(u^{-2})\Big)$$

$$\times\exp\Big[-iu\int_\pi^t\Big(1+u^{-2}\frac{5-t'}{t'-1}+O(u^{-3})\Big)^{1/2}dt'\Big]. \tag{B5c}$$

Here the integration constants $\mathbf{Pe}_1(\pi)$ have been taken equal to 0.

Expanding quantities such as

$$\frac{\exp\{-iu\int_\pi^t[t'+u^{-2}\frac{5}{16}(t')^{-2}]^{1/2}dt'\}}{\exp[-iu\int_\pi^t(t')^{1/2}dt']}$$

as power series in $u^{-1}$, one recovers from Eqs. (B5) the series expansions found for this example in Ref. 1, Sec. 4. (Note that the contribution from $p_{s+1}$ is of the same order as that from $\mathbf{Pe}_s$; cf. Sec. 6.) Similarly, one can show that $A\,\mathbf{h}_1^{(1)}+B\,\mathbf{h}_1^{(2)}$, with $|A|^2+|B|^2=1$, is equal (but only up through the terms of the indicated orders) to

$$(2u)^{-1/2}[t+O(u^{-2})]^{-1/4}\Big(A\mathbf{b}_1^{(1)}+B\mathbf{b}_1^{(2)}+2iu^{-1}t^{1/2}$$

$$\times(1-t)^{-1}B\mathbf{b}_2+4iu^{-1}|B|^2\int_\pi^t\frac{(t')^{1/2}}{1-t'}dt'(A\mathbf{b}_1^{(1)}$$

$$+B\mathbf{b}_1^{(2)})+\tfrac{1}{2}iu^{-1}AB\int_\pi^t(t')^{-1/2}\frac{1+3t'}{1-t'}dt'$$

$$\times(-B^*\mathbf{b}_1^{(1)}+A^*\mathbf{b}_1^{(2)})+O(u^{-2})\Big)$$

$$\times\exp\Big\{-iu\int_\pi^t\Big[t'+u^{-2}\Big(|B|^2\frac{5t'-1}{1-t'}$$

$$+\frac{5}{16}(t')^{-2}\Big)+O(u^{-3})\Big]^{1/2}dt'\Big\}, \tag{B6}$$

which is the approximate solution corresponding to $\mathbf{e}_0=A\mathbf{b}_1^{(1)}+B\mathbf{b}_1^{(2)}$. Of course, the power-series expansions of oscillatory functions involved here are not uniformly valid in $t$.

Incidentally, it is instructive to try to manipulate a linear combination of expansions corresponding to two *different* eigenvalues, $p$ and $q$, into a single expression like Eq. (B6). One encounters in the amplitude factors like $\exp[-iu^{+1}\int(p^{1/2}-q^{1/2})\,dt']$, which cannot be expanded in power series in $u^{-1}$. This shows why the *assumption* (12) forces Eq. (15) to be true, even though the representation (9) is valid for *any* nonvanishing function satisfying $(\mathbf{h}',\mathbf{h})-(\mathbf{h},\mathbf{h}')=i$.

Finally, let us attempt to find purely local adiabatic expansions by the approach of Sec. 5. In the first nontrivial order, we must solve

$$\mathbf{e}_0'=\mathbf{P}'\mathbf{e}_0+u^{-1}(\mathbf{e}_0')_1, \tag{B7}$$

where $(\mathbf{e}_0')_1$ is given by Eq. (46) with Eqs. (40) and (47). Since

$$\mathbf{P}=\mathbf{P}_1\equiv\sum_{j=1}^{2}b_1^{(j)}\otimes b_1^{(j)*}, \tag{B8}$$

we have

$$\mathbf{P}'=\mathbf{b}_2\otimes\mathbf{b}_1^{(2)*}+\mathbf{b}_1^{(2)}\otimes\mathbf{b}_2^*, \tag{B9}$$

$$\mathbf{P}''+(\mathbf{P}')^2=3\mathbf{b}_2\otimes\mathbf{b}_2^*-\mathbf{b}_1^{(2)}\otimes\mathbf{b}_1^{(2)*}. \tag{B10}$$

If $\mathbf{e}_0(\pi)=\mathbf{b}_1^{(1)}(\pi)$, then $\mathbf{e}_0(t)=\mathbf{b}_1^{(1)}(t)$ is a solution of Eq. (B7), and hence one simply obtains Eq. (B5a) as the type-B expansion for this essentially one-dimensional problem.

If $\mathbf{e}_0(\pi)=\mathbf{b}_1^{(2)}(\pi)$, then the ansatz that $\mathbf{e}_0(t)$ equals $\mathbf{b}_1^{(2)}(t)$ up to a phase leads to a solution. One verifies that the $\mathbf{P}-\mathbf{e}_0\otimes\mathbf{e}_0^*$ term in Eq. (46) vanishes, since the bracketed vector is orthogonal to $\mathbf{b}_1^{(1)}$, and so the problem reduces to calculating [cf. Eq. (45)]

$$(\mathbf{e}_0,(\mathbf{e}_0')_1)=4ip^{1/2}(\mathbf{b}_2,(\mathbf{M}-p)^{-1}\mathbf{b}_2)=4it^{1/2}(1-t)^{-1}$$

and solving Eq. (B7):

$$\mathbf{e}_0=\mathbf{b}_1^{(2)}\exp\Big(4iu^{-1}\int_\pi^t\frac{(t')^{1/2}}{1-t'}dt'+O(u^{-2})\Big).$$

Thus one obtains the expansion

$$\mathbf{h}=(2u)^{-1/2}[t+O(u^{-2})]^{-1/4}\Big[[\mathbf{b}_1^{(2)}+2iu^{-1}t^{1/2}$$

$$\times(1-t)^{-1}\mathbf{b}_2]\exp\Big(4iu^{-1}\int_\pi^t\frac{(t')^{1/2}}{1-t'}dt'+O(u^{-2})\Big)$$

$$+ O(u^{-2})\Big]\exp\Big\{-iu\int_\pi^t\Big[t' + u^{-2}\Big(\frac{5t'-1}{1-t'} + \frac{5}{16}(t')^{-2}\Big) + O(u^{-3})\Big]^{1/2}dt'\Big\}, \quad \text{(B11)}$$

whose relation to Eq. (B5b) is obvious. The case $\mathbf{e}_0(\pi) = \mathbf{b}_2(\pi)$ is of course similar.

These problems were so easily solvable only because the matrix (B1) is a direct sum of two nondegenerate matrices. The complications of the general case may be glimpsed by considering the initial condition

$$\mathbf{e}_0(\pi) = A\mathbf{b}_1^{(1)}(\pi) + B\mathbf{b}_1^{(2)}(\pi), \quad |A|^2 + |B|^2 = 1. \quad \text{(B12)}$$

If one sets

$$\mathbf{e}_0(t) = A(t)\mathbf{b}_1^{(1)}(t) + B(t)\mathbf{b}_1^{(2)}(t), \quad \text{(B13)}$$

then a calculation using Eqs. (B9)–(B10) and

$$\mathbf{P} - \mathbf{e}_0\otimes\mathbf{e}_0^* = |B|^2\mathbf{b}_1^{(1)}\otimes\mathbf{b}_1^{(1)*} + |A|^2\mathbf{b}_1^{(2)}\otimes\mathbf{b}_1^{(2)*} - AB^*\mathbf{b}_1^{(1)}\otimes\mathbf{b}_1^{(2)*} - A^*B\mathbf{b}_1^{(2)}\otimes\mathbf{b}_1^{(1)*}$$

yields the differential equations

$$A' = \tfrac{1}{2}iu^{-1}t^{-1/2}\frac{5t-1}{1-t}A(1-|A|^2), \quad \text{(B14a)}$$

$$B' = 2iu^{-1}t^{1/2}(1-t)^{-1}B(1+|B|^2) + \tfrac{1}{2}iu^{-1}t^{-1/2}B(1-|B|^2), \quad \text{(B14b)}$$

which we shall not attempt to solve.


[1]S.A. Fulling, J. Math. Phys. **16**, 875 (1975).
[2]S.A. Fulling (paper still in preparation).
[3]F.W.J. Olver, Proc. Cambridge Philos. Soc. **57**, 790 (1961); see also Ref. 1, Appendix A.
[4]N. Fröman, Arkiv Fys. **32**, 541 (1966).
[5]N. Fröman, Ann. Phys. (N.Y.) **61**, 451 (1970).
[6]J.A. Campbell, J. Comput. Phys. **10**, 308 (1972).
[7]B. Chakraborty, J. Math. Phys. **14**, 188 (1973).
[8]S.F. Feshchenko, N.I. Shkil', and L.D. Nikolenko, *Asymptotic Methods in the Theory of Linear Differential Equations* (American Elsevier, New York, 1967).
[9]Correspondingly, the type-C expansions for second-order coupled equations in Ref. 8, which apply only to the case of nondegenerate eigenvalues, must be closely related to the type-B expansions of the present paper, specialized to that case. See Appendix B of Ref. 1.
[10]J.D. Cole, *Perturbation Methods in Applied Mathematics* (Blaisdell, Waltham, Massachusetts, 1968), pp. 4–7, 79–80.
[11]L. Parker and S.A. Fulling, Phys. Rev. D **9**, 341 (1974); T.S. Bunch, S.M. Christensen, and S.A. Fulling, Phys. Rev. D **18**, 4435 (1978); S.A. Fulling, "Remarks on Positive Frequency and Hamiltonians in Expanding Universes," Gen. Rel. Grav. (in press).
[12]B.L. Hu, S.A. Fulling, and L. Parker, Phys. Rev. D **8**, 2377 (1973).
[13]Close examination shows that Eq. (27) of Ref. 1, for the integrand $\mathbf{P}[(\mathbf{Pa}_1)']$, contains all the terms of our Eq. (28) for $\mathbf{P}[(\mathbf{Pe}_1)']$ plus additional terms whose effects are taken into account by $p_2$ in the present approach.
[14]Reference 8, pp. 66–73, 85–90.
[15]See, e.g., S.G. Krein, *Linear Differential Equations in Banach Space* (American Mathematical Society, Providence, Rhode Island, 1971).
[16]KFA FORMAC, Version 2A (S.I.U., 1973).

# Asymptotic forms of radial wavefunctions and Jost functions for cutoff potentials

W. J. Romo[a)]

*Department of Physics, Carleton University, Ottawa, Ontario, K1S 5B6*
(Received 21 September 1978)

Rigorous asymptotic forms for the radial wavefunctions and the Jost functions are obtained for a broad class of cutoff potentials as the momentum $k$ tends to infinity in the complex $k$ plane. The asymptotic positions of the zeros of the Jost functions ($S$ matrix poles) are also determined. It is found that the solutions of $f_l(-k)=0$ for which $|k|$ is very large, form a finite sequence of families of the form $k_N^{(m)} \approx \alpha_m N - i\beta_m \log N$ with $N \geq N_0 \gg 1$ and $m = 1$ to max with max a finite integer greater than or equal to one which is determined by the potential parameters.

## 1. INTRODUCTION

As is well known[1] the zeros of the Jost function $f_l(-k)$ in the complex $k$ (momentum) plane give rise to poles of the $S$ matrix. Those zeros of $f_l(-k)$ that lie in the upper-half $k$ plane fall on the imaginary $k$ axis and are associated with bound states of the system which have binding energy $-\hbar^2 k^2/2m$. If the system can be described by a central potential that vanishes beyond a finite radius, then it is also known[1] that $f_l(-k)$ will have an infinite number of zeros in the lower-half $k$ plane. Only a finite number of zeros can lie along the negative imaginary axis. Such zeros correspond to virtual bound states of the system. In addition to the virtual bound state zeros $f_l(-k)$ will have an infinite number of zeros in the lower-half $k$ plane for which both Re$k$ and Im$k$ are nonzero. By convention one associates these zeros with resonances of the system.

We are currently involved in a study of the conditions under which matrix elements of the form $\langle\psi_1|P_a|\psi_2\rangle$, where $P_a$ is the projection operator

$$P_a\psi(\mathbf{r}) = \begin{cases}\psi(\mathbf{r}), & \text{if } |\mathbf{r}|\leqslant a,\\ 0, & \text{if } |\mathbf{r}|>a,\end{cases}$$

can be expanded in terms of the infinite discrete set of bound, virtual, and resonant states of a Hamiltonian $H$ that describes a system in which a spinless nonrelativistic particle moves in a central potential of finite range. In two previous papers, we first[2] established the validity of such an expansion for the $s$-wave channel of a surface delta-function potential $V(r)=\lambda\delta(r-R)$ for all $a\leqslant R$, and next,[3] as a first step in extending the study to a wider class of potentials and to all partial waves, we established the convergence conditions for a Mittag–Leffler expansion of matrix elements of the type

$$I_{12} = \langle\psi_1|P_a(E+i\epsilon-H)^{-1}P_a|\psi_2\rangle.$$

In principle the conditions under which $\langle\psi_1|P_a|\psi_2\rangle$ can be expanded in terms of bound, virtual, and resonant states then follows from a careful examination of the limit

$$\lim_{E\to\infty}(EI_{12}).$$

However, in order to complete the study, in addition to the results of Ref. 3 one needs more accurate asymptotic expressions for the positions of the large momentum zeros and the resonant state wave functions associated with them, than can be obtained from the literature, for example as in Ref. 1. Since the mathematics involved in deriving the needed asymptotic expressions is both complex and lengthy, and the subject matter is somewhat divorced from the main line of study outlined above, we shall confine the subject matter of this paper to the derivation of the needed asymptotic expressions, and shall complete the program in a separate paper.

In Sec. 2 we define the class of potentials that are to be considered and derive asymptotic equations for the radial wave functions and the Jost functions as $k\to\infty$. The application that we have described above requires that all asymptotic expansions be carried out at least to terms of order $k^{-1}$ relative to the leading term for functions of $k$ only, such as $f_l(\pm k)$, and to terms of order $k^{-2}$ relative to the leading term of the expansion for functions of $k$ and $r$, such as $\phi_l(k,r)$. In Sec. 3 the asymptotic expressions for $f_l(\pm k)$ are used to determine the positions of all large momentum zeros of the Jost functions. The results are discussed in Sec. 4.

## 2. DERIVATION OF ASYMPTOTIC FORMS

The regular radial wavefunctions, whose asymptotic forms are to be determined are solutions of the radial Schrödinger equation

$$\frac{d^2\phi_l(k,r)}{dr^2} + \left(k^2 - \frac{l(l+1)}{r^2} - V(r)\right)\phi_l(k,r) = 0, \quad (2.1)$$

where $k^2$ in appropriate units is the energy and $l$ is the angular momentum. The regular solution $\phi_l(k,r)$ also satisfies the boundary condition

$$\lim_{r\to 0}(2l+1)!!\,r^{-l-1}\phi_l(k,r) = 1. \quad (2.2)$$

It will be assumed that the potential $V(r)$ satisfies each of the following list of properties:

a)Supported in part by a research grant from the National Research Council of Canada.

 0022-2488/79/061210-15$01.00  

(a) $V(r) = 0$ for $r > R$, where $R$ is a fixed positive radius.

(b) Immediately below the cutoff radius, i.e., for $R(1-\delta) \leq r \leq R$ with $0 < \delta \ll 1$, it will be assumed that $V(r) = (1 - r/R)^{\sigma} P_0(1 - r/R)$, where $\sigma \geq 0$ and $P_0(t) \equiv \Sigma_{j=1}^{\infty} v_j t^j$ is a power series that converges for all $t$ in the closed interval $[0,\delta]$.

(c) Near the origin, $V(r) = r^{-\gamma} Q_0(r)$, where $\gamma < 2$ and $Q_0(r) \equiv \Sigma_{j=0}^{\infty} \tau_j r^j$ is a power series that converges for all $r$ in the closed interval $[0,\epsilon R]$ with $0 < \epsilon \ll 1$.

(d) $V(r)$ is piecewise continuous and bounded and all of its derivatives up to some order $L \geqslant \sigma + 2$ are also piecewise continuous and bounded in the open interval $(0,R)$.

(e) $d^L V(r)/dr^L$ has $n$ points of discontinuity in the open interval $(0,R)$. One defines $R_i$, with $\epsilon R < R_1 < R_2 < \cdots < R_n < R(1-\delta)$, to be such a point, and further defines $m_i$ to be the smallest integer for which $d^{m_i} V(r)/dr^{m_i}$ is discontinuous at $R_i$.

(f) Finite right and left hand limits of $d^L V(r)/dr^L$ exist at each of the points $R_i$.

Property (a) implies that

$$\phi_l(k,r) = \tfrac{1}{2} i k^{-l-1} [f_l(-k) w_l(kr) - (-1)^l f_l(k) w_l(-kr)] \tag{2.3}$$

for $r > R$, where $w_l(\pm kr)$ are Riccati–Hankel functions[4]

$$w_l(z) = -i\sqrt{\pi z/2}\, H^{(2)}_{l+1/2}(z) \tag{2.4}$$

and $f_l(\pm k)$ are the Jost functions, whose asymptotic forms will also be determined in this section. We have defined $\phi_l(k,r)$ and $f_l(\pm k)$ so that they agree with the definitions given in Ref. 1.

In order to determine the asymptotic expansion of $\phi_l(k,r)$, we first define a radius $\rho$ by

$$\rho \equiv \mu/\log|k|, \tag{2.5}$$

where $\mu$ is an arbitrary real positive number, and next we divide the line segment $[0,R]$ into $n+3$ subsegments $\Sigma_j$ $(j = -1,0,1,\dots,n+1)$ with $\Sigma_{-1} \equiv [0,\rho]$, $\Sigma_0 \equiv [\rho, R_1]$, $\Sigma_j \equiv [R_j, R_{j+1}]$ for $j = 1, n-1$, $\Sigma_n = [R_n, R - |k|^{-1}]$ and $\Sigma_{n+1} \equiv [R - |k|^{-1}, R]$. (The reason that $R - |k|^{-1}$ was chosen as the boundary between $\Sigma_n$ and $\Sigma_{n+1}$ will be discussed in Appendix A.) For $r$ in $\Sigma_j$ with $j \geqslant 0$, $\phi_l(k,r)$ can be written as a linear combination of irregular solutions of the radial Schrödinger equation and the asymptotic form of the irregular solutions can be determined by the techniques outlined in Sec. 4.3 of the monograph by Erdélyi.[5] These techniques can only be employed if $kr \to \infty$ as $k \to \infty$, so a different approach must be used when $r$ is in $\Sigma_{-1}$.

To determine the asymptotic form of $\phi_l(k,r)$ when $r \in \Sigma_{-1}$, we shall employ the Born series expansion:

$$\phi_l(k,r) = \sum_{j=0}^{\infty} \phi_l^{(j)}(k,r), \tag{2.6}$$

where

$$\phi_l^{(0)}(k,r) = k^{-l-1} u_l(k,r) \tag{2.7}$$

and

$$\phi_l^{(j)}(k,r) = \int_0^r dr'\, g_l(k,r,r') V(r') \phi_l^{(j-1)}(k,r') \tag{2.8}$$

for $j \geqslant 1$, with

$$g_l(k,r,r') = \frac{i(-1)^l}{2k} [w_l(kr) w_l(-kr') - w_l(-kr) w_l(kr')]. \tag{2.9}$$

In the above equation $u_l(kr)$ is the Riccati–Bessel function[4]

$$u_l(z) = \sqrt{\pi z/2}\, J_{l+1/2}(z). \tag{2.10}$$

The Born series is absolutely convergent for[1] every value of $r$ and $k$. However, we shall only be concerned with determining the asymptotic forms of $\phi_l(k,\rho)$ and $d\phi_l(k,r)/dr$ at the boundary point $\rho$, where it will be matched with a linear combination of the two irregular solutions associated with the segment $\Sigma_0$. As $k \to \infty$ in $S$, where $S$ consists of all points in the fourth quadrant of the $k$ plane for which $|k| \geqslant \kappa \gg R^{-1}$, $\rho$ will tend to zero. Consequently $V(r)$ can be represented by the leading terms of its small $r$ series expansions, given in (c), for all $r$ in $\Sigma_{-1}$. On the other hand $\rho$ is sufficiently large so that $|k|\rho \to \infty$ as $k \to \infty$ in $S$.

If one defines $\mathscr{I}_l^{(j)}(k,r)$ by

$$\mathscr{I}_l^{(j)}(k,r) = (2ik)^{-1} \int_0^r dr'\, w_l(kr') V(r') \phi_l^{(j-1)}(k,r') \quad \text{for } j \geqslant 1, \tag{2.11a}$$

$$\mathscr{I}_l^{(0)}(k,r) = (2ik^{l+1})^{-1}, \tag{2.11b}$$

then by Eqs. (8) and (9) and the fact[1] that $\phi_l^{(j-1)}(k,r)$ is a symmetric function of $k$, one has

$$\phi_l^{(j)}(k,r) = (-1)^l [w_l(-kr) \mathscr{I}_l^{(j)}(k,r) + w_l(kr) \mathscr{I}_l^{(j)}(-k,r)]. \tag{2.12}$$

For the class of potentials being considered

$$|V(r)| \leqslant C_1 r^{-\gamma} \text{ and } |V(r) - \tau_0 r^{-\gamma}| \leqslant C_2 r^{1-\gamma} \tag{2.13}$$

for all values of $r$. Using these bounds and the bounds on $|\phi_l^{(j)}(k,r)|$ and $|w_l(\pm kr)|$ given by Newton,[1] one can easily show that[6]

$$|\mathscr{I}_l^{(j)}(\pm k,\rho)| \leqslant |k|^{-l-1} [cf(|k|) + o(e^{\mp 2|\nu|\rho}/|k|\rho)] \times [cf(|k|)]^{j-1}/(j-1)! \tag{2.14a}$$

for all $j \geqslant 1$ as $k \to \infty$ in $S$, where $\nu = \operatorname{Im} k$ and

$$f(|k|) = \begin{cases} \log|k|/|k|, & \text{if } \gamma \leqslant 1, \\ |k|^{\gamma-2}, & \text{if } 1 < \gamma < 2. \end{cases} \tag{2.14b}$$

If one confines $k$ to a subdomain $D$ of $S$ consisting of all points in $S$ for which

$$-\operatorname{Im}(2kR) \leqslant \Lambda \log|k| \equiv \frac{2+\sigma}{1 - R_n/R} \log|k|, \tag{2.15}$$

then $|\nu|\rho$ will be finite and the term $o(e^{\mp 2|\nu|\rho}/|k|\rho)$ in Eq. (14a) can be ignored. Restricting our consideration at this time to $k$ in $D$ and employing Eqs. (11) and (14) gives us

$$\left|\sum_{j=m+1}^{\infty} \mathscr{I}_l^{(j)}(\pm k,\rho)\right| < O(|\mathscr{I}_l^{(0)}(k,\rho)|[f(|k|)]^{m+1}) \tag{2.16}$$

as $k\to\infty$ in $D$. If the integer $m$ in Eq. (16) is chosen to satisfy

$$(\gamma-1)(2-\gamma)^{-1} < m \leqslant (2-\gamma)^{-1} \quad \text{for } \gamma>1 \tag{2.17}$$

$$m=1 \quad \text{for } \gamma=1, \quad \text{or } m=0 \quad \text{for } \gamma<1,$$

then the sum in Eq. (2.16) will be smaller than the product of $|\mathscr{I}_l^{(0)}(\pm k,\rho)|$ times a factor of order $o(|k|^{-1})$. Since the only terms in the asymptotic expansions of $\Sigma_j \mathscr{I}_l^{(j)}(\pm k,\rho)$ that we must examine in detail are those whose magnitudes exceed $O(|\mathscr{I}_l^{(0)}||k|^{-1})$, it follows that we need only consider the terms for which $j\leqslant m$. To shed further light on these terms we first note that the kernel of Eq. (8) can be expressed as $(2ik)^{-1}$ times a function of the two variable $kr$ and $kr'$, namely

$$g_l(k;r,r') = h(kr,kr')/2ik, \tag{2.18a}$$

where

$$h(x,y) = (-1)^l[w_l(-x)w_l(y) - w_l(x)w_l(-y)]. \tag{2.18b}$$

Iterating the defining equation for $\phi_l^{(j)}(k,r)$ $j$ times, changing variables to $x_i = kr_i$ and employing Eq. (13) gives

$$\mathscr{I}_l^{(j)}(k,\rho) = (2ik^{l+1})^{-1}(2ik)^{j(\gamma-2)}[\mathscr{A}(l,j,k\rho) + O(k^{-1})] \tag{2.19a}$$

as $k\to\infty$ in $D$, where

$$\mathscr{A}(l,1,z) = -4\tau_0\int_0^z dx\, w_l(x)(2ix)^{-\gamma}u_l(x), \tag{2.19b}$$

and

$$\mathscr{A}(l,j,z) = 2i(2i\tau_0)^j\int_0^z dx_1\, w_l(x_1)(2ix_1)^{-\gamma} \times \prod_{i=1}^{j-1}\left(\int_0^{x_i} dx_{i+1}\, h(x_i,x_{i+1})(2ix_{i+1})^{-\gamma}\right)u_l(x_j), \tag{2.19c}$$

for $j>1$. Combining Eqs. (5), (12), and (14)–(19) gives

$$\phi_l(k,\rho) = (2ik^{l+1})^{-1}\Big\{(-1)^l w_l(-k\rho) \times\Big[\sum_{j=0}^{m}(2ik)^{j(\gamma-2)}\mathscr{A}(l,j,k\rho) + O(k^{-1})\Big] - w_l(k\rho)\Big[\sum_{j=0}^{m}(-2ik)^{j(\gamma-2)}\mathscr{A}(l,j,-k\rho) + O(k^{-1})\Big]\Big\}. \tag{2.20}$$

Comparing Eqs. (19) and (14a) shows us that

$$|\mathscr{A}(l,j,\pm k\rho)| \leqslant \text{const} \quad (\gamma\neq 1),$$

or

$$|\mathscr{A}(l,j,\pm k\rho)| \leqslant \text{const}\times\log|k| \quad (\gamma=1) \tag{2.21}$$

as $k\to\infty$ in $D$, for all $0<j<m$. The bounds given in Eq. (21) will be all of the information required if $j>1$ or $\gamma<1$. For the $j=1$ case the integral on the right hand side of Eq. (19b) can be evaluated and the result expressed in terms of functions with known asymptotic forms, then as $k\to\infty$ in $D$, with $\rho$ given by Eq. (5) one finds that

$$\mathscr{A}(l,1,\pm k\rho) = b(l,1) + \tau_0(\pm 2ik\rho)^{1-\gamma}/(1-\gamma) + O([k\rho]^{-\gamma}) \tag{2.22a}$$

if $1<\gamma<2$, and

$$\mathscr{A}(l,1,\pm k\rho) = \tau_0\{\log(\pm 2ik\rho) - \psi(l+1) + O([k\rho]^{-1}\} \tag{2.22b}$$

if $\gamma=1$, where

$$b(l,1) = (-1)^{l+1}\tau_0 \times \sum_{m,n=0}^{l}\frac{(l+m)!(l+n)!\Gamma(1-m-n-\gamma)}{(l-m)!(l-n)!n!m!}, \tag{2.22c}$$

and $\psi(l+1)$ is a digamma function.[4]

Applying $d/dr$ to Eq. (8) and following the same procedure that lead to Eq. (19) gives

$$\frac{d\phi_l(k,\rho)}{d\rho} = (2ik^{l+1})^{-1}\Big((-1)^l\frac{dw_l(-k\rho)}{d\rho}\Big[\sum_{j=0}^{m}(2ik)^{j(\gamma-2)} \times \mathscr{A}(l,j,k\rho) + O(k^{-1})\Big] - \frac{dw_l(k\rho)}{d\rho} \times\Big[\sum_{j=0}^{m}(-2ik)^{j(\gamma-2)}\mathscr{A}(l,j,-k\rho) + O(k^{-1})\Big]\Big). \tag{2.23}$$

Next we shall determine the asymptotic expansion of $\phi_l(k,r)$ for $r>\rho$. Using a theorem given by Erdélyi[5] one finds that in any region of space $b_1\leqslant r\leqslant b_2$ for which the potential $V(r)$ has continuous derivatives up to order $N$, the radial Schrödinger equation, Eq. (1), possesses a fundamental system of solutions $y_{-1}(k,r)$ and $y_{+1}(k,r)$ such that

$$y_\alpha(k,r) = Y(\alpha k,r)[1+O(k^{-N})], \tag{2.24a}$$

and

$$\frac{dy_\alpha(k,r)}{dr} = \frac{dY(\alpha k,r)}{dr}[1+O(k^{-N})], \tag{2.24b}$$

uniformly in $r$ and $\arg k$ as $k\to\infty$ in $S$ for $\alpha=\pm 1$, where

$$Y(k,r) = e^{ikr}\sum_{j=0}^{N}a_j(r)k^{-j}, \tag{2.24c}$$

$$a_0(r) = 1, \tag{2.24d}$$

$$\frac{da_j(r)}{dr} = (2i)^{-1}\Big(\mathscr{V}(r)a_{j-1}(r) - \frac{d^2a_{j-1}(r)}{dr^2}\Big) \quad \text{for } j\geqslant 1, \tag{2.24e}$$

$$\mathscr{V}(r) = l(l+1)r^{-2} + V(r). \tag{2.24f}$$

(For simplicity the $l$ dependence has been suppressed in these equations!) Clearly, $d^LV(r)/dr^L$ and all lower order derivatives will be continuous and bounded within $\Sigma_j$ ($j=1$, to $n$). Thus, pairs of solutions of the radial Schrödinger equation can be obtained by applying Eqs. (24a)–(24f) with $N = L$ to the segments $\Sigma_j$ with $j = 1$ to $n$. The proof given by Erdélyi assumes that $\mathscr{V}(r)$ and all its derivatives up to the $N$th order are $O(1)$ as $k\to\infty$, so the theorem as it stands cannot be applied to the segment $\Sigma_0$, since $\mathscr{V}^{(j)}(\rho) = O([\log|k|]^{j+2})$, nor to the segments $\Sigma_n$ and $\Sigma_{n+1}$, when the potential parameter $\sigma$ is not an integer. However, by a slight modification of Erdélyi's proof, asymptotic solutions similar to those given by Eqs. (24) can be constructed in segments $\Sigma_0$, $\Sigma_n$, and $\Sigma_{n+1}$, as is done in Appendices A for $\Sigma_0$ and B for $\Sigma_n$ and $\Sigma_{n+1}$.

Since the set of all derivatives of $V(r)$ up to order $L$ changes from one segment to the next, the set of coefficients $a_j(r)$ will also change, and thus the pair of solutions determined in one segment will not be the same as those determined in the next. On the other hand, a second-order differential equation only has two linearly independent solutions, so that each of the solutions associated with $\Sigma_{j-1}$ will be a linear combination of the pair associated with $\Sigma_j$. The expansion coefficients can be determined by matching the functions and their derivatives at $r = R_j$. Let $y_{\alpha j}(k,r)$, with $\alpha = \pm 1$, be the pair of solutions associated with the segment $\Sigma_j$, then

$$y_{\alpha j-1}(k,r) = \sum_{\beta=-1}^{1} \mathscr{M}^j_{\alpha\beta}(k) y_{\beta j}(k,r) \tag{2.25a}$$

with

$$\mathscr{M}^j_{\alpha\beta}(k) = W[y_{-\beta,j}(k,r), y_{\alpha,j-1}(k,r)] \times \{W[y_{-\beta,j}(k,r), y_{\beta,j}(k,r)]\}^{-1} \tag{2.25b}$$

for $\alpha$ and $\beta = \pm 1$, where $W[\phi(r),\psi(r)]$ is the Wronskian, defined by

$$W[\phi(r),\psi(r)] = \phi(r)\frac{d\psi(r)}{dr} - \frac{d\phi(r)}{dr}\psi(r).$$

Since the Wronskians are independent of the choice of $r$, one can choose $r = R_j$ to be the point of evaluation and use the asymptotic expansions of $y_{\alpha j}(k,r)$ and $dy_{\alpha j}(k,r)/dr$, which are valid for $r = R_j + \epsilon$ and those for $y_{\alpha j-1}(k,r)$ and $dy_{\alpha j-1}(k,r)/dr$ which are valid for $r = R_j - \epsilon$ as $\epsilon\to 0^+$ in both cases to determine the asymptotic expansions of the Wronskians. In Appendix C it is shown that if $d^j\mathscr{V}(r)/dr^j$ is continuous in some domain containing a point $r_0$ for all $j\leqslant N$, and if $d^{N+1}\mathscr{V}(r)/dr^{N+1}$ has a discontinuity at $r_0$, then solutions to the system of equations given by Eq. (24d)–(24f) can be constructed such that $a_j(r)$ and $da_j(r)/dr$ are continuous throughout the whole domain for $j < N + 2$. It necessarily follows from this choice, as is shown in the Appendix, that $da_{N+2}(r)/dr$ will be discontinuous at $r_0$. Consequently if one is given the solutions to Eqs. (24e) in the segment $\Sigma_{j-1}$, call them $a_i^{j-1}(r)$, he can construct solutions $a_i^j(r)$ for $\Sigma_j$ such that

$$a_i^j(R_j) = a_i^{j-1}(R_j) \quad \text{for } i\leqslant m_j + 1,$$

and

$$\left.\frac{da_i^j(r)}{dr}\right|_{R_j} = \left.\frac{da_i^{j-1}(r)}{dr}\right|_{R_j} \quad \text{for } i\leqslant m_j, \tag{2.26}$$

while

$$\left.\frac{da^j_{m_j+1}(r)}{dr}\right|_{R_j} \neq \left.\frac{da^{j-1}_{m_j+1}(r)}{dr}\right|_{R_j}.$$

We shall choose solutions $a_i^j(r)$ such that Eq. (26) is satisfied at each boundary point. For the three lowest coefficients one then has

$$a_1^j(r) = -(2i)^{-1}\int_r^\infty \mathscr{V}(r')dr' = -(2i)^{-1}\left[l(l+1)r^{-1} + \int_r^R dr'\,V(r')\right], \tag{2.27a}$$

$$a_2^j(r) = -\tfrac{1}{8}\left[\int_r^\infty ds\,\mathscr{V}(s)\right]^2 + \frac{\mathscr{V}(r)}{4}, \tag{2.27b}$$

$$a_3^j(r) = \frac{1}{48i}\left[\int_r^\infty ds\,\mathscr{V}(s)\right]^3 + \frac{i}{8}\left(\int_r^\infty ds\,\mathscr{V}^2(s) + \mathscr{V}(r) \times\int_r^\infty ds\,\mathscr{V}(s) + \frac{d\mathscr{V}(r)}{dr}\right), \tag{2.27c}$$

for $r$ in $\Sigma_j$ and $j = 0,1,\ldots,n-1$. With this choice one finds that

$$\begin{aligned}W[y_{-\alpha,j}(k,r),y_{\alpha,j-1}(k,r)] &= W[Y_j(-\alpha k,r),Y_{j-1}(\alpha k,r)]_{r=R_j} + O(k^{-L+1})\\ &= W[Y_j(-\alpha k,r),Y_j(\alpha k,r)]|_{r=R_j} + \sum_{P=m_j+1}^{L-1}\beta_p[\alpha k]^{-P} + O(k^{-L+1})\end{aligned} \tag{2.28a}$$

for $\alpha = \pm 1$, where

$$\beta_p = \sum_{q=m_j+2}^{p+1}(-1)^q a^j_{p+1-q}(r)\left(2i[a_q^j(r) - a_q^{j-1}(r)] + \frac{da_q^j(r)}{dr} - \frac{da_q^{j-1}(r)}{dr}\right) + (-1)^{p+1}\sum_{q=2}^{P}(-1)^q \times[a^j_{p+1-q}(r) - a^{j-1}_{p+1-q}(r)]\left.\frac{da^j_{q-1}(r)}{dr}\right|_{r=R_j}, \tag{2.28b}$$

where the last sum vanishes for $p<2$. One also finds that

$$W[y_{\alpha j}(k,r),y_{\alpha j-1}(k,r)] = \exp(2i\alpha kR_j)\left(\sum_{p=m_j+1}^{L-1}\gamma_p(\alpha k)^{-P} + O(k^{-L+1})\right) \tag{2.28c}$$

with

$$\gamma_p = \sum_{q=1}^{P}\left(a^j_{p-q}(r)\frac{da_q^{j-1}(r)}{dr} - a^{j-1}_{p-q}(r)\frac{da_q^j(r)}{dr}\right)_{r=R_j}, \tag{2.28d}$$

and therefore

$$\mathscr{M}^j_{\alpha\alpha}(k) = 1 + O(k^{-\mu_j - 2})$$

and

$$\mathscr{M}^j_{\alpha+\alpha}(k) = e^{2i\alpha kR_j}O(k^{-\mu_j-2}) \qquad (2.28e)$$

for $j = 1$ to $n - 1$, where $\mu_j = \min[m_j, L - 2]$. Applying Eq. (25a) repeatedly and then using Eqs. (2.28), (A21), and (A24) gives

$$y_{\alpha,o}(k,r) = \sum_{\beta=-1}^{+1} [\mathscr{M}^1(k)\mathscr{M}^2(k)\cdots\mathscr{M}^j(k)]_{\alpha\beta}\, y_{\beta,j}(k,r),$$

or with

$$[\mathscr{M}^1(k)\cdots\mathscr{M}^j(k)]_{\alpha\beta} = \Big\{1 + \sum_{p=1}^{j} O(k^{-\mu_p-2}) + \sum_{\substack{p,q=1\\(p>q)}}^{j} \times O(\exp[-2i\alpha k(R_p - R_q)]\, k^{-\mu_p-\mu_q-4})\Big\}\delta_{\alpha\beta}$$
$$+ \sum_{p=1}^{j} O[\exp(2i\alpha kR_{p_k})k^{-\mu_p-2}] \times \Big\{1 + \sum_{\substack{q,r=1\\(q>r)}}^{P-1} O(\exp[-2i\alpha k(R_q - R_r)] \times k^{\mu_q-\mu_r-4})\Big\}\delta_{\alpha,-\beta}, \qquad (2.29)$$

for $j = 1$ to $n + 1$. Since asymptotic expansions of the functions $y_{\alpha,j}(k,r)$ of the form given by Eqs. (24a)–(24f) with $N = L$ and $a_p(r) = a^j_p(r)$ are only valid for $r \in \Sigma_j$, Eq. (29) enables one to write an asymptotic expansion of $y_{\alpha,0}(k,r)$ that is valid in any segment $\Sigma_j$. The reason that this is useful is that one can express $\phi_l(k,r)$ as a linear combination of $y_{+1,0}(k,r)$ and $y_{-1,0}(k,r)$, since they are solutions of the same second order linear differential equation as $\phi_l(k,r)$. The functions will be matched at the point $r = \rho$, which lies at the boundary of $\Sigma_{-1}$ and $\Sigma_0$, where Eqs. (20) and (23) are valid and where $y_{\alpha,0}(k,r)$ and $dy_{\alpha,0}(k,r)/dr$ can be replaced by $Y_0(\alpha k,r)$ and $dY_0(\alpha k,r)/dr$, respectively.

Writing $\phi_l(k,r)$ as a linear combination of the $y_{\alpha,0}(k,r)$'s, i.e.,

$$\phi_l(k,r) = \sum_{\alpha=-1}^{+1} B_{\alpha,0}(k) y_{\alpha,0}(k,r), \qquad (2.30a)$$

and determining the expansion coefficients by the usual Wronskian techniques gives

$$B_{\alpha,0}(k) = W[y_{-\alpha,0}(k,\rho),\phi_l(k,\rho)] \times\{W[y_{-\alpha,0}(k,\rho),y_{\alpha,0}(k,\rho)]\}^{-1}. \qquad (2.30b)$$

Employing Eqs. (20), (23), (B7) and (27) one can calculate the asymptotic forms of the Wronskians on the right-hand side of Eq. (30b). After a considerable amount of algebra one finds that

$$B_{\alpha,0}(k) = \tfrac{1}{2}(i\alpha k)^{-l-1}\Big\{1 + (2i\alpha k)^{\gamma-2}[\mathscr{A}(l,1,\alpha k\rho) + \tau_0(2i\alpha k\rho)^{1-\gamma}/(\gamma-1)] + \sum_{j=2}^{m}(2i\alpha k)^{j(\gamma-2)}\mathscr{A}(l,j,\alpha k\rho) + O(k^{-1})\Big\}$$
$$\text{if } 1 < \gamma < 2 \qquad (2.31a)$$

$$B_{\alpha,0}(k) = \tfrac{1}{2}(i\alpha k)^{-l-1}[1 + \tau_0\log(2i\alpha k)/2i\alpha k + O(k^{-1})] \quad \text{if } \gamma = 1, \qquad (2.31b)$$

$$B_{\alpha,0}(k) = \tfrac{1}{2}(i\alpha k)^{-l-1}[1 + O(k^{-1})] \quad \text{if } \gamma < 1, \qquad (2.31c)$$

as $k\to\infty$ in $D$. Since $B_{\alpha,0}(k)$ is independent of the choice of $\rho$, a fact that follows immediately from the definition of $B_{\alpha,0}(k)$ given by Eq. (30b), the various $k\rho$ dependences on the right-hand side of Eq. (31a) must cancel. Furthermore, since this equation must hold for a wide range of values of $k$, the cancellations must occur independently for each power of $k^{\gamma-2}$. On the other hand, there may be cancellations between any one of these terms and those terms which are lumped into $O(k^{-1})$. One concludes that

$$\mathscr{A}(l,1,k\rho) = b(l,1) - \tau_0(2ik\rho)^{1-\gamma}/(\gamma-1) + o[(k\rho)^{1-\gamma}], \qquad (2.32a)$$

$$\mathscr{A}(l,j,k\rho) = b(l,j) + o[(k\rho)^{j(2-\gamma)-1}] \quad \text{for } 2\leqslant j\leqslant m, \qquad (2.32b)$$

where $b(l,j)$ with $j = 1$ to $m$ is a constant. We note that Eq. (32a) agrees with the explicit calculation of $\mathscr{A}(l,1,k\rho)$ given by Eq. (22a). Combining Eqs. (31) and (32) gives us

$$B_{\alpha,0}(k) = \tfrac{1}{2}(i\alpha k)^{-l-1}[1 + K_l(\alpha k) + O(k^{-1})], \qquad (2.33a)$$

with

$$K_l(k) = \begin{cases} \sum_{j=1}^{m}(2ik)^{j(\gamma-2)}b(l,j), & \text{for } 1<\gamma<2\\ \tau_0\log(2ik)/2ik, & \text{for } \gamma = 1,\\ 0, & \text{for } \gamma<1,\end{cases} \qquad (2.33b)$$

as $k\to\infty$ in $D$, where the integer $m$ that appears in the definition of $K_l$ in the case for which $1<\gamma<2$ was defined in Eq. (17).

The asymptotic form of $\phi_l(k,r)$ as $k\to\infty$ in $D$ with $r$ in $\Sigma_0$ is obtained by substituting Eqs. (33) and (B7a) into Eq. (30a). The asymptotic form of $d\phi_l(k,r)/dr$ is obtained by applying $d/dr$ to Eq. (30a) and substituting Eqs. (33) and (B7b) into the resulting equation.

To obtain asymptotic forms for $\phi_l(k,r)$ and $d\phi_l(k,r)/dr$ as $k\to\infty$ in $D$ and $r$ in $\Sigma_j$ with $j\geqslant 1$, one first defines coefficients $B_{\alpha,j}(k)$ by

$$\phi_l(k,r) = \sum_{\alpha=-1}^{1} B_{\alpha,j}(k) y_{\alpha,j}(k,r). \qquad (2.34)$$

From Eqs. (29), (30a), and (33) it can be seen that

$$B_{\alpha,j}(k) = \sum_{\beta=-1}^{+1} B_{\beta,0}(k)[\mathscr{M}^1(k)\times\cdots\times\mathscr{M}^j(k)]_{\beta\alpha}$$
$$= \tfrac{1}{2}(i\alpha k)^{-l-1}\Big\{1 + K_l(\alpha k) + O(k^{-1}) + \sum_{q=1}^{j} O[\exp(-2i\alpha kR_{q_k})k^{-\mu_q-2}]\Big\} \qquad (2.35)$$

as $k\to\infty$ in $D$. Combining this equation with the asymptotic forms of $y_{\alpha_j}(k,r)$ and $dy_{\alpha_j}(k,r)/dr$ previously derived yields the following result: For every $r$ in $\Sigma_j$ with $1 \leqslant j \leqslant n+1$ and for $p=0$ and 1

$$\frac{d^P\phi_l(k,r)}{dr^P} = \tfrac{1}{2}(-i\alpha k)^{-l-1}\left\{1 + K_l(\alpha k) + O(k^{-1}) + \sum_{q=1}^{j} O[\exp(-2i\alpha k R_{q_k})k^{-\mu_q-2}]\right\} \times \frac{d^P[\exp(i\alpha kr)\Sigma_{s=0}^{M_j} a_s^j(r)(\alpha k)^{-s}]}{dr^P} \times [1 + O(k^{-X_j})] \quad (2.36)$$

uniformly for $r$ in $\Sigma_j$ as $k \rightarrow \infty$ in $D$, where $M_j = X_j = L$ for $j=1$ to $n-1$, $M_n = L$, $M_{n+1} = [\sigma]+1$ and $X_n = X_{n+1} = \sigma + 2$. {By the zeroth derivative we mean the function itself and $[\sigma]$ = largest integer $\leqslant \sigma$.}

Next we shall determine the asymptotic forms of the two Jost functions as $k \rightarrow \infty$ in $D$. To carry out this program we shall use the integral representations[1] for $f_l(\pm k)$,

$$f_l(\pm k) = 1 + (\pm k)^l \int_0^R dr\, w_l(\pm kr)V(r)\phi_l(k,r). \quad (2.37)$$

If one defines $I_1^\pm(k,\rho)$ and $I_2^\pm(k,\rho)$ by

$$I_1^\pm(k,\rho) = \int_0^\rho dr\, w_l(\pm kr)V(r)\phi_l(k,r), \quad (2.38a)$$

$$I_2^\pm(k,\rho) = \int_\rho^R dr\, w_l(\pm kr)V(r)\phi_l(k,r), \quad (2.38b)$$

one sees that $f_l(\pm k)$ is proportional to $\Sigma_{j=1}^2 I_j^\pm(k,\rho)$. Furthermore, by examining Eqs. (6), (11a), and (38a) one finds that

$$I_1^\pm(k,\rho) = \pm 2ik \sum_{j=1}^\infty \mathscr{I}^{(j)}(\pm k,\rho).$$

It then follows from Eqs. (19a), (22), (32), and (33) that

$$I_1^\pm(k,\rho) = (\pm k)^{-l}[K_l(\pm k) + \mathscr{F}_l(\pm k,\rho) + O(k^{-1})], \quad (2.39a)$$

where $K_l$ was given by Eq. (33b) and

$$\mathscr{F}_l(\pm k,\rho) = \begin{cases} \tau_0 \rho^{1-\gamma}/2ik(1-\gamma), & \text{for } 1<\gamma<2,\\ \tau_0 \log \rho/2ik, & \text{for } \gamma = 1,\\ 0, & \text{for } \gamma < 1. \end{cases} \quad (2.39b)$$

To evaluate the integrals $I_2^\pm(k,\rho)$ one combines Eqs. (38b) and (34) to obtain

$$I_2^\pm(k,\rho) = \sum_{j=0}^{n} \sum_{\alpha=-1}^{+1} B_{\alpha j}(k)\mathscr{J}_{\alpha j}^\pm(k), \quad (2.40a)$$

with $B_{\alpha j}(k)$ as defined in Eq. (35) and

$$\mathscr{J}_{\alpha j}^\pm(k) = \int_{R_j}^{R_{j+1}} dr\, w_l(\pm kr)V(r)y_{\alpha j}(k,r), \quad \text{for } j=0 \text{ to } n-1, \quad (2.40b)$$

$$\mathscr{J}_{\alpha,n}^\pm(k) = \int_{R_n}^{R} dr\, w_l(\pm kr)V(r)y_{\alpha,n}(k,r). \quad (2.40c)$$

Using the asymptotic form of $y_{\alpha,j}(k,r)$ for $r$ in $\Sigma_j$ and the expansion of $w_l(\pm kr)$ given in Ref. 4 one finds that

$$\mathscr{J}_{\alpha,j}^\pm(k) = i^l \sum_{P=0}^{L} \sum_{q=0}^{l} (\mp 1)^q \alpha^P k^{-p-q} \int_{R_j}^{R_{j+1}} dr\, e^{i(\alpha \mp 1)kr} \times V(r)a_P^j(r)b_q(r) + \int_{R_j}^{R_{j+1}} dr\, w_l(\pm kr) \times V(r)Y_j(\alpha k,r)O(k^{-L}) \quad (2.41a)$$

for $j=1$ to $n-1$, where

$$b_q(r) = (l + \tfrac{1}{2},q)(-2ir)^{-q}. \quad (2.41b)$$

[The $j=0$ and $j=n$ cases are dealt with in Appendices B and A, respectively.] Using the bound on $|w_l(Z)|$ given in Ref. 1, along with Eq. (13) and the following bound on $|Y_j(\alpha k,r)|$ which holds as $k \rightarrow \infty$ in $D$

$$|Y_j(\alpha k,r)| \leqslant \text{const}\, e^{\alpha|v|r},$$

one sees that the final integral in Eq. (41a) is bound by a term of order $O[\exp(2|v|R_{j+1})|k|^{-L}]$ for $\mathscr{J}_{+1,j}^{-}(k)$, $O[\exp(-2|v|R_j)|k|^{-L}]$ for $\mathscr{J}_{-1,j}^{+}(k)$, and $O[|k|^{-L}]$ for the two remaining possibilities. The functions $V(r)$, $a_p(r)$, and $b_q(r)$ are $L$, $L-p+2$, and $\infty$ times continuously differentiable and bounded in $\Sigma_j$. Therefore, the function $\Phi_{pq}^j(r)$ defined by

$$\Phi_{pq}^j(r) = V(r)a_p^j(r)b_q(r) \quad (2.42a)$$

is $N_p$ times continuously differentiable and bounded in $\Sigma_j$ with

$$N_p = \begin{cases} L, & \text{if } p \leqslant 2,\\ L-p+2, & \text{if } p>2. \end{cases} \quad (2.42b)$$

Integrating the $pq$th term by parts $N_p$ times gives

$$\int_{R_j}^{R_{j+1}} dr\, e^{2i\alpha kr}V(r)a_p^j(r)b_q(r) = (2i\alpha k)^{-1} \sum_{s=0}^{N_p-1} (-2i\alpha k)^{-s} e^{2i\alpha kr} \frac{d^s\Phi_{pq}^j(r)}{dr^s}\bigg|_{R_j}^{R_{j+1}} + (-2i\alpha k)^{-N_p} \int_{R_j}^{R_{j+1}} dr\, e^{2i\alpha kr} \frac{d^{N_p}\Phi_{pq}^j(r)}{dr^{N_p}}. \quad (2.43)$$

Since the $N_p$th derivative of $\Phi_{pq}^j(r)$ is bounded, the last integral on the right-hand side of Eq. (43) is bounded by $C\exp(2|v|R_{j+1})|k|^{-N_p}$ for $\alpha = +1$ and $C\exp(-2|v|R_j) \times |k|^{-L}$ for $\alpha = -1$. Hence

$$\mathscr{J}_{+1,j}^{-}(k) = i^l(2ik)^{-1} \sum_{p=0}^{L} \sum_{q=0}^{l} k^{-p-q} \sum_{s=0}^{N_p-1} (-2ik)^{-s} \times e^{2ikr} \frac{d^s\Phi_{pq}^j(r)}{dr^s}\bigg|_{R_j}^{R_{j+1}} + O[k^{-L}\exp(2ikR_{j+1})]$$

(2.44a)

$$\mathscr{J}^{\pm}_{-1,j}(k) = j^l(-2ik)^{-1}\sum_{p=0}^{L}\sum_{q=0}^{l}(-k)^{-p-q}\sum_{s=0}^{N_p-1}(2ik)^{-s} \times e^{-2ikr}\frac{d^s\Phi^j_{pq}(r)}{dr^s}\Big|_{R_j}^{R_{j+1}} + O[k^{-L}\exp(-2ikR_j)]. \tag{2.44b}$$

For the two remaining integrals one easily obtains

$$\mathscr{J}^{+}_{+1,j}(k) = \mathscr{J}^{-}_{-1,j}(k) = i^l\int_{R_j}^{R_{j+1}} dr\, V(r)[1+O(k^{-1})] \tag{2.44c}$$

[Eq. (2.44c) is not intended to imply exact equality between $\mathscr{J}^{+}_{+1,j}$ and $\mathscr{J}^{-}_{-1,j}$. The two will differ in terms of order $k^{-1}$.] Equations (44a)–(44c) are valid for $j = 1$ to $n-1$ as $k\to\infty$ in $S$. The asymptotic expressions for $\mathscr{J}^{\pm}_{\alpha,0}(k)$ and $\mathscr{J}^{\pm}_{\alpha,n}(k)$ are given by Eqs. (B11a)–(B11d) and (A26a)–(A26d), respectively. Combining these equations with Eqs. (26), (29), (35), and (40a) yields

$$\begin{aligned} I_2^{\pm}(k,\rho) &= (\mp 2ik)^{-1}i^l\Big\{B_{\mp 1,n}(k)v_0\Gamma(\sigma+1)e^{\mp 2ikR}(\mp 2ikR)^{\sigma} \\ &- \sum_{j=1}^{n} B_{\mp 1,j-1}(k)\Delta_j\exp(\mp 2ikR_j)(\pm 2ik)^{-m_j} \\ &+ \sum_{j=1}^{n}[B_{\mp 1,j-1}(k) - B_{\mp 1,j}(k)]V_j^+\exp(\mp 2ikR_j) \\ &- B_{\mp 1,0}(k)e^{\mp 2ik\rho}\sum_{\substack{p=0,L\\ q=0,l}}\sum_{s=0}^{N_p-1}\left(\frac{i}{2}\right)^s\frac{d^s\Phi_{pq}(\rho)}{d\rho^s} \\ &\times[1+O(k^{-1})] + i^lB_{\pm 1,0}(k)\int_\rho^R dr\,V(r)\left[1+O\left(\frac{1}{k\rho}\right)\right] \\ &+ i^l\sum_{j=1}^{n}[B_{\pm 1,j}(k) - B_{\pm 1,0}(k)]\int_{R_j}^{R_{j+1}} dr\,V(r), \end{aligned} \tag{2.45}$$

where

$$\Delta_j \equiv \lim_{\epsilon\to 0+}\frac{d^{m_j}V(r)}{dr^{m_j}}\Big|_{R_j-\epsilon}^{R_j+\epsilon},\quad V_j^+ \equiv \lim_{\epsilon\to 0+}V(R_j+\epsilon),$$

and $R_{n+1}\equiv R$.

From Eqs. (35) and (28e) one sees that

$$\begin{aligned} B_{\alpha,j}(k) - B_{\alpha,j-1}(k) &= B_{\alpha,j-1}(k)[\mathscr{M}^j_{\alpha\alpha}(k) - 1] + B_{-\alpha,j-1}(k)\mathscr{M}^j_{-\alpha\alpha}(k) \\ &= B_{+1,0}(k)O(\exp[i(1-\alpha)kR_j]k^{-\mu_j-2}). \end{aligned} \tag{2.46}$$

If one now inserts Eq. (46) into Eq. (25), replaces the $\rho$ dependent functions by their small $\rho$ expansions, and observes from Eqs. (33a) and (35) that

$$B_{\alpha,j-1}(k)\exp(2i\alpha kR_j) = B_{\alpha,0}(k)\{\exp(2i\alpha kR_j) + \delta_{\alpha,+1}o[k^{-1}\exp(2ikR_j)] + \delta_{\alpha,-1}o(k^{-1})\},$$

he obtains

$$\begin{aligned} I_2^{\pm}(k,\rho) &= (\mp 2ik)^{-1}i^lB_{\mp 1,0}(k)\Big[v_0\Gamma(\sigma+1)e^{\mp 2ikR}(\mp 2ikR)^{-\sigma} \\ &- \sum_{j=1}^{n}\Delta_j\exp(\mp 2ikR_j)(\pm 2ik)^{-m_j}\Big][1+0(k^{-1})] \\ &+ i^lB_{\pm 1,0}(k)[C+\tau_0\psi(\rho)][1+O(1/k\rho)], \end{aligned} \tag{2.47a}$$

where

$$\psi(\rho) = \begin{cases} -\log\rho & \text{if } \gamma = 1, \\ \dfrac{\rho^{1-\gamma}}{\gamma-1} & \text{otherwise.}\end{cases} \tag{2.47b}$$

Finally, if one adds $I_1^{\pm}(k,\rho)$ to $I_2^{\pm}(k,\rho)$ using Eqs. (33b), (39), and (47b), he obtains the asymptotic form of the integral on the right-hand side of Eq. (37). Hence,

$$\begin{aligned} f_l(\pm k) &= 1 + (-1)^l[1 + K_l(\mp k) + O(k^{-1})] \\ &\Big[v_0R^2\Gamma(\sigma+1)e^{\mp 2ikR}(\mp 2ikR)^{-\sigma-2} \\ &- \sum_{j=1}^{n}\Delta_j\exp(\mp 2ikR_j)(\pm 2ik)^{-2-m_j}\Big] \\ &+ K_l(\pm k) + O(k^{-1}) \end{aligned} \tag{2.48}$$

as $k\to\infty$ in $D$.

Although the expression for the asymptotic form of $\phi_l(k,r)$ which we have derived only holds for the domain $D$, this is adequate for our purposes as outlined in Sec. 1, since, as will be shown in the next section, the domain $D$ contains all of the zeros of $f_l(-k)$ in the right half-plane that lie beyond a distance $\kappa \ll R^{-1}$ of the origin. However, to complete our analysis of the asymptotic form of $f_l(\pm k)$ and thereby to prove the above assertion about the location of the zeros, we must also determine the asymptotic form of $f_l(\pm k)$ as $k\to\infty$ in $S$ with $-\mathrm{Im}k < (2R)^{-1}\Lambda\log|k|$, that is in the complement of $D$ in $S$. Fortunately, most of the work needed has already been done. For example, the asymptotic forms derived for $\mathscr{J}^{\pm}_{\alpha,j}(k)$ with $\alpha = \pm 1$ and $j = 0$ to $n$ hold as $k\to\infty$ in $S$. In fact, only the functions $\mathscr{J}^{(0)}_l(-k,\rho)$ given by Eqs. (19), (22), and (32), and the asymptotic forms of the related functions $B_{-1,0}(k)$ and $I_1^-(k,\rho)$ were restricted to $k\to\infty$ in $D$.[7] However, rather than determining the asymptotic form of those functions in the complement of $D$, we simply employ the bound on the magnitude of $\mathscr{J}^{(0)}_l(k,\rho)$ given by Eq. (14), which holds for all of $S$, to obtain bounds on $|B_{-1,0}(k)|$ and $|I_1^-(k,\rho)|$ as $k\to\infty$ in $S$. From the resulting bounds one finds that the asymptotic expansions of $f_l(+k)$ given by Eq. (48) holds uniformly in arg$k$ as $k\to\infty$ in $S$, and that Eq. (48) also holds for $f_l(-k)$, uniformly in arg$k$ as $k\to\infty$ in $S$ if a term that is bounded by $[Cf(|k|) + o(e^{2|\eta|\rho}/|k|\rho)]$ is added to the right-hand side of the equation. But such a term is down in magnitude by at least a factor of $|k|^{-1}$ from the magnitude of the leading term of the expansion, namely $v_0R^2\Gamma(\sigma+1)e^{2ikR}(2ikR)^{-\sigma-2}$, when $-2\nu R > \Lambda\log|k|$, i.e., when $k$ lies in the space complementary to $D$ in $S$. Hence, the added term can be absorbed into terms already present on the right-hand side of Eq. (48) as long as $-2\nu R > A\log|k|$. Combining this result with the previous derivation for the subdomain $D$ reveals that Eq. (48) holds uniformly in $\arg k$ as $k \to \infty$ in $S$. Similar arguments show that Eq. (3.6) is valid uniformly in $\arg k$ and $r$ in $\Sigma_j$ with $j = 1$ to $n+1$ as $k \to \infty$ in $S$.

From Eq. (4.8) it follows that $k \to \infty$ with $k$ real

$$f_l(\pm k) = 1 + K_l(\pm k) + O(k^{-1}),$$

and since $S_l(k) = \exp[2i\delta_l(k)] = f_l(k)/f_l(-k)$ one concludes that

$$\delta_l(k) = (2i)^{-1}[K_l(k) - K_l(-k) + O(k^{-1})],$$

or employing Eq. (2.33b) that

$$\delta_l(k) = \begin{cases} (2i)^{-1} \sum_{l=1}^{m} b(l,j)[(2ik)^{j(\gamma-2)} \\ \quad - -(2ik)^{j(\gamma-2)}] + O(k^{-1}) & \text{if } \gamma > 1, \\ -\tau_0 \log/2k + O(k^{-1}), & \text{if } \gamma = 1, \\ \quad O(k^{-1}), & \text{if } \gamma < 1. \end{cases}$$

One thus sees that the high energy phase shifts are entirely determined up to terms of order $k^{-1}$ by the factor $\gamma$, that is by the behavior of the potential near $r = 0$.

## 3. DISTRIBUTION OF ZEROS OF $f_l(-k)$

In this section we shall determine the distribution of zeros of the Jost function $f_l(-k)$ in the domain of the complex $k$ plane for which $|k| \geqslant \kappa \gg R^{-1}$. To simplify the derivation we shall change variables from $k$ to $z$ with

$$z = 2kR. \tag{3.1}$$

In terms of $z$ Eq. (2.48) can be rewritten as

$$f_l(\mp k) \equiv F_0(\pm z) = \sum_{j=0}^{n+1} A_j e^{\pm i\beta_j z}(\pm z)^{-\alpha_j}[1 + o(1)], \tag{3.2}$$

where

$$A_0 = 1, \quad \alpha_0 = \beta_0 = 0;$$
$$A_{n+1} = (-i)^{\sigma+2}(-1)^l v_0 R^2 \Gamma(\sigma+1),$$
$$\alpha_{n+1} = \sigma + 2, \quad \beta_{n+1} = 1;$$
$$A_j = (-1)^{l+1}\Delta_j(iR)^{\alpha_j}, \quad \alpha_j = 2 + m_j, \quad \beta_j = R_j/R \quad \text{for } 1 \leqslant j \leqslant n.$$

Property (c) of the potential given in Sec. 2 implies that the $\beta_j$'s are ordered as follows,

$$0 = \beta_0 < \beta_1 < \cdots < \beta_n < \beta_{n+1} = 1. \tag{3.3}$$

Equation (3.2) holds for all $z$ in the fourth quadrant of the complex $z$ plane for which $|z| \gg 1$. Thus, the solutions of $F_0(z) = 0$ yields the positions of those zeros of the Jost function $f_l(-k)$ which lie in the fourth quadrant far from the origin, while the solution of the equation $F_0(-z) = 0$ yields the positions of the zeros of $f_l(-k)$ that lie in the second quadrant, again far from the origin. The positions of the corresponding zeros in the first and third quadrant can then be obtained by the well known reflection symmetry of the zeros,[1] i.e., if $k_N$ is a zero of $f_l(z)$, then $-k_N^*$ is also a zero.

An examination of $F_0(-z)$ reveals that the $j = 0$ term of the sum is identically equal to one, while other terms are $o(1)$. Hence, $F_0(-z)$ has no zeros in the fourth quadrant of the $z$ plane for which $z \to \infty$. This result is simply a confirmation of the fact that $f_l(-k)$ only has a finite number of zeros in the upper half $k$ plane and each of these zeros lies along the positive imaginary axis within a finite distance of the origin.[1]

Next we wish to find the zeros of $F_0(z)$ that lie in that portion of the fourth quadrant of the $z$ plane for which the asymptotic formula applies, that is, for $|z| \geqslant \kappa R \gg 1$. Let us write $z$ in polar form, i.e.,

$$z = xe^{-i\theta}. \tag{3.4}$$

If $\sin\theta > 0$ is fixed and $x \to \infty$, then the $n+1$th term of the sum on $j$ will dominate all others and $|F_0(z)| \to \infty$. Thus, we see that $F_0(z)$ cannot vanish in the limit $x \to \infty$ unless $\theta \to 0$ simultaneously. To simplify the discussion of this region of the $z$ plane we shall introduce a new parameter $\lambda$ given by

$$\lambda = x\sin\theta/\log x,$$

in terms of which

$$z = x\exp[-i\sin^{-1}(\lambda\log x/x)].$$

For any fixed value of $\lambda$ the argument of the arcsin tends to zero as $x \to \infty$ and one gets

$$z = x - i\lambda\log x. \tag{3.5}$$

In terms of $x$ and $\lambda$ the magnitude of the $j$th term of Eq. (3.2) is given by

$$|A_j e^{iz\beta_j} z^{-\alpha_j}| = |A_j| x^{\lambda\beta_j - \alpha_j}. \tag{3.6}$$

When $\lambda < 2$ the exponent of $x$ will be negative for all $j \geqslant 1$, so that $F_0(z) \to 1$ as $x \to \infty$. Therefore, $F_0(z)$ can have no zeros in this region of the $z$ plane. The smallest possible value of $\lambda$ for which the conditions for a root are possible corresponds to the smallest value of $\lambda$ for which one of the terms in the sum with $j > 0$ attains unit magnitude. This first occurs when $\lambda = \lambda_1$ with

$$\lambda_1 = \lim_{j(j>0)} \alpha_j/\beta_j \equiv \alpha_{N(1)}/\beta_{N(1)}, \tag{3.7}$$

where $N(1)$ is the value of $j$ that yields the minimum. For simplicity, it shall be assumed that $N(1)$ is unique. (With a slight adjustment of the potential parameters $\alpha_j$ and $\beta_j$ a degeneracy can be removed.) From Eqs. (6) and (7) it is seen that the magnitude of the $N(1)$th term is $|A_{N(1)}|$, which is $O(1)$, while the magnitudes of all other terms with $j > 0$ are $o(1)$ at $\lambda = \lambda_1$. Thus, the condition for a zero becomes

$$e^{iz\beta_{N(1)}} = A_{N(1)}^{-1} z^{\alpha_{N(1)}}[1 + o(1)], \tag{3.8}$$

as $x \to \infty$ and $\lambda$ is in the neighborhood of $\lambda_1$. A straightforward calculation then gives

$$z_M^1 = 2M\pi\beta_{N(1)}^{-1} - i\alpha_{N(1)}\beta_{N(1)}^{-1}\log M + O(1), \quad \text{or}$$

$$= 2M\pi\beta_{N(1)}^{-1} - i\lambda_1\log M + O(1) \tag{3.9}$$

for all integers $M$ for which $2M\pi\beta_{N(1)}^{-1} \geqslant \kappa R \gg 1$. Equation (3.9) is seen to be a family of zeros. The superscript 1 labels $z_M^1$ to be a member of the first family.

Since the magnitude of the $N(1)$th term tends to infinity with $x$ for $\lambda > \lambda_1$, it is convenient to consider a new function $F_1(z)$ defined by

$$F_1(z) = A_{N(1)}^{-1}\exp(-iz\beta_{N(1)})z^{\alpha_{N(1)}}F_0(z)$$

$$= \sum_{j=0}^{N+1}(A_j/A_{N(1)})\exp[iz(\beta_j-\beta_{N(1)})]$$

$$\times z^{\alpha_j-\alpha_{N(1)}}[1+o(1)], \qquad (3.10)$$

rather than $F_0(z)$ when $\lambda > \lambda_1$. $F_1(z)$ has the same zeros as $F_0(z)$ and it is better behaved in this region. The magnitude of the $j$th term of the sum on the right-hand side of Eq. (3.10) is

$$|(A_j/A_{N(1)})\exp[iz(\beta_j-\beta_{N(1)})]z^{\alpha_{N(1)}-\alpha_j}|$$

$$\equiv M(j,1;\lambda,x) = |A_j/A_{N(1)}|x^{P(j,1,\lambda)} \qquad (3.11)$$

where

$$P(j,1,\lambda) = \lambda(\beta_j-\beta_{N(1)}) - \alpha_j + \alpha_{N(1)}$$

$$= \left(\lambda - \frac{\alpha_j-\alpha_{N(1)}}{\beta_j-\beta_{N(1)}}\right)(\beta_j-\beta_{N(1)}). \qquad (3.12)$$

One first notes that $P(j,1,\lambda)$ is a decreasing function of $\lambda$ if $j<N(1)$, an increasing function of $\lambda$ if $j>N(1)$ and that

$$P(N(1),1,\lambda) = 0 \qquad (3.13)$$

for all $\lambda$.

Let us examine the cases in which $j \neq N(1)$ in more detail. For $j=0$ one has

$$P(0,1,\lambda) = -\lambda\beta_{N(1)} + \alpha_{N(1)} = (\lambda_1-\lambda)\beta_{N(1)}$$

so that

(i) $P(0,1,\lambda_1)=0$ and $P(0,1,\lambda)<0$ for $\lambda>\lambda_1$.

For $0<j<N(1)$ and $\lambda \geqslant \lambda_1$, one has

(ii) $P(j,1,\lambda) \leqslant P(j,1,\lambda_1) = \left(\frac{\alpha_{N(1)}}{\beta_{N(1)}} - \frac{\alpha_j}{\beta_j}\right)\beta_j \leqslant 0,$

where the final result follows from Eq. (7).

To facilitate the discussion of the $j>N(1)$ case one defines two new quantities $\lambda_2$ and $\bar{\lambda}_2$,

$$\lambda_2 = \min_{j\,[j>N(1)]}\frac{\alpha_j-\alpha_{N(1)}}{\beta_j-\beta_{N(1)}} \equiv \frac{\alpha_{N(2)}-\alpha_{N(1)}}{\beta_{N(2)}-\beta_{N(1)}}, \qquad (3.14a)$$

where $N(2)$ is the value of $j$ (assumed to be unique) for which the minimum is achieved, and

$$\bar{\lambda}_2 = \min_{j\,[j>N(1),\,j\neq N(2)]}\frac{\alpha_j-\alpha_{N(1)}}{\beta_j-\beta_{N(1)}}. \qquad (3.14b)$$

By the uniqueness assumption one immediately sees that

$$\lambda_2 < \bar{\lambda}_2. \qquad (3.15)$$

Furthermore, by Eqs. (7) and (14)

$$\lambda_2 - \lambda_1 = \frac{\alpha_{N(2)}-\alpha_{N(1)}}{\beta_{N(2)}-\beta_{N(1)}} - \frac{\alpha_{N(1)}}{\beta_{N(1)}}$$

$$= \left(\frac{\alpha_{N(2)}}{\beta_{N(2)}} - \frac{\alpha_{N(1)}}{\beta_{N(1)}}\right)\frac{\beta_{N(2)}}{\beta_{N(2)}-\beta_{N(1)}} > 0. \qquad (3.16)$$

For $j=N(2)$ one has by Eqs. (12) and (14a), $P[N(2),1,\lambda] = (\lambda-\lambda_2)(\beta_{N(2)}-\beta_{N(1)})$, so that

(iii) $P(N(2),1,\lambda)>0\,(<0)$ for $\lambda>\lambda_2\,(<\lambda_2)$ and $P(N(2),1,\lambda_2)=0$.

Finally, for $j>N(1)$ but $j\neq N(2)$ and $\lambda<\bar{\lambda}_2$,

$$P(j,1,\lambda) = \left(\lambda - \frac{\alpha_j-\alpha_{N(1)}}{\beta_j-\beta_{N(1)}}\right)(\beta_j-\beta_{N(1)})$$

$$< \left(\bar{\lambda}_2 - \frac{\alpha_j-\alpha_{N(1)}}{\beta_j-\beta_{N(1)}}\right)(\beta_j-\beta_{N(1)}).$$

Hence, from Eq. (14b) one has

(iv) $P(j,1,\lambda)<0$, for every $\lambda<\bar{\lambda}_2$ and every $j>N(1)$ except $j=N(2)$.

Statements (i)–(iv) imply that all terms in the sum on the right-hand side of Eq. (10) are $o(1)$ in the interval $\lambda_1<\lambda<\bar{\lambda}_2$ except the $j=N(2)$ term and the $j=N(1)$ term, which is identically equal to one. Thus the zeros of $F_1(z)$ that lie in this interval are given by

$$\exp[iz(\beta_{N(2)}-\beta_{N(1)})]$$

$$= (A_{N(1)}/A_{N(2)})z^{\alpha_{N(2)}-\alpha_{N(1)}}[1+o(1)]. \qquad (3.17)$$

Rather than determining the family of zeros that result from Eq. (17) at this time, we shall pass on to the general case. In analogy to Eqs. (14a) and (14b) one defines a sequence of $\lambda_m$'s and $\bar{\lambda}_m$'s by

$$\lambda_m = \min_{j\,[j>N(m-1)]}\left(\frac{\alpha_j-\alpha_{N(m-1)}}{\beta_j-\beta_{N(m-1)}}\right) \equiv \frac{\alpha_{N(m)}-\alpha_{N(m-1)}}{\beta_{N(m)}-\beta_{N(m-1)}}, \qquad (3.18a)$$

where $N(m)$ is the $j$ (assumed unique) for which the minimum exists, and

$$\bar{\lambda}_m = \min_{j\,[j>N(m-1),\,j\neq N(m)]}\left(\frac{\alpha_j-\alpha_{N(m-1)}}{\beta_j-\beta_{N(m-1)}}\right), \qquad (3.18b)$$

with $m\geqslant 2$ in both definitions. For $m=2$ one regains Eqs. (14a) and (14b). From Eq. (18b) it follows that $\bar{\lambda}_m>\lambda_m$, while from Eq. (18a) one sees that

$$(\alpha_{N(m+1)}-\alpha_{N(m-1)})/(\beta_{N(m+1)}-\beta_{N(m-1)})$$

$$> (\alpha_{N(m)}-\alpha_{N(m-1)})/(\beta_{N(m)}-\beta_{N(m-1)}) = \lambda_m.$$

Multiplying both sides of this inequality by the positive factor $f=(\beta_{N(m+1)}-\beta_{N(m-1)})(\beta_{N(m)}-\beta_{N(m-1)})$, then subtracting $(\alpha_{N(m)}-\alpha_{N(m-1)})(\beta_{N(m)}-\beta_{N(m-1)})$ from each side of the resulting inequality, and finally dividing the result of the previous operation by the factor $f$ yields

$$\lambda_{m+1} = \frac{\alpha_{N(m+1)}-\alpha_{N(m)}}{\beta_{N(m+1)}-\beta_{N(m)}} > \frac{\alpha_{N(m)}-\alpha_{N(m-1)}}{\beta_{N(m)}-\beta_{N(m-1)}} = \lambda_m, \qquad (3.19)$$

which along with Eq. (16) implies that $\lambda_{m+1}>\lambda_m$ for any $m\geqslant 1$.

One next defines

$$F_m(z) = A^{-1}_{N(m)} e^{-iz\beta_{N(m)}} z^{\alpha_{N(m)}} F_0(z)$$
$$= \sum_{j=0}^{n+1} \left(\frac{A_j}{A_{N(m)}}\right) \exp[iz(\beta_j - \beta_{N(m)})]$$
$$\times z^{\alpha_{N(m)} - \alpha_j}[1 + o(1)]. \quad (3.20)$$

The magnitude of the $j$th term of the sum is given by

$$M(j,m;\lambda,x) = |A_j/A_{N(m)}| x^{P(j,m,\lambda)} \quad (3.21a)$$

with

$$P(j,m,\lambda) = \lambda(\beta_j - \beta_{N(m)}) - \alpha_j + \alpha_{N(m)}. \quad (3.21b)$$

We are now in a position to prove a theorem that will play a key role in the analysis.

*Theorem*: The following statements hold for every $m \geqslant 1$:

(i) $P(N(m-1),m,\lambda_m) = 0$ and $P(N(m-1),m,\lambda) < 0$ for $\lambda > \lambda_m$.

(ii) For every $j < N(m), j \neq N(m-1)$, $P(j,m,\lambda) < o$ for $\lambda \geqslant \lambda_m$.

(iii) $P(N(m+1),m,\lambda) > 0$ for $\lambda > \lambda_{m+1}$, $P(N(m+1), m,\lambda_{m+1}) = 0$ and $P(N(m+1),m,\lambda) < 0$ for $\lambda < \lambda_{m+1}$.

(iv) For every $j > N(m)$, but $j \neq N(m+1)$, $P(j,m,\lambda) < 0$ for $\lambda < \overline{\lambda}_{m+1}$.

*Proof*: Clearly all four statements hold for $m = 1$ if one simply defines $N(0) \equiv 0$. It will be assumed that they also hold for some arbitrary $m$ greater than unity and the validity of (i)–(iv) for the $m + 1$ case will be examined.

From the definition of $P(j,m+1,\lambda)$, Eq. (21b), it can be seen that

$$P(j,m+1,\lambda) = P(j,m,\lambda) - P(N(m+1),m,\lambda), \quad (3.22)$$

for every $j$ and $\lambda$. Setting $j = N(m)$ yields

$$P(N(m),m+1,\lambda)$$
$$= P(N(m),m,\lambda) - P(N(m+1),m,\lambda).$$

The first term on the right-hand side of this equation is identically equal to zero, as is easily seen by Eq. (21b), while by statement (ii) of the induction hypothesis the second term vanishes if $\lambda = \lambda_{m+1}$ and it is negative if $\lambda > \lambda_{m+1}$. This establishes the validity of (i) for the $m + 1$ case.

Next one considers all values of $j$ less than $N(m+1)$ that differ from $N(m)$. From Eqs. (3) and (21b) it follows that $P(j,m+1,\lambda)$ is a decreasing function of $\lambda$. Hence for $\lambda \geqslant \lambda_{m+1}$ and $j$ as previously defined

$$P(j,m+1,\lambda) \leqslant P(j,m+1,\lambda_{m+1})$$
$$= P(j,m,\lambda_{m+1}) - P(N(m+1),m,\lambda_{m+1}).$$

The second term on the right-hand side of the equal sign vanishes by statement (iii) of the induction hypothesis and the first term is always negative, which follows from statement (i) of the induction hypothesis if $j = N(m-1)$, from statement (ii) if $j < N(m)$ but $j \neq N(m-1)$ and from (iv) if $N(m) < j < N(m+1)$. Thus, statement (ii) is also valid for the $m + 1$ case.

To test statement (iii) we note that by Eqs. (18a) and (21b)

$$P(N(m+2),m+1,\lambda)$$
$$= \lambda(\beta_{N(m+2)} - \beta_{N(m+1)}) - \alpha_{N(m+2)} + \alpha_{N(m+1)}$$
$$= (\lambda - \lambda_{m+2})(\beta_{N(m+2)} - \beta_{N(m+1)}).$$

The second term in brackets in the last expression is positive by Eq. (3). Therefore, $P(N(m+2),m+1,\lambda)$ is positive (negative) if $\lambda$ is greater (lesser) than $\lambda_{m+2}$ and $P(N(m+2), m+1,\lambda_{m+2}) = 0$. This establishes the validity of (iii) for the $m + 1$ case.

For $j > N(m+1)$ but $j \neq N(m+2)$ one has

$$P(j,m+1,\lambda) = [\lambda - (\alpha_j - \alpha_{N(m+1)})$$
$$\div (\beta_j - \beta_{N(m+1)})](\beta_j - \beta_{N(m+1)}).$$

Since $\beta_j - \beta_{N(m+1)}$ is positive and $(\alpha_j - \alpha_{N(m+1)})/(\beta_j - \beta_{N(m+1)}) \geqslant \overline{\lambda}_{m+2}$ for the values of $j$ being considered, one sees that $P(j,m+1,\lambda) < 0$ for $\lambda < \overline{\lambda}_{m+1}$. Hence, (iv) is valid for the $m + 1$ case. This completes the induction proof.

As an immediate consequence of the theorem it follows that in the range of $\lambda$ given by $\lambda_{m-1} < \lambda < \overline{\lambda}_m$, the $j$th term of $F_{m-1}(z)$ is $o(1)$ for every $j$ except $j = N(m-1)$ and $j = N(m)$. Thus, the condition for a zero of $F_m(z)$, and hence of $F_0(z)$, for which $\lambda$ lies between $\lambda_{m-1}$ and $\overline{\lambda}_m$ can be expressed as

$$\{1 + (A_{N(m)}/A_{N(m-1)}) \exp[iz(\beta_{N(m)} - \beta_{N(m-1)})]$$
$$\times z^{\alpha_{N(m-1)} - \alpha_{N(m)}}\}[1 + o(1)] = 0,$$

so that

$$z^m_M = 2M\pi/(\beta_{N(m)} - \beta_{N(m-1)}) - i\lambda_m \log M + O(1) \quad (3.23)$$

for all integers $M$ for which $2M\pi/(\beta_{N(m)} - \beta_{N(m-1)}) \geqslant \kappa R \gg 1$. Each value of $m$ corresponds to a given family of zeros of $F_0(z)$. A smooth curve can be drawn through each of these families. Equations (16) and (19) reveal that these lines move further and further away from the real $z$ axis as $m$ increases. Each time $m$ is increased at least one value of $j$ is removed from the set of $j$ values over which the minimum in Eq. (18a) is taken. Thus, the process must terminate after a finite number of steps $m = \max$. Clearly $N(\max) = n + 1$ and $\max \leqslant n + 1$. The case for which $\max = 1$, so that a single family of zeros results, corresponds to the class of potentials considered by Humblet.[8]

Re-expressing our results in terms of the wavenumber $k$ gives us all of the fourth quadrant zeros of $f_l(-k)$ for which $|k| \to \infty$,

$$k^m_M = M\pi/(R_{N(m)} - R_{N(m-1)}) - i\lambda_m \log M/2R + O(1) \quad (3.24)$$

with $m = 1$ to max, $M \to \infty$ and where $R_0 = 0$. The third quadrant zeros are obtained by reflecting these zeros across the imaginary axis as previously mentioned.

Finally, if we cast Eq. (2.15) into the notation of this section, we see that

$$\Lambda = \frac{\alpha_{n+1}}{1 - \beta_n} > \frac{\alpha_{n+1} - \alpha_{N(\max - 1)}}{1 - \beta_{N(\max - 1)}} = \lambda_{\max}.$$

Thus, all of the zeros of $f_l(-k)$ that can be found in the fourth quadrant of the $k$ plane with $|k|\geqslant\kappa$ will lie in the subdomain $D$ as asserted in the previous section.

## 4. COMMENTS AND CONCLUSIONS

The results of the previous section are somewhat surprising. The article and book by Newton,[1] which are standard references on the subject, report a single family of zeros of the form

$$k_n = n\pi/R - i[(\sigma+2)/2R]\log|n| + O(1) \quad \text{as} \quad n\to\infty \tag{4.1}$$

for all potentials that satisfy properties (a) and (b) of Sec. 2 and have finite first and second absolute moments, i.e.,

$$\int_0^R dr\, r^n |V(r)| < \infty \quad \text{for} \quad n = 1,2.$$

This is a much broader class of potentials than the one we have considered; in fact, it includes the class considered in this paper. Newton cites two references for his result, a paper by Humblet[8] published in 1952, and one by Regge[9] in 1958. Both Humblet and Regge obtained Eq. (4.1), but they are considering somewhat different classes of potentials than Newton. Expressed in our notation, Humblet was considering potentials that satisfy properties (a) through (f) with $\gamma\leqslant 0$, $n = 0$ and $\sigma$ an integer, i.e., the potential was assumed to be finite everywhere and to have continuous derivatives up to order $L\geqslant\sigma+2$ for the half closed interval $[0,R)$. Clearly this corresponds to a subclass of the class we have considered and it is seen that Humblet's result agrees with ours for that subclass. Regge was considering a class of potentials that satisfy (a) and (b) and have finite zeroth absolute moments, i.e., $\int_0^R dr\,|V(r)| < \infty$. This class of potentials contains all of those that we considered for which $\gamma < 1$, where $\gamma$ is the potential parameter defined under property (c). Since our prediction of the possibility of a number of families of zeros of $f_l(-k)$ did not depend on the value of $\gamma$, Regge's result is at variance with ours. The source of the disagreement can be traced to an approximation that Regge makes in his derivation, but one which, it should be stated, he clearly indicates he is making. He derives the asymptotic form of $f_l(-k)$ for $k$ belonging to the complement of $D$ in $S$, where it is approximately given by $V_0 R^{\,2}\Gamma(\sigma+1)e^{2ikR}(2ikR)^{-\sigma-2}$. Although he acknowledges that this form is not valid near the real axis (in the domain $D$), he takes as an approximation for $f_l(-k)$ in $D$ a sum of the above term plus the zeroth order Born term $[f_l^{(0)}(-k) = 1]$. Equation (4.1) follows immediately. Finally, returning to Ref. 1, one finds that the same approximation employed by Regge is employed, but without the clear warning given by Regge.

The class of potentials can be slightly extended without requiring major revisions of the derivation. For example one could easily accomodate a finite number of delta function terms in $V(r)$ between 0 and $R$ of the form $\xi\delta(r-b)$. This would result in the inclusion in the sum over $j$ on the right-hand side of Eq. (2.48) a term with $\Delta_j = -\xi, R_j = b$, and $m_j = -1$. One could also relax the constraints imposed by properties (d) and (f) to allow for potentials for which

$$\frac{d^{m_j}V(r)}{dr^{m_j}} \approx \begin{cases} a_j(R_j - r)^{b_j} & \text{for } r<R_j,\\ c_j(r-R_j)^{d_j} & \text{for } r>R_j \end{cases}$$

for $r$ in the neighborhood of $R_j$, where $a_j, b_j, c_j$, and $d_j$ are constants and at least one of the powers $b_j$ and $d_j$ is less than zero but greater than minus one for one of the $j$. The only modification this would entail is that in deriving the asymptotic forms of $y_{\alpha j-1}(k,r)$, $y_{\alpha j}(k,r)$, $\mathscr{F}^{\pm}_{\alpha j-1}(k)$, and $\mathscr{F}^{\pm}_{\alpha j}(k)$, one would have to use techniques similar to those given in Appendix A. The asymptotic forms of the two Jost functions would be unchanged.

In conclusion, we have derived expressions for the asymptotic forms of $\phi_l(k,r)$ in $D$, $f_l(\pm k)$ in $S$ and for the zeros $k_n$ of $f_l(-k)$ for which $k_n \geqslant\kappa\gg R^{-1}$. From these results one can easily determine the asymptotic forms of $\phi_l(k_n,r)$, $f_l(k_n)$, and $df_l(-k)/dk\,|_{k=k_n}$ as $k_n\to\infty$, which is the information that is needed to complete the study of the completeness properties of the resonance sates which was outlined in the Introduction. Furthermore, we showed that the zeros of $f_l(-k)$ that lie closest to the real $k$ axis, are not necessarily those which are generated by the discontinuity of $V(r)$, or one of its derivatives, at the cutoff radius $R$. Thus, if one truncates a potential with an exponentially decreasing tail, it does not necessarily follow that the dominant $s$ matrix poles, i.e., those closest to the real axis, will always be changed.

## APPENDIX A

We have assumed that

$$V(r) = (1 - r/R)^{\sigma} P_0(1 - r/R) \tag{A1}$$

throughout a $\delta$ neighborhood of $R$ given by $R-\delta\leqslant r\leqslant R$, where $P_0(t)$ is a convergent power series, $\sigma = \mu + N$, $0\leqslant\mu < 1$, and $N$ is a nonnegative integer. It then follows that

$$\begin{aligned}\mathscr{V}(r) &\equiv V(r) + l(l+1)r^{-2}\\ &= (1 - r/R)^{\sigma}P_0(1 - r/R) + P_1(1 - r/R)\end{aligned} \tag{A2}$$

in the same $\delta$ neighborhood, where $P_1$ is a power series expansion of the centrifugal potential about the point $r = R$. A straightforward induction argument that employs the recurrence relation for the expansion coefficients $a_j(r)$, Eq. (2.24e), and general properties of convergent power series[10] then gives

$$a_j(r) = P_{j0}(t) + \sum_{P=1}^{j} t^{P(\sigma+2)-j}P_{jP}(t), \tag{A3}$$

for $r$ in the $\delta$ neighborhood, where $t = 1 - r/R$ and the $P_{jP}(t)$ with $p = 0$ to $j$ are a set of convergent power series. From Eq. (A3) and its derivatives with respect to $r$ it follows that

$$\left|\frac{d^J a_j(r)}{dr^J}\right| \leqslant \text{const}[1 + (1 - r/R)^{\sigma+2-j-J}], \tag{A4}$$

for all $J\geqslant 0$, and $r$ in the $\delta$ neighborhood, where $d^0 a_j(r)/dr^0 \equiv a_j(r)$.

We now wish to determine a bound on the error made in approximating the two solutions $y_\alpha(k,r)$, with $\alpha = \pm 1$, of the radial Schrödinger equation, cf. Eq. (2.24a), by the asymptotic approximation

$$Y(\alpha k,r)=e^{i\alpha kr}\sum_{j=0}^{M}a_j(r)(\alpha k)^{-j}, \tag{A5}$$

where $M$ and the range of $r$ over which Eq. (A5) is defined are to be determined. Defining $z_\alpha(k,r)$ by

$$y_\alpha(k,r)=Y(\alpha k,r)z_\alpha(k,r) \tag{A6}$$

one has

$$y_\alpha(k,r)-Y(\alpha k,r)=Y(\alpha k,r)[z_\alpha(k,r)-1]. \tag{A7}$$

Hence, a bound on the error follows immediately if one canot construct a bound on $|z_\alpha(k,r)-1|$. Inserting Eq. (A6) into the radial Schrödinger equation yields a differential equation for $z_\alpha(k,r)$. After two successive integrations and a suitable choice of integration constants one then obtains the integral equation[5]

$$z_\alpha(k,r)=1-\int_a^r K(\alpha k,r,t)F(\alpha k,t)z_\alpha(k,t)\,dt, \tag{A8}$$

where

$$K(\alpha k,r,t)=\int_t^r Y^2(\alpha k,t)Y^{-2}(\alpha k,s)\,ds, \tag{A9}$$

and

$$F(\alpha k,r)=-2i\left(\frac{da_{M+1}(r)}{dr}\right)e^{i\alpha kr}(\alpha k)^{-M}Y^{-1}(\alpha k,r). \tag{A10}$$

The lower limit $a$ of the integral on the right-hand side of Eq. (A8) must be chosen with care. It will differ in the two cases $\alpha=-1$ and $+1$. We shall examine the $\alpha=+1$ case in some detail.

Since $|Y(k,t)/Y(k,s)|$ is bounded for all $t<s$ and $k$ in $S$, if one chooses $a\leqslant r$ for all values of $r$ considered, then $t\leqslant r$ for all values of $t$ in Eqs. (A8) and (A9). Hence

$$|K(k;r,t)|<\text{const}(r-t) \tag{A11}$$

for all $s$ and $t$ for which $a\leqslant t<s\leqslant r$ and $k\to\infty$ in $S$. Setting

$$z_{+1}(k,r)=\sum_{j=0}^{\infty}z^{(j)}(r) \tag{A12}$$

with $z^{(0)}(r)=1$, it follows that

$$z^{(j)}(r)=-\int_a^r K(k;r,t)F(k,t)z^{(j-1)}(t)\,dt \tag{A13}$$

and thus

$$|z^{(j)}(r)|\leqslant\int_a^r |K(k;r,t_1)F(k,t_1)|dt_1\times\cdots$$

$$\times\int_a^{t_{j-1}}|K(k;t_{j-1},t_j)F(k,t_j)|dt_j. \tag{A14}$$

From Eqs. (A10), (A4), and (A5) it follows that for $k\to\infty$ in $S$ and $r$ in the $\delta$ neighborhood of $R$

$$|F(k,r)|\leqslant C[1+(1-r/R)^{\mu+N+M}]|k|^{-M}\equiv\mathscr{F}(|k|,r). \tag{A15}$$

Equation (A15) holds for all $r$ for which $R_n\leqslant r\leqslant R$ as long as $M\leqslant L$, since $V(r)$ was assumed to have continuous bounded derivatives up to order $L$ for $R_n\leqslant r<R$, which implies that $|da_{M+1}(r)/dr|$ is bounded over the same interval. Combining Eqs. (A14), (A15), and (A11) yields

$$|z^{(j)}(r)|\leqslant[B(|k|,r)]^j/j! \tag{A16}$$

with

$$B(|k|,r)=\int_a^r(r-t)\mathscr{F}(|k|,t)\,dt\leqslant C|k|^{-M}\left(\frac{r^2+a^2}{2}-ra-\frac{(R-r)^{\mu+N+2-M}-(R-a)^{\mu+N+2-M}}{(\mu+N+2-M)R^{\mu+N-M}}\right) \tag{A17}$$

and thus,

$$|z_{+1}(k,r)-1|\leqslant\sum_{j=1}^{\infty}|z^{(j)}(r)|\leqslant B(|k|,r)\exp[B(|k|,r)]. \tag{A18}$$

If one takes $M=N+1$, then $a_j(r)$, $da_j(r)/dr$, with $j\leqslant M$, and the term in large parens on the right-hand side of Eq. (A17) are finite for all values of $r\in[R_n,R]$, even when $r\to R$. They are also finite for all values of $M\leqslant L$ if $r$ is bounded away from $R$. It is therefore convenient to divide the segment $[R_n,R]$ into two subsegments $\Sigma_n$ and $\Sigma_{n+1}$ as defined in Sec. 2, i.e., $\Sigma_n=[R_n,R-|k|^{-1}]$ and $\Sigma_{n+1}=[R-|k|^{-1},R]$. Setting $a=R_n$ and $M=L\geqslant\sigma+2$ in $\Sigma_n$, and $a=R-|k|^{-1}$ and $M=N+1$ in $\Sigma_{n+1}$, then gives two separate solutions $y_{+1j}(k,r)$ with $j=n$ and $n+1$ for which

$$|z_{+1j}(k,r)-1|\leqslant B(|k|,r)\exp[B(|k|,r)]\leqslant\text{const}\times|k|^{-\mu-N-2} \tag{A19}$$

uniformly for $r$ in $\Sigma_j$ and $k$ in $S$ and for both $j=n$ and $n+1$. The constants of integration implicit in the definitions of the expansion coefficients can be chosen so that $a_j(r)$ is continuous for all $r$ and $j$ for which $R_n\leqslant r\leqslant R$ and $j\leqslant N+1$.

The case $\alpha=-1$ can be treated in a similar manner, the principal difference being that the fixed limit of integration in the integral equation for $z_{-1j}(k,r)$ is taken as $R-|k|^{-1}$ for $j=n$ and $R$ for $j=n+1$, while $M=L$ in $\Sigma_n$ and $M=N+1$ in $\Sigma_{n+1}$, just as in the $\alpha=+1$ case. One finds that Eq. (A19) also holds when $z_{+1j}$ is replaced by $z_{-1j}$. Hence,

$$y_{\alpha j}(k,r)=Y_j(\alpha k,r)z_{\alpha j}(k,r)=Y_j(\alpha k,r)[1+O(k^{-\sigma-2})] \tag{A20}$$

for $\alpha=1,2$ and $j=n,n+1$ uniformly for $r$ in $\Sigma_j$ and $\arg k$, as $k\to\infty$ in $S$. Adapting the last two equations of Sec. 4.3 of Ref. 5 to this case yields a bound on the asymptotic approximation to the derivative, namely

$$\frac{dy_{\alpha j}(k,r)}{dr}=\frac{dY_j(\alpha k,r)}{dr}[1+O(k^{-\sigma-2})], \tag{A20b}$$

again, uniformly for $r\in\Sigma_j$ and $\arg k$ as $k\to\infty$ in $S$.

From Eqs. (A20a), (A20b), and the asymptotic expansions of $y_{\alpha,n-1}(k,r)$ and $dy_{\alpha,n-1}(k,r)/dr$, which were derived in Sec. 2, one can determine the coefficients $\mathscr{M}^n_{\alpha\beta}(k)$ [cf. Eq. (2.25a)]. The derivation is precisely the same as the one leading up to Eqs. (2.28a)–(2.28e), only in this case the fractional errors made in replacing $y_{\alpha j}(k,r)$ and $dy_{\alpha j}(k,r)/dr$ with $j=n$ by their asymptotic expansions are $O(k^{-\sigma-2})$ rather than $O(k^{-L})$, as was the case for $j<n$. As a result,

one finds that Eq. (2.28e) also holds for $j = n$ if $\mu_n$ is defined as follows,

$$\mu_n = \min[m_n,\sigma]. \tag{A21}$$

Next we shall determine the asymptotic forms of the integrals

$$\mathscr{J}^{\pm}_{\alpha,n}(k) = \int_{R_n}^{R} dr\, w_l(\pm kr)V(r)y_{\alpha,n}(k,r), \quad \text{with } \alpha = \pm 1. \tag{A22}$$

First we note that $y_{\alpha,n}(k,r) = \Sigma_{\beta=-1}^{+1} \mathscr{M}^{n+1}_{\alpha\beta}(k) y_{\beta,n+1}(k,r)$ with

$$\mathscr{M}^{n+1}_{\alpha\beta} = W[y_{-\beta,n+1}(k,r),y_{\alpha,n}(k,r)] \times \{W[y_{-\beta,n+1}(k,r),y_{\beta,n+1}(k,r)]\}^{-1}. \tag{A23}$$

Choosing $r = R - |k|^{-1}$ as the point at which the Wronskians are to be evaluated, employing Eqs. (A20a), (A20b), and (A4) gives

$$\mathscr{M}^{n+1}_{\alpha\alpha}(k) = 1 + O(k^{-\sigma-2})$$

and

$$\mathscr{M}^{n+1}_{\alpha,-\alpha}(k) = O[k^{-\sigma-2}\exp(2i\alpha kR)]. \tag{A24}$$

Combining Eqs. (2.41b), (A5), (A20a), and (A20b) leads to

$$\mathscr{J}^{-}_{+1,n}(k) = \sum_{j=n}^{n+1}\Bigg[i^l \sum_{h=0}^{l}\sum_{g=o}^{G(j)} k^{-h-g}\int_{R_j}^{R_{j+1}} dr\, V(r)b_h(r) \times a_g(r)\exp(2ikr) + \int_{R_j}^{R_{j+1}} dr\, 'w_l(-k,r)V(r)Y_j(k,r) \times (z_{+1,j}(k,r)-1)\Bigg](1+O(k^{-\sigma-2}))$$
$$+\Bigg\{i^l \sum_{h=0}^{l}\sum_{g=0}^{N+2} k^{-h-g}(-1)^g\int_{R-|k|^{-1}}^{R} dr \times V(r)b_h(r)a_g(r) + \int_{R-|k|^{-1}}^{R} dr\, w_l(-k,r) \times V(r)Y_{n+1}(-k,r)[z_{-1,n+1}(k,r)-1]\Bigg\}O(k^{-\sigma-2}e^{2ikR}), \tag{A25}$$

where $G(n)\equiv L$, $G(n+1)\equiv N+2$, $R_{n+1}\equiv R-|k|^{-1}$, and $R_{n+2}\equiv R$.

From Eqs. (A1) and (A3) it can be seen that $e^{2ikr}V(r)a_g(r)\,b_h(r)$ can be written as $e^{2ikr}(r-R)^{\lambda-1}\phi(r)$, where $0<\lambda<1$ and $\phi(r)$ is $N_g$ times continuously differentiable for $R_n \leqslant r \leqslant R$, where $N_g$ is defined in Eq. (2.42b). The first integral on the right-hand side of Eq. (A25) can be evaluated by using Eqs. 2.8(2) and 2.8(3) of Ref. 5 and the third integral can be determined by applying Eqs. 2.8(11) and 2.8(12) of the same reference.[11] Bounds can easily be constructed on the contributions of the second, fourth, fifth, and sixth integrals, using Eqs. (A1), (A4), and (A19). A somewhat tedious but straightforward calculation then gives

$$\mathscr{J}^{-}_{+1,n}(k) = i^l v_0 R e^{2ikR}\Gamma(\sigma+1)(2ikr)^{-\sigma-1}[1+O(k^{-1})] - i^l(2ik)^{-1}\sum_{g=0}^{l}\sum_{h=0}^{L} k^{-g-h}\sum_{j=0}^{N_h}(-2ik)^{-j} \times \exp(2ikR_n)\frac{d^j[V(r)a_h(r)b_g(r)]}{dr^j}\bigg|_{r=R_n} O(e^{2|v|R}|k|^{-\sigma-2}) \tag{26a}$$

uniformly in arg$k$ as $k\rightarrow\infty$ in $S$. Proceeding in a similar manner, one then finds that

$$\mathscr{J}^{+}_{+1,n}(k) = i^l\int_{R_n}^{R} dr\, V(r)[1+O(k^{-1})], \tag{A26b}$$

$$\mathscr{J}^{-}_{-1,n}(k) = i^l\int_{R_n}^{R} dr\, V(r)[1+O(k^{-1})], \tag{A26c}$$

$$\mathscr{J}^{+}_{-1,n}(k,r) = i^l v_0 R e^{-2ikR}\Gamma(\sigma+1)(-2ikR)^{-\sigma-1} \times[1+O(k^{-1})] + i^l(2ik)^{-1}\sum_{g=0}^{l}\sum_{h=0}^{L}(-k)^{-g-h} \times\sum_{j=0}^{N_h}(2ik)^{-j}\exp(-2ikR_n) \times\frac{d^j[V(r)a_h(r)b_g(r)]}{dr^j}\bigg|_{r=R_n} + O[\exp(-2|v|R_n)|k|^{-L}] \tag{A26d}$$

uniformly in arg$k$ as $k\rightarrow\infty$ in $S$.

## APPENDIX B

The derivation of the asymptotic forms of $\mathscr{J}^{\pm}_{\alpha,0}(k)$ is quite similar to the derivation of the asymptotic forms of $\mathscr{J}^{\pm}_{\alpha,n}(k)$ carried out in Appendix A. This analogy will be exploited to abbreviate the derivation in this appendix.

We have assumed that [cf. item(c) in the list of potential properties]

$$V(r) = r^{-\gamma}\sum_{j=0}^{\infty}\tau_j r^j \equiv r^{-\gamma}P_0(r) \tag{B1}$$

for $r\leqslant\epsilon R$, where $\epsilon$ is a small positive number and $P_0(r)$ is a uniformly and absolutely convergent power series, and we have further assumed that $V(r)$ has bound and continuous derivatives up to order $L$ for $\epsilon R\leqslant r\leqslant R_1$. Adding the centrifugal potential to $V(r)$ gives $\mathscr{V}(r)$, which is then used in conjunction with Eqs. (2.24d) and (2.24e) to determine the form of the expansion functions $a^0_j(r)$ in $\Sigma_0$. An induction argument then yields the following form for $a^0_j(r)$ for $r\leqslant\epsilon R$,

$$a^0_j(r) = \sum_{s=0}^{j-1}(\log r)^s r^{s-j}\Big[\mathscr{P}^{(j)}_s(r) + \sum_{t=1}^{j-s} r^{t(2-\gamma)}P^{(j)}_s(r)\Big] + C^{(j)}(\log r)^j, \tag{B2}$$

where $\mathscr{P}^{(j)}_s$ is a $(j-s)$th order polynomial and $P^{(j)}_s(r)$ is an absolutely convergent power series. [The polynomials and power series will of course depend on the parameter of $\mathscr{V}(r)$ such as $l$, $\gamma$, $\tau_0$, etc. For example, the logarithmic terms will all vanish if $\gamma$ is an irrational number, and the polynomial $\mathscr{P}_s^{(j)}(r)$ can be written as $r$ times a $(j-s-1)$th order polynomial for $s$ waves.] From Eq. (B2) it follows that

$$\left|\frac{d^J a_j(r)}{dr^J}\right| \leqslant \text{const} \times [1+(r/R)^{-j-J}] \tag{B3}$$

for all $J \geqslant 0$ and $r < \epsilon R$. Since $d^J \mathscr{V}(r)/dr^J$ is assumed to be continuous and bounded for all $J \leqslant L$ if $\epsilon R \leqslant r \leqslant R_1$, it follows that Eq. (B3) also holds for all $r \in \Sigma_0$ as long as $j + J \leqslant L + 2$.

The relationship between the exact solutions to the radial Schrödinger equation denoted by $y_{\alpha,0}(k,r)$ with $\alpha = +1$ and their asymptotic approximations $Y_0(\alpha k,r)$ is obtained by substituting $L$ for $M$ and adding the subscript 0 where appropriate in Eqs. (A5)–(A10). If one next specializes to the $\alpha = \pm 1$ case, sets the lower limit of integration $a$ in Eq. (A8) equal to $\rho$, and restricts $r$ to lie in $\Sigma_0$, then Eqs. (A11)–(A14) also apply to the present case. From the bounds given by Eq. (B3) and the definition of $F(k,r)$ given by Eq. (A10) one finds that

$$|F(k,r)| \leqslant C[1+(r/R)^{-L-2}]|k|^{-L} \equiv \mathscr{F}(|k|,r) \tag{B4}$$

for $\rho \leqslant r \leqslant R_1$ as $k \to \infty$ in $S$. Combining Eqs. (B4), (C11), and (C14), with $a$ replaced by $\rho$, yields

$$\begin{aligned} |z^{(j)}(r)| &\leqslant \left\{ c\int_\rho^r dt\,(r-t)[1+(t/R)^{-L-2}]|k|^{-L}\right\}^j/j! \\ &\leqslant [c'(\rho/R)^{-L-1}|k|^{-L}]^j/j! \\ &= [c''(\log|k|)^{L+1}|k|^{-L}]^j/j! \end{aligned} \tag{B5}$$

for any $r$ in $\Sigma_0$ and $k$ in $S$. Thus,

$$|z_{+1,0}(k,r) - 1| = \left|\sum_{j=1}^{\infty} z^{(j)}(r)\right| \leqslant \text{const}(\log|k|)^{L+2}|k|^{-L} \tag{B6}$$

for $r$ in $\Sigma_0$ and $k$ in $S$. The analog of Eq. (A20) becomes

$$y_{\alpha,0}(k,r) = Y_0(\alpha k,r)[1+o(|k|^{-L+1})] \tag{B7a}$$

uniformly in $r$ and $\arg k$ as $k \to \infty$ in $S$. With the usual substitutions the counterpart of Eq. (A20b) becomes

$$\frac{dy_{\alpha,0}(k,r)}{dr} = \frac{dY_0(\alpha k,r)}{dr}[1+o(|k|^{-L+1})] \tag{B7b}$$

uniformly in $r$ and $\arg k$ as $k \to \infty$ in $S$.

Although we have only established Eqs. (D7a) and (D7b) for the $\alpha = +1$ case it can easily be established for the $\alpha = -1$ case as well. One need only choose the parameter $a$ to be $R_1$, then it follows that

$$|K(-k;s,t)| \leqslant \text{const} \times (t-r) \tag{B8}$$

for all $s$ and $t$ for which $a = R_1 \geqslant t > s \geqslant r$. The rest of the derivation is almost identcial to that of the $\alpha = +1$ case and it will not be repeated.

Next we consider

$$\mathscr{J}_{\alpha,0}^{\pm}(k) = \int_\rho^{R_1} dr\, w_l(\pm kr)V(r)y_{\alpha,0}(k,r). \tag{B9}$$

Substituting Eq. (B7a) for $y_{\alpha,0}(k,r)$ and replacing $w_l(\pm kr)$ and $Y_0(\alpha k,r)$ by their expanded forms gives

$$\begin{aligned} \mathscr{J}_{\alpha,0}^{\pm}(k) &= \sum_{j=0}^{L}\sum_{h=0}^{l} k^{-j-h}\int_\rho^{R_1} dr\, e^{ik(\alpha\mp 1)r}V(r)a_j^0(r)b_h(r) \\ &\quad + \int_\rho^{R_1} dr\, w_l(\pm kr)V(r)Y_0(\alpha k,r)[z_\alpha(k,r)-1]. \end{aligned} \tag{B10}$$

The employment of bounds on $|Y_0(\alpha k,r)|$ and $|z_{\alpha,0}(k,r)-1|$ obtained from Eqs. (A5), (B3), and (B6) along with Eq. (2.13) and the bounds on $|w_l(\pm kr)|$ given in Ref. 1, yields a bound on the last integral on the right-hand side of Eq. (B10). To discuss the first terms on the right-hand side we consider two cases separately. For $\mathscr{J}_{+1,0}^{-}(k)$ and $\mathscr{J}_{-1,0}^{+}(k)$ the integrals are of the form $e^{\pm 2ikr}\Phi_{jh}(r)$, where $\Phi_{jh}(r) \equiv V(r)a_j^0(r)b_h(r)$ is $N_j$ times continuously differentiable in $\Sigma_0$ and $N_j$ is given by Eq. (2.42b). Thus, the $(jh)$th term of the sum can be integrated by parts $N_j$ times and the same bounds referred to above can be employed to place bounds on the remainder terms. One finds that

$$\begin{aligned} \mathscr{J}_{+1,0}^{-}(k) &= i^l(2ik)^{-1}\sum_{j=0}^{L}\sum_{h=0}^{l} k^{-j-h}\sum_{s=0}^{N_j} e^{2ikr}(-2ik)^{-s} \\ &\quad \times \left.\frac{d^s\Phi_{jh}(r)}{dr^s}\right|_\rho^{R_1} + o[k^{-L+1}\exp(2ikR_1)], \end{aligned} \tag{B11a}$$

$$\begin{aligned} \mathscr{J}_{-1,0}^{+}(r) &= -i^l(2ik)^{-1}\sum_{j=0}^{L}\sum_{h=0}^{l}(-k)^{-j-h}\sum_{s=0}^{N_j} e^{-2ikr}(2ik)^{-s} \\ &\quad \times \left.\frac{d^s\Phi_{jh}(r)}{dr^s}\right|_\rho^{R_1} + o[k^{-L+1}\exp(-2ik\rho)] \end{aligned} \tag{B11b}$$

uniformly in $\arg k$ as $k \to \infty$ in $S$.

A much less detailed expansion of $\mathscr{J}_{-1,0}^{-}(k)$ and $\mathscr{J}_{+1,0}^{+}(k)$ will suffice. One easily obtains

$$\mathscr{J}_{-1,0}^{-}(k) = i^l\int_\rho^{R_1} dr\, V(r)[1+O(1/k\rho)], \tag{B12a}$$

$$\mathscr{J}_{+1,0}^{+}(k) = i^l\int_\rho^{R_1} dr\, V(r)[1+O(1/k\rho)], \tag{B12b}$$

again, uniformly in $\arg k$ as $k \to \infty$ in $S$.

## APPENDIX C

*Theorem*: Given that $a_0(x) \equiv 1$, that $a_n'(x) = (2i)^{-1} \times [\mathscr{V}(x)a_{n-1}(x) - a_{n-1}''(x)]$ for $n \geqslant 1$, and that $\mathscr{V}^{(j)}(x)$ is continuous at every point $x$ on a line segment $L$ and for every $j = 0$ to $N$, then it is possible to construct solutions $a_n(x)$ such that $a_n^{(m)}(x)$ is continuous at every $x$ on $L$ if $0 \leqslant m \leqslant N+2-n$ and $n \leqslant N+2$. [In this Appendix we define $F^{(0)}(x) \equiv F(x)$ and $F^{(m)}(x) \equiv d^mF(x)/dx^m$ for $m \geqslant 1$.]

*Proof*: As a preliminary to a proof by induction an essential equation is derived. Applying the defining equation for $a_n'(x)$ repeatedly to eliminate all terms in which second derivatives of $a$'s occur yields

$$a_n'(x) = (2i)^{-1}\mathscr{V}(x)a_{n-1}(x) - (2i)^{-2}[\mathscr{V}(x)a_{n-2}(x)]'$$

$$+\cdots-(-2i)^{-n}\mathscr{V}^{(n)}(x)$$

$$=-\sum_{j=1}^{n}(-2i)^{-j}[\mathscr{V}(x)a_{n-j}(x)]^{(j-1)}.$$

Next, taking higher order derivatives of this equation then gives

$$a_n^{(m)}(x)=-\sum_{j=1}^{n}(-2i)^{-j}[\mathscr{V}(x)a_{n-j}(x)]^{(j+m-2)}$$

$$=-\sum_{j=1}^{n}(-2i)^{-j}\sum_{k=0}^{j+m-2}\binom{j+m-2}{k}$$

$$\times\mathscr{V}^{(j+m-2-k)}(x)a_{n-j}^{(k)}. \qquad \text{(C1)}$$

We now observe that the theorem holds for $n=0$ and we assume it holds for all $a_s^{(t)}(x)$ with $s\leqslant n-1$ where $1\leqslant n\leqslant N+2$, that is we assume that $a_s^{(t)}(x)$ is continuous for all $x$ in $L$ as long as $s+t\leqslant N+2$ and $s\leqslant n-1$. Examining the $s=n$ case we see by Eq. (C1) that the highest order derivative of $\mathscr{V}(x)$ that appears in the expansion of $a_n^{(m)}(x)$ occurs in the term in which $j=n$ and $k=0$ for which $j+m-2-k=n+m-2$. Thus, as long as $m\leqslant N+2-n$, no derivative of $\mathscr{V}(x)$ of order greater than $N$ will appear in the expansion. We next note that the sum of the upper and lower indices of $a_{n+1-j}^{(k)}(x)$ equals $n-j+k\leqslant n+m-2$. Again, as long as $m\leqslant N+2-n$ this sum will be smaller than $N+2$ so that all of the factors $a_{n-j}^{(k)}(x)$ that appear in the expansion will be continuous for $x$ in $L$ by the induction hypothesis as long as $m\leqslant N+2-n$. Thus, $a_n^{(m)}(x)$ will be continuous for all $x$ on $L$ if $1\leqslant m\leqslant N+2-n$. Now clearly if $a_n^{(1)}(x)\equiv a_n'(x)$ is continuous for $x$ in $L$, then $a_n^{(0)}(x)\equiv a_n(x)$ will be continuous as well. So the theorem holds for the $s=n$ case and the proof by induction is complete.

*Corollary*: Suppose that the conditions of the theorem hold and $a_n(x)$ with $n\leqslant N+2$ is constructed so that $a_n^{(m)}(x)$ is continuous for all $x$ on $L$ and $m\leqslant N+2-n$, then if $\mathscr{V}_n^{(N+1)}(x)$ is discontinuous at a point $x_0$ in $L$ it follows that $a_n^{(N+3-n)}(x)$ is discontinuous at $x_0$.

*Proof*: From Eq. (C1) one has

$$a_n^{(N+3-n)}(x)$$

$$=-\sum_{j=1}^{n}(-2i)^{-j}\sum_{k=0}^{j+N+1-n}\binom{j+N+1-n}{k}$$

$$\times\mathscr{V}^{(j+N+1-n-k)}(x)a_{n-j}^{(k)}(x).$$

Since $(n-j)+k\leqslant N+1$, $a_{n-j}^{(k)}(x)$ will be continuous at the point $x_0$ by the previous theorem for every $k$ in the sum. Next we note that $j+N+1-n-k\leqslant N+1$. Thus, no term in the double sum except the one with $j=n$ and $k=0$ involves a derivative of as high an order as $N+1$, and consequently all terms for which $(j,k)\neq(n,0)$ will be continuous while the $(n,0)$th term equals

$$-(2i)^{-n}\mathscr{V}^{(N+1)}(x),$$

which is discontinuous by our basic hypothesis.


[1]R. Newton, J. Math. Phys. **1**, 319 (1960); *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[2]W.J. Romo, Nucl. Phys. A **237**, 275 (1975).
[3]W.J. Romo, Nucl. Phys. A **302**, 61 (1978).
[4]M. Abromowitz and A. Stegun, Eds., *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1965).
[5]A. Erdélyi, *Asymptotic Expansions* (Dover, New York, 1956).

[6]By $\psi(x)=o[\phi(x)]$ we mean that $\psi(x)/\phi(x)$ tends to zero as $x$ tends to infinity, by $\psi(x)=O[\phi(x)]$ we shall mean that $\psi(x)/\phi(x)$ tends to a finite nonzero limit as $x\to\infty$.
[7]In reviewing the derivation of the asymptotic forms of $\mathscr{F}^{(\wp)}(+k,\rho)$, $B_{+1,0}(k)$ and $I_1^+(k,\rho)$ one finds that the restriction of $k$ in $D$ was unnecessary since the term $O(e^{-2|\nu|\rho}/|k|\rho)$ is negligible compared with $f(|k|)$ for all $k$ in $S$. Thus, the asymptotic forms of these functions hold uniformly in $\arg k$ as $k\to\infty$ in $S$.

[8]J. Humblet, Mém. Soc. Roy. Sci. Liège **4**, 12 (1952).
[9]T. Regge, Nuovo Cimento **8**, 671 (1958).
[10]W. Kaplan, *Advanced Calculus* (Addison-Wesley, Reading, Massachusetts, 1957), pp. 349–62.

[11]Since $k$ will generally have an imaginary part in our application, while the variable $x$ which plays the role of $k$ in the theorems of Ref. 5 was assumed to be real, the equations referred to above cannot be directly applied. However, a straightforward rederivation of these equations with $x$ complex shows that all of these equations remain valid if the last term on the right-hand side of Eq. 2.8(2) and 2.8(11) are replaced by $O(e^{ix\beta}x^{-N})$ if $\mathrm{Im}x<0$ and by $O(e^{ix\alpha}X^{-N})$ if $\mathrm{Im}x>0$, where $\alpha$ and $\beta$ are the lower and upper limits of integration, respectively, in the two theorems and it is assumed that $\alpha<\beta$.

# Yet another formulation of the Einstein equations for stationary axisymmetry[a)]

Dale Cox

*Department of Mathematics, Physics, and Astronomy, Sul Ross State University, Alpine, Texas 79830*

William Kinnersley

*Physics Department, Montana State University, Bozeman, Montana 59715*

(Received 18 July 1978; revised manuscript received 14 November 1978)

We rewrite Einstein's equations for stationary axially symmetric gravitational fields, using a pair of noncanonical, intrinsically-defined coordinates. We show that both field equations of the Ernst formulation of this problem can be solved identically, by means of a new superpotential $K$. One more field equation remains to be satisfied, however. It can be expressed as a single fourth-order equation for $K$, or as a pair of coupled second-order equations. The approach works equally well for the wave metrics one can get from the stationary case via complex coordinate transformations. We illustrate our method by using it to derive a new class of wave solutions.

## I. INTRODUCTION

This paper is concerned with the formulation and solution of Einstein's equations, for the important case in which the field is stationary and axially symmetric. The results will apply equally well to the two types of wave metrics, which can be obtained from the stationary case by means of complex coordinate transformations.[1] We depart from the traditional approach by not using "canonical" coordinates. Instead we use intrinsic coordinates, defining them directly in terms of the gravitational Newtonian potential $f$ and the gravitational twist potential $\Omega$.

We *are* aware of the main drawback of intrinsic coordinate systems: they are poorly adapted to a discussion of any limit in which the field is weak. In our case this will lead to a complicated condition for asymptotic flatness. Moreover, the static (nonrotating) case $\Omega \equiv 0$ will require special treatment as a limiting case. For certain other solutions, $f$ and $\Omega$ happen to be functionally dependent, and our coordinate system becomes completely degenerate. However, all solutions for which this happens are already well known.[1]

Making up for these drawbacks is the considerably simplification that we have been able to achieve in the field equations. Of the two field equations which constitute the Ernst formulation,[2] we are able to solve both identically, in terms of a new "superpotential" $K$. This leaves only one more field equation remaining to be satisfied (one that had been an identity in canonical coordinates). It produces a single nonlinear fourth-order equation which must be obeyed by $K$. Alternatively, if one prefers, the equation can be cast as a pair of coupled second-order Monge–Ampère equations. The latter version is particularly convenient when it is actually the related wave metrics that one is dealing with. The Monge–Ampère equations are then hyperbolic, and the method of characteristics can then be used to advantage to look for new solutions.

Of course the hope in reformulating the vacuum equations is that we will eventually better understand their structure. In the meantime, there may be particular solutions which are simple in one formulation but complicated in another. In Sec. V we show how our method can in fact be used to find a new class of wave solutions.

## II. FIELD EQUATIONS

We will consider first the general stationary axially symmetric metric in the canonical form introduced by Lewis[3]

$$ds^2 = f(dt + \omega d\phi)^2 - f^{-1}[e^{2\gamma}(d\rho^2 + dz^2) + \rho^2 d\phi^2], \tag{2.1}$$

where $f$, $\omega$, $\gamma$ are functions only of the two coordinates $\rho$, $z$. The main Einstein vacuum equations may be written as two divergence equations for $f$, $\omega$[1]:

$$\nabla\cdot[\rho^{-2} f^2 \nabla(\rho^2 f^{-2} - \omega^2)] = 0, \quad \nabla\cdot[\rho^{-2} f^2 \nabla\omega] = 0. \tag{2.2}$$

Here the gradient operator and the "dot" are covariant derivative and scalar product with respect to a 3-metric,

$$\begin{aligned} d\sigma^2 &= g_{\alpha\beta}\, dx^\alpha dx^\beta \\ &= h_{AB}\, dx^A dx^B + \rho^2 d\phi^2 \\ &= e^{2\gamma}(d\rho^2 + dz^2) + \rho^2 d\phi^2, \\ &\alpha,\beta = 1,2,3, \quad A,B = 1,2. \end{aligned} \tag{2.3}$$

That is,

$$\nabla\cdot\mathbf{V} = g^{-1/2}(g^{1/2} g^{\alpha\beta}\mathbf{V}_\alpha)_{,\beta}. \tag{2.4}$$

In fact, Eqs. (2.2) are conformally invariant, in the sense that they do not depend on $\gamma$. The remaining field equations require that

$$\begin{aligned} \gamma_\rho &= [\rho f^{-2}(f_\rho^2 - f_z^2) - \rho^{-1} f^2(\omega_\rho^2 - \omega_z^2)]/4, \\ \gamma_z &= [\rho f^{-2} f_\rho f_z - \rho^{-1} f^2 \omega_\rho \omega_z]/2. \end{aligned} \tag{2.5}$$

[a)]Supported in part by NSF Grant GP-43844-X.

 0022-2488/79/061225-04$01.00  

As shown by Harrison[4] and Ernst,[2] the equations may also be formulated in terms of a twist potential $\Omega$, defined by

$$\nabla\Omega = \rho^{-1} f^2 \mathbf{e}_\phi \times \nabla\omega, \qquad (2.6)$$

where $\mathbf{e}_\phi$ is the unit vector in the $\phi$ direction. Eliminating $\omega$ in favor of $\Omega$ we can obtain an equivalent set of equations:

$$\nabla\cdot[f^{-2}\nabla(f^2+\Omega^2)] = 0,$$

$$\nabla\cdot[f^{-2}\nabla\Omega] = 0, \qquad (2.7)$$

$$\gamma_\rho = [\rho f^{-2}(f_\rho^2 - f_z^2 + \Omega_\rho^2 - \Omega_z^2)]/4,$$

$$\gamma_z = [\rho f^{-2}(f_\rho f_z + \Omega_\rho \Omega_z)]/2. \qquad (2.8)$$

Equations (2.7) are what we call the Ernst equations. Once they are solved, $\gamma$ may be found up to an additive constant from Eqs. (2.8).

## III. SOLUTION OF THE ERNST EQUATIONS

We will now choose intrinsic coordinates $(x^1,x^2)$ and solve the Ernst equations. Define the conformal metric

$$\bar{h}^{AB} = h^{1/2} h^{AB}. \qquad (3.1)$$

Then Eqs. (2.7) written out are

$$[\rho f^{-2}\bar{h}^{AB}(f^2+\Omega^2)_{,B}]_{,A} = 0, \quad (\rho f^{-2}\bar{h}^{AB}\Omega_{,B})_{,A} = 0. \qquad (3.2)$$

An advantageous choice of coordinates is $(x^1,x^2) = (f^2+\Omega^2,\Omega)$. This choice is possible as long as $f$ and $\Omega$ are functionally independent, which we assume. Eqs. (3.2) then become an equation for $\bar{h}^{AB}$

$$(\rho f^{-2}\bar{h}^{AB})_{,A} = 0, \quad B = 1,2. \qquad (3.3)$$

The most general solution is

$$\rho f^{-2}\bar{h}^{AB} = \epsilon^{AC}\epsilon^{BD}K_{,CD}, \qquad (3.4)$$

where $\epsilon^{AB} = \pm 1$ is the alternating symbol, and $K$ is an arbitrary scalar superpotential.

Taking account of the fact that $\bar{h}^{AB}$ has determinant unity, we find

$$\bar{h}^{AB} = \rho^{-1} f^2 K_{,AB}, \qquad (3.5)$$

$$\rho^2 = f^4 \det(K_{,AB}). \qquad (3.6)$$

The function $\omega$ can now be determined from Eq. (2.6):

$$\rho^{-1} f^2 \bar{h}^{AB}\omega_{,A} = \epsilon^{BC}\Omega_{,C},$$

$$\omega_{,A} = \rho f^{-2}\bar{h}_{AB}\epsilon^{B2},$$

$$= K_{,1A},$$

$$\omega = K_{,1} + \text{const.}$$

The constant can be removed by a coordinate transformation of the form $t \to t + a\phi$. Thus, we take

$$\omega = K_{,1}. \qquad (3.7)$$

To obtain this simple relation between $K$ and $\omega$ was why we chose to start with the field equations in the Ernst form, Eq. (2.7), rather than the original version, Eq. (2.2) (for which a superpotential could also be defined).

From the remaining field equations we find

$$\bar{\nabla}^2\rho = 0,$$

$$\bar{\nabla}_A\bar{\nabla}_B\rho - \rho_{,C}[\gamma_{,A}\delta^C_B + \gamma_{,B}\delta^C_A - \gamma_{,D}\bar{h}^{CD}\bar{h}_{AB}]$$
$$= -\rho f^{-2}[f_{,A}f_{,B} + \Omega_{,A}\Omega_{,B} - \bar{h}_{AB}\bar{h}^{CD} \times (f_{,C}f_{,D} + \Omega_{,C}\Omega_{,D})/2]/2, \qquad (3.8)$$

where $\bar{\nabla}$ is the covariant derivative for the metric $\bar{h}_{AB}$.

Summarizing our results to this point, we have

$$ds^2 = f(dt + K_{,1}d\phi)^2 - f^{-1}[e^{2\gamma}\bar{h}_{AB}dx^A dx^B + \rho^2 d\phi^2], \quad (3.9)$$

where

$$f^2 = x^1 - (x^2)^2, \quad \bar{h}_{AB} = \Delta^{-1}K_{,AB}, \qquad (3.10)$$

$$\Delta \equiv \det{}^{1/2}(K_{,AB}), \quad \rho = \Delta f^2. \qquad (3.11)$$

The remaining field equation is

$$\bar{\nabla}^2\rho = 0, \qquad (3.12)$$

and $\gamma$ is determined by Eq. (3.8).

## IV. DISCUSSION

Our main result is the pair of equations (3.11) and (3.12). Written out in detail, they are

$$\Delta^2 = K_{,11}K_{,22} - (K_{,12})^2, \qquad (4.1)$$

$$RK_{,11} + 2SK_{,12} + TK_{,22} = 0, \qquad (4.2)$$

$$R = 2[f^2(\ln\Delta)_{,2}]_{,2} - 4, \qquad (4.3)$$

$$S = [f^2(\ln\Delta)_{,1}]_{,2} + [f^2(\ln\Delta)_{,2}]_{,1}, \qquad (4.4)$$

$$T = 2[f^2(\ln\Delta)_{,1}]_{,1}. \qquad (4.5)$$

If all of these are combined, one nonlinear fourth-order equation will result. An alternative method is to try to solve Eqs. (4.1) and (4.2) simultaneously. For example, if $\Delta$ is assumed to be a "known" function of $x^1$ and $x^2$, then Eqs. (4.1) and (4.2) are an overdetermined system of equations for the unknown function $K$. The equations are then two examples of the Monge–Ampère differential equation.

The general Monge–Ampère equation has the form

$$U[K_{,11}K_{,22} - (K_{,12})^2] + RK_{,11} + 2SK_{,12} + TK_{,22} + V = 0, \qquad (4.6)$$

where $U,R,S,T,$ and $V$ are all specified functions of $x^1,x^2,K,K_{,1}$, and $K_{,2}$. This equation plays an important and distinguished role in the classical theory of partial differential equations. In the method of characteristics, for example, a general nonlinear second-order partial differential equations may be reduced to an equivalent set of eight ordinary differential equations.[5,6] The Monge–Ampère equation requires only five such equations in its characteristic set. For a further discussion of its exceptional nature, see Forsyth.[5] Previous applications of Monge–Ampère equations have occurred in differential geometry[7] and nonlinear elasticity.[8] Our case is a relatively simple one if we make the assumption that $\Delta(x^1,x^2)$ is known, since all the coefficients $U,R,S,T,$ and $V$ for Eq. (4.1) are then only functions of $x^1,x^2$, and Eq. (4.2) regarded as an equation for $K$ is actually linear.

We should make some further remarks at this point on the origin of Eq. (3.12). This equation was used originally in the reduction of the metric to its canonical form. Eq. (2.1), as a necessary condition to insure the existence of the coordinate $\rho$. In canonical coordinates it is therefore satisfied identically. However, when noncanonical coordinates are used, it must be explicitly retained as a field equation

For example, Hoffman[9] has considered the case when $\rho = \rho_0$ is constant, and therefore cannot be used as one of the coordinates. It is easy to show that the solution of Eqs. (4.1) and (4.2) in this case is

$$\rho = \rho_0 \qquad \Delta = \rho_0 f^{-2}, \qquad K = 2\rho_0 f^{-1}, \tag{4.7}$$

and that this is equivalent to Hoffman's solution.

As another example, we have found the superpotential for the Kerr solution to be

$$K = -a[\Omega^2 + (1-f)^2] - (4m^2\Omega^2/a)[\Omega^2 + (1-f)^2]^{-1}. \tag{4.8}$$

We will use it to illustrate the difficulties mentioned earlier that one encounters in the static limit. Taking the limit $a \to 0$ with $f, \Omega$ held fixed, we find that $K$ and all its derivatives tend to infinity.

The obvious source of the trouble is that in the standard description $\Omega$ itself is proportional to $a$. In the static limit the entire solution actually concentrates itself in a "boundary layer" along the coordinate line $\Omega = 0$. We can therefore treat solutions with *slow* rotation by rescaling the twist potential,

$$\hat{\Omega} = \Omega / a, \tag{4.9}$$

and taking the limit with $f, \hat{\Omega}$ held fixed. To keep all quantities finite, it is necessary at the same time to rescale

$$\hat{K} = K/a, \quad \hat{\Delta} = a\Delta. \tag{4.10}$$

When this limit is applied to Eqs. (4.1)–(4.5) the equations which result are unchanged in form, with only one exception. In the rescaled set,

$$R = 2[f^2(\ln\hat{\Delta})_{,\hat{2}}]_{,\hat{2}} \tag{4.11}$$

lacks the second term.

## V. THE CASE Δ = CONSTANT

We now consider in detail the solution of Eqs. (4.1)–(4.5) for the particular case $\Delta$ = constant. The coefficients are readily found to be

$$R = -4, \quad S = T = 0.$$

Then Eq. (4.2) reduces to

$$K_{,11} = 0. \tag{5.1}$$

Letting $(x^1, x^2) = (x, y)$ for convenience, the general solution of Eq. (5.1) is

$$K = ax + b, \tag{5.2}$$

where $a,b$ are arbitrary functions of $y$. Using this result in Eq. (4.1) we find

$$\Delta^2 = -(a')^2 \tag{5.3}$$

where the prime denotes differentiation with respect to $y$. This equation has no real solution except the trivial one $a = \Delta = 0$. We therefore look for possible transformations to alter the reality of some of our variables.

Consider the complex coordinate transformation[1]

$$t = i\hat{\rho}, \quad \rho = i\hat{t}, \quad \phi = i\hat{\phi}. \tag{5.4}$$

In canonical coordinates the metric becomes

$$ds^2 = f^{-1}e^{2\hat{\gamma}}(d\hat{t}^2 - dz^2) - f(d\hat{\rho} + \omega d\hat{\phi})^2 - f^{-1}\hat{t}^2 d\hat{\phi}^2, \tag{5.5}$$

where $f$, $\omega$, $\hat{\gamma}$ are functions only of $z$, $\hat{t}$, and $\hat{\gamma} = \gamma + i\pi/2$. Physically the metric represents waves propagating in both directions along the $z$ axis.

The line element in the form we use, Eq. (3.9), becomes

$$ds^2 = -f(d\rho + K_{,1}d\phi)^2 - f^{-1}[e^{2\gamma}\bar{h}_{AB}dx^A dx^B + \Delta^2 f^4 d\phi^2], \tag{5.6}$$

where all hats have been dropped for convenience. The 2-metric $\bar{h}_{AB}$ is now one with an indefinite signature,

$$\det(\bar{h}_{AB}) = -1.$$

After the transformation is performed, Eq. (5.3) becomes

$$-\Delta^2 = -(a')^2,$$

with the solution

$$a = \pm y\Delta + \beta, \quad \beta = \text{const.} \tag{5.7}$$

The superpotential $K$ is then

$$K = \pm xy\Delta + \beta x + b(y). \tag{5.8}$$

Using the gauge transformation

$$K \to K + Ax^1 + Bx^2 + C, \quad \rho \to \rho + \Delta\phi, \tag{5.9}$$

we can set $\beta = 0$. The reflection

$$y \to -y, \quad b(y) \to b(-y),$$

permits us to choose the ambiguous sign in $K$ to be positive without loss of generality, and a rescaling,

$$\phi \to \Delta\phi, \tag{5.10}$$

allows us to set $\Delta = 1$. In terms of the function

$$B(y) \equiv b''(y), \tag{5.11}$$

the complete solution for the case $\Delta$ = constant is given by

$$ds^2 = -f(d\rho + yd\phi)^2 - f^{-1}[e^{2\gamma}(2dxdy + Bdy^2) + f^4 d\phi^2], \tag{5.12}$$

where

$$f^2 = x - y^2 \tag{5.13}$$

$$\gamma = \frac{\ln(f^2)}{16} + \frac{\ln(B + 4y)}{2} - \int [B(y) + 4y]^{-1}\frac{dy}{2}. \tag{5.14}$$

Note that the coordinate $y$ actually turns out to be a null coordinate, since $g^{22} = 0$. From Eq. (5.14), we cannot permit $B(y) + 4y \equiv 0$. Other than this restriction, $B(y)$ is an arbitrary function of the null coordinate $y$, and so our solution has the appearance of a wave of arbitrary profile.

The situation is somewhat clearer when we return to canonical coordinates. The metric is then

$$ds^2 = f^{-1}e^{2\bar{\gamma}}(dt^2 - dz^2) - f(d\rho + Hd\phi)^2 - f^{-1}t^2 d\phi^2, \tag{5.15}$$

where

$$f^2 = t, \tag{5.16}$$

$$\bar{\gamma} = \frac{\ln(f^2)}{16} - \int [H'(u)]^2 \frac{du}{2}, \tag{5.17}$$

and $H(u)$ is the arbitrary function. The coordinate transformation from $(x,y)$ to $(z,t)$ is given by Eqs. (5.13), (5.16), and by

$$u = \int [B(y) + 4y] dy = z - t. \tag{5.18}$$


[1]W. Kinnersley, "Recent Progress in Exact Solutions" in *Proceedings of the Seventh International Conference on General Relativity and Gravitation*, edited by G. Shaviv and J. Rosen (Wiley, New York, 1975).
[2]F. Ernst, Phys. Rev. **167**, 1175 (1968).
[3]T. Lewis, Proc. Roy. Soc. (London) Ser. A **136**, (1932).
[4]B.K. Harrison, J. Math. Phys. **9**, 1744 (1968).
[5]A.R. Forsyth, *Theory of Differential Equations* (Dover, New York, 1959), Vol. VI, p. 200.
[6]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley, New York, 1966), Vol. II, p. 496.
[7]C. Weatherburn, *Differential Geometry in Three Dimensions* (Cambridge U.P., Cambridge, 1930), Vol. II, p. 201.
[8]J.J. Stoker, *Nonlinear Elasticity* (Gordon and Breach, New York, 1968), p. 47.
[9]R. Hoffman, J. Math. Phys. **10**, 953 (1969).

# A relativistic Lee model[a)]

Bernard M. de Dormale

*Départment de Physique-Mathématiques, Université de Moncton, Moncton, N.B.,Canada*
(Received 4 August 1978; revised manuscript received 5 December 1978)

We construct a fully relativistic version of the Lee model in the $S(1,1)$ sector.

## 1. INTRODUCTION

One of the main problems of relativistic quantum theory has always been the complete lack of well-defined relativistic models exhibiting nontrivial phenomena as particle creation and annihilation. The attempts to provide such models have only succeeded in two- or three-dimensional theories, and the way those models are constructed is extremely involved (see, e.g., Ref. 1). On the other hand, there exists quite simple models of quantum field theory, like the Lee model,[2] which are easy to handle, but are not relativistic.

The aim of this paper is to present a modified version of the Lee model which is both relativistic and simple (finite renormalization, etc.). The construction of this model follows ideas introduced several years ago by Coester[3] and others (see, e.g., Ref. 4 and references therein).

## 2. STRUCTURE OF UNITARY REPRESENTATIONS OF THE POINCARÉ GROUP

As it is well known,[5] every irreducible unitary representation (IUR) $U_{m,j}$ of the Poincaré group $\mathscr{P}$ with $m>0$ and $2j\in\mathbb{N}$ can be realized as acting in $\mathscr{H}_{mj} = L^2(\mathbb{R}^3)\otimes C^{2j+1}$ in the following way:

$$(U(\Lambda,a)\Psi)_\alpha(\mathbf{p}) = (\omega_{\mathbf{p}'}/\omega_{\mathbf{p}})^{1/2}\exp[i(\omega_{\mathbf{p}}a_0 - \mathbf{p}\cdot\mathbf{a})] \times \sum_{\alpha'} D_{\alpha'\alpha}(R(\Lambda,\mathbf{p}))\Psi_{\alpha'}(\mathbf{p}'), \tag{1}$$

where

$$p' = (\omega_{\mathbf{p}'},\mathbf{p}') = \Lambda^{-1}p, \quad p = (\omega_{\mathbf{p}},\mathbf{p}), \omega_{\mathbf{p}} = (\mathbf{p}^2+m^2)^{1/2}, \tag{2}$$

$D$ being the $(2j+1)$-dimensional IUR of SU(2) and $R$, a rotation depending on $\Lambda$ and $\mathbf{p}$.

It is also well known that every unitary representation (UR) $U$ of $\mathscr{P}$ in a Hilbert space $\mathscr{H}$ can be decomposed in a direct integral of IUR,

$$U(\Lambda,a) = \int^{\oplus} U_{m,j,\zeta}(\Lambda,a)\, d\mu, \tag{3}$$

with corresponding decomposition of $\mathscr{H}$,

$$\mathscr{H} = \int^{\oplus} \mathscr{H}_{m,j,\zeta}\, d\mu, \tag{4}$$

where $\zeta$ is a variable used to label the different representations with same $m$ and $j$ (i.e., $\mathscr{H}_{m,j,\zeta} = \mathscr{H}_{m,j}$ and $U_{m,j,\zeta} = U_{m,j}$). Now, for $m>0$,

$$\mathscr{H}_{m,j,\zeta} = L^2(\mathbb{R}^3)\otimes V_{m,j,\zeta}, \quad V_{m,j,\zeta} = \mathbb{C}^{2j+1}. \tag{5}$$

Thus, if 0 is not an eigenvalue of the mass operator $M$ of $U$, one has[6]

$$\mathscr{H} = L^2(\mathbb{R}^3)\otimes \int^{\oplus} V_{m,j,\zeta}\, d\mu = L^2(\mathbb{R}^3)\otimes \mathscr{H}_{\text{c.m.}} \tag{6}$$

We shall call $\mathscr{H}_{\text{c.m.}}$ the "center of mass" space.

Let us now make some remarks about the decomposition (6):

(i) The generators of the translations (momentum) $\mathbf{P}$ acts in the first factor of the tensor product (6), i.e., its components can be written as $P_i\otimes 1$.

(ii) The mass operator acts in the second factor $\mathscr{H}_{\text{c.m.}}$, i.e., is of the form $1\otimes M$.

(iii) If we restrict $U$ to the rotation subgroup, one has

$$U(R) = U_1(R)\otimes U_2(R). \tag{7}$$

With $U_1$ acting in $L^2(\mathbb{R}^3)$ in the usual way, while the action of $U_2$ in $\mathscr{H}_{\text{c.m.}}$, leaves $M$ invariant; $U_2$ can be decomposed in direct integral of IUR of SU(2) acting in $V_{m,j,\zeta}$.

We are thus in a situation analogous to the usual Galilean decomposition in total momentum and center of mass system, $M$ playing here the role of the internal energy. But the relativistic situation is more involved since one cannot decompose $U$ as a tensor product (this decomposition is valid for the Euclidian subgroup only).

The most interesting feature about the decomposition (6) is that the representation $U$ is completely determined by $U_2$ and $M$. As a matter of fact, we have the following:

*Theorem*: Let $U_2$ be any UR of SU(2) in a Hilbert space $\mathscr{H}_{\text{c.m.}}$, and let $M$ be a self-adjoint operator invariant under $U_2$ such that $\langle\Phi,M\Phi\rangle>0$, $\forall\Phi\in D_M$, $\Phi\neq0$. Then, there exists a unique UR $U$ of $\mathscr{P}$ on $L^2(\mathbb{R}^3)\otimes\mathscr{H}_{\text{c.m.}}$, such that:

(i) When restricted to the Euclidean subgroup, $U$ can be decomposed as a tensor product

$$U(R,\mathbf{a}) = U_1(R,\mathbf{a})\otimes U_2(R); \tag{8}$$

(ii) the mass operator of $U$ is $1\otimes M$.

*Proof*: $M$ being invariant under $U_2$, we have a decomposition of $\mathscr{H}_{\text{c.m.}}$ as a direct integral

$$\mathscr{H}_{\text{c.m.}} = \int^{\oplus} V_{m,j,\zeta}\, d\mu, \quad V_{m,j,\zeta} = C^{2j+1} \tag{9}$$

such that

[a)]Work supported by a grant of the C.N.R.C.

  

$$M = \int^{\oplus} m \, d\mu \tag{10}$$

with $\mu(\{0\}) = 0$, while $U_2$ decomposes in a direct integral of IUR of SU(2) in $V_{m,j,\zeta}$. We can use formulas (1) and (2) to define on $L^2(\mathbb{R}^3) \otimes V_{m,j,\zeta}$ an IUR. $U_{m,j,\zeta}$ of $\mathscr{P}$ and

$$U = \int^{\oplus} U_{m,j,\zeta} \, d\mu \tag{11}$$

will be a UR of $\mathscr{P}$ on

$$L^2(\mathbb{R}^3) \otimes \mathscr{H}_{\text{c.m.}} = \int^{\oplus} L^2(\mathbb{R}^3) \otimes V_{m,j,\zeta} \, d\mu. \tag{12}$$

It is then trivial to show that $U$ has the claimed properties; unicity is immediate too.

This theorem will enable us to construct very easily relativistic models, since the problem reduces to the construction of a rotation invariant mass operator in $\mathscr{H}_{\text{c.m.}}$.

The most natural way to do it is first to start with free particles and then to introduce an interaction, that is,

(i) to take a free representation $U_0$ of $\mathscr{P}$ in the Fock space of one or several massive particles, then to restrict $U_0$ to the subspace $\mathscr{H}$ orthogonal to the vacuum,

(ii) to achieve the decomposition $\mathscr{H} = L^2(\mathbb{R}^3) \otimes \mathscr{H}_{\text{c.m.}}$ and to compute the free mass operator $M_0$ in $\mathscr{H}_{\text{c.m.}}$.

(iii) to add to $M_0$ a rotation invariant perturbation $M_I$ to obtain the total mass operator $M = M_0 + M_I$. Of course, $M$ has to be a positive self-adjoint operator in $\mathscr{H}_{\text{c.m.}}$.

We shall see an example of such a construction in the next section.

## 3. COVARIANT LEE MODEL: BASIC INTERACTION

As in the original model,[3] we want to introduce a model where three scalar bosons, called, $N$, $\theta$, and $V$, interact according to the following scheme:

$$N + \theta \longleftrightarrow V. \tag{13}$$

Let $\mathscr{F}$ be the Fock space obtained by tensoring IUR of $\mathscr{P}$ corresponding to the respective masses $m_1$, $m_2$, and $m_3$ of the $N$, the $\theta$ and the $V$ (of course, the tensor products have to be symmetrized in the proper way; this will be tacitly assumed in the following). The usual creators and annihilators will be written as $N^*(\mathbf{p})$, $N(\mathbf{P})$, $\theta^*(\mathbf{p})$, etc. They satisfy the canonical commutation relations. Let us define

$$N_1 = \int N^*(\mathbf{p}) N(\mathbf{p}) \, d\mathbf{p} + \int V^*(\mathbf{p}) V(\mathbf{p}) \, d\mathbf{p},$$
$$N_2 = \int \theta^*(\mathbf{p}) \theta(\mathbf{p}) \, d\mathbf{p} + \int V^*(\mathbf{p}) V(\mathbf{p}) \, d\mathbf{p}, \tag{14}$$

and the $(n_1,n_2)$ sector $S(n_1,n_2)$ will be the common eigenspace of $N_1$ and $N_2$ corresponding to eigenvalues $n_1$ and $n_2$.

We have know to introduce our representation $U$ of $\mathscr{P}$. Because of the form of the interaction (13), each sector will be invariant under $U$, so that it is sufficient to define $U$ in $S(n_1,n_2)$ for each pair of values $(n_1,n_2)$. Let us start with

$$S(1,1) = L^2(\mathbb{R}^3) \oplus L^2(\mathbb{R}^6). \tag{15}$$

The first space occurring in S(1,1) is the space of one $V$ state while the second is the space of one $N$ and one $\theta$ state. Let us represent such a state by the square integrable function $\Psi(\mathbf{p}_1,\mathbf{p}_2)$, where $\mathbf{p}_1$ is the momentum of the $N$ and $\mathbf{p}_2$, the momentum of the $\theta$. If we define

$$\Omega_1 = (\mathbf{p}_1^2 + m_1^2)^{1/2}, \quad \Omega_2 = (\mathbf{p}_2^2 + m_2^2)^{1/2},$$
$$\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2, \quad \Omega = \Omega_1 + \Omega_2, \tag{16}$$

then the restriction of the free mass operator to S(1,1) will be given by

$$M_0 = m_3 \oplus \omega, \tag{17}$$

where

$$\omega = (\Omega^2 - \mathbf{P}^2)^{1/2}. \tag{18}$$

Let us now make in $L^2(\mathbb{R}^6)$ the change of variables

$$(\mathbf{p}_1,\mathbf{p}_2) \rightarrow (\mathbf{P},\mathbf{q}), \tag{19}$$

where

$$\mathbf{q} = \frac{1}{\omega}\left[\Omega_2 \mathbf{p}_1 - \Omega_1 \mathbf{p}_2 + \frac{\omega - \Omega}{\mathbf{P}^2} \mathbf{P} \times (\mathbf{p}_1 \times \mathbf{p}_2)\right] \tag{20}$$

is the momentum of the $N$ in the c.m. frame of the $N$–$\theta$ system. The inverse transformation is given by[5]

$$\mathbf{p}_1 = \frac{\omega_1}{\omega}\mathbf{P} + \mathbf{q} + \frac{\Omega - \omega}{\mathbf{P}^2 \omega}(\mathbf{P}\cdot\mathbf{q})\,\mathbf{P},$$
$$\mathbf{p}_2 = \frac{\omega_2}{\omega}\mathbf{P} - \mathbf{q} - \frac{\Omega - \omega}{\mathbf{P}^2 \omega}(\mathbf{P}\cdot\mathbf{q})\,\mathbf{P}, \tag{21}$$

$$\omega_1 = (\mathbf{q}^2 + m_1^2)^{1/2}, \qquad \omega_2 = (\mathbf{q}^2 + m_2^2)^{1/2}, \tag{22}$$

and the Jacobian of the transformation can be easily computed:

$$J = \left|\frac{\partial(\mathbf{p}_1,\mathbf{p}_2)}{\partial(\mathbf{p},\mathbf{q})}\right| = \frac{\omega}{\Omega}\frac{\Omega_1\Omega_2}{\omega_1\omega_2}. \tag{23}$$

The transformation $W$ defined by

$$\Psi'(\mathbf{P},\mathbf{q}) = (W\Psi)(\mathbf{P},q) = J^{1/2}\Psi(\mathbf{p}_1,\mathbf{p}_2) \tag{24}$$

is a unitary operator

$$W : L^2(\mathbb{R}^6) \rightarrow L^2(\mathbb{R}^3) \otimes L^2(\mathbb{R}^3). \tag{25}$$

We thus have an isomorphism

$$1 \oplus W : S(1,1) \rightarrow L^2(\mathbb{R}^3) \otimes (L^2(\mathbb{R}^3)),$$
$$\Psi_1(\mathbf{P}) + \Psi_2(\mathbf{p}_1,\mathbf{p}_2) \rightarrow \Psi_1(\mathbf{P}) \oplus \Psi_2'(\mathbf{P},\mathbf{q}). \tag{26}$$

Since $\omega = \omega_1 + \omega_2$ is a function of $\mathbf{q}$, this isomorphism clearly achieves the decomposition

$$S(1,1) = L^2(\mathbb{R}^3) \otimes S_{\text{c.m.}}(1,1). \tag{27}$$

We have now to introduce an interaction term $M_I$. Let us write it symbolically as

$$M_I = \lambda \int V^*(\mathbf{P}) \varphi(\mathbf{q}) N\theta(\mathbf{P},\mathbf{q}) \, d\mathbf{P} \, d\mathbf{q} + \text{h.c.}; \tag{28}$$

in terms of $\mathbf{p}_1$ and $\mathbf{p}_2$, we shall have

$$M_I = \lambda \int V^*(\mathbf{p}_1 + \mathbf{p}_2)$$

$$\times \varphi\left(\frac{1}{\omega}\left[\Omega_2\mathbf{p}_1 - \Omega_1\mathbf{p}_2 + \frac{\omega\Omega}{\mathbf{P}^2}\mathbf{P}\times(\mathbf{p}_1\times\mathbf{p}_2)\right]\right)$$

$$\times\left(\frac{\Omega\omega_1\omega_2}{\omega\Omega_1\Omega_2}\right)^{1/2} N(\mathbf{p}_1)\theta(\mathbf{p}_2)\, d\mathbf{p}_1, d\mathbf{p}_2 + \text{h.c.} \tag{29}$$

Thus, the total mass operator in $S_{c.m.}(1,1)$ will be defined on $\Psi = \Psi_1 \oplus \Psi_2(\mathbf{q})$ by

$$M\Psi = \left[m_3\Psi_1 + \lambda\int \varphi(\mathbf{q}')\Psi_2(\mathbf{q}')\, d\mathbf{q}'\right]$$

$$\oplus\, [\omega(\mathbf{q})\Psi_2(\mathbf{q}) + \lambda\Psi_1\bar{\varphi}(\mathbf{q})]. \tag{30}$$

Since $\omega(\mathbf{q}) \geqslant m_1 + m_2$, it is clear that $\langle\Psi, M\Psi\rangle > 0$ if $\Psi \neq 0$ provided $\lambda\, \|\varphi\|_2 < \min(m_1 + m_2, m_3)$. Moreover, $H_I$ is bounded so that $M$ will be self-adjoint. According to the theorem of the preceding section, $M$ is thus the mass operator of a representation $U$ of $\mathscr{P}$ in S(1,1).

We are now able to solve the "one-body problem": Let us suppose that $m_3 < m_1 + m_2$. $M_0$ has then an isolated eigenvalue and so will have $M$ for sufficiently small $\lambda$ because $M_I$ is bounded.[7] If we call this eigenvalue $m'_3$, we shall then have $m'_3 < m_1 + m_2$ and we can check that

$$m'_3 = m_3 + \lambda^2 \int \frac{|\varphi(\mathbf{q})|^2}{m'_3 - \omega(\mathbf{q})}\, d\mathbf{q} < m_3 \tag{31}$$

while the corresponding eigenvector is

$$\Phi = Z_\lambda\left[1 \oplus \frac{\bar{\varphi}(\mathbf{q})}{m'_3 - \omega(\mathbf{q})}\right] \tag{32}$$

with

$$Z_\lambda^2 = 1 + \int \frac{|\varphi(\mathbf{q})|^2}{[m'_3 - \omega(\mathbf{q})]^2}\, d\mathbf{q}. \tag{33}$$

As in the corresponding Galilean model,[8] it is possible to compute generalized eigenvectors $\Phi_{\mathbf{q}'}^{\pm}$ such that

$$M\,\Phi_{\mathbf{q}'}^{\pm} = \omega(\mathbf{q}')\,\Phi_{\mathbf{q}'}^{\pm}, \tag{34}$$

$\Phi$ and $\{\Phi_{\mathbf{q}'}^{\pm} \mid \mathbf{q}' \in \mathbb{R}^3\}$ being a complete set of generalized eigenvectors. It is then possible to prove that $M$ is unitarily isomorphic to the mass operator

$$M' = m'_3 \oplus \omega \tag{35}$$

of a direct sum of two free representations, the first one corresponding to a mass $m'_3$ and the other one to masses $m_1$ and $m_2$.

## 4. INTERACTION IN HIGHER SECTORS

We would now have to define an interaction in the higher sectors. Unfortunately, the work of Mutze[4] proves that this cannot be achieved in any reasonable way using our kind of approach. The problem is the next one: Suppose that we want to define the interaction in[9]

$$S(2,1) = S(1,1) \otimes L^2(R^3), \tag{36}$$

i.e., the space of two $N - 1$ $\theta$ or one $N - 1$ $V$ states. The more natural way to do it is

(i) First suppose that there is no interaction between the supplementary $N$ and the rest of the system and compute the mass operator corresponding to the previous interaction in S(1,1) plus a free $N$.

(ii) Then perform some kind of symmetrization on this mass operator to take the interaction with this $N$ into account.

But Mutze has proved that if we achieve the decomposition

$$S(2,1) = L^2(R^3) \otimes S_{c.m.}(2,1) \tag{37}$$

corresponding to free particles, the mass operator $M''$ obtained in step (i) cannot be tensorized in the same way unless the $S$ matrix in S(1,1) is the identity operator. Of course, $M''$ will be a mass operator since it corresponds to the tensor product (36) of two representations of $P$, but we are unable to tell if the operator we shall obtain in step (ii) will be the mass operator of some representation of $P$. Even worse, if there was both more than one $N$ and one $\theta$, there would be no cannonical way to perform the operation described in step (ii).

It seems thus that the problem of interaction in higher sectors cannot be given a satisfactory answer at the present stage.

## 5. CONCLUSION

Even if we were unable to extend the interaction to higher sectors, we have at least succeeded in defining a truly relativistic Lee model in S(1,1) and this has been achieved in a particularly simple way.

The asymptotic theory of this model is, of course, very limited. Nevertheless, it seems to us that it deserves to be carefully studied, since it is the first time that objects like Feynmann diagrams, etc., could be computed in a nontrivial relativistic model with finite renormalization.

## ACKNOWLEDGMENTS

This paper was written while the author was host of the Centre de Recherches Mathématiques de l'Université de Montréal, whose hospitality is greatly acknowledged. The author specially thanks Professor M. Perroud for several very useful discussions on the subject of this paper.

[1]N.N. Bogoliubov, A.A. Logunov, and I.T. Todosov, *Introduction to Axiomatic Quantum Field Theory* (Benjamin, New York, 1975).
[2]T.D. Lee, Phys. Rev. **95**, 1329 (1954).
[3]F. Coester, Helv. Phys. Acta **38**, 7 (1965).
[4]U. Mutze, J. Math. Phys. **19**, 231 (1978).
[5]F.R. Halpern, *Special Relativity and Quantum Mechanics* (Prentice-Hall, Englewood Cliffs, N.J., 1968).
[6]J. Dixmier, *Les algèbres d'opérateurs dans l'espace Hilbertien* (Gauthier-Villars, Paris, 1957).
[7]T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, New York, 1966).
[8]J.M. Levy-Leblond, Commun. Math. Phys. **4**, 157 (1967).
[9]We neglect the symmetrization for the sake of simplicity.

# Soliton solutions of nonlinear Dirac equations

K. Takahashi

*Department of Physics, Tohoku University, Sendai 980, Japan*
(Received 26 October 1978)

Dirac equations with fourth order self-couplings are investigated in one time and three space dimensions. Both stringlike and ball-like soliton solutions carrying nontopological quantum numbers are shown to exist, depending on the symmetries taken into account. The energies (per unit length) of stringlike solutions turn out to be always smaller than those of plane wave solutions. For ball-like solutions, it is shown that the region of values of the nontopological quantum numbers exists in which the total energies of solitons are smaller than those of plane waves.

## 1. INTRODUCTION

Soliton solutions of nonlinear field equations have been studied in detail in the last few years.[1] They have finite spatial extensions and finite energies and are regarded as candidates for models of elementary particles. These soliton solutions are classified into two categories: topological and nontopological solitons.[2,3] The stability of topological solitons is expected to some degree from their topological quantum numbers, while U(1) charges play an important role for the stability of nontopological solitons. We investigate nontopological solitons in this article.

Models are given by Lagrangians (1.1) or (3.1), which describe systems of self-interacting Dirac spinors. The space–time is four-dimensional. Analogous models in one time and one space dimension have already been studied[4] and exact solutions are known. Also, in four-dimensional space–time, exact solutions can be obtained by introducing a peculiar interaction.[5] In the present case, however, we shall have recourse to computer calculations at the final step of the investigation.

We start our discussions by considering plane wave solutions for the model

$$\mathscr{L}_1 = \bar\psi(i\partial\!\!\!/ - m)\psi + \frac{g^2}{2}(\bar\psi\psi)^2, \tag{1.1}$$

where $\psi$ is a four component Dirac spinor. Throughout this paper we are concerned with the case in which the force is attractive (i.e., $g^2 > 0$) when $\psi$ is regarded as a quantized Fermi field. The equation of motion is

$$(i\partial\!\!\!/ - m)\psi + g^2(\bar\psi\psi)\psi = 0. \tag{1.2}$$

The energy and charge density are given by

$$\mathscr{H}_1 = \bar\psi\Big(-i\gamma^i\frac{\partial}{\partial x^i} + m\Big)\psi - \frac{g^2}{2}(\bar\psi\psi)^2, \tag{1.3}$$

$$\rho = \psi^\dagger\psi. \tag{1.4}$$

The lowest energy plane wave of the form

$$\psi = e^{-i\omega t}\binom{u}{v}, \tag{1.5}$$

where $u$ and $v$ are constant two component spinors, satisfies (1.2) if

$$u^\dagger u = \frac{m-\omega}{g^2}, \quad v = 0. \tag{1.6}$$

Here we have confined the whole system into a large cube of volume $\Omega$ and imposed the periodic boundary condition. The total energy and the total charge are given by

$$E = \mathscr{H}_1\Omega = \frac{m^2-\omega^2}{2g^2}\Omega, \tag{1.7}$$

$$Q = \rho\Omega = \frac{m-\omega}{g^2}\Omega. \tag{1.8}$$

Fixing $Q$ and taking the limit $\Omega\to\infty$, we obtain the expected result

$$E = Qm. \tag{1.9}$$

Thus, for a given $Q$, the energy of a stable soliton must be smaller than $Qm$.

In Sec. 2, we shall see that there exist stringlike solutions for the U(1) invariant model (1.1). Given a charge $Q$ per unit length along the string, the energy per unit length is always smaller than $Qm$. In Sec. 3, ball-like solutions with finite spatial extensions in the three-dimensional space are found. The associated group is SO(3). Computer calculations in Sec. 4 show that there is a region of $Q$ in which the energies of solitons are smaller than $Qm$.

## 2. STRINGLIKE SOLUTIONS

### A. Equation of motion

In this section we solve Eq. (1.2) and seek cylindrically symmetric solutions, whose axis lies along the $x_3$ direction, under the ansatz

$$\psi(\mathbf{p},t) = e^{-i\nu mt + in\phi}\Big(i\frac{x\!\!\!/}{\rho}\psi_1(\rho) + \psi_2(\rho)\Big),$$

$$x\!\!\!/ \equiv -\sum_{i=1}^{2}\gamma^i x^i = \sum_{i=1}^{2}\gamma^i x_i, \tag{2.1}$$

where $\nu$ is a real number whose sign we set to be positive. $\phi$ is the azimuthal angle around the $x_3$ axis. In order that $\psi$ is single-valued, $n$ must be an integer. $\rho$ is the distance from the $x_3$ axis: $\rho = (x_1^2 + x_2^2)^{1/2}$. $\psi_1$ and $\psi_2$ are real functions of $\rho$ only.

 0022-2488/79/061232-07$01.00  

For $\gamma$ matrices, we adopt the notation of Bjorken and Drell.[6] One may recognize that the appearance of the phase factor $\exp in\phi$ in (2.5) is similar to the situation for the Higgs field in the Ginzburg–Landau type equation studied by Nielsen and Olesen.[2] However, we are interested in solutions which vanish at infinity so that no topological quantum number will emerge.

The substitution of (2.1) in Eq. (1.2) gives a complex coupled equation in $\psi_1$ and $\psi_2$. However, if $\psi_1$ and $\psi_2$ are eigenfunctions of $\gamma^0$ with the same eigenvalue, $-P$, Eq. (1.2) reduces to a simpler form owing to the equalities $\bar{\psi}_k\gamma^i\psi_l = 0$. In fact, requiring the coefficients of $\exp i(-\nu mt + n\phi)$ and $\exp i(-\nu mt + n\phi)\not{x}$ in Eq. (1.2) to vanish independently, we obtain a set of equations

$$\frac{d}{d\rho}\psi_1 + \frac{1}{\rho}(1 + n\Sigma_3)\psi_1 - m(1 + \nu P)\psi_2 + g^2(\bar{\psi}\psi)\psi_2 = 0, \tag{2.2}$$

$$\frac{d}{d\rho}\psi_2 - \frac{n}{\rho}\Sigma_3\psi_2 - m(1 - \nu P)\psi_1 + g^2(\bar{\psi}\psi)\psi_1 = 0. \tag{2.3}$$

Here $\Sigma_3 = (i/2)[\gamma^1,\gamma^2]$ and $\bar{\psi}\psi = -\bar{\psi}_1\psi_1 + \bar{\psi}_2\psi_2$.

The simultaneous presence of the second terms of (2.2) and (2.3) will, in general, hinder us from finding a solution nonsingular at $\rho = 0$ and nondivergent at $\rho = \infty$. Therefore, we have to impose as a condition either $n\Sigma_3\psi_1 = -\psi_1$ or $n = 0$. Since we see that the interchange of conditions $\{n\Sigma_3 = -1, P\} \leftrightarrow \{n = 0, -P\}$ simply means the interchange of $\psi_1$ and $\psi_2$, it is sufficient to treat the case $n\Sigma_3 = -1$ for both $P = +1$ and $P = -1$. For definiteness, we set $n = +1$. This means $\Sigma_3 = -1$ (note that $\psi_1$ and $\psi_2$ must have the same eigenvalue of $\Sigma_3$) and the forms of $\psi_k$ are

$$\psi_1(\rho) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ f(\rho) \end{bmatrix}, \quad \psi_2(\rho) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ h(\rho) \end{bmatrix}, \quad \text{for } P = +1, \tag{2.4}$$

$$\psi_1(\rho) = \begin{bmatrix} 0 \\ f(\rho) \\ 0 \\ 0 \end{bmatrix}, \quad \psi_2(\rho) = \begin{bmatrix} 0 \\ h(\rho) \\ 0 \\ 0 \end{bmatrix}, \quad \text{for } P = -1, \tag{2.5}$$

## B. Special limit

First, consider the case $P = +1$. Equations (2.2) and (2.3) can be rewritten in terms of $f$ and $h$ as

$$\frac{d}{d\rho}f - m(1 + \nu)h + g^2(f^2 - h^2)h = 0, \tag{2.6}$$

$$\frac{d}{d\rho}h + \frac{1}{\rho}h - m(1 - \nu)f + g^2(f^2 - h^2) = 0. \tag{2.7}$$

Since it is not manifest whether or not these equations have nontrivial solutions, we investigate this point by adopting the tactics used in Ref. 3, i.e., we assume that solutions of Eqs. (2.6) and (2.7), if any, approach the plane wave solution as $\xi = (m(1 - \nu))^{1/2} \to 0$. The following expansion in $\xi$ turns out to be of much use for our purpose:

$$f = \xi\alpha(\tau) + \ldots,$$

$$h = (\xi^2/\sqrt{2m})\beta(\tau) + \ldots, \tag{2.8}$$

where $\tau \equiv \sqrt{2m}\,\xi\rho$. Substituting (2.8) into (2.6), one finds

$$\frac{d\alpha}{d\tau} = \beta, \tag{2.9}$$

and from (2.7)

$$\frac{d\beta}{d\tau} + \frac{1}{\tau}\beta - \alpha + g^2\alpha^3 = 0. \tag{2.10}$$

From (2.9) and (2.10), one obtains a differential equation for $\alpha$,

$$\frac{d^2\alpha}{d\tau^2} + \frac{1}{\tau}\frac{d\alpha}{d\tau} - \alpha + g^2\alpha^3 = 0. \tag{2.11}$$

As mentioned in Ref. 3, Eq. (2.11) can be looked upon as describing the motion of a point particle in a potential

$$V = -\frac{1}{2}\alpha^2 + \frac{g^2}{4}\alpha^4, \tag{2.12}$$

and under the action of a frictional force

$$F_f = -\frac{1}{\tau}\frac{d\alpha}{d\tau}.$$

Soliton solution must satisfy the boundary conditions

$$\frac{d\alpha}{d\tau} = 0, \quad \text{at } \tau = 0 \tag{2.13}$$

and

$$\alpha \to 0, \quad \text{when } \tau \to \infty. \tag{2.14}$$

Note that (2.13) together with (2.9) requires $\beta$ to vanish at $\tau = 0$ and assures the regularity of $\psi$ given by (2.1). (2.14) may be necessary for the sake of Lorentz invariance of the vacuum.

An infinite number of solutions are known to exist, which are specified by the number of radial nodes. The solution with the lowest energy (per unit length) has no radial nodes. The total energy per unit length of the soliton is the integration of the Hamiltonian density:

$$E = \int d^2x\,\mathscr{H}_1$$

$$= Qm + \frac{\pi}{m}\xi\int_0^\infty \tau d\tau\left[\left(\frac{d\alpha}{d\tau}\right)^2 - \frac{g^2}{2}\alpha^4\right] + O(\xi^2), \tag{2.15}$$

where $Q$ is the total charge per unit length,

$$Q = \int d^2x\rho = \frac{\pi}{m}\int_0^\infty \tau d\tau\,\alpha^2 + O(\xi^2). \tag{2.16}$$

The second term on the right-hand side of Eq. (2.15) vanishes identically. This can be seen in the following way. First, note that Eq. (2.10) is just the condition that the quantity $A$ defined by

$$A = \int_0^\infty \tau d\tau\left[\frac{1}{2}\left(\frac{d\alpha}{d\tau}\right)^2 + \frac{1}{2}\alpha^2 - \frac{g^2}{4}\alpha^4\right] \tag{2.17}$$

should be stationary under small variations of $\alpha$,

$$\frac{\delta A}{\delta \alpha} = 0. \qquad (2.18)$$

Now, we scale $\alpha$: $\alpha(\tau) \to a^P \alpha(a\tau)$. Then (2.17) becomes

$$A(a) = a^{2p}K + a^{2p-2}L_1 - a^{4p-2}L_2. \qquad (2.19)$$

$K$, $L_1$, and $L_2$ are defined by

$$K = \frac{1}{2}\int_0^\infty \tau d\tau\left(\frac{d\alpha}{d\tau}\right)^2,$$

$$L_1 = \tfrac{1}{2}\int_0^\infty \tau d\tau \alpha^2, \quad L_2 = \frac{g^2}{4}\int_0^\infty \tau d\tau \alpha^4.$$

Since $A(a)$ is stationary at $a = 1$,

$$\left.\frac{dA}{da}\right|_{a=1} = 2p(K + L_1 - 2L_2) - 2(L_1 - L_2) = 0. \qquad (2.20)$$

Requiring (2.20) to hold for any $p$, we obtain the virial relation

$$K = L_1 = L_2. \qquad (2.21)$$

Thus, the above statement is proved. At any rate, we see that $E$ and $Q$ approach definite quantities. It is these observations that motivate further numerical calculations in the region of finite $\xi$. The results are given in Sec. 4.

Finally, we coment on the case $P = -1$. Equations (2.2) and (2.3) are rewritten as

$$\frac{d}{d\rho}f - m(1-\nu)h + g^2(-f^2 + h^2)h = 0, \qquad (2.22)$$

$$\frac{d}{d\rho}h + \frac{1}{\rho}h - m(1+\nu)f + g^2(-f^2 + h^2)f = 0. \qquad (2.23)$$

For $\xi \to 0_+$, nontrivial solutions may be obtained from presumptive expansions

$$f = (\xi^2/\sqrt{2m})\alpha(\tau) + \ldots, \quad h = \xi\beta(\tau) + \ldots. \qquad (2.24)$$

The equations for $\alpha$ and $\beta$ are

$$\frac{d\alpha}{d\tau} - \beta + g^2\beta^3 = 0, \qquad (2.25)$$

$$\frac{d\beta}{d\tau} + \frac{1}{\tau}\beta - \alpha = 0. \qquad (2.26)$$

A differential equation obtained from these equations is

$$\frac{d^2\beta}{d\tau^2} + \frac{1}{\tau}\frac{d\beta}{d\tau} - \frac{1}{\tau^2}\beta - \beta + g^2\beta^3 = 0. \qquad (2.27)$$

(2.27) is analogous to (2.11) except for the third term in the left-hand side. However, the existence of this term is essential. $\beta(0) = 0$ for the sake of regularity of the solutions. On the other hand, $(d\beta/d\tau)|_{\tau=0}$ need not vanish because singularities cancel each other in the combination

$$\frac{1}{\tau}\frac{d\beta}{d\tau} - \frac{1}{\tau^2}\beta.$$

Thus, a point particle starts from $\beta = 0$, the local maximum of the potential, with some initial velocity. It falls down and next proceeds to rise up the potential wall. After changing its direction of motion, it returns to $\beta = 0$ at $\tau = \infty$ if the magnitude of the initial velocity is appropriate. This solution has no nodes. The ones with nodes will appear if the initial velocity is varied. For each solution, $\alpha(0) = 0$ and $\alpha(\infty) = 0$. Contrary to the solutions for $P = -1$, $f$ is a small component of $\psi$ when $\xi$ is small, as can be seen from (2.24).

## 3. BALL-LIKE SOLUTIONS

In this section we investigate a model which yields ball-like soliton solutions. Consider the model invariant under the group SO(3) as well as U(1),

$$\mathcal{L}_2 = \bar{\psi}^a(i\partial\!\!\!/ - m)\psi^a + \frac{g^2}{2}(\bar{\psi}^a\psi^a)^2. \qquad (3.1)$$

The index $a$ runs from 1–3 and the summation over repeated index is implied. The equation of motion is

$$(i\partial\!\!\!/ - m)\psi^a + g^2(\bar{\psi}^b\psi^b)\psi^a = 0. \qquad (3.2)$$

The Hamiltonian density is

$$\mathcal{H}_2 = \bar{\psi}^a\left(-i\gamma^i\frac{\partial}{\partial x^i} + m\right)\psi^a - \frac{g^2}{2}(\bar{\psi}^a\psi^a)^2. \qquad (3.3)$$

Although we have seen that (3.2) will have stringlike solutions if two of $\psi^a$ identically vanish, we seek another type of solution by setting up an ansatz

$$\psi^a(\mathbf{r},t) = \frac{1}{\sqrt{3}}e^{-i\nu m t}\left(\gamma^a\psi_1(r) - 2i\frac{x^a}{r}\psi_2(r)\right.$$

$$\left. + i\gamma^a\frac{x\!\!\!/}{r}\psi_3(r) - \frac{x^a x\!\!\!/}{r^2}\psi_4(r)\right). \qquad (3.4)$$

Here $x\!\!\!/ \equiv -\Sigma_{i=1}^3 \gamma^i x^i$. $\psi_k$ $(k = 1\sim 4)$ are assumed to be real functions of $r$ only, $r = (x_1^2 + x_2^2 + x_3^2)^{1/2}$. $\nu$ is again a real positive parameter. We further assume that all $\psi_k$ have the same eigenvalue, $-P$, of $\gamma^0$ for the reason mentioned in Sec. 2. After substituting (3.4) into the left-hand side of (3.2), we set terms proportional to $\gamma^a$, $x^a$, $\gamma^a x\!\!\!/$, and $x^a x\!\!\!/$ equal to zero, respectively. Then we obtain a set of equations:

$$\frac{d}{dr}\psi_1 - \frac{1}{r}\psi_4 - m(1-\nu P)\psi_3 + g^2(\bar{\psi}^a\psi^a)\psi_3 = 0, \qquad (3.5)$$

$$\frac{d}{dr}\psi_1 + \frac{1}{2}\frac{d}{dr}\psi_4 + \frac{1}{2r}\psi_4 - m(1-\nu P)\psi_2 + g^2(\bar{\psi}^a\psi^a)\psi_2 = 0, \qquad (3.6)$$

$$\frac{d}{dr}\psi_3 + \frac{2}{r}\psi_2 - m(1+\nu P)\psi_1 + g^2(\bar{\psi}^a\psi^a)\psi_1 = 0, \qquad (3.7)$$

$$\frac{d}{dr}\psi_2 - \frac{d}{dr}\psi_3 - \frac{1}{r}\psi_2 + \frac{1}{r}\psi_3 + \frac{m}{2}(1+\nu P)\psi_4$$

$$- \frac{g^2}{2}(\bar{\psi}^a\psi^a)\psi_4 = 0, \qquad (3.8)$$

where

$$\bar{\psi}^a\psi^a = \bar{\psi}_1\psi_1 + \tfrac{4}{3}\bar{\psi}_2\psi_2 + \bar{\psi}_3\psi_3 - \tfrac{1}{3}\bar{\psi}_4\psi_4$$

$$- \tfrac{1}{3}\bar{\psi}_1\psi_4 - \tfrac{1}{3}\bar{\psi}_4\psi_1 - \tfrac{2}{3}\bar{\psi}_2\psi_3 - \tfrac{2}{3}\bar{\psi}_3\psi_2.$$

Although these equations are rather complicated to solve in a general way, we can simplify them greatly if we deal with the case in which $\psi_4$ does vanish. Then, from (3.8), toghether with the boundary conditions $\psi_2 = \psi_3 = 0$ at $r = \infty$, we have

$$\psi_3 = \psi_2, \quad \text{when } \psi_4 = 0. \tag{3.9}$$

From (3.9), we immediately recognize that Eqs. (3.5) and (3.6) become identical. In the end, two equations remain:

$$\frac{d}{dr}\psi_1 - m(1 + \nu P)\psi_2 + g^2(\bar{\psi}^a\psi^a)\psi_2 = 0, \tag{3.10}$$

$$\frac{d}{dr}\psi_2 + \frac{2}{r}\psi_2 - m(1 - \nu P)\psi_1 + g^2(\bar{\psi}^a\psi^a)\psi_1 = 0, \tag{3.11}$$

where $\bar{\psi}^a\psi^a = -\bar{\psi}_1\psi_1 + \bar{\psi}_2\psi_2$.

One may see that these equations resemble (2.2) and (2.3) with $\Sigma_3\psi_1 = -\psi_1$. The only difference is that the coefficient of $\psi_2/r$ in (3.11) is double that in (2.7). Of course, this is a reflection of the difference of the effective dimension of space. By an method similar to that presented in Sec. 2 B, we see the possibility of the existence of soliton solutions, whose total energy and total charge now diverge as $m(1 \pm \nu P) \equiv \xi^2 \to 0$. These observations will be confirmed in Sec. 4 by numerical calculations.

Before closing this section, let us consider the total angular momentum of the system

$$J_i = \int d^3x\, \psi^{a\dagger} \mathscr{J}_i \psi^a,$$

$$\mathscr{J}_i = \frac{i}{2}\epsilon_{ijk}\left(-x^j\frac{\partial}{\partial x^k} + \tfrac{1}{4}[\gamma^j,\gamma^k]\right) \equiv M_i + \tfrac{1}{2}\Sigma_i. \tag{3.12}$$

$M_i$ and $\frac{1}{2}\Sigma_i$ are the orbital and spin angular-momentum operator, respectively. With the form (2.4) for $\psi_1$ and $\psi_2$, we can show by straightforward calculations,

$$J_1 = J_2 = 0, \quad J_3 = \tfrac{1}{6}Q, \tag{3.13}$$

where $Q$ is the total U(1) charge $\int d^3x\, \psi^{a\dagger}\psi^a$. Now, the Sommerfeld's quantum condition is expressed in terms of $\psi^a$ and its canonically conjugate momentum $\pi^a = i\psi^{a\dagger}$ as

$$\frac{1}{2\pi}\oint dt \int d^3x\, \pi^a \frac{d\psi^a}{dt} = \int d^3x\, \psi^{a\dagger}\psi^a = n_a \tag{3.14}$$

($a$ is not summed).

$\oint dt$ stands for the integration over one period of time $2\pi/\nu m$. $n_a$ is an integer. On the other hand, since we can also show for our soliton solution that $n_1 = n_2 = n_3$, (3.14) and (3.13) mean that $Q = 3n$ ($n$ is an integer) and that the magnitude of the total angular momentum will take values of integer or half-integer.

## 4. NUMERICAL CALCULATIONS

We first solve equations which correspond to $P = +1$. The case of $P = -1$ can be treated in an analogous way mentioned below.

### A. Stringlike solutions

Equations to be solved are (2.6) and (2.7). The following redefinition of variables is convenient:

$$z = m\rho, \quad F = \frac{g}{\sqrt{m}}f, \quad H = \frac{g}{\sqrt{m}}h. \tag{4.1}$$

Equations (2.6) and (2.7) become

$$\frac{d}{dz}F - (1 + \nu)H + (F^2 - H^2)H = 0, \tag{4.2}$$

$$\frac{d}{dz}H + \frac{1}{z}H - (1 - \nu)F + (F^2 - H^2)F = 0. \tag{4.3}$$

The energy and charge per unit length are

$$E_s = 2\pi\int_0^\infty \rho d\rho \mathscr{H}_1 = \frac{2\pi}{g^2}\int_0^\infty zdz\Big[H\frac{dF}{dz} - F\frac{dH}{dz} - \tfrac{1}{2}FH + F^2 - H^2 - \tfrac{1}{2}(F^2 - H^2)^2\Big], \tag{4.4}$$

$$Q_s = 2\pi\int_0^\infty \rho d\rho \psi^\dagger\psi = \frac{2\pi}{mg^2}\int_0^\infty zdz(F^2 + H^2). \tag{4.5}$$

Note that Eqs. (4.2) and (4.3) can be derived from the variational principle $\delta L_s = 0$ with $\nu$ fixed, where

$$L_s = \frac{\pi}{g^2}\int_0^\infty zdz\Big(F\frac{dH}{dz} - H\frac{dF}{dz} + \frac{1}{z}FH - (1 - \nu)F^2 + (1 + \nu)H^2 + \tfrac{1}{2}(F^2 - H^2)^2\Big). \tag{4.6}$$

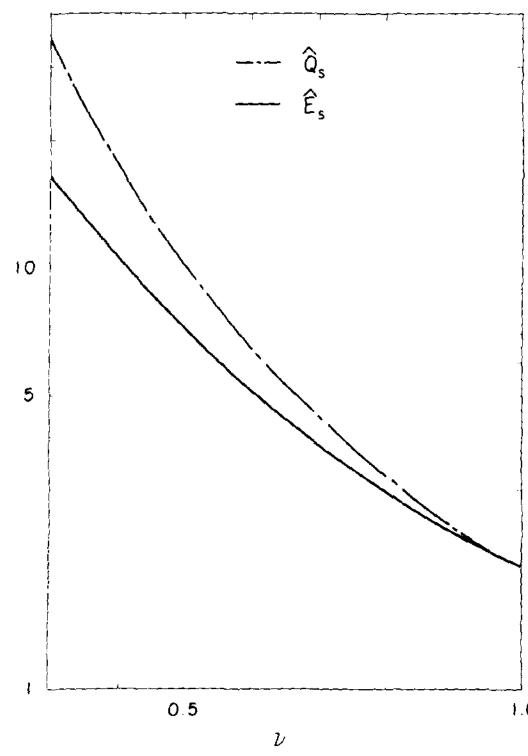


FIG. 1. $\hat{E}_s \equiv (g^2/\pi)E_s$ and $\hat{Q}_s \equiv (mg^2/\pi)Q_s$ for solutions of (2.6) and (2.7) ($P = +1$).

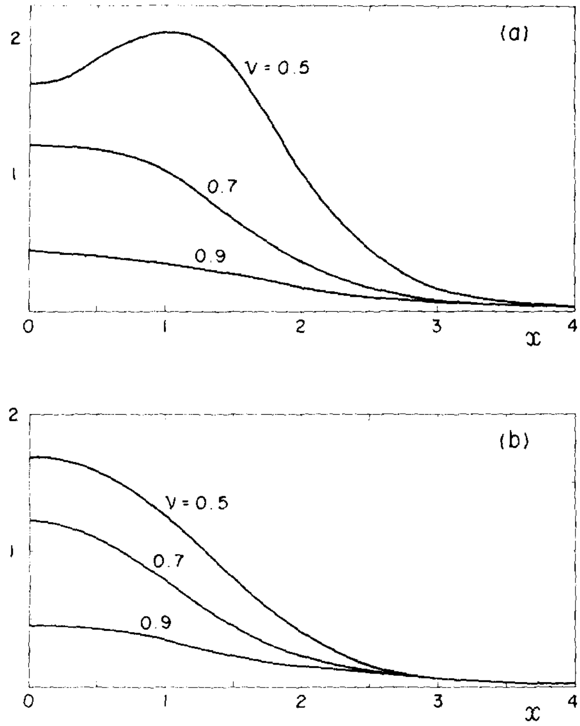


FIG. 2. (a) $(g^2/m)\psi^\dagger\psi$ and (b) $(g^2/m)\bar{\psi}\psi$ for $\nu = 0.5, 0.7$, and $0.9$ ($P = +1$).

By the same argument used in deriving (2.21), we have the virial relation

$$K^{(s)} = -2V_2^{(s)}, \quad V_1^{(s)} = 0, \tag{4.7}$$

where

$$K^{(s)} = \frac{\pi}{g^2}\int_0^\infty zdz\left(F\frac{dH}{dz} - H\frac{dF}{dz} + \frac{1}{z}FH\right),$$

$$V_1^{(s)} = \frac{\pi}{g^2}\int_0^\infty zdz[-(1-\nu)F^2 + (1+\nu)H^2],$$

$$V_2^{(s)} = \frac{\pi}{g^2}\int_0^\infty zdz\tfrac{1}{2}(F^2 - H^2)^2.$$

Thus, the energy (4.4) is expressed as

$$E_s = \nu m Q_s + 2V_2^{(s)}. \tag{4.8}$$

Now, in solving Eqs. (4.2) and (4.3), we seek solutions which have no radial node. We start with an initial value of $F$ at $z = 0$ and investigate numerically whether $F$ and $H$ approach zero at large $z$ or not. If they do and (4.7) is satisfied simultaneously with appropriate precision, we adopt them as solutions. Calculations were performed by the Runge–Kutta method with precision of about $10^{-4}$. In Fig. 1, we show the behaviors of $\widehat{E}_s \equiv (g^2/\pi)E_s$ and $\widehat{Q}_s \equiv (g^2/\pi)mQ_s$ vs. $\nu$. In the region $\nu < 1$, $\widehat{E}_s < \widehat{Q}_s$, and both $\widehat{E}_s$ and $\widehat{Q}_s$ approach the same value as $\nu \to 1$, as expected, which implies that our solutions have lower energies than that of plane wave solutions once $Q_s$ is fixed. We illustrate $(m/g^2)\psi^\dagger\psi$ and $(m/g^2)\bar{\psi}\psi$ for some values of $\nu$ in Fig. 2 (a) and (b). It is interesting to note that $\psi^\dagger\psi$ has a peripheral structure for small $\nu$ while $\bar{\psi}\psi$ is always central. For large $\nu$, the quantities $\psi^\dagger\psi$ and $\bar{\psi}\psi$ have similar behaviors.

## B. Ball-like solutions

The equations to be solved are (3.10) and (3.11). We again restrict ourselves to the case $P = +1$. Furthermore, we can choose $\psi_1$ and $\psi_2$ to be eigenfunctions of say, $\Sigma_3$ $\{\Sigma_i = (i/2)\epsilon^{ijk}[\gamma^j,\gamma^k]\}$ due to the invariance of (3.10) and (3.11) under the operation $\exp(i/2)\boldsymbol{\omega}\cdot\boldsymbol{\Sigma}$ on $\psi_k$. Let us set $\Sigma_3\psi_k = \psi_k$. Then $\psi_k$ are expressed in a similar manner to (2.4).

After changing variables

$$z = mr, \quad F = \frac{g}{\sqrt{m}}f, \quad H = \frac{g}{\sqrt{m}}h. \tag{4.9}$$

Equations (3.10) and (3.11) become

$$\frac{d}{dz}F - (1+\nu)H + (F^2 - H^2)H = 0, \tag{4.10}$$

$$\frac{d}{dz}H + \frac{2}{z}H - (1-\nu)F + (F^2 - H^2)F = 0. \tag{4.11}$$

The energy and charge are

$$E_B = 4\pi\int_0^\infty r^2dr\mathscr{H}_2 = \frac{4\pi}{mg^2}\int_0^\infty z^2dz\left(H\frac{dF}{dz} - F\frac{dH}{dz} - \frac{2}{z}F \times H + F^2 - H^2 - \tfrac{1}{2}(F^2 - H^2)^2\right), \tag{4.12}$$

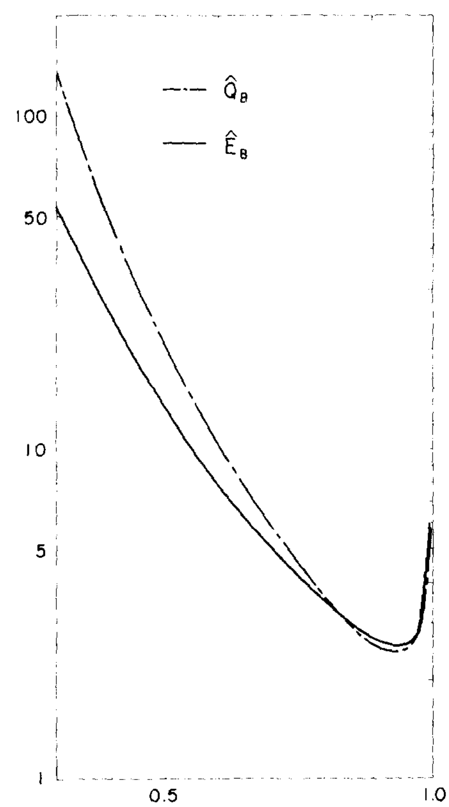


FIG. 3. $\widehat{E}_B \equiv (mg^2/6\pi)E_B$ and $\widehat{Q}_B \equiv (m^2g^2/6\pi)Q_B$ for solutions of (3.10) and (3.11) ($P = +1$).

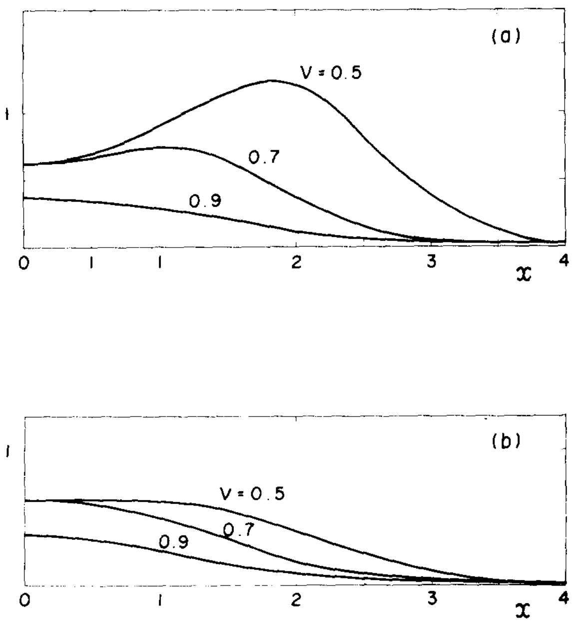


FIG. 4. (a) $(g^2/3m)\psi^{a\dagger}\psi^a$ and (b) $(g^2/3m)\bar{\psi}^a\psi^a$ for $\nu = 0.5, 0.7$, and 0.9 $(P = +1)$.

$$Q_B = 4\pi \int_0^\infty r^2 dr \psi^{a\dagger}\psi^a = \frac{4\pi}{m^2 g^2}\int_0^\infty z^2 dz (F^2 + H^2). \tag{4.13}$$

As in Sec. 4 A, it is useful to derive the virial relation for the

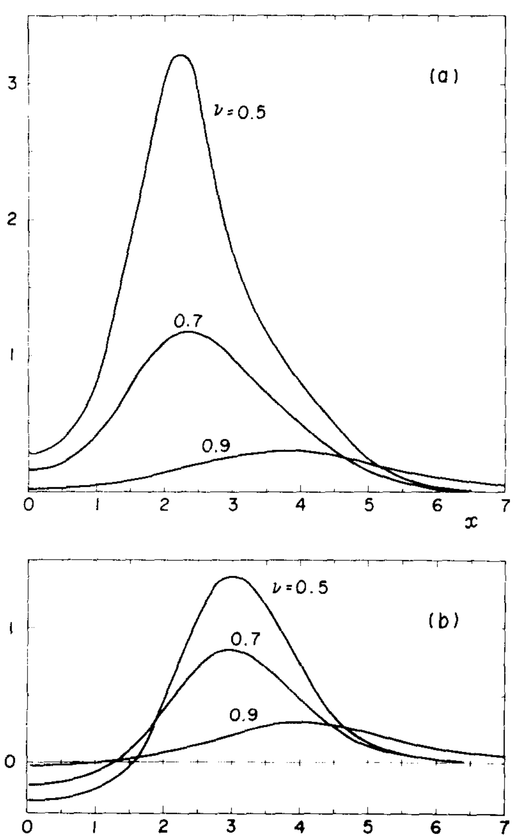


FIG. 5. $\widehat{E}_s \equiv (g^2/\pi)E_s$ and $\widehat{Q}_s \equiv (mg^2/\pi)Q_s$ for solutions of (2.23) and (2.24) $(P = -1)$.

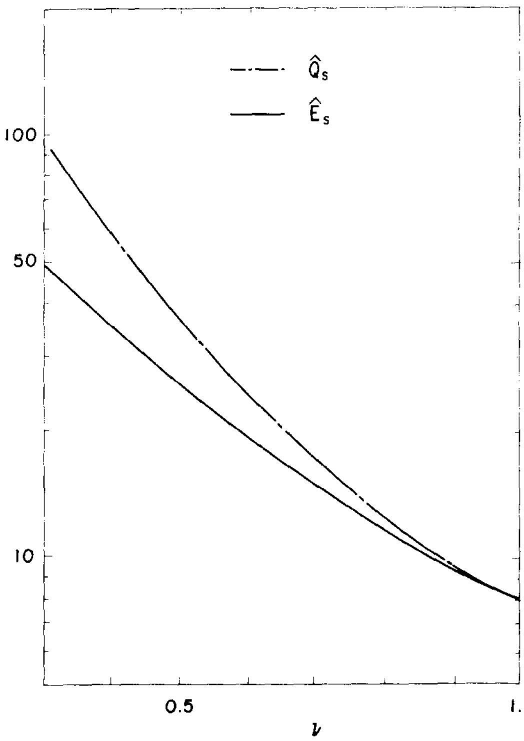


FIG. 6. (a) $(g^2/m)\psi^\dagger\psi$ and (b) $(g^2/m)\bar{\psi}\psi$ for $\nu = 0.5, 0.7$, and 0.9 $(P = -1)$.

present case. The variational principle which produces (4.10) and (4.11) is $\delta L_B = 0$ with $\nu$ fixed, where

$$L_B = \frac{2\pi}{mg^2}\int_0^\infty z^2 dz\Big(F\frac{dH}{dz} - H\frac{dF}{dz} + \frac{2}{z}FH - (1-\nu)F^2 + (1+\nu)H^2 + \tfrac{1}{2}(F^2 - H^2)^2\Big). \tag{4.14}$$

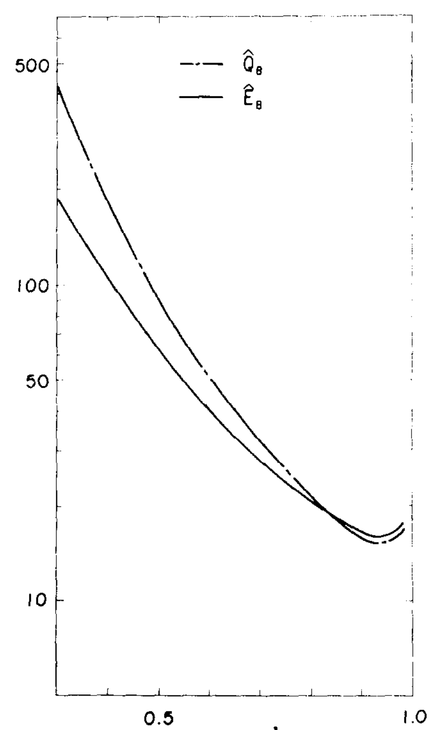


FIG. 7. $\widehat{E}_B \equiv (mg^2/6\pi)E_B$ and $\widehat{Q}_B \equiv (m^2g^2/6\pi)Q_B$ for solutions of (3.10) and (3.11) $(P = -1)$.

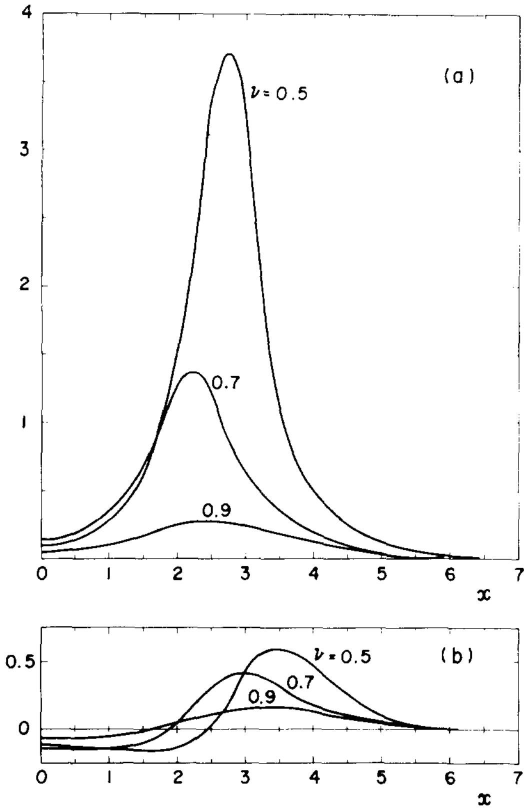


FIG. 8. (a) $(g^2/3m)\psi^{a\dagger}\psi^a$ and (b) $(g^2/3m)\bar{\psi}^a\psi^a$ for $\nu = 0.5, 0.7$, and $0.9$ ($P = -1$).

Define

$$K^{(B)} = \frac{2\pi}{mg^2}\int_0^\infty z^2 dz\left(F\frac{dH}{dz} - H\frac{dF}{dz} + \frac{2}{z}FH\right),$$

$$V_1^{(B)} = \frac{2\pi}{mg^2}\int_0^\infty z^2 dz\,[-(1-\nu)F^2 + (1+\nu)H^2],$$

$$V_2^{(B)} = \frac{2\pi}{mg^2}\int_0^\infty z^2 dz\,\tfrac{1}{2}(F^2 - H^2)^2.$$

The following relation holds for these quantities:

$$K^{(B)} = -3V_2^{(B)}, \quad V_1^{(B)} = V_2^{(B)}. \tag{4.15}$$

Therefore, the energy (4.12) is related to $Q_B$ by

$$E_B = \nu m Q_B + 2V_2^{(B)}. \tag{4.16}$$

The numerical calculations are performed in a similar way to that mentioned in Sec. 4 A. The criterion is now (4.15). In Fig. 3 the curve $\widehat{E}_B \equiv (mg^2/6\pi)E_B$ vs. $\nu$ and the curve $\widehat{Q}_B \equiv (m^2g^2/6\pi)Q_B$ vs $\nu$ are shown. They intersect at about $\nu = 0.82$. $\widehat{E}_B < \widehat{Q}_B$ for $\nu < 0.82$ and $\widehat{E}_B > \widehat{Q}_B$ for $\nu > 0.82$. These curves have minima and they occur at the same value of $\nu$, since $(dE_B/d\nu) = \nu m(dQ_B/d\nu)$.[3] From this result we recognize that soliton solutions for $\nu < 0.82$ are stable in the sense that their energies are smaller than that of plane wave solutions. The quantities $\psi^{a\dagger}\psi^a$ and $\bar{\psi}^a\psi^a$ are illustrated in Fig. 4 (a) and (b), respectively. $\psi^{a\dagger}\psi^a$ has again a peripheral structure for small $\nu$, whereas $\bar{\psi}^a\psi^a$ is always central.

The results of calculations for the case of $P = -1$ are depicted in Figs. 5 and 6 (stringlike solutions), and Figs 7 and 8 (ball-like solutions). Let us compare, as an example, Figs 3 and 7 for ball-like solutions. We see that in the common region of $\widehat{Q}_B$ ($\widehat{Q}_B < 15, \nu < 0.9$), $\widehat{E}_B(P = -1)$ is always larger than $\widehat{E}_B(P = +1)$, which means solitons of $P = -1$ are unstable. We see also that a similar situation occurs for stringlike solutions.

## ACKNOWLEDGMENTS

The author would like to express his thanks to G. Takeda and M. Yoshimura for discussions. He also thanks the High Energy Theory Group of Tohoku University for providing him with facilities for computer calculations.

[1] For a review on solitons, see R. Jackiw, Rev. Mod. Phys. **49**, 681 (1977); W. Marciano and H. Pagels, Phys. Rep. **36**, 137 (1978).
[2] H.B. Nielsen and P. Olesen, Nucl. Phys. B **61**, 45 (1973); G.'t Hooft, Nucl. Phys. B **79**, 276 (1974); A.M. Polyakov, JETP Lett. **20**, 194 (1974).
[3] R. Friedberg, T.D. Lee, and A. Sirlin, Phys. Rev. D **13**, 2739 (1976); Nucl. Phys. B **115**, 1, 32 (1976).
[4] S.J. Chang, S.D. Ellis, and B.W. Lee, Phys. Rev. D **11**, 3572 (1975); S.Y. Lee, T.K. Kuo, and A. Gavrielides, Phys. Rev. D **12**, 2249 (1975).
[5] J. Werle, Phys. Lett. B **71**, 357 (1977).
[6] J.D. Bjorken and S.D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).

# Conserved densities for nonlinear evolution equations. I. Even order case

L. Abellanas and A. Galindo

*Department of Theoretical Physics, Universidad Complutense de Madrid, Madrid-3, Spain*
(Received 4 August 1978)

This paper is a first attempt to analyze in detail the number and structure of nontrivial polynomial conserved densities for a nonlinear evolution equation $u_t = P$, $P$ an arbitrary polynomial in the spatial derivatives of $u$. Our attention is here focused on the even order case, where stronger conclusions can be derived. Several criteria for nonexistence of conserved densities are afforded. The coexistence of conserved densities is shown to severely restrict both the evolution equation and the functional structure of such densities. Finally for the case of second order equations the problem is completely worked out.

## I. INTRODUCTION

The discovery of some amazing properties of the Korteweg–de Vries (KdV) equation, with the soliton phenomena on one side[1] and the existence of an infinite number of conserved densities on the other,[2] has led to growing interest in the subject of the conservation laws for nonlinear partial differential equations, particularly of the evolution type. In a sense, the KdV equation seems rather exceptional; the simple consideration of the heat equation $u_t = u_{xx}$, for which the only nontrivial conserved density $\rho(u,u_x,\cdots)$ is $\rho = \lambda u$, leads one to suspect that the existence of infinitely many conserved polynomial quantities (with its physical implications) is expectedly a peculiar situation. This feeling however bears more on the extraordinary solitonlike properties of the KdV equation than on a profound knowledge of the role played by the conserved densities in the theory of nonlinear processes. A systematic analysis in that direction is still lacking.

Most of the efforts devoted lately to this question have dealt with the relationship between conserved currents and invariance properties of the equations. The revival of Lie methods[3] and their so called Lie–Bäcklund generalization[4] to include dependence of the transformations on derivatives has actually led to relevant results in this field. For instance, Ibragimov[5] has proved that any conserved current is noetherian for a suitable weak Larangian density. However this remarkable theorem does not seem useful enough from a practical viewpoint. For instance, the Lie invariance group of the KdV equation has only four parameters and yields four conservation laws. Even though the Ibragimov theorem and previous work by Kumei *et al.*[6] guarantee that all KdV conserved currents in the Gardner *et al.* series[2] can in principle be obtained through the Lie–Bäcklund method, the actual calculation becomes too complicated with increasing order of derivatives to be efficient except for a few densities.

To get further insight into the number and structure of the conserved densities for an evolution equation it might prove to be a natural path to attack frontally the problem. A previous analysis of the simplest situation, namely linear evolution equations, led us by means of the Gel'fand–Dikii[7] algebraic algorithm to the exhaustive enumeration of all their polynomial conserved densities.[8] This analysis revealed a striking difference between the odd and even cases, which suggests that we explore both instances separately.

The aim of this paper is to analyze in some detail the question of polynomial conserved densities for general evolution equations of even order. In Sec. II we provide necessary conditions for the existence of conserved densities, which are strong enough to limit drastically our search, first to evolution equations linear in the leading derivative, and second, to conserved densities depending essentially on a small fixed number of spatial derivatives. Section III analyzes in detail the remaining possibilites, and several criteria are given for nonexistence of conserved densities. The general form of evolution equations which conserve a given density is afforded in Sec. IV, which incidentally shows that the bound on the order of derivatives found in Sec. II is optimal. In Sec. V we prove that the coexistence of two inequivalent conserved densities imposes severe restrictions not only on the evolution equation but also on the functional structure of two such densities. Finally the case of second-order equations is completely analyzed in Sec. VI.

Work on the odd-order evolution equations is in progress and will be the subject of a forthcoming publication.

### A. Notation

Given a real-valued function $u = u(x,t)$, $x,t\in\mathbb{R}$, with partial derivatives $u_1\equiv\partial u/\partial x,\ldots,u_M\equiv\partial^M u/\partial x^M$, we shall denote by $\mathscr{F}_M$ the set of sufficiently smooth functions $F(u,u_1,\ldots,u_M)$. $\mathscr{F}$ will stand for the union $\cup_M\mathscr{F}_M$. Given a polynomial $P\in\mathscr{F}_M$, $d(P)$ will stand for its degree in the variable $u_M$.

Similarly

$$\mathscr{F}'_M\equiv\left\{F\in\mathscr{F}_M : F_{u_M}\equiv\frac{\partial F}{\partial u_M}\neq 0\right\},$$

$$\mathscr{F}''_M\equiv\left\{F\in\mathscr{F}_M : F_{u_M u_M}\equiv\frac{\partial^2 F}{\partial u_M^2}\neq 0\right\}.$$

A function $\rho\in\mathscr{F}_N$ such that $\rho(0,0,\ldots,0) = 0$, is called a conserved density for $u_t = F(u,\ldots,u_M)$, $F\in\mathscr{F}_M$, if $\rho_t \equiv(d/dt)\rho\equiv\Sigma_j\rho_{u_j}D^jF$ is a total derivative, i.e., if $\exists Q\in\mathscr{F}_{N-1}$ such that $\rho_t = DQ$. We shall write indistinctly $(d/dx)f\equiv Df\equiv f_1$ for any $f\in\mathscr{F}$.

 0022-2488/79/061239-05$01.00  

Two conserved densities $\rho, \hat{\rho}$ are said to be equivalent, in symbols $\rho \sim \hat{\rho}$, whenever $\rho - \hat{\rho}$ is trivial, i.e., $\rho - \hat{\rho} = Q_1$, for some $Q \in \mathscr{F}$. It is known that $\rho \sim \hat{\rho}$ iff

$$\frac{\delta \rho}{\delta u} = \frac{\delta \hat{\rho}}{\delta u}, \quad \text{where } \frac{\delta}{\delta u} \equiv \sum_{j=0}^{\infty} (-1)^j D^j \frac{\partial}{\partial u_j}.$$

Henceforth, assertions concerning existence or nonexistence of conserved densities should be understood modulo trivial densities. A somewhat different convention will be used through Sec. V.

Given $u_t = F$, $F \in \mathscr{F}$, we shall write $C_N(F)$ for the set of its nontrivial polynomial conserved densities $\rho \in \mathscr{F}''_N$, and $C(F) = \cup_N C_N(F)$.

## II. NONEXISTENCE OF CONSERVATION LAWS IN ARBITRARILY HIGH-ORDER DERIVATIVES

Let us consider the generic (polynomial) evolution equation of even order $M \geqslant 2$:

$$u_t = P(\cdots u_M) \tag{1}$$

with $P \in \mathscr{F}'_M$ a polynomial in the variables $u = u(x,t), \ldots, u_M \equiv \partial^M u / \partial x^M$. The aim of this section is to show that all conserved $\rho \in C_N(P)$ for Eq. (1) can only depend on derivatives of fixed bounded order. This is the content of the next theorem, which makes a strong difference with the well-known behavior of the (third-order) KdV equation.

*Theorem* 1: $\rho \in C_N(P) \Rightarrow N < M/2$

*Proof*: We must prove that every conserved $\rho \in \mathscr{F}_N$, where $N \geqslant M/2$, is equivalent to some $\hat{\rho} \in \mathscr{F}_{N-1}$ of lower order in the derivatives. First of all we observe that

$$\rho_{u_j} P_j \sim (-1)^{M/2 - N + j} (\rho_{u_j})_{M/2 - N + j} P_{N - M/2},$$
$$\forall j \ni N - M/2 \leqslant j \leqslant N.$$

Thus we have

$$\rho_t = \rho_u P + \rho_{u_1} P_1 + \cdots + \rho_{u_N} P_N$$

$$\sim \sum_{0}^{N - M/2 - 1} \rho_{u_j} P_j + \Big[ \sum_{N - M/2}^{N} (-1)^{M/2 - N + j}$$

$$\times (\rho_{u_j})_{M/2 - N + j} \Big] P_{N - M/2}. \tag{2}$$

Now if $\rho$ is to be conserved by (1), it is necessary that (2) equals a total derivative. Hence the derivative of maximal order in (2), i.e., $u_{N + M/2}$, can only appear linearly. Therefore, the coefficient in (2) of $u^2_{N + M/2}$ must be zero, namely $\rho_{u_N u_N} P_{u_M} = 0$. So we conclude $\rho_{u_N u_N} = 0$ (by hypothesis $P_{u_M} \neq 0$). Integrating by parts we can write $\rho \sim \hat{\rho}(\cdots u_{N-1})$.
(Q.E.D)

A similar argument allows us to show that, generically speaking, equation (1) will not have any nontrivial conservation law. Indeed, the existence of $\rho \in C_N(P)$, $N < M/2$, would require $\rho_t \sim (\delta \rho / \delta u) P$ to be a total derivative, let us say

$$\frac{\delta \rho}{\delta u} P = Q_1, \tag{3}$$

where $Q \in \mathscr{F}'_{M-1}$, since $\rho \in \mathscr{F}_N \Rightarrow \delta\rho/\delta u \in \mathscr{F}_{2N}$, and $2N < M$.

If we compare the powers of $u_M$ in both members of (3) we arrive at the following conclusion:

*Theorem* 2: $P_{u_M u_M} \neq 0 \Rightarrow \nexists$ (nontrivial) conserved $\rho$ for (1).

Therefore, in order to avoid the obstructions pointed out by these two theorems, we must restrict in the sequel our attention to those polynomials $P \in \mathscr{F}'_M - \mathscr{F}''_M$ and to densities $\rho \in C_N(P)$ with $N < M/2$.

It may be interesting to note that both theorems are optimal in the sense that there exist equations (1) conserving densities $\rho \in C_{M/2-1}(P)$, and on the other hand one can exhibit examples of $P$ linear in $u_M$ for which conserved densities do really exist. For instance we will show later (Sec. IV) that the equation $u_t = (u^2 + 2u_2)u_4 + (4u_3 + 1)(u_3 + uu_1)$ has $\rho = u_1^2 - u^3/3$ as a conserved density.

## III. SOME PRACTICAL CRITERIA FOR THE EXISTENCE OF CONSERVED DENSITIES

Let us now analyze in some detail the case of equations (1) with $P = a(\cdots u_{M-1})u_M + b(\cdots u_{M-1})$, with $a, b \in \mathscr{F}_{M-1}$. We are primarily interested in deriving some simple necessary conditions on the form of $P$ for the existence of a conserved $\rho(\cdots u_N)$, $N < M/2$.

As we already know, conservation of $\rho$ implies the existence of $Q \in \mathscr{F}'_{M-1}$ such that $Q_1/P = \delta\rho/\delta u$. Since $Q_1/P$ is in the range of the variational derivative, it must satisfy[9]:

$$K_\lambda(Q_1/P) = 0, \quad \forall \lambda \in \mathbb{R}, \tag{4}$$

where

$$K_\lambda \equiv \sum_{j=0}^{\infty} [\lambda^j - (-1)^j (D + \lambda)^j] \partial_{u_j}.$$

But as $\delta\rho/\delta u$ depends only on $u, \cdots, u_{M-2}$ the summation in $K_\lambda(\delta\rho/\delta u)$ will run from $j = 0$ to $j = M - 2$. A trivial calculation yields

$$K_\lambda\left(\frac{Q_1}{P}\right) = \lambda^{M-3} [2\partial_{u_{M-3}} - (M-2) D \partial_{u_{M-2}}] \frac{Q_1}{P}$$

$$+ O(\lambda^{M-4}). \tag{5}$$

Thus we derive the following necessary conditions on $f(\cdots u_{M-2}) \equiv Q_1/P = \delta\rho/\delta u$:

$$f_{u_M} = 0 \Rightarrow Q_{u_{M-1}} = af, \tag{6}$$

$$f_{u_{M-1}} = 0 \Rightarrow Q_{u_{M-2}} = fP_{u_{M-1}} - (af)_1, \tag{7}$$

$$2f_{u_{M-3}} = (M-2) Df_{u_{M-2}}. \tag{8}$$

If we differentiate (7) with respect to $u_{M-1}$, we get

$$fP_{u_{M-1} u_{M-1}} = 2(fa)_{u_{M-2}} + (fa_{u_{M-1}})_1. \tag{9}$$

Finally from (8) it follows that

$$\left(\frac{M}{2} - 1\right) [P(Q_{u_{M-2}})_2 - P_1(Q_{u_{M-2}})_1 - P_1 Q_{u_{M-3}}$$

$$+ P(Q_{u_{M-3}})_1 - P_{u_{M-2}}Q_2 - Q_1(P_{u_{M-2}})_1 + 2fP_1P_{u_{M-2}}]$$
$$= P(Q_{u_{M-3}})_1 + PQ_{u_{M-4}} - Q_1P_{u_{M-3}}. \qquad (10)$$

The coefficient of $u_{M+1}$ in both sides of equation (10) reproduces (9). To obtain new information we look at the coefficient of $u_M^4$, which gives the condition:

$$a_{u_{M-1}}[fa_{u_{M-2}} - af_{u_{M-2}}] = faa_{u_{M-1},u_{M-2}}. \qquad (11)$$

Another useful piece of information can be obtained by reexpressing $P$ in the form

$$P = au_M + b = A_1 + B, \qquad (12)$$

where $A \equiv \int_0^{u_{M-1}} a\, du_{M-1}$ and $B \equiv P - A_1$. Clearly $A$, $B$ are polynomials in $u,\ldots,u_{M-1}$. With this notation we have

$$\rho_t \sim \frac{\delta\rho}{\delta u}P \sim \frac{\delta\rho}{\delta u}B - \left(\frac{\delta\rho}{\delta u}\right)_1 A. \qquad (13)$$

Writing out the conservation condition for $\rho$, i.e, $\delta\rho_t/\delta u = 0$, the vanishing of the coefficient of $u_{2M-2}$ leads to

$$\left(\ln\frac{\delta\rho}{\delta u}\right)_1 = \frac{B_{u_{M-1},u_{M-1}}}{A_{u_{M-1},u_{M-1}}}, \quad \forall\ \rho(\cdots u_{M/2-2}). \qquad (14)$$

An alternative formulation of the conservation of $\rho$ in this context is to require

$$\left(\frac{\delta\rho}{\delta u}\right)B - \left(\frac{\delta\rho}{\delta u}\right)_1 A = Q_1,\ \ Q(\cdots u_{M-2}). \qquad (15)$$

Comparing the dominant powers of $u_{M-1}$ in both sides of (15) yields the following two consequences:

$$\text{(a)}\ \exists\rho\in C_{M/2-1}(P) \Rightarrow d(B) = d(A) + 1, \qquad (16)$$

$$\text{(b)}\ \exists\rho(\cdots u_{M/2-2}) \Rightarrow \text{either}\ \ 0\neq d(B) = d(A)\ \ \text{or}\ \ d(A) = 1,\ d(B) = 0.$$

To close this section we list below some elementary criteria which follows immediately from the above stated conditions. In practice they can be of great utility in order to rule out the existence of conserved densities for a given equation of type (1).

### A. Criteria

(1) $\left.\begin{matrix} a(\cdots u_{M-3}) \\ b_{u_{M-1},u_{M-1}} \neq 0\end{matrix}\right\} \Rightarrow \nexists\rho(\cdots u_{M/2-2})$,

(2) $a_{u_{M-1}}a_{u_{M-2}} \neq aa_{u_{M-1},u_{M-2}} \Rightarrow \nexists\rho(\cdots u_{M/2-2})$,

(3) $a_{u_{M-1}} \neq 0 = a_{u_{M-2}} \Rightarrow \forall\rho \sim \rho(\cdots u_{M/2-2})$,

(4) $A_{u_{M-1},u_{M-1}} \neq 0 = B_{u_{M-1},u_{M-1}} \Rightarrow \forall\rho(\cdots u_{M/2-2}) \sim \lambda u$,

(5) $A_{u_{M-1},u_{M-1}} = 0 \neq B_{u_{M-1},u_{M-1}} \Rightarrow \nexists\rho(\cdots u_{M/2-2})$,

(6) $0 \neq d(B) \neq d(A) \Rightarrow \nexists\rho(\cdots u_{M/2-2})$

$d(B) \neq d(A) + 1 \Rightarrow \forall\rho \sim \rho(\cdots u_{M/2-2})$

$0 \neq d(B) - d(A) \neq \pm 1 \Rightarrow \nexists\rho.$

The reader may check that they are obvious consequences of (9), (11), (11), (14), (14), and (16), respectively.

*Examples*:

(i) $\nexists$ conserved $\rho$ for $u_t = u_4 + u_3^3$ [see (6)],

(ii) The only $\rho$ conserved by $u_t = u_3u_4$ are of the form $\rho(u)$ [see (3)],

(iii) $\nexists\rho(u,u_1)$ conserved by the equations $u_t = u_6 + u_5^2$ [see (1)], $u_t = (u_4 + u_5)u_6$ [see (2)], $u_t = u_5(u_1 + u_6)$ [see (4)] or $u_t = u_4u_6 + uu_5^2$ [see (5)].

(iv) The only $\rho(u)$ conserved by $u_t = u_3(u_4 + 1)$ are of the form $\lambda u$ [see (4)].

## IV. GENERAL FORM OF THE EQUATIONS WHICH CONSERVE A GIVEN $\rho$

Let us consider again the decomposition (12), and define $F \equiv B - [\ln(\delta\rho/\delta u)]_1 A$. Then we have $\rho_t \sim (\delta\rho/\delta u)F$. In other words, $\rho$ is a conserved density for the associated equation $u_t = F(\cdots u_{M-1})$. Hence we deduce the following:

*Theorem* 3: Let $\rho\in C_s(F)$, $F\in\mathscr{F}_r$, $s < M/2$, $r < M$. Then $\rho\in C_s(G)$, where

$$G(u,\cdots,u_M) \equiv F + A_1 + \left(\ln\frac{\delta\rho}{\delta u}\right)_1 A, \quad \forall A\in\mathscr{F}'_{M-1}. \qquad (17)$$

Conversely each equation $u_t = G$ conserving $\rho$ is of the form (17).

*Corollary*: Let $M$ be an even integer $\geqslant 2$. Then there exists some $G\in\mathscr{F}'_M$ such that (1) admits a conserved $\rho\in C_{M/2-1}(G)$.

*Proof*: For $M > 2$ it is a consequence of the well-known series of conservation laws of the Korteweg-deVries equation. The result is obvious for $M = 2$. (Q.E.D.)

*Examples*: $u_t = (u^2 + 2u_2)u_4 + (u_3 + uu_1)(4u_3 + 1)$ is a 4th-order equation which conserves $\rho = u_1^2 - u^3/3$. It is obtained by applying the algorithm proposed in Theorem 3 with the choices $F = u_3 + uu_1$ (KdV), $\rho = u_1^2 - u^3/3$ (conserved by KdV equation) and $A = (u^2 + 2u_2)u_3$.

*Corollary*: Let $\rho(\cdots u_s)$, $s < M/2$, $M$ even, be given. Then the most general $M$th-order evolution equation conserving $\rho$ has the form

$$u_t = \rho\left[\ln\left(f\frac{\delta\rho}{\delta u}\right)\right]_1, \quad f\in\mathscr{F}'_{M-1}. \qquad (18)$$

*Proof*: It is sufficient to write $F(\delta\rho/\delta u) = Q_1$ (conservation of $\rho$ under $u_t = F$). The function $f \equiv A + [Q/(\delta\rho/\delta u)]$ leads to the form (18). (Q.E.D.)

For instance the equation of the previous example can be expressed in the form (18) with $f = (u^2 + 2u_2)(u_3 + \frac{1}{4}) + \lambda/(u^2 + 2u_2)$, $\lambda\in\mathbb{R}$ arbitrary. Indeed $\delta\rho/\delta u = -(u^2 + 2u_2)$ in this case.

## V. UPPER BOUNDS TO THE NUMBER OF (INEQUIVALENT) CONSERVATION DENSITIES FOR $M$ EVEN $\geqslant 4$

As we already know from Sec. III, the existence of at least one (nontrivial) conserved $\rho$ imposes severe restrictions on the form of the polynomial $P$ in (1). Hence we expect that the existence of two or more conservations laws for (1) will be compatible with a very narrow family of polynomials $P$. Throughout this section we restrict ourselves to analyze the case of only two simultaneous conserved densities.

More explicitly we are going to investigate some necessary conditions on $P$ for (1) to admit two conserved densities, say $\rho$ and $\bar{\rho}$, inequivalent in the sense that $\bar{\rho} \not\sim \lambda\rho$, $\lambda \in \mathbb{R}$. Thus let us assume that there exist $F, G \in \mathscr{F}'_{M-1}$, such that

$$P = \frac{F_1}{\delta\rho/\delta u} = \frac{G_1}{\delta\bar{\rho}/\delta u}. \tag{19}$$

The function $\alpha(\cdots u_{M-2}) \equiv (\delta\bar{\rho}/\delta u)/(\delta\rho/\delta u)$ ($\neq$ constant by hypothesis) satisfies the equation

$$G_1 = \alpha F_1. \tag{20}$$

If we define $g(\cdots u_{M-2}) \equiv \alpha F - G$, it follows from (20) that

$$\alpha_1 F = g_1. \tag{21}$$

Comparing the terms in $u_{M-1}$ in both sides of (21) we obtain

$$\alpha_{u_{M-2}} = 0, \quad \text{i.e.,} \quad \alpha = \alpha(\cdot\cdot u_{M-3}) \tag{22}$$

and moreover $F$ has to be linear in $u_{M-1}$. Thus we have the following:

*Proposition* 1: $\exists\, \rho, \bar{\rho}$ inequivalent $\in C(P) \Rightarrow P_{u_M u_{M-1}} = 0$.

That is a way of saying that whenever $P_{u_M u_{M-1}} \neq 0$ in (1), there exists at most one conserved density (essentially unique up to a constant factor). In the remainder of this section we implicitly assume $P_{u_M u_{M-1}} = 0$.

*Remark*: For the sake of brevity we simply say "$\exists \rho, \bar{\rho}$" to mean the existence of a pair of inequivalent densities in the sense explained above.

Let us decompose $F = S(\cdots u_{M-2})u_{M-1} + T(\cdots u_{M-2})$, $P = a(\cdots u_{M-2})u_M + b(\cdots u_{M-1})$. Then the condition $F_1 = P(\delta\rho/\delta u)$ reads

$$a\frac{\delta\rho}{\delta u} = S, \tag{23}$$

$$b\frac{\delta\rho}{\delta u} = S_{u_{M-2}}u_{M-1}^2 + T_{u_{M-2}}u_{M-1} + H(\cdot\cdot u_{M-2}). \tag{24}$$

*Proposition* 2:

$$\left\{\begin{matrix}\exists\rho(\cdots u_{M/2-1}) \in C_{M/2-1}(P), \bar{\rho} \in C(P) \\ P_{u_M u_{M-2}} = 0\end{matrix}\right\} \Rightarrow P_{u_{M-1}u_{M-1}} \neq 0.$$

*Proof*: Since $\delta\rho/\delta u$ is linear in $u_{M-2}$, with nonvanishing coefficient, it follows from (23) that $S_{u_{M-2}} \neq 0$. On the other hand (24) shows that in consequence $P_{u_{M-1}u_{M-1}}$ must be different from zero. (Q.E.D.)

*Proposition* 3:

(i)

$$a_{u_{M-2}} \neq 0 = b_{u_{M-1}u_{M-1}} \Rightarrow \nexists\rho \in C_{M/2-1}(P), \quad \bar{\rho} \in C(P),$$

(ii)

$$a_{u_{M-2}} = 0 \neq b_{u_{M-1}u_{M-1}} \Rightarrow \nexists\rho \in C_{M/2-1}(P), \quad \bar{\rho} \in C(P).$$

*Proof*: In both cases the values of $S_{u_{M-2}}$ derived from (23), (24) are contradictory. (Q.E.D.)

Turning back to (22), if we use the explicit form

$$\alpha = \frac{\bar{\rho}_{u_{M/2-1}u_{M/2-1}}u_{M-2} + \bar{R}}{\rho_{u_{M/2-1}u_{M/2-1}}u_{M-2} + R} \tag{25}$$

we get at once

$$\frac{\bar{\rho}_{u_{M/2-1}u_{M/2-1}}}{\rho_{u_{M/2-1}u_{M/2-1}}} = \frac{\bar{R}}{R} = \alpha \tag{26}$$

from which we conclude that

$$\alpha \in \mathscr{F}_{M/2-1}. \tag{27}$$

For $M > 4$ the condition (27) is stronger than (22).

On the other hand, given a density $\rho(\cdots u_{M/2-2})$ we have $\alpha = R^{-1}(\bar{\rho}_{u_{M/2-1}u_{M/2-1}} + \bar{R})$. Therefore, (22) requires $\bar{\rho}_{u_{M/2-1}u_{M/2-1}}$ to be zero. Hence it follows

*Proposition* 4: $\exists\, \rho(\cdots u_{M/2-2}), \bar{\rho} \Rightarrow \bar{\rho} \sim \hat{\rho}(\cdots u_{M/2-2})$.

In other words, there are just two possibilities: either $C_{M/2-1}(P) = \emptyset$ or $C_{M/2-1}(P) = C(P)$.

The criteria previously established are valid for $M$ even $\geqslant 4$. The next two results, however, hold only for $M$ even $\geqslant 6$. Let us start by noting that

$$\frac{\delta\rho}{\delta u} - [\rho_{u_n u_n}u_{2n} + n(\rho_{u_n u_n})_1 u_{2n-1}] \in \mathscr{F}_{2n-2},$$
$$\forall\rho(\cdots u_n), n \geqslant 3$$

Using this expression we may write

$$\alpha = \frac{\bar{\rho}_{u_{M/2-1}u_{M/2-1}}u_{M-2} + (M/2-1)(\bar{\rho}_{u_{M/2-1}u_{M/2-1}})_1 u_{M-3} + \bar{\theta}}{\rho_{u_{M/2-1}u_{M/2-1}}u_{M-2} + (M/2-1)(\rho_{u_{M/2-1}u_{M/2-1}})_1 u_{M-3} + \theta}$$

Thus, if $M \geqslant 8$, it follows from (26) and (27) that

$$\bar{\rho}_{u_{M/2-1}u_{M/2-1}} = \lambda\rho_{u_{M/2-1}u_{M/2-1}}, \quad \lambda \in \mathbb{R}.$$

*Proposition* 5:

$$\left\{\begin{matrix}\exists\rho, \bar{\rho} \\ M \ \text{even} \geqslant 6\end{matrix}\right\} \Rightarrow \bar{\rho} - \lambda\rho \in \mathscr{F}_{M/2-2} \ \text{for some} \ \lambda \in \mathbb{R}.$$

The case $M = 6$ is handled by direct calculation [use $\alpha_{u_3} = 0$ according to (27)].

Since the set $C(P)$ of conserved densities for (1) is a linear space, the last two propositions imply together:

*Corollary*:

$$\left\{\begin{matrix}\exists\rho \in C_{M/2-1}(P), \bar{\rho} \\ M \ \text{even} \geqslant 6\end{matrix}\right\} \Rightarrow \bar{\rho} = \lambda\rho \quad \text{for some} \ \lambda \in \mathbb{R}.$$

We remark that this result does not admit generalization to $M \geqslant 4$, as it can be seen in the case of $u_t = u_2 u_4 + 2u_3^2$, which conserves $\rho = u_1^2$ and $\bar{\rho} = u_1^3$.

## VI. MORE DETAILS ON THE SECOND-ORDER EVOLUTION EQUATIONS

Let us consider in more detail the case $M = 2$,

$$u_t = a(u, u_1)u_2 + b(u, u_1). \tag{28}$$

We know from Theorem 1 that every $\rho$ conserved by equation (28) is of the form $\rho = \rho(u)$. Furthermore the conservation of $\rho(u)$ requires $(\delta/\delta u)(\rho' P) = 0$, where $\rho' \equiv \partial\rho/\partial u$. More explicitly $\rho$ is conserved if and only if

$$u_2[\rho'(2a + a_{u_1}u_1)]_u - \rho' b_{u_1u_1}u_2 + u_1^2[a\rho']_{uu} + [\rho'(b - u_1 b_{u_1})]_u = 0. \quad (29)$$

Separating the $u_2$ terms from the rest this is equivalent to the following pair of equations:

$$(\rho'[2a + a_{u_1}u_1])_u = \rho' b_{u_1u_1}, \quad (30)$$

$$u_1^2[\rho' a]_{uu} = [\rho'(u_1 b_{u_1} - b)]_u. \quad (31)$$

Integration of (31) yields $u_1^2(\rho' a)_u = \rho'(u_1 b_{u_1} - b) + h(u_1)$, $h$ arbitrary. Now we differentiate this equality with respect to $u_1$ and use (30) to conclude $h(u_1) = \alpha$ (constant). In other words the pair (30)–(31) is equivalent to the single equation

$$u_1^2(\rho' a)_u = \rho'(u_1 b_{u_1} - b) + \alpha. \quad (32)$$

Let $a(u,u_1) = \Sigma a_n(u)u_1^n$, $b(u,u_1) = \Sigma b_n(u)u_1^n$ be the explicit polynomial forms of $a,b$. Making these subsititutions in (32) yields:

$$b_0(u)\rho'(u) = \alpha, \quad (33)$$

$$(a_n\rho')_u = (n+1)b_{n+2}\rho'. \quad (34)$$

Since $a$, $b$, $\rho$ are polynomials, (33) implies $\rho = \lambda u$ whenever $\alpha \neq 0$. Moreover in this case it follows from (32) that

$$u_1^2 a_u + b - u_1 b_{u_1} = \text{constant} \ (\neq 0). \quad (35)$$

Conversely, if $a$, $b$ do not verify (35) then $\alpha = 0$. Therefore the only possible situations for a given $P = a(u,u_1)u_2 + b(u,u_1)$ are:

(i) $P$ satisfies (35) $\Rightarrow$ Every conserved $\rho$ is of the form $\rho = \lambda u$, $\lambda \in \mathbb{R}$ (conserved if and only if $P$ is an exact derivative).

(ii) $P$ does not satisfy (35) $\Rightarrow$ Every conserved $\rho$ must be a solution of the system:

$$(\rho' a_n)_u = (n+1)b_{n+2}\rho', \quad \forall n \geqslant 0. \quad (36)$$

Moreover $b_0$ has to be zero, see (33).

Generically speaking, equation (28) does not admit (polynomial) conservation laws, because of the obstruction imposed by (36).

To close this section we want to answer the following question: given a function $\rho(u)$, what is the most general equation (28) which has $\rho$ as a conserved density? Since the special case $\rho = \lambda u$, $\lambda \in \mathbb{R}$, has already been solved in (i), we only deal here with those $\rho$ such that $\rho'' \neq 0$.

Let $N/D$ be the irreducible fractional form of $\rho''/\rho'$. From equations (36) it follows that

$$D[(n+1)b_{n+2} - a_n'] = N a_n, \quad \forall n \geqslant 0, \quad (37)$$

where $a_n' \equiv \partial_u a_n$. But $N$, $D$ are (by definition) mutually prime polynomials. Hence there exist polynomials $q_n$ such that $a_n = Dq_n$, and in consequence

$$b_{n+2} = [1/(n+1)][Nq_n + (Dq_n)_u].$$

Finally, then, the most general equation (28) conserving $\rho$ is of the form:

$$u_t = Du_2 \sum_0^\infty q_n(u)u_1^n + \sum_0^\infty \frac{u_1^{n+2}}{n+1}[Nq_n + (Dq_n)_u], \quad (38)$$

where $q_n(u)$ are arbitrary polynomials in $u$.

*Example*: Let us take $\rho = u^3(u-1)^3$. Then $N = 10u^2 - 10u + 2$, $D = u(u-1)(2u-1)$. The choice $q_0 = 1$, $q_n = 0$, $\forall n > 0$ leads to the equation

$$u_t = u(u-1)(2u-1)u_2 + (16u^2 - 16u + 3)u_1^2$$

which conserves $\rho$.

## ACKNOWLEDGMENTS

The authors are grateful to Professor G. García Alcaine, Professor F. Guil, and Professor L. Martínez for many discussions and useful suggestions. The financial support of the Instituto de Estudios Nucleares, J.E.N., is also acknowledged.

[1]A.C. Scott, F.Y.F. Chu, and D. McLaughlin, Proc. IEEE **61**, 1443–83 (1973).
[2]R.M. Miura, C.S. Gardner, and M.D. Kruskal, J. Math. Phys. **9**, 1204–09 (1968).
[3]L.V. Ovsjannikov: *Group Properties of Differential Equations* (Novosibirsk, Moscow, 1962); G.W. Bluman and J.D. Cole, *Similarity Methods for Differential Equations* (Springer Verlag, Berlin, 1974).
[4]R.L. Anderson, S. Kumei, and C.E. Wulfman, Phys. Rev. Lett. **28**, 988 (1972); Rev. Mex Fis. **21**, 1, 35 (1972); J. Math. Phys. **14**, 1527 (1973).
[5]N.H. Ibragimov, Lett. Math. Phys. **1**, 423–28 (1977).
[6]S. Kumei, J. Math. Phys. **18**, 256 (1977).
[7]I.M. Gel'fand and L.A. Dikii, Russ. Math. Surveys **30**, 77–113 (1975).
[8]L. Abellanas and A. Galindo, Lett. Math. Phys. **2**, 399–404 (1978).
[9]A. Galindo (to be published in Anal. Física).

# Integral representation for the dimensionally regularized massive Feynman amplitude

M. C. Bergère and F. David

*Physique Théorique CNRS, Cen-Saclay, Boite Postale No. 2, 91190 Gif-S/-Yvette, France*

(Received 6 June 1978)

A convergent integral representation for the massive Feynman amplitude in complex dimension $D$ is defined away from Re$D$ equal to some rationals. The Feynman integrand is modified according to the technique of Zimmermann's forests but each subgraph is subtracted at zero external momenta and zero internal masses; the order of the Taylor subtractions depends upon Re$D$. The so defined regularized Feynman amplitude is a meromorphic function of $D$ with multiple poles at some rationals, which satisfies field equations and Ward identities. These amplitudes may be used in the construction of a bare Lagrangian field theory.

## I. INTRODUCTION

The existence of a dimensionally regularized massive Feynman amplitude has been proved by many authors.[1-9] This kind of regularization is used extensively to study various properties of gauge field theories and appears to be of practical interest in critical phenomena. A great advantage of this regularization is that it is the only one which preserves the field equations and satisfies trivially the symmetries of the system (at the exception of the well-known anomalies)[10]; this property is due to an adequate definition of the so-called "$D$-dimensional covariants," i.e., the analytic continuation in $D$ of the scalar products and of the tensorial contractions over Lorentz indices.[7,9,11] Such regularized amplitudes are meromorphic functions of $D$ [1,4] and are defined as analytic continuation in $D$ of an integral which converges for Re$D$ less than the first ultraviolet pole $D_0$.

The purpose of this paper is to define an absolutely convergent integral representation for the dimensionally regularized massive Feynman amplitude and for any complex dimension $D$ away from Re$D$ equal to the poles. Such a representation can be used directly to calculate any bare amplitude for Re$D > D_0$ (away from the uv poles) without having to isolate the complete structure at the poles, that is, to perform explicitly the analytic continuation in $D$.

This convergent integral representation is obtained by applying a new subtraction $R$ operator upon the integrand. The subtractions are performed in a similar way than for BPHZ subtractions, but the main difference is that the subgraphs are subtracted not only at zero external momentum but also at zero internal masses; of course, the number of subtractions depends upon the value of $D$.

As expected, field equations and Ward identities, preserved by dimensional regularization are not violated by this subtraction procedure; this is due to the fact that external momentum and internal masses are treated on the same footing. Although it is beyond the scope of this paper let us mention as an application of our result, the definition, for the dimensionally renormalized amplitude, of an absolutely convergent integral representation at $D = 4$. Such a result defines for any graph a compact computable form for the so-called 't Hooft[3] minimal renormalization which is known to preserve field equations and Ward identities.

To introduce the reader to our purpose, let us mention the well-known example of the Euler $\Gamma(x)$ function. The Euler $\Gamma(x)$ function is defined through an integral representation

$$\Gamma(x) = \int_0^\infty d\lambda\, \lambda^{x-1} e^{-\lambda}. \tag{I.1}$$

The integral (I.1) is absolutely convergent for Re$x > 0$. There, we may split the integral into two parts and write

$$\Gamma(x) = \int_1^\infty d\lambda\, \lambda^{x-1} e^{-\lambda} + \int_0^1 d\lambda\, \lambda^{x-1}(e^{-\lambda} - 1) + \frac{1}{x}. \tag{I.2}$$

Away from the value $x = 0$, the representation (I.2) analytically continue the function $\Gamma(x)$ up to Re$x > -1$. Moreover, for $-1 < \text{Re}x < 0$, we may use

$$\int_1^\infty d\lambda\, \lambda^{x-1} = \frac{-1}{x}, \tag{I.3}$$

to write an integral representation of $\Gamma(x)$ for $-1 < \text{Re}x < 0$,

$$\Gamma(x) = \int_0^\infty d\lambda\, \lambda^{x-1}(e^{-\lambda} - 1). \tag{I.4}$$

More generally, for $-(n+1) < \text{Re}(x) < -n$ $(n = 0,1,2,\cdots)$, the integral representation of $\Gamma(x)$ is

$$\Gamma(x) = \int_0^\infty d\lambda\, \lambda^{x-1}(1 - T_\lambda^n) e^{-\lambda}, \tag{I.5}$$

where $T_\lambda^n$ is the subtraction Taylor operator of order $n$, around $\lambda = 0$. In this publication, we define a subtraction operator $R$ which does for a massive Feynman amplitude as a function of the dimension $D$ what $(1 - T_\lambda^n)$ does for the Euler $\Gamma(x)$ function.

The end of this introduction is devoted to notations and definitions. In Sec. II, we establish the meromorphy in the dimension $D$ of a Feynman amplitude. In Sec. III, we introduce a set of subtraction operators and we study their properties in Sec. IV. Finally, the integral representation for the dimensionally regularized Feynman amplitude is defined in Sec. V. Everything is performed in Euclidean space although it may be extended without difficulty to Minkowsky space via a Wick rotation.

Given a Feynman graph $G$ with $d_a$ derivative couplings $\{k_a^{\mu_1},\ldots,k_a^{\mu_{d_a}}\}$ on each line $a$ ($\mu$ are the vector indices), we

 0022-2488/79/061244-12$01.00  

associate in dimension $N = 4$, a Schwinger integral of the type

$$\int_0^\infty \prod_{a=1}^{l} dk_a \left(\exp - \sum_{a=1}^{l} \alpha_a m_a^2\right) \times \prod_{a=1}^{l} \prod_{i=1}^{da} \left(\frac{-1}{\alpha_a^{1/2}} \frac{\partial}{\partial z_a^{\mu_i}}\right) Z_G(\alpha_a, p_i, z_a)|_{z_a=0}, \tag{I.6}$$

where $l$ is the number of lines of the graph $G$ and the function $Z_G$ is

$$Z_G(\alpha,p,z_a) = [P_G(\alpha)]^{-N/2} \exp\left(\sum_{a=1}^{l} \frac{z_a^2}{4}\right) \times\left[\exp - \sum_{i,j=1}^{n-1}\left(p_i + \sum_{a=1}^{l} \frac{\epsilon_{ia} z_a}{2\sqrt{\alpha_a}}\right) \times [d_G^{-1}(\alpha)]_{ij}\left(p_j + \sum_{b=1}^{l} \frac{\epsilon_{jb} z_b}{2\sqrt{\alpha_b}}\right)\right] \tag{I.7}$$

In (I.7), the incidence matrix $\epsilon_{ia}$ and the Symanzik functions $P_G(\alpha)$ and $[d_G^{-1}(\alpha)]_{ij}$ are characteristic of the topology of the graph. A quadrivector $z_a$ is attached to each line $a$ in such a way that the derivative couplings may be taken care of by performing the corresponding $\partial/\partial z$ derivatives at $z$ equal to zero. A Feynman amplitude is obtained by associating $\gamma$ and internal group matrices with expression (I.6), by contracting some Lorentz indices (for instance $\not\partial$ or $\Box$) and internal indices (trace of products of internal or $\gamma$ matrices), and by summing over different combinations of derivative couplings. In such a way, we obtain an integral of the form

$$I_G(p_i,m_a,\tau) = \int_0^\infty \prod_{a=1}^{l} d\alpha_a \exp\left(-\sum_{a=1}^{l} \alpha_a m_a^2\right) S_G(p_i,\tau,\alpha_a) \times \exp\{-p_i[d_G^{-1}(\alpha)]_{ij} p_j\} P_G(\alpha)^{-N/2}, \tag{I.8}$$

where $\tau$ remains the external dependence in Lorentz and internal group indices. The function $S_G$ depends explicitly of the dimension $N = 4$ of the space because of the presence of external variable like $p_i$, of the Lorentz indices in $\tau$ and implicitly by the various contractions over Lorentz indices ($g^\mu{}_\mu = 4$ for instance), and by the different relations between the $\gamma$ matrices (see Appendix A).

We remind the reader in Sec. II that absolute convergence of (I.8) around the variables $\alpha \sim 0$ (ultraviolet divergences) lies in the behavior of $P_G(\alpha)$. A regulator may be introduced in (I.8) by replacing $[P_G(\alpha)]^{-N/2}$ by $P_G(\alpha)]^{-D/2}$, where $D$ is complex and Re$D$ small enough to ensure the convergence of(I.8). There is no need to change anything to the function $S_G$ to ensure convergence. However, we may do so in several ways provided that at $D = 4$ we recover the original function $S_G$. The delicate problem at that stage is to know whether or not one could define such a function $S_G(D)$ without breaking the field equations and their consequences: Ward identities. This problem was solved in the past by several authors[7,9,11]: We discuss briefly in Appendix A how to construct such a function $S_G(D)$. We thus define the analytic continuation of $I_G(p_i,m_a,\tau,D)$:

$$I_G(p_i,m_a,\tau,D) = \int_0^\infty \prod_{a=1}^{l} d\alpha_a \exp\left(-\sum_{a=1}^{l} \alpha_a m_a^2\right) S_G(p_i,\tau_i,\alpha_a,D) \times \exp\{-p_i[d_G^{-1}(\alpha)]_{ij} p_j\} P_G(\alpha)^{-D/2}. \tag{I.9}$$

## II. MEROMORPHIC STRUCTURE OF THE DIMENSIONALLY REGULARIZED FEYNMAN AMPLITUDE

To show the absolute convergence of (I.9), it is convenient to decompose the domain of integration ($\alpha_a \geqslant 0$) into $l!$. Hepp's sectors[12] defined as

$$0 \leqslant \alpha_{a_1} \leqslant \cdots \leqslant \alpha_{a_l} \tag{II.1}$$

where $(a_1,\ldots,a_l)$ is an ordering of the $l$ lines of $G$. We perform the change of variables

$$\alpha_{a_i} = \prod_{j=i}^{l} \beta_j^2, \tag{II.2a}$$

$$d\alpha_{a_i} = \prod_{j=1+1}^{l} \beta_j^2 \cdot 2\beta_i d\beta_i, \tag{II.2b}$$

so that the contribution of this sector $\sigma$ to integral (I.9) becomes[13]

$$I_G^\sigma(p_i,m,\tau,D) = 2^l \int_0^\infty d\beta_l \int_0^1 \prod_{i=1}^{l-1} d\beta_i \times \prod_{i=1}^{l} \beta_i^{2l(R_i) - DL(R_i) - d(R_i) - 1} \frac{S_G(p,\tau,\beta_i^2,D)}{[1+Q(\beta_i^2)]^{D/2}} \times\left\{\exp - \beta_l^2\left[m_{a_l}^2 + \sum_{j=1}^{l-1} m_{a_j}^2 \beta_j^2 \cdots \beta_{l-1}^2 + \frac{p_i R_{ij}(\beta^2) p_j}{1+Q(\beta^2)}\right]\right\}, \tag{II.3}$$

where $R_{ij}(\beta^2)$ and $Q(\beta^2)$ are nonnegative polynomials of $\beta_1^2,\ldots,\beta_{l-1}^2$, $S_G(p,\tau,\beta_i^2,D)$ is a continuous function of $\beta_1,\ldots,\beta_l$, polynomially bounded in $\beta_l$ when $\beta_l \to \infty$ and has a simultaneous Taylor series in the $\beta$'s around $\beta_i = 0$. The subgraphs $R_i$ are defined by the lines $\{a_1,\ldots,a_i\}$ and have $l(R_i)(=i)$ lines, $L(R_i)$ independent loops. In (II.3) we use the homogeneity property of the function $S_G(p,\tau,\alpha_a,D)$ as expressed in Appendix B. The quantity $d(R_i)$ is taken to be (Appendix B)

$$d(R_i) = 0 \quad \text{if } L(R_i) = 0,$$

$$d(R_i) = 2E\left(\frac{\delta(R_i)}{2}\right) \quad \text{if } L(R_i) \neq 0, \tag{II.4}$$

where $\delta$ is the maximum number of derivative couplings on the lines of connected one particle irreducible parts of $R_i$ and

$E(x)$ means integer part of $x$ smaller or equal to Re$x$. We note that for a scalar amplitude, $d(R_i)$ is null and $S_G$ is equal to 1.

Provided that $m^2_{a_i} > 0$, the integral (II.3) is convergent when $\beta_l \to +\infty$. The convergence at $\beta_i \sim 0$ is ensured if

$$\mathrm{Re}[DL(R_i) + d(R_i) - 2l(R_i)] < 0 \quad \forall\ i \in \{1, l\}. \quad \text{(II.5)}$$

Let us note $D^\sigma_0$:

$$D^\sigma_0 = \inf_{R_i : L(R_i) \neq 0} \left\{ \frac{2l(R_i) - d(R_i)}{L(R_i)} \right\}. \quad \text{(II.6)}$$

In every compact domain $\mathscr{D} \subset \{D : \mathrm{Re}D < D^\sigma_0\}$, the integrand of (II.3) can easily be bounded in modulus by an integrable function independent of $D$. Since the integrand is an analytic function of $D$, the integral (II.3) is proved to be an analytic function of $D$ in every domain $\mathscr{D}$; hence in the domain

$$\mathscr{D}^\sigma = \{D : \mathrm{Re}D < D^\sigma_0\}. \quad \text{(II.7)}$$

We now prove that the function $I^\sigma_G(p,m,\tau,D)$ so defined in $\mathscr{D}^\sigma$ can be analytically continued in a meromorphic function of $D$,[9] extensively noted $I^\sigma_G(p,m,\tau,D)$, in the whole complex plane, with poles only at some rationals greater or equal to $D^\sigma_0$. We introduce the operator

$$\mathbb{1} = \sum_{I \subset \{1,\ldots,l\}} \prod_{i \in I} (T^{n_i}_{\beta_i^2}) \prod_{j \notin I} (1 - T^{n_j}_{\beta_j^2}), \quad \text{(II.8)}$$

(where $I$ also contains the empty element) which acts upon the function

$$F(p,\beta_i,D) = \frac{S_G(p,\tau,\beta^2_i,D)}{[1 + Q(\beta^2_i)]^{D/2}} \times \exp\left\{ -\beta^2_l \left[ \sum_{j=1}^{l-1} m^2_a \beta^2_j \cdots \beta^2_{l-1} + \frac{p_i R_{ij}(\beta^2) p_j}{1 + Q(\beta^2)} \right] \right\}, \quad \text{(II.9)}$$

inside the integral (II.3). In (II.8), $T^{n_i}_{\beta^2_i}$ is the Taylor operator of degree $n_i$. The integrand of (II.3) is decomposed in

$$\sum_{I \subset \{1,l\}} \sum_{q_i = 0}^{n_i} \prod_{i \in I} \beta_i^{2l(R_i) + 2q_i - DL(R_i) - d(R_i) - 1} \times \prod_{j \notin I} \beta_j^{2l(R_j) + 2n_j - DL(R_j) - d(R_j) + 1} e^{-\beta^2_l m^2_{a_l}} F_{\{q_i,n_j\}}(p,\tau,\beta^2_j,D), \quad \text{(II.10)}$$

where the $F_{\{q_i,n_j\}}$'s are continuous functions of $\{\beta_j ; j \notin I\}$ and satisfy the same properties as $S_G(p,\tau,\beta^2_j,D)$.

The integration over the $\beta_i$'s $(i \in I)$ defines several meromorphic functions of $D$ in $\mathbb{C}$ with poles like

$$\tfrac{1}{2} m_{a_l}^{DL(G) + d(G) - 2q_l - 2l(G)} \times \Gamma\left( \frac{2l(G) + 2q_l - DL(G) - d(G)}{2} \right)$$

if $l \in I$, (II.11)

and

$$[2l(R_i) + 2p_i - d(R_i) - DL(R_i)]^{-1} \quad \text{for } i \in I,\ i \neq l. \quad \text{(II.12)}$$

The integration over the $\beta_j$'s $(j \notin I)$ is possible provided that

$$\mathrm{Re}[DL(R_j) + d(R_j) - 2l(R_j) - 2n_j - 2] < 0 \quad \forall\ j \notin I, \quad \text{(II.13)}$$

and gives, as shown for the integral (II.3), an analytic function of $D$ in the domain

$$\mathrm{Re}D < \inf_{\substack{j \notin I \\ L(R_j) \neq 0}} \left\{ \frac{2l(R_j) + 2n_j + 2 - d(R_j)}{L(R_j)} \right\}. \quad \text{(II.14)}$$

By taking the $n_j$'s great enough, one could obtain explicitly an analytic continuation of $I^\sigma_G(p,m,\tau,D)$ for every complex $D$.

So $I^\sigma_G(p,m,\tau,D)$ is proved to be a meromorphic function of $D$ in the complex plane, with poles only at rationals of the form

$$D = \frac{2l(R_i) - d(R_i) + 2p}{L(R_i)}, \quad p \in \mathbb{N},\ L(R_i) \neq 0. \quad \text{(II.15)}$$

If we sum the contributions of each sector $\sigma$, we deduce finally, since each subgraph $\varphi \subset G$ appears in at least one sector $\sigma$, that the integral $I_G(p,m,\tau,D)$ defined by (I.9) is absolutely convergent and defines an analytic function of $D$ in the domain

$$\mathscr{D} = \{D : \mathrm{Re}D < D_0\}, \quad \text{(II.16)}$$

where

$$D_0 = \inf_{\substack{\varphi \subset G \\ L(\varphi) \neq 0}} \left\{ \frac{2l(\varphi) - d(\varphi)}{L(\varphi)} \right\}; \quad \text{(II.17)}$$

this function defines by analytic continuation a meromorphic function of $D$, which is noted $I_G(p,m,\tau,D)$, in the complex plane, with poles only at rationals of the form

$$D = \frac{2l(\varphi) - d(\varphi) + 2p}{L(\varphi)}, \quad \text{(II.18)}$$

where $\varphi$ is any subgraph of $G$ with $L(\varphi) \neq 0$, and $p$ is a positive integer. We prove in the next section that in fact only connected, one particle irreducible (1PI) subgraphs $\varphi$ give poles of the form (II.18), so that the other poles found here must cancel out between the various Hepp's sectors.

## III. THE $\mu$-SUBTRACTED AMPLITUDE

In this section we define a class of subtracted amplitudes which are well defined for any dimension $D$. It may be shown that such subtractions satisfy Bogoliubov. Parasiuk scheme of renormalization.[12,14] The introduction of this class of subtracted amplitudes allows us in Sec. V to define the integral representation for the dimensionally regularized amplitude as a limit.

*Definition*: We define the $\mu$-subtracted amplitude as

$$I^\mu_G(p_i,m_a,\tau,D) = \int_0^\infty \prod_{a=1}^{l} d\alpha_a \exp\left( -\sum_{a=1}^{l} \alpha_a \mu^2_a \right)$$

$$\times\left(\left[\exp - \sum_{a=1}^{l} \alpha_a(m_a^2 - \mu_a^2)\right] S_G(p_i\tau,\alpha_a,D)\right.$$

$$\times \exp\{-p_i[d_G^{-1}(\alpha)]_{ij}p_j\}\times [P_G(\alpha)]^{-D/2}. \qquad \text{(III.1)}$$

The subtraction operator $R$ is defined in Ref. 15 as

$$R = \prod_{\varphi\subseteq G} (1 - \tau_\varphi^{-2l(\varphi)}), \qquad \text{(III.2)}$$

where the product over $\varphi$ runs over the $(2^l - 1)$ subgraphs of $G$. The operator $\tau$ is a generalized Taylor operator defined as follows: given a function $f(x)$ such that $x^{-\nu}f(x)$ (where $\nu$ may be complex) is infinitely differentiable at $x = 0$, then we define $\tau_x^n$ as

$$\tau_x^n f(x) = x^{-\lambda-\epsilon} T_x^{n+\lambda}\{x^{\lambda+\epsilon}f(x)\}, \qquad \text{(III.3)}$$

where $\lambda \geqslant -E'(\nu)$ is an integer, $E'(\nu)$ is the smallest integer $\geqslant \text{Re}\nu$, and $\epsilon = E'(\nu) - \nu$. The operator $T_x$ is the usual Taylor operator at $x = 0$, and the above definition is $\lambda$ independent. We also define for a given subgraph $\varphi$

$$\tau_\varphi^n f(\alpha) = [\tau_\rho^n f(\alpha)]_{\alpha_a = \rho\alpha_a\ \forall\ a\in\varphi}\big|_{\rho=1}. \qquad \text{(III.4)}$$

Although the $\tau$ operators do not commute (overlapping subgraphs), they do commute if the graphs are disjoint or nested, and it has been shown[15] that the product in (III.2) is independent of the order of application of the $\tau$ operators. In fact, it has been shown that

$$R = \left[1 + \sum_{\mathscr{F}} \prod_{\varphi\in\mathscr{F}} (-\tau_\varphi^{-2l(\varphi)})\right], \qquad \text{(III.5)}$$

where the sum runs over all forests of connected, one particle irreducible subgraphs (even if they are not generalized vertices).

The presence of the subtraction operator $R$ in (III.1) makes the integral absolutely convergent for any $D$ when the $\alpha$'s$\sim 0$, and whatever product of $\tau$ operators is applied, there is always an exponential damping (provided that $m_a$ and $\mu_a$ are different from zero) which ensures the convergence of (III.1) for large $\alpha$'s (see Appendix C). In momentum space representation, such an operator $R$ would subtract the subgraphs at zero external momentum and at $m_a^2 = \mu_a^2$. It is important to note that the function $I_G^\mu$ is not continuous in $D$. Let us find out the nature of the discontinuities. Given a forest $\mathscr{F} = \{\varphi\}$, if we dilate $\alpha_a$ by

$$\alpha_a \to \alpha_a \prod_{\varphi\ni a} \rho_\varphi^2, \qquad \text{(III.6)}$$

we get

$$P_G(\alpha_a) \to \prod_\varphi \rho_\varphi^{2L(\varphi)} \widetilde{P}_G(\alpha_a,\rho_\varphi), \qquad \text{(III.7)}$$

where $\widetilde{P}_G(\alpha_a,0)\neq 0$. Then, from (III.3)

$$\prod_{\varphi\in\mathscr{F}} (-\tau_\varphi^{-2l(\varphi)}) \left\{\exp\left[-\sum_{a=1}^{l} \alpha_a(m_a^2-\mu_a^2)\right]\right.$$

$$\left.\times S_G(p_i,\tau,\alpha_a,D) e^{-p_i[d_G^{-1}(\alpha)]_{ij}p_j} P_G(\alpha)^{-D/2}\right\} \qquad \text{(III.8)}$$

is equal to

$$\prod_{\varphi\in\mathscr{F}} (-\tau_{(\varphi)}^{-2l+e(DL(\varphi))}) \{\alpha_a,\rho_\varphi^2\}\big|_{\rho_\varphi=1}, \qquad \text{(III.9)}$$

where the curly bracket $\{\ \}$ now contains $\widetilde{P}_G$ instead of $P_G$ and admits a Laurent expansion in $\rho_\varphi^2$, the coefficients of which are the only dependence in $D$. Because the curly bracket in (III.9) is a function of $\rho_\varphi^2$, it is clear that (III.9) is analytic in $D$ as long as $E(DL(\varphi)/2)$ is a constant for any $\varphi\in\mathscr{F}$. Consequently, $I_G^\mu$ is analytic in $D$ as long as $E(DL(\varphi)/2)$ is a constant for all connected, one particle irreducible subgraphs $\varphi$. If we note that the operator $R$ is equal to 1 for $\text{Re}D < D_0'$, where

$$D_0' = \inf_\varphi \left\{\frac{2l(\varphi) - d(\varphi)}{L(\varphi)}\right\}, \qquad \text{(III.10)}$$

and inf runs over all connected, one particle irreducible subgraphs $\varphi$, then $I_G^\mu$ is continuous in $D$ for $\text{Re}D < D_0'$, and has discontinuities for each value of Re$D$ equal to $[2l(\varphi) - d(\varphi) + 2p]/L(\varphi)$ for any subgraph $\varphi\subseteq G$ connected, one particle irreducibe, and for any nonnegative integer $p$.

Let $B_\Delta$ be the band defined as

$$B_\Delta = \left\{D\in\mathbb{C}: E\left(\frac{DL(\varphi)}{2}\right) = E\left(\frac{\Delta L(\varphi)}{2}\right)\right. \qquad \text{(III.11)}$$

for every connected one particle irreducible subgraph $\varphi$,

$$\left.\Delta \leqslant \text{Re}D\right\}.$$

We recall that

$$\Delta' = \sup\{\text{Re}D,\ D\in B_\Delta\}.$$
(III.12)

Then $I_G^\mu$ and its integrand are analytic functions of $D$ in $B_\Delta$, and can be analytically continued in the complex plane (see Appendix C):

—the integrand is an analytic function of $D$, integrable if $\text{Re}D < \Delta'$.

—$I_G^\mu(p,m,\tau,D)$ in $B_\Delta$ can be extended into a meromorphic function of $D$, noted $I_G^{\mu,\Delta}(p,m,\tau,D)$, with poles only at some rationals $\geqslant\Delta'$, and which coincides with the integral of the analytic continuation of the integrand for $\text{Re}D < \Delta'$. The function $I_G^{\mu,4}(p,m,\tau,4)$ have all the necessary properties of a renormalized Feynman amplitude and for strictly renormalizable field theory may be generated from a Langrangian by adequate $\mu$ dependent counterterms. Especially, when all masses $\mu_a$ are taken to be equal to $m_a$, we recover the usual $R$ operation of subtractions at zero external momentum.

## IV. EXTRACTION OF POLES BETWEEN TWO SUCCESSIVE BANDS

In order to compare the analytically continued function $I_G(p,m,\tau,D)$ with $I_G^\mu(p,m,\tau,D)$, it is convenient to evaluate the difference between the functions $I_G^{\mu,\Delta}(p,m,\tau,D)$ and $I_{G(p,m,\tau,D)}^{\mu,\Delta'}$ originally defined in two consecutive bands.

Let $B_\Delta$ and $B_\Delta$ be two consecutive bands defined as

$$B_\Delta = \{D{:}\Delta \leqslant \mathrm{Re}D < \Delta'\},$$
$$B_{\Delta'} = \{D{:}\Delta' \leqslant \mathrm{Re}D < \Delta''\}. \tag{IV.1}$$

The difference between the integrands of $I_G^{\mu,\Delta}$ and of $I_G^{\mu,\Delta'}$ evaluated for $D \in B_{\Delta'}$ is

$$\exp\Big(-\sum_a \alpha_a\mu_a^2\Big)\sum_{\mathscr{F}}\Big[\prod_{\varphi\in\mathscr{F}}(-\tau_\varphi^{-2l(\varphi)-\Delta(\varphi)}) - \prod_{\varphi\in\mathscr{F}}(-\tau_\varphi^{-2l(\varphi)})\Big]\Big\{\exp\Big[\sum_a \alpha_a(\mu_a^2-m_a^2)\Big]$$
$$\times S_G(p,\tau,\alpha_a,D)e^{-pd_G^{-1}(\alpha)p}P_G(\alpha)^{-D/2}\Big\}, \tag{IV.2}$$

where the sum runs above all nonempty forests of connected 1PI subgraphs in $G$. The quantity $\Delta(\varphi)$ is defined as

$$\Delta(\varphi) = 2 \tag{IV.3}$$

if $\varphi$ gives a pole at $\Delta'$, i.e., if $[\Delta' L(\varphi) + d(\varphi) - 2l(\varphi)]$ is a nonnegative even integer, and $\Delta(\varphi)$ is zero otherwise. We use a generalization of the technique used in the past[13,16] to prove Zimmermann's identity; we write, for each forest $\mathscr{F}$,

$$\Big[\prod_{\varphi\in\mathscr{F}}(-\tau_\varphi^{-2l(\varphi)-\Delta(\varphi)}) - \prod_{\varphi\in\mathscr{F}}(-\tau_\varphi^{-2l(\varphi)})\Big]$$
$$= \sum_{\{\varphi_i\}}\prod_i(\tau_{\varphi_i}^{-2l(\varphi_i)} - \tau_{\varphi_i}^{-2l(\varphi_i)-\Delta(\varphi_i)})\prod_{\substack{\varphi\subset\varphi_i\\ \varphi\in\mathscr{F}}}(-\tau_\varphi^{-2l(\varphi)-\Delta(\varphi)})\prod_{\varphi\in[\mathscr{F}]_{\{\varphi_i\}}}(-\tau_\varphi^{-2l(\varphi)}). \tag{IV.4}$$

The sum runs over all nonempty families $\{\varphi_i\}$ of disjoint elements of $\mathscr{F}$ such that $\Delta(\varphi_i)\neq 0$, and $[\mathscr{F}]_{\{\varphi_i\}}$ is the forest of all elements $\varphi$ of $\mathscr{F}$ such that, for all $i$, $\varphi\neq\varphi_i$ and $\varphi\cap\varphi_i$ is $\phi$ or $\varphi_i$. We sum (IV.4) over all forests containing $\{\varphi_i\}$; we thus reconstruct the $\mathscr{F}_i$ of the subgraphs $\varphi_i$ and the forests $[\mathscr{F}]_{\{\varphi_i\}}$ of the reduced graph $[G/\cup\varphi_i]$. In the sum over the forests $\mathscr{F}_i$, the forests containing the graph $\varphi_i$ itself contribute zero to the sum. We obtain

$$\sum_{\{\varphi_i\}}\prod_i\Big\{\Big[(\tau_{\varphi_i}^{-2l(\varphi_i)}) - \tau_{\varphi_i}^{-2l(\varphi_i)-\Delta(\varphi_i)}\Big[1+\sum_{\mathscr{F}_i}\prod_{\varphi\in\mathscr{F}_i}(-\tau_\varphi^{-2l(\varphi)-\Delta(\varphi)})\Big]\Big\}\Big[1+\sum_{[\mathscr{F}]_{\{\varphi_i\}}}\prod_{\varphi\in[\mathscr{F}]_{\{\varphi_i\}}}(-\tau_\varphi^{-2l(\varphi)})\Big]. \tag{IV.5}$$

Since the Taylor operators in (IV.5) commute, we evaluate in a first step

$$(\tau_{\varphi_i}^{-2l(\varphi_i)} - \tau_{\varphi_i}^{-2l(\varphi_i)-2})\Big\{\exp\Big[\sum_a(\mu_a^2-m_a^2)\Big]S_G(p,\tau,\alpha_a,D)e^{-pd_G^{-1}(\alpha)p}P_G(\alpha)^{-D/2}\Big\}. \tag{IV.6}$$

It is convenient to decompose (IV.6) as

$$\sum_{\eta_a=0}^{N_a}\prod_{a\in\varphi_i}(\mu_a^2-m_a^2)^{\eta_a}\exp\Big[\sum_{a\in G/\varphi_i}\alpha_a(\mu_a^2-m_a^2)\Big](\tau_{\varphi_i}^{-2l(\varphi_i)} - \tau_{\varphi_i}^{-2l(\varphi_i)-2})\Big[\prod_{a\in\varphi_i}\alpha_a^{\eta_a}S_G(p,\tau,\alpha,D)e^{-pd_G^{-1}(\alpha)p}P_G(\alpha)^{-D/2}\Big], \tag{IV.7}$$

where $N_a$ is great enough to make the action of the $\tau$ operators to be zero. We now use the theorem (B2) of Appendix B and write (IV.7) as

$$\sum_{\eta_a=0}^{N_a}\prod_{a\in\varphi_i}[(\mu_a^2-m_a^2)\alpha_a]^{\eta_a}\eta_a\exp\Big[\sum_{a\in G/\varphi_i}\alpha_a(\mu_a^2-m_a^2)\Big]\sum_{\chi_K}y_{\varphi_i}^{\chi_K}(\alpha,D)y_{[G/\varphi_i]_{\chi_K}}(p,\alpha,D), \tag{IV.8}$$

where

$$K_i = L(\varphi_i)\Delta' - 2l(\varphi_i) + \delta'(\varphi_i) - 2\sum_{a\in\varphi_i}\eta_a \geqslant 0. \tag{IV.9}$$

The notations $\chi_K$, $y_\varphi^{\chi_K}$, and $y_{[G/\varphi]_{\chi_K}}$ are explained in Appendix B. If we apply the Taylor opertors relative to all subgraphs $\varphi_i$ in (IV.5), we get

$$\sum_{\eta_a=0}^{N_a}\sum_{\{\chi_K\}}\prod_i\Big\{\prod_{a\in\varphi_i}[(\mu_a^2-m_a^2)\alpha_a]^{\eta_a}y_{\varphi_i}^{\chi_K}\alpha,D)\Big\}\exp\Big[\sum_{a\in G/\cup\varphi_i}\alpha_a(\mu_a^2-m_a^2)\Big]y_{[G/\cup\varphi_i]_{\{\chi_K\}}}(p,\alpha,D). \tag{IV.10}$$

In (IV.10), $\Big[\prod_{a\in\varphi_i}\alpha_a^{\eta_a}y_{\varphi_i}^{\chi_K}(\alpha,D)\Big]$ is homogeneous in $\alpha$ of degree

$$\tfrac{1}{2}[L(\varphi)(\Delta'-D) - 2l(\varphi)]. \tag{IV.11}$$

Now, we can evaluate the action of the other Taylor operators of (IV.5) on each term of (IV.10). With the preceding homogeneity relation, it is easy to see from

$$\tau^n\{x^\lambda x^\mu f(x)\} = x^\lambda \tau^{n + E'(\mu) - E'(\lambda+\mu)}\{x^\mu f(x)\}, \qquad \text{(IV.12)}$$

where $f(x)$ has a Taylor expansion around $x=0$, that if $D$ satisfies

$$\Delta' \leqslant \mathrm{Re}D < \inf(\Delta'', \Delta'''), \qquad \text{(IV.13)}$$

where $\Delta'''$ is the first singularity in $D$ (larger than $\Delta'$) of the reduced graph $[G/\cup\varphi_i]$, the product $\Pi_i[\Pi_{a\in\varphi_i}\alpha_a^{\eta_a} y_{\varphi_i}^{\chi_K}(\alpha,D)]$ passes through the Taylor operators $\tau_\varphi^{-2I(\varphi)}$ with $\varphi\in[\mathscr{F}]_{\{\varphi_i\}}(\varphi\supset\varphi_i)$, to give $(\tau_{[\varphi/\cup\varphi_i]}^{-2I(\varphi/\cup\varphi_i)})$ acting upon the part of (IV.10) relative to $[G/\cup\varphi_i]$.. The $\tau$ operators acting upon the part of (IV.10) relative to $\varphi_i$ is nothing but the $R$ operator defined in the band $B_\Delta$, while the $\tau$ operators acting upon the part of (IV.10) relative to $[G/\cup\varphi_i]$ is the $R$ operator in the band (IV.13). We may now integrate over the variables $\alpha$ for $\mathrm{Re}D<\Delta'$ and we obtain the following theorem.

*Theorem*:

$$I_G^{\mu,\Delta}(p,m,\tau,D) - I_G^{\mu,\Delta'}(p,m,\tau,D)$$
$$= \sum_{\{\varphi_i\}} \sum_{\eta_a}^{N_a} \sum_{0\{\chi_K\}} \prod_i I_{\varphi_i}^{\Delta,\eta_a,\chi_K}(\mu,m,D) I_{[G/\cup\varphi_i]_{\{\chi_K\}}}^{\mu,\Delta'}(p,m,\tau,D), \qquad \text{(IV.14)}$$

where $\{\varphi_i\}$ is any family of disjoint connected, 1PI subgraphs of $G$ giving a pole of $I_G(p,m,\tau,D)$ at $D=\Delta'$. In (IV.14)

$$I_{\varphi_i}^{\Delta,\eta_a,\chi_K}(\mu,m,D) = \prod_{a\in\varphi_i}(\mu_a^2 - m_a^2)^{\eta_a}\int_0^\infty \prod_{a\in\varphi_i} d\alpha_a$$
$$\times \exp\Big(-\sum_a \alpha_a \mu_a^2\Big) R\Big\{\prod_a \alpha_a^{\eta_a} y_{\varphi_i}^{\chi_K}(\alpha,D)\Big\}, \qquad \text{(IV.15)}$$

with the operator $R$ defined in the band $B_\Delta$; it is a meromorphic function of $D$ with poles at rationals $\geqslant\Delta'$. It is important to note that, from the homogeneity of the integrand (IV.15),

$$\lim_{\{\mu_a\}\to 0} I_{\varphi_i}^{\Delta,\eta_a,\chi_K}(\mu,m,D) = 0 \quad \text{for } \mathrm{Re}D>\Delta', \qquad \text{(IV.16)}$$

the convergence being uniform for all $m$ in a compact set of $]0,\infty[$ and for $D$ in a compact set of $\{\mathrm{Re}D>\Delta', D$ away from poles of $I_{\varphi_i}^{\Delta,\eta_a,\chi_K}(\mu,m,D)\}$. Of course, as we already said in Sec. III, $I_{[G/\cup\varphi_i]_{\{\chi_K\}}}^{\mu,\Delta'}(p,m,\tau,D)$ is a meromorphic function of $D$ with poles at rationals $\geqslant \inf(\Delta'',\Delta''')$. In (IV.14), the poles at $\Delta'''$ are spurious and are absent on the left-hand side. This theorem is used recurrently in Sec. V to describe the analytic continuation of $I_G(p,m,\tau,D)$.

## V. INTEGRAL REPRESENTATION FOR THE DIMENSIONALLY REGULARIZED AMPLITUDES

We now use the theorem of decomposition (IV.14) and we extract recursively the poles of any amplitude $I_G(p,m,\tau,D)$ between the band $B^{-\infty}$ and any band $B_\Delta = \{\Delta\leqslant \mathrm{Re}D<\Delta'\}$. We have, of course,

$$(I_G^{\mu,-\infty}(p,m,\tau,D) = I_G(p,m,\tau,D). \qquad \text{(V.1)}$$

We use the following notations: for any forest $\mathscr{F}$ of connected, 1PI subgraphs of $G$, and for any subgraph $\varphi$ of $\mathscr{F}$, we note by $[\varphi]_{\mathscr{F}}$ the reduced subgraph obtained from $\varphi$ by shrinking into points all the maximum subgraphs $\varphi'$ of $\mathscr{F}$ such that $\varphi'\subset\varphi$. In the same way, $[G]_{\mathscr{F}}$ is the graph obtained by reducing all maximum subgraphs $\varphi$ of $\mathscr{F}$ in $G$ (if $G\in\mathscr{F}$, $[G]_{\mathscr{F}}$ is reduced to one vertex). From now on, it is understood that at each reduced vertex, we associate the appropriate set of derivative couplings characteristic of the graph which is reduced.

*Theorem I*: For any Feynman integral, and for any rational $\Delta$

$$I_G(p,m,\tau,D)$$
$$= \sum_{\mathscr{F}}\sum_{\{\eta,\chi\}}\prod_{\varphi\in\mathscr{F}} I_{[\varphi]_{\mathscr{F}}}^{\Delta_{[\varphi]_{\mathscr{F}}},\{\eta,\chi\}}(\mu,m,D) I_{[G]_{\mathscr{F}}}^{\mu,\Delta'}(p,m,\tau,D), \qquad \text{(V.2)}$$

where we sum over all possible sets of indices $\eta_a$ and $\chi_{K_i}$ obtained from (IV.14) over all forests of connected, 1PI subgraphs of $G$ which are responsible for poles at $\mathrm{Re}D<\Delta$.

The proof of this theorem is easy; we introduce a complete set of poles given by all rationals $2k/L$ $[1\leqslant L\leqslant L(G)]$, where $k$ is an integer such that $2k/L\geqslant D_0'$ [the lower pole for $G$ defined in (III.10)]. These rationals take into account all the poles of the subgraphs and of the reduced graphs of $G$. We then apply (IV.14) again and again from band to band by transforming

$I_{[G]_{\mathscr{F}}}^{\mu,2k/L}(p,m,\tau,D)$. We note that:

(i) $\Delta_{[\varphi]_{\mathscr{F}}}$ attached to $I_{[\varphi]_{\mathscr{F}}}^{\Delta_{[\varphi]}}$ means that the $R$ subtraction operator is defined in the band $B_{\Delta_{[\varphi]}}$, that $[\varphi]_{\mathscr{F}}$ is responsible for poles in $D$ the lowest being at a rational on the right boundary of $B_{\Delta_{[\varphi]}}$;

(ii) $\Delta$ attached to $I_{[\varphi]_{\mathscr{F}}}^{\mu,\Delta}$ means similarly that the $R$ subtraction operator is defined in the band $B_\Delta$ and that $I_{[\varphi]_{\mathscr{F}}}^{\mu,\Delta}$ has poles for $\mathrm{Re}D>\Delta$ only;

(iii) the family $\{\Delta_{[\varphi]_{\mathscr{F}}}\}$ is strictly increasing with respect to the inclusion order in $\mathscr{F}$, and, of course,

$$D_0'\leqslant\Delta_{[\varphi]_{\mathscr{F}}}<\Delta. \qquad \text{(V.3)}$$

We could from (V.2) extract all the poles of $I_G(p,m,\tau,D)$ between $D_0'$ and $\Delta$ and we would obtain the expression of $I_G(p,m,\tau,D)$ for $\mathrm{Re}D<\Delta$ and for $\mathrm{Re}D\in B_\Delta$ but different from some rationals (and this for any $\Delta$). It is important to note that the right-hand side of (V.2) is necessarily $\mu$ independent, and we intend to use this property to calculate the limit of $\mu\to 0$ of (V.2) and prove the following theorem.

*Theorem II*: For any Feynman integral $I_G(p,m,\tau,D)$ of the form (I.9)

(a) $I_G^{\mu,\Delta}(p,m,\tau,D)$ converges uniformly when $\{\mu_a^2\}\to 0$ to $I_G(p,m,\tau,D)$ for $m$ in any compact set of $]0,\infty[$ and for $D$ in any compact set in $B_\Delta$; the band $B_\Delta$ is defined as Re$D$ between two consecutive poles ($\Delta$ being the lower one) of the type $[2l(\varphi)-d(\varphi)+2p]/L(\varphi), p\geqslant 0$, for any connected, 1PI subgraphs $\varphi$ of $G$.

(b) Consequently $I_G(p,m,\tau,D)$ is meromorphic with poles at only the above rationals

(c) We obtain the integral representation

$$I_G(p,m,\tau,D)=\int_0^\infty \prod_{a=1}^{l} d\alpha_a \times R\left[\frac{\exp\left(-\sum_{a=1}^{l}\alpha_a m_a^2\right) S_G(p,\tau,\alpha,D) e^{-p_i[d_G^{-1}(\alpha)]_{ij}p_j}}{P_G(\alpha)^{D/2}}\right]. \tag{V.4}$$

The above integral is absolutely convergent for any Re$D$ away from the above rationals.

*Proof*: In a first step, we prove the convergence property described in (a) for $D\in B_\Delta$ and different from the rationals $2k/L$ defined in the proof of Theorem I (these extra poles are the spurious poles coming from reduced subgraphs). We prove the theorem by recurrence on the number of loops of the graph $G$. For a tree graph, $I_G^{\mu,\Delta}(p,m,\tau,D)$ is $\mu$ independent, and coincides everywhere with $I_G(p,m,\tau,D)$.

We suppose the convergence property true for all graphs with less than $L(G)$ loops. Then, from Theorem I,

$$I_G(p,m,\tau,D)=I_G^{\mu,\Delta}(p,m,\tau,D) + \sum_{\mathscr{F}}\sum_{\{\eta,\chi\}}\prod_{\varphi\in\mathscr{F}'} I_{[\varphi]_{\mathscr{F}'}}^{\Delta_{[\varphi]_{\mathscr{F}'}},\{\eta,\chi\}}(\mu,m,D)\, I_{[G]_{\mathscr{F}'}}^{\mu,\Delta}(p,m,\tau,D), \tag{V.5}$$

where $\mathscr{F}'$ is a nonempty forest. From (IV.16), (V.3) and the recurrence for $I_{[G]_{\mathscr{F}'}}^{\mu,\Delta}(p,m,\tau,D)$, we prove (a) for $D\in B_\Delta$ but different from the rationals $2k/L$. We now prove that the convergence property (a) is valid in the entire band $B_\Delta$. We just state the following lemma: Let $\{f_\lambda;\lambda\in[0,1]\}$ be a family of analytic functions of $D$ defined in an open set $\Omega$ of the complex plane $D$, and $D_0$ be a single point in $\Omega$. If $f_\lambda$ converges uniformly in every compact set in $\Omega-\{D_0\}$ towards a function $f$ when $\lambda\to 0$, then $f_\lambda$ converges uniformly in every compact of $\Omega$, and the limit function (which coincides with $f$ in $\Omega-\{D_0\}$) is analytic in $\Omega$.

Since $I_G^{\mu,\Delta}(p,m,\tau,D)$ is shown in Sec. III to be analytic in $D$ for $D\in B_\Delta$, we apply the above lemma and prove (a) in the entire band $B_\Delta$.

Now, this proof may be performed for any band $B_\Delta$. We already proved in Sec. II that the singularities for $D=\Delta$ are poles. This achieves part (b). Moreover, we show in Appendix C that the integral (V.4) is absolutely convergent inside any band $B_\Delta$. It remains to prove the equality between both sides of (V.4). We give the following arguments: from the uniform convergence of $I_G^{\mu,\Delta}(p,m,\tau,D)$ to $I_G(p,m,\tau,D)$ for any finite nonzero $m$ and from the continuity in $m$ of $I_G^{\mu,\Delta}(p,m,\tau,D)$, we deduce that

$$\lim_{\mu\to 0} I_G^{\mu,\Delta}(p,(m^2+\mu^2)^{1/2},\tau,D)=I_G(p,m,\tau,D). \tag{V.6}$$

If we denote by $A^\mu(m^2,\alpha)$ the integrand of $I_G^{\mu,\Delta}$ given in (III.1),

$$|A^\mu(m^2+\mu^2,\alpha)|\leqslant|A^0(m^2,\alpha)|, \tag{V.7}$$

and of course $\lim_{\mu\to 0}A^\mu(m^2+\mu^2,\alpha)=A^0(m^2,\alpha)$, for any nonzero $\alpha_s'$. Then by Lebesgue's bounded convergence theorem, and since $A^0(m^2,\alpha)$ is integrable for $D\in B_\Delta$,

$$\lim_{\mu\to 0} I_G^{\mu,\Delta}(p,(m^2+\mu^2)^{1/2},\tau,D)=\int_0^\infty\prod_{a=1}^{l} d\alpha_a\, A^0(m^2,\alpha), \tag{V.8}$$

which proves with (V.6) the equal sign of (V.4).

The expression (V.4) is the integral representation of $I_G(p,m,\tau,D)$ for all Re$D$ away from some rationals; the subtraction operator $R$ subtracts as many times as needed depending of Re$D$. Each subgraph is subtracted at zero external momenta and zero internal masses. Let us note that in each open band $B_\Delta=\{D:\Delta<\text{Re}D<\Delta'\}$, when we make $D\to\Delta'$, the integral (V.4) diverges because of ultraviolet singularities at $\alpha\sim 0$, and when $D\to\Delta$, the integral diverges because of infrared singularities at $\alpha\sim\infty$. So, it is remarkable that any pole $\Delta$ of $I_G(p,m,\tau,D)$ appears in (V.4) as corresponding to UV divergences if $D\to\Delta-$, and as IR divergences if $D\to\Delta+$. We face, in that peculiar case, a very general type of equivalence between UV and IR singularities.

## VI. CONCLUSION

Most of what is known about analytic continuation in dimension of Feynman amplitudes is based upon its very existence, its respect of field equations and of Ward identities (except for well-known anomalies), and upon a recursive procedure which extracts the singularities at dimension four without breaking the symmetries of the system. We have given here, in a simple form, a convergent integral representation for the massive Feynman amplitudes, which gives a sense perturbatively to the bare Langrangian field theory, at any dimension away from some rationals.

It is expected that the recursive procedure of extraction of poles at dimension four, may be replaced by a subtraction operator, so that the so-called 't Hooft dimensional renormalization may be formulated in compact integral representation. Of course, for the applications to the gauge field theories, we need a generalization of (V.4) to massless Feynman amplitudes[17] (some results may be found in Ref. 18).

The various renormalizations $(\mu,\Delta)$ at dimension four $(\mu\neq 0, \Delta=4)$, can also be considered as a regularization of the bare theory when $\mu\to 0$. This approach is also useful to study the implications of the renormalization group for a dimension between two and four.

We finally mention that the result of this paper is not only applicable to Feynman amplitudes but to any "similar" integrals.

## ACKNOWLEDGMENTS

We wish to express our gratitude to Doctor Y. M. P.

Lam for helpful discussions during the early stage of this work.

## APPENDIX A

The purpose of this appendix is to make a convenient choice for the function $S_G(p_i,\tau,\alpha_a,D)$ which analytically continue for all $D$ the function $S_G(p_i,\tau,\alpha_a)$ introduced at dimension four in (I.8). As we already mentioned in Sec. I, $S_G(p_i,\tau,\alpha_a)$ itself could have been an easy candidate since the integral (I.8) with $N$ replaced by $D$ is ultraviolet convergent for Re$D$ small enough. In fact, it is the very structure of Feynman integrands which requires for consistency another function.

First let us recall how we obtain $S_G(p_i,\tau,\alpha_a)$. We first perform the derivatives $\partial/\partial z$ and set $z$ equal to zero, then we use algebraic relations (denoted $\mathscr{R}_1$) between objects like $\gamma$ matrices (essentially anticommunication relations) to obtain some Lorentz tensor $g^{\mu\nu}$. Then we contract dummy indices (using peculiarly $g^{\mu}{}_{\mu} = 4$).

However, in momentum space, several integrands describe the same integral because of the following relations:

$$\mathscr{R}_2: \quad p_i + \sum_a \epsilon_{ia} k_a = 0, \tag{A1}$$

which expresses the conservation of energy-momentum at each vertex

$$\mathscr{R}_3: \quad \frac{k_a^2 + m_a^2}{k_a^2 + m_a^2} = 1, \tag{A2}$$

which simplifies the numerator and the denominator and has the effect of introducing a new graph $[G/a]$ obtained from $G$ by shrinking the line $a$ into a point.

Following Ref. 11, we know that, for convergent amplitudes, the field equations and the Ward identities are satisfied, at each order of perturbation, provided that the relations $\mathscr{R}_1$,$\mathscr{R}_2$, and $\mathscr{R}_3$ are true at dimension four.

Our first requirement for the choice of $S_G(p,\tau,\alpha,D)$ is such that whatever use we make of relations $\mathscr{R}_2$ and $\mathscr{R}_3$, the functions $S_G(p,\tau,\alpha,D)$ describe the same integral. The function $S_G(p,\tau,\alpha)$ satisfies this requirement because we may verify that

$$p_i + \sum_a \epsilon_{ia} \left( \frac{-1}{\sqrt{\alpha_a}} \frac{\partial}{\partial z_a} \right) Z_G(\alpha_a,p_i,z_a) = 0 \tag{A3}$$

and

$$\left[ \frac{1}{\alpha_a} \frac{\partial^2}{\partial z_a^2} - \frac{z_a}{2\alpha_a} \frac{\partial}{\partial z_a} + m_a^2 + \frac{d}{d\alpha_a} \right] \times \exp\left( -\sum_a \alpha_a m_a^2 \right) Z_G(\alpha_a,p_i,z_a) = 0, \tag{A4}$$

where $Z_G(\alpha_a,p_i,z_a)$ is given in (I.7). By application of $((-1/\sqrt{\alpha})\,\partial/\partial z)$ upon (A3) and (A4) at $z = 0$, we obtain

$$\left[ p_i S_G(p_i,\alpha_a) + \sum_a \epsilon_{ia} S_{G_a}(p_i,\alpha_a) \right] e^{-p_i d_G^{-1}(\alpha) p_j} P_G(\alpha)^{-N/2} = 0 \tag{A5}$$

where $G_a$ is the graph obtained from $G$ by inserting on the line $a$ a derivative coupling $k_a$, and

$$[S_{G_{aa}}(p_i,\alpha_a) + m_a^2 S_G(p_i,\alpha_a)] e^{-\Sigma \alpha_a m_a^2} e^{-p_i d_G^{-1}(\alpha) p_j} P_G(\alpha)^{-N/2} + \frac{d}{d\alpha_a} [S_G(p_i,\alpha_a) e^{-\Sigma \alpha_a m_a^2} e^{-p_i d_G^{-1}(\alpha) p_j} P_G(\alpha)^{-N/2}] = 0, \tag{A6}$$

where $G_{aa}$ is the graph obtained from $G$ inserting on the line $a$ two derivative couplings $k_a^{\mu} \cdot k_{a\mu}$ and by summing over $\mu$ from 1 to 4. Equations (A3) and (A5) remain valid when $N$ is replaced by the complex number $D$; on the other hand (A4), (A6) are not valid anymore because $(d/d\alpha_a)[P_G(\alpha)^{-D/2}]$ gives a multiplicative factor $D$ while the summation over $\mu$ of $k_a^{\mu} k_{a\mu}$ still gives a factor 4. This situation forces us to define the analytic continuation $S_{G_{aa}}(p_i,\alpha_a,D)$ of the function $S_{G_{aa}}(p_i,\alpha_a)$, by analytically continuing the trace $g^{\mu}{}_{u}$ into $D$.

However, it is difficult to define an "analytic continuation" at $D \neq 4$ of the algebraic relations $\mathscr{R}_1$. Following the previous rule, one would write $\gamma^{\mu}\gamma_{\mu}$ as $D \cdot \mathbb{1}$, as it is well known that such a requirement is enough to ensure Ward identities, provided that there are no $\gamma^5$ matrices. We shall show that such a rule is enough, provided that we give a specific definition of what are dummy and external indices in the graph. We define the following procedure to choose $S_G(p,\tau,\alpha,D)$:

(i) We split the set of Lorentz indices which appear in the expresssion of any graph G into two parts:

(a) dummy indices (denoted by Greek letters $\mu,\nu,\cdots$), which are:

(1) indices summed inside the T product defining the Green's function corresponding to the graph; an example is given by $\langle TA(\varphi)\partial_{\mu}\partial^{\mu}\varphi B(\varphi)\rangle$;

(2) indices given by internal propagators and vertex contributions (for instance $g^{\mu\nu}, k^{\mu}k^{\nu}, \gamma^{\mu}, \gamma^{\mu}k_{\mu}$, with the exception of the matrix $\gamma^5$.

(b) external indices (denoted by Latin letters a,b,$\cdots$), which are:

(1) nonsummed indices inside the T product (like external momentum);

(2) indices introduced by replacing all $\gamma^5$ matrices by

$$\gamma^5 = \frac{1}{4!} \epsilon^{a_1 a_2 a_3 a_4} \gamma_{a_1} \gamma_{a_2} \gamma_{a_3} \gamma_{a_4}, \tag{A7}$$

which is nothing more than its definition at dimension four. The tensor $\epsilon^{a_1 a_2 a_3 a_4}$ is the totally antisymmetric tensor. One could make summations about some external indices, for instance in $g^{ab}\langle TA(\varphi)\partial_a\partial_b\varphi B(\varphi)\rangle$, the summation been made explicitly for $a = 1\cdots4$. Nonsummed external indices belong to the set $\tau$.

(ii) We make the derivations $\partial/\partial z_a$, and we take $z = 0$.

(iii) We let the $\gamma$'s disappear in order to obtain the maximum number of g's by using algebraic relations between $\gamma$ matrices and also by using

$$\epsilon^{a_1 a_2 a_3 a_4} \epsilon_{b_1 b_2 b_3 b_4} = -\prod_{\sigma \in \sigma_4} \epsilon(\sigma) \prod_{i=1}^{4} g^{a_i}{}_{b_{\sigma(i)}}, \tag{A8}$$

where $\sigma_4$ is the set of all permutations in $\{1,...,4\}$ and $\epsilon(\sigma)$ is the signature of the permutation,

(iv) We contract dummy indices (without contracting external indices) via the relations

$$g^{\mu\nu}T_{\nu...} = T^{\mu}{}_{...}, \quad (A9)$$

$$g^{a\nu}T_{\nu...} = T^{a}{}_{...}, \quad (A10)$$

where $T_{\nu...}$ is any tensorial object, so that the only dependence in dummy indices is reported in $g^{\mu}{}_{\mu}$.

(v) We make the final contraction

$$g^{\mu}{}_{\mu} = D, \quad (A11)$$

which defines the analytic continuation, and, of course, we also make contractions about external indices, with

$$g^{a}{}_{a} = 4, \quad (A12)$$

$$\mathrm{Tr}\,\mathbb{1} = 4, \quad (A13)$$

$$\gamma_a p_i^a = \not{p}_i, \quad (A14)$$

$$p_i^a p_{ja} = p_i \cdot p_j, \quad (A15)$$

which are perfectly well defined upon four-dimensional objects. Such a procedure defines in a unique way a function $S_G(p_i,\tau,\alpha_a,D)$. The above procedure gives exactly the same results as those obtained by Speer[9] in momentum space and by Breitenlohner and Maison[7] in the $\alpha$ space; these results ensure that we do not break field equations and Ward identities, with the exception of the "anomalies" like the axial current anomaly in Gauge theory or the following "trace anomaly," which shows the difference between summation over internal and external indices: if we consider a graph $G$ with derivative couplings $k_a^{a_1}$ on line $a$ and $k_b^{a_2}$ on line $b$, we define the corresponding tensorial function $S_G^{a_1,a_2}(p_i,\tau,\alpha_a,D)$ which usually contains the tensor $g^{a_1 a_2}$; then, by taking the trace $S_G^{aa}(p_i,\tau,\alpha_a,D)$, we get a factor 4 which is certainly different from the factor $D$ obtained when $k_a^{\mu}k_{b\mu}$ is considered as a trace upon internal indices. The analytic continuation of a trace is in general different from the trace of the analytic continuation.

## APPENDIX B: DILATATION PROPERTIES OF DIMENSIONALLY REGULARIZED FEYNMAN INTEGRANDS IN THE α-PARAMETRIC REPRESENTATION

We give or establish in this appendix some results needed in Secs. II and IV. The regularized integrand defined by the procedure of Appendix A is a combination of scalar integrands

$$y_G(p,\alpha,D) = S_G(p,\alpha,D)e^{-p_i d_G^{ij}(\alpha)p_j}P_G(\alpha)^{-D/2}, \quad (B1)$$

with coefficients which are matrices and external momentum denoted by $\tau$. The expression (B1) at $D = N$ integer $\geqslant 4$ coincides with the corresponding Feynman integrand-$y_G(p,\alpha)_N$ calculated in a space of dimension $N$ with the same scalar products $p_i \cdot p_j$. It is clear that $S_G(p,\alpha)_N$ is a sequence of $N$, depending polynomially of $N$ while $S_G(p,\alpha,D)$ is a polynomial in $D$.

Let $\varphi$ be a subgraph of $G$. We call dilatation relative to $\varphi$ the change of variable

$$\alpha'_a = \alpha_a\rho^2 \quad \text{if } a\in\varphi,$$
$$\alpha'_a = \alpha_a \quad \text{if } a\notin\varphi. \quad (B2)$$

For any function of the $\alpha$'s, $Z(\alpha)$, we obtain a new function $Z(\alpha,\rho)$. We have in Ref, 13, the fundamental expansion theorem.

*Theorem B.*1: For any integer $N\geqslant 4$ and any subgraph $\varphi\subseteq G$, after dilatation relative to $\varphi$, the functions $y_G(p,\alpha,\rho)_N$ have the expansion $\rho$

$$y_G(p,\alpha,\rho)_N = \sum_{k=0}^{\infty}\rho^{-NL(\varphi)-\delta'(\varphi)+K}\sum_{\chi_K} y_\varphi^{\chi_K}(\alpha)_N y_{[G/\varphi]_{\chi_K}}(p,\alpha)_N, \quad (B3)$$

where $\delta'(\varphi)$ is the number of derivative couplings on the internal lines of $\varphi$, where the sum runs over all families $\chi_K$ of $k$ external momenta to $\varphi$, where for the sequence $\chi_k = \{k_{i_1}^{\mu_1},...,k_{i_K}^{\mu_K}\}$

$$y_\varphi^{\chi_K}(\alpha)_N = \left.\frac{\partial^k}{\partial k_{i_1}^{\mu_1}\cdots\partial k_{i_K}^{\mu_K}}y_\varphi(k,\alpha)_N\right|_{k=0}, \quad (B4)$$

[in ($B$ 4),$k$ denotes external momentum to $\varphi$ which may be external or internal to $G$], and where $y_{[G/\varphi]_{\chi_K}}(\alpha,p)_N$ is the Feynman integrand for the reduced graph $[G/\varphi]$ with $K$ derivative couplings $\{k_{i_1}^{\mu_1},...,k_{i_K}^{\mu_K}\}$ around the reduced vertex.

When we set $k = 0$ in (B4), $y_\varphi^{\chi_K}(\alpha)_N$ is found to be a sum over all contractions of tensorial indices (product of $g_{\mu\nu}$'s) times a scalar function characteristic of each set of contractions. It is then convenient to introduce the various sets of $g_{\mu\nu}$'s into $y_{[G/\varphi]_{\chi_K}}$ and to reinterpret (B3) in terms of scalar quantities $y_\varphi^{\chi_K}$ and $y_{[G/\varphi]_{\chi_K}}$ the summations over Lorentz indices being made in $[G/\varphi]$. Let us now extend Thereom (B1) for any complex $D$.

*Theorem B.*2: For any subgraph $\varphi\subseteq G$, after dilatation relative to $\varphi$, the function $y_G(p,\alpha,D)$ has the expansion in $\rho$

$$y_G(p,\alpha,D) = \sum_{k=0}^{\infty}\rho^{-DL(\varphi)-\delta'(\varphi)+K}\sum_{\chi_K}y_\varphi^{\chi_K}(\alpha,D) \times y_{[G/\varphi]_{\chi_K}}(p,\alpha,D). \quad (B5)$$

*Proof*: The functions $y_\varphi^{\chi_K}(\alpha)_N$ and $y_{[G/\varphi]_{\chi_K}}(p,\alpha)_N$ are of the form (B1), and, using Appendix A, we may in a unique way define their analytic continuation in $D$:$y_\varphi^{\chi_K}(\alpha,D)$ and $y_{[G/\varphi]_{\chi_K}}(p,\alpha,D)$. We now prove (B5). The expansion in $\rho$ of $y_G(p,\alpha,D,\rho)$ is necessarily of the form

$$\frac{e^{-p_i d_{[G/\varphi]}^{ij}(\alpha)p_j}}{P_{[G/\varphi]}(\alpha)^{D/2}P_\varphi(\alpha)^{D/2}} \times\left[\sum_{k=0}^{\infty}\rho^{-DL(\varphi)-\delta'(\varphi)+K}F_K(p,\alpha,D)\right] \quad (B6)$$

where $F_K(p,\alpha,D)$ is a polynomial in $D$. From Theorem B.1, for any integer $N\geqslant 4$

$$F_K(p,\alpha,D=N)=\sum_{\chi_K} S_\varphi^{\chi_K}(\alpha)_N S_{[G/\varphi]_{\chi_K}}(p,\alpha)_N. \tag{B7}$$

Hence, for $D$ equal to any integer $N\geqslant 4$,

$$F_K(p,\alpha,D)=\sum_{\chi_K} S_\varphi^{\chi_K}(\alpha,D)_N S_{[G/\varphi]_{\chi_K}}(p,\alpha,D). \tag{B8}$$

Since both sides of (B8) are polynomial in $D$ and are equal for an infinite number of values of $D=N$, they are equal for any complex $D$. This proves the theorem.

To establish the expression (II.4) for $d(\varphi)$, we have to make the two following remarks:

$$y_\varphi^{\chi_K}(\alpha,D)=0 \quad \text{if } \delta'(\varphi)-K \text{ is odd}, \tag{B9}$$

since we contract the vectorial indices by pair, and

$$y_\varphi^{\chi_K}(\alpha,D)=0 \tag{B10}$$

if $K$ is smaller than the number $\nu$ of derivative couplings on lines which are not inside a connected, 1PI component of $\varphi$, since such derivative couplings generates in $y_\varphi(p,\alpha,D)$ a homogeneous polynomial of degree $\nu$ in the external momentum $p$. Thus, the maximum value of$[\delta'(\varphi)-K]$ which gives a nonzero contribution to (B5) is given by $d(\varphi)$ defined in (II.4) $[\delta(\varphi)=\delta'(\varphi)-\nu]$.

## APPENDIX C: THEOREMS OF ABSOLUTE CONVERGENCE

### 1. Introduction

The purpose of this introduction is to remind the reader how to express the subtraction operator $R$ in Hepp's sector (II.1,II.2). The $R$ operator acts upon the $\alpha$ variables of each subgraph and does not recognize the subgraphs (except for the subgraphs $R_i$) when the integrand is expressed in the $\beta$ variables. Consequently, before performing the change of variables (II.2), we must introduce new variables which allow the generalized Taylor operators $\tau_\varphi$ to recognize its subgraph $\varphi$ after the change of variable $\alpha\rightarrow\beta$ is performed. The $R$ operator is expressed as a sum over all nests $\mathscr{N}$ of products of $\tau$ operators. If we consider only one nest $\mathscr{N}$, the corresponding $\beta$ integrals diverge, but we know[15] how to construct, for each sector $\sigma$, equivalent classes of nests $\Gamma$ such that, for the sum over all nests in $\Gamma$, the corresponding $\beta$ integrals converge. We now resume here the main features of this construction.

(i) Each equivalent class $\Gamma$ is characterized by its maximal nest $\mathscr{G}$ and its minimal nest $\mathscr{K}\subseteq\mathscr{G}$.

(ii) every nest $\mathscr{N}$ such that $\mathscr{K}\subseteq\mathscr{N}\subseteq\mathscr{G}$ belongs to $\Gamma$; each nest belongs only to one equivalent class $\Gamma$ and the sum over all equivalent classes reconstruct the sum over all nests.

(iii) The subgraphs of any nest $\mathscr{N}$ which belong to $\Gamma$ can be partitioned into the subgraphs of $\mathscr{K}$ and some subgraphs of $\mathscr{H}=\mathscr{G}-\mathscr{K}$. Consequently,

$$\sum_{\mathscr{N}\in\Gamma}\prod_{\varphi\in\mathscr{N}}(-\tau_\varphi^{-2l(\varphi)})=\prod_{\varphi\in\mathscr{K}}(-\tau_\varphi^{-2l(\varphi)})\prod_{\varphi\in\mathscr{H}}(1-\tau_\varphi^{-2l(\varphi)}). \tag{C1}$$

At this point we consider a given sector $\sigma$ and a given equivalent class of nests $\Gamma$.

In the construction of $\mathscr{K}$ and $\mathscr{H}$ [15] we define from the subgraphs $R^i$ the subnests $\mathscr{K}^i$ and $\mathscr{H}^i$ for $i=1,\dots,l$ such that $\cup\mathscr{K}^i=\mathscr{K}$ and $\cup\mathscr{H}^i=\mathscr{H}$ and we label the subgraphs of $\mathscr{K}^i$ and $\mathscr{H}^i$ by $K_j^i$ and $H_j^i$ for $j=1,\dots,r_i-1$. Moreover, $K_j^i\subset H_j^i\subset K_{j+1}^i\subset H_{j+1}^i\subset\cdots$ and

$$H_j^i=K_{j+1}^i\cap(R^i\cup K_j^i), \tag{C2}$$

$$\sum_{j=1}^{r_i-1}[l(H_j^i)-l(K_j^i)]=l(R^i) \quad \text{for } i=1,\dots l. \tag{C3}$$

Let us remind the reader that $\mathscr{H}^i$ is never empty. We now define the new variables upon which the $\tau$ operators act. Given a line $a\in K_j^i$, we dilate the varible $\alpha_a\rightarrow\alpha_a(\sigma_j^i)^2$, and given a line $a\in H_j^i$, we dilate $\alpha_a\rightarrow\alpha_a(\chi_j^i)^2$. Then, we perform the change of variables (II.2).

*Theorem*: We denote by $(\sigma\Gamma)$ the transformation of a function $Z(\alpha_a)$ into a function $Z^{\sigma\Gamma}(\beta_i,\sigma_j^i,\chi_j^i)$. Then, the function $Z^{\sigma\Gamma}$ is of the form $Z^{\sigma\Gamma}(\sigma_j^i/\beta_i,\beta_i\chi_j^i)$. The proof is given in Ref. 15.) Consequently, in the $(\sigma\Gamma)$ transformation

$$P_G(\alpha)\rightarrow\prod_{K_j^i\in\mathscr{K}}(\sigma_j^i/\beta_i)^{2l(K_j^i)}\prod_{H_j^i\in\mathscr{H}}(\beta_i\chi_j^i)^{2L(H_j^i)}[1+Q(\sigma/\beta,\beta\chi)], \tag{C4}$$

$$S_G(\alpha)\rightarrow\prod_{K_j^i\in\mathscr{K}}(\sigma_j^i/\beta_i)^{-d(K_j^i)}\prod_{H_j^i\in\mathscr{H}}(\beta_i\chi_j^i)^{-d(H_j^i)}\widetilde{S}_G(\sigma/\beta,\beta\chi), \tag{C5}$$

where $d(\varphi)$ is defined in (II.4),

$$pd_G^{-1}(\alpha)p\rightarrow(\beta_1\chi_1^1)^2\mathscr{D}(p,\sigma/\beta,\beta\chi), \tag{C6}$$

$$m_a^2\alpha_a\rightarrow m_a^2\prod_{K_j^i\ni a}(\sigma_j^i/\beta_i)^2\prod_{H_j^i\ni a}(\beta_i\chi_j^i)^2. \tag{C7}$$

The functions $Q,\widetilde{S}_G,\mathscr{D}$ have a "simultaneous Taylor expansion" in the variables $\sigma/\beta$ and $\beta\chi$ around zero. We now apply to the integrand (III.1) the operators $\tau_{\sigma_j^i}^{-2l(K_j^i)}$ corresponding to subgraphs of $\mathscr{K}$, by following the rules given in (III.3). We get a sum of terms of the form

$$\prod_{i=1}^{l}\beta_i^{\sum_{j=1}^{r_i-1}[DL(K_j^i)+d(K_j^i)-DL(H_j^i)-d(H_j^i)-a_j^i]}$$

$$\times\prod_{H_j^i\in\mathscr{H}}(1-T_{\chi_j^i}^{\omega_\Delta(H_j^i)})A_{\{a_j^i\}}(\beta\chi), \tag{C8}$$

where $0\leqslant a_j^i\leqslant\omega_\Delta(K_j^i)$ provided that all graphs $K_j^i$ are divergent (otherwise we get zero for this equivalent class) and where the function $A_{\{a_j^i\}}(\beta\chi)$ has a "simultaneous Taylor series" in $\beta\chi$ around zero. In (C8) the degree of subtraction $\omega_\Delta(\varphi)$ is defined, relative to a band $B_\Delta=\{\Delta\leqslant\mathrm{Re}D<\Delta', E[DL(\varphi)]=E[\Delta L(\varphi)]$ for every graph $\varphi\}$, inside which the $R$-operator subtracts, as

$$\omega_\Delta(\varphi)=\Delta L(\varphi)-2P(\varphi)+d(\varphi). \tag{C9}$$

A subgraph $\varphi$ is said to be divergent if $\omega_\Delta(\varphi)\geqslant 0$. We now use the integral representation for the rest of the Taylor series relative to the divergent subgraphs $H^i_j$, and we transform (C8) into

$$\prod_{i=1}^{l}\beta_i^{p_i}\int_0^1 \prod_{\substack{H^i_j\in\mathscr{H}\\ \omega_\Delta(H^i_j)\geqslant 0}}\left[d\chi^i_j\frac{(1-\chi^i_j)^{\omega_\Delta(H^i_j)}}{\omega_\Delta(H^i_j)!}\left(\frac{\partial}{\partial(\beta_i\chi^i_j)}\right)^{\omega_\Delta(H^i_j)+1}\right]$$

$$\times \Lambda_{\{a'_i\}}(\beta\chi)\Big|_{\substack{\chi^i_j=1\\ \omega_\Delta(H^i_j)<0}}\Big\}, \tag{C10}$$

with

$$p_i=\sum_{j=1}^{r_i-1}[DL(K^i_j)+d(K^i_j)-DL(H^i_j)-d(H^i_j)-a^i_j]$$

$$+\sum_{\substack{H^i_j\\ \omega_\Delta(H^i_j)\geqslant 0}}[\omega_\Delta(H^i_j)+1]. \tag{C11}$$

## 2. Absolute convergence of $I_G^{\mu,\Delta}$ for $\mu_a>0$, $m_a>0$, Re$D\leqslant\Delta$ or $D$ in the band $B_\Delta$

To simplify the formalism, we consider in this appendix equivalent classes of nests instead of equivalent classes of forests of connected, one-particle irreducible subgraphs. As a consequence, the bands

$$B_\Delta=\{D\in\mathbb{C}:E[DL(\varphi)]$$
$$=E[\Delta L(\varphi)] \text{ for every subgraph } \varphi, \Delta\leqslant \text{Re}D\}, \tag{C12}$$

considered in this appendix are in fact subbands of the bands defined in Sec. III. For each sector, the use of equivalent classes of nests makes the subtractions different in each subband although cancellation occurs between equivalent classes and between sector so that the complete $R$ operator subtracts exactly the same quantitites for all subbands of the same band of Sec. III. Everything which is proved here in the subbands $B_\Delta$ using equivalent classes of nests could be proved in the band of Sec. III using the heavier formalism of equivalent classes of forests of connected, one-particle irreducible subgraphs.

The function $I_G^\mu$ defined in (III.1) is a discontinuous function which depends upon the band inside which the $R$ operator is defined; we call $I_G^{\mu,\Delta}$ the function $I_G^\mu$ with the $R$ operator defined in the subband $B_\Delta$. Using the introduction of this appendix, we may decompose $I_G^{\mu,\Delta}$ into contributions corresponding to each sector $\sigma$ and to each equivalent class of nests $\Gamma$. We need to calculate

$$\prod_{K^i_j}\left[\frac{\partial}{\partial(\sigma^i_j/\beta_i)}\right]^{a^i_j}\prod_{H^i_j}\left[\frac{\partial}{\partial(\beta_i\chi^i_j)}\right]^{\omega_\Delta(H^i_j)+1}\Big\{\ \Big\}\Big|_{\sigma^i_j=0}$$

$$=\phi(p_i,m_a^2-\mu_a^2,\beta\chi)\exp\Big[-\sum_{a\in[G/\cup K^i_j]}(m_a^2-\mu_a^2)$$

$$\times\prod_{H^i_{r_i-1}\ni a}(\beta\chi^i_{r_i-1})^2\Big]\exp[-\beta\chi^l_1\mathscr{D}(p,0,\beta\chi)] \tag{C13}$$

where the curly bracket { } is the curly bracket in (II.1) when $P_G$ and $S_G$ are replaced by $[1+Q]$ and $\widetilde{S}_G$, and expressed in terms of $\beta_i\chi^i_j$ and $\sigma^i_j/\beta_i$. The function $\phi$ is a continuous function of the variables $\beta_i$ and is polynomially bounded when $\beta_l\to\infty$. We thus obtain a finite sum of integrals of the form

$$\int_0^\infty d\beta\beta_l^{2l+p_l-1}\prod_{i=1}^{l-1}\int_0^1 d\beta\beta_i^{2i+p_i-1}$$

$$\times\prod_{\substack{H^i_j\\ \omega_\Delta(H^i_j)\geqslant 0}}\int_0^1 d\chi^i_j(1-\chi^i_j)^{\omega_\Delta(H^i_j)}$$

$$\times\exp[-\beta_l^2M(\mu_a,m_a,\chi^l_1)]\phi'(p_i,m_a^2-\mu_a^2,\beta\chi). \tag{C14}$$

In (C14), the function $\phi'$ has the same properties than the function $\phi$ in (C13). The absolute convergence of (C14) is proved if $M(\mu_a,m_a,\chi^l_1)$ and Re$(2i+p_i)$ for $i=1,\ldots,l$ are strictly positive. In the mass term, we focus on that line $a_l$ which belongs to $R^l=G$ but is outside $R^{l-1}$ and which corresponds to the largest $\alpha$ in the sector. If this line belongs to $(\cup K^i_j)$, then $M(\mu_a,m_a,\chi^l_1)$ is $\mu_{a_l}^2>0$. If this line does not belong to $(\cup K^i_j)$ from (C2), it does not belong either to any $H^i_{r_i-1}$ for $i\neq l$; then, $M(\mu_a,m_a,\chi^l_1)$ is equal to $[m_{a_l}^2(\chi^l_1)^2+\mu_{a_l}^2(1-\chi^l_1)^2]$ which is always positive if $m_{a_l}$ and $\mu_{a_l}$ are positive. Now, by (C3) and (C9),

$$\text{Re}(2i+p_i)\geqslant\sum_{\substack{j=1\\ \omega_\Delta(H^i_j)<0}}^{r_i-1}[-\omega_\Delta(H^i_j)]+\sum_{\substack{j=1\\ \omega_\Delta(H^i_j)\geqslant 0}}^{r_i-1}(1)$$

$$+\sum_{j=1}^{r_i-1}[L(H^i_j)-L(K^i_j)](\Delta-\text{Re}D). \tag{C15}$$

For Re$D\leqslant\Delta$, since $H^i_j\supset K^i_j$ that is $L(H_j)\geqslant L(K^i_j)$, we have clearly (from the fact that $\mathscr{H}^i$ is never empty)

$$\text{Re}(2i+p_i)>0. \tag{C16}$$

Now, for $D\in B_\Delta$, the right-hand side of (C15) is larger or equal to

$$\sum_{\substack{j=1\\ \omega_\Delta(H^i_j)\geqslant 0}}^{r_i-1}[1+L(H^i_j)(\Delta-\text{Re}D)]+\sum_{\substack{j=1\\ \omega_\Delta(H^i_j)<0}}^{r_i-1}[-\omega_D(H^i_j)]. \tag{C17}$$

Now the band $B_D$ is such that $E[DL(\varphi)]=E[\Delta L(\varphi)]$ for all subgraphs $\varphi$, that is, for divergent subgraphs

$$\text{Re}D\cdot L(H^i_j)<\Delta L(H^i_j)+1, \tag{C18a}$$

and for convergent subgraphs

$$\omega_D(H^i_j)<0. \tag{C18b}$$

Again, $\mathscr{H}^i$ nonempty makes Re$(2i+p_i)>0$. This achieves the proof of absolute convergence of $I_G^{\mu,\Delta}$ for Re$D<\Delta$ and for $D\in B_\Delta$.

## 3. Absolute convergence of $I_G^{\mu=0,\Delta}$ for $D$ in the band $B_\Delta$ and $m_a>0$

We consider now the contribution to $I_G^{\mu=0,\Delta}$ of one sector $\sigma=\{\alpha_{a_1}<\alpha_{a_2}<\cdots<\alpha_{a_l}\}$4 and of one equivalent class of nests $\Gamma$. We now suppose that the lines $\alpha_l,\alpha_{l-1},\ldots,a_{k+1}$ belong to the $(\cup_{ij}K^i_j)$ and that $a_k$ does not belong to this union. By construction of the sectors, $R^{i<K}$ does not contain the line $a_k$ and since $K^{i<K}_{r_i-1}$ does not contain $a_K$, $H^{i<K}_{r_i-1}$ does not

contain $a_K$ either. On the other hand,

$$R^{i>K}\cup_{ij}K^i_j = G, \qquad \text{(C19)}$$

and also, from the convention $K^i_{r_i} = G' \supset G$ [15] and from (C2),

$$R^i \cup H^i_{r_i-1} = H^i_{r_i-1}. \qquad \text{(C20)}$$

Since the line $a_K \in R^{i>K}$, $a_K \in H^{i>K}_{r_i-1}$. Because all the subgraphs $H^i_j$ and $K^i_j$ form a nest, we conclude that $H^{i>K}_{r_i-1} \supset (\cup_{ij}K^i_j)$ and then, by (C19), (C20), $H^{i>K}_{r_i-1}$ is the graph $G$ itself. So, we have from (C2)

$$R^{i>K}\cup H^{i>K}_{r_i-1} = G, \qquad \text{(C21)}$$

and then $K^{i>K}_{r_i-1}$ is nonempty if $i \neq l$. The $\beta$ integrals may now be written

$$\int_0^\infty d\beta_l \beta_l^{2l+p_l-1} \prod_{i=1}^{l-1}\int_0^1 d\beta_i \beta_i^{2i+p_i-1} \times \prod_{\omega_\Delta(H^i_j)\geqslant 0}\int_0^1 d\chi^i_j(1-\chi^i_j)^{\omega_\Delta(H^i_j)}\phi'(p,m,\beta\chi) \times \exp\left\{-\prod_{i=K}^{l}(\beta_i\chi^i_{r_i-1})^2[m^2_{i_K}+\mathscr{E}(m,p,\beta\chi)]\right\}, \qquad \text{(C22)}$$

where $\mathscr{E}(m,p,\beta,\chi)$ is nonnegative and is continuous in the variables $\beta\chi$. From (C11) and for $i = l$, since $\mathscr{H}^l = \{G\}$ and $\mathscr{K}^l = \{G',\phi\}$,[15]

$$\text{Re}(2l+p_l) = 2l(G) - \text{Re}DL(G) - d(G) > 0$$

$$\text{if } G \text{ is } CV, \qquad \text{(C23a)}$$

$$1 > \text{Re}(2l+p_l) = L(G)[\Delta - \text{Re}D] + 1 > 0$$

$$\text{if } G \text{ is } DV. \qquad \text{(C23b)}$$

The integration over $\beta_l$ may be performed; we get

$$\prod_{i=1}^{l-1}\int_0^1 d\beta_i \prod_{i=K}^{l-1}\beta_i^{2i+p_i-2l-p_l-1}\prod_{i<K}\beta_i^{2i+p_i-1} \times \prod_{\omega_\Delta(H^i_j)\geqslant 0}\int_0^1 d\chi^i_j(1-\chi^i_j)^{\omega_\Delta(H^i_j)}\prod_{i=K}^{l}(\chi^i_{r_i-1})^{-(2l+p_l)} \times\phi'(p,m,\beta\chi)[m^2_{i_K}+\mathscr{E}(m,p,\beta\chi)]^{-(2l+p_l)/2}. \qquad \text{(C24)}$$

The $\chi^i_j$ integrals which occur only for divergent subgraphs $H^i_j$ (since $\chi^i_j = 1$ for convergent subgraphs $H^i_j$) are convergent by (C.23b). The $\beta_i$ integrals for $i < K$ are convergent by (C15), (C17), (C18). Now, the $\beta_i$ integrals for $K \leqslant i \leqslant l-1$ are convergent if $[\text{Re}(2i+p_i) - \text{Re}(2l+p_l)]$ is strictly positive; we have, since $H^{i>K}_{r_i-1} = R^l = G$,

$$\text{Re}(2i+p_i) - \text{Re}(2l+p_l) \geqslant -\sum_{j=1}^{r_i-1}L(K^i_j)(\Delta - \text{Re}D) + \left\{\sum_{\substack{j=1\\ \omega_D(H^i_j)<0}}^{r_i-2}[-\omega_\Delta(H^i_j)] + \sum_{\substack{j=1\\ \omega_\Delta(H^i_j)\geqslant 0}}^{r_i-2}[1+L(H^i_j)(\Delta-\text{Re}D)]\right\}. \qquad \text{(C25)}$$

By (C18), the curly bracket { } in (C25) is positive, or null if $\mathscr{H}^i = \{H^i_{r_i-1}\}$. On the other hand, $K^i_{r_i-1}$ is never empty for $K \leqslant i < l$ [see (C21)]; since all $K^i_j$ are divergent, $L(K^i_j)$ is $> 0$ [otherwise by (II.4), $\omega_\Delta(K^i_j) = -2l(K^i_j) < 0$]. Consequently, for $D \in B_\Delta$,

$$\text{Re}(2i+p_i) - \text{Re}(2l+p_l) > 0. \qquad \text{(C26)}$$

This achieves the proof of absolute convergence of $I_G^{\mu=0,\Delta}$ for $D \in B_\Delta$. It is interesting to note that the ultraviolet conditions of convergence $[\text{Re}(2i+p_i) > 0]$ are conditions on the right boundary of $B_\Delta$, while the infrared conditions of convergence $\{\text{Re}(2l+p_l) < 1, [\text{Re}(2i+p_i) - \text{Re}(2l+p_l)] > 0\}$ are conditions on the left boundary ($\text{Re}D = \Delta$) of $B_\Delta$.


[1]E.R. Speer and M.J. Westwater, Ann. Inst. H. Poincaré A **14**, 1 (1971).
[2]C.G. Bollini and J.J. Giambiagi, Phys. Lett. B **40**,566 (1972).; Acta Phys. Austr., **38**, 211 (1973).
[3]G.'t Hooft and M. Veltman, Nucl. Phys **B44**, 189 (1972).
[4]E.R. Speer, J. Math. Phys. **15**, 1 (1974).
[5]H.J. de Vega and F.A. Schaposnik, J. Math. Phys. **15**, 1998 (1974).
[6]J.C.Collins, Nucl. Phys. B **92**, 477 (1975).
[7]P.Breitenlohner and D.Maison, Commun. Math. Phys. **52**, 11 (1977).
[8]J. F. Ashmore, Commun. Math. Phys. **29**, 177 (1973).
[9]E.R. Speer, Commun. Math. Phys. **37**, 83, (1974); E.R. Speer, Dimensional and Analytic Renormalization," in the *Erice Summer School on "Renormalization Theory,"* edited by G.Velo and A.S. Wightman (1975).
[10]S.L. Adler, Phys. Rev. **177**, 2426 (1969); W. Bardeen, Phys. Rev. **184**, 1848 (1969).
[11]Y. M. P. Lam, Phys. Rev. D **6**, 2154, 2161 (1972) and private communication.
[12]K.Hepp, Commun. Math. Phys. **2**, 301 (1966).
[13]M.C. Bergére and Y.M.P. Lam, Commun. Math. Phys. **39**, 1 (1974); *"Asymptotic expansion of Feynman amplitudes, part II—the divergent case,"* Freie Universität Berlin, Preprint HEP, May 74/9 (1974).
[14]N.N. Bogoliubov and G. Parasiuk, Acta. Math. **97**, 227 (1957).
[15]M.C. Bergére and J.B. Zuber, Commun. Math. Phys. **35**, 113 (1974); M.C. Bergére and Y.M.P. Lam, J. Math. Phys. **17**, 1546 (1976).
[12]W. Zimmermann, Ann. Phys. **77**, 536 (1973).
[17]P. Breitenlohner and D. Maison, Commun. Math. Phys. **52**, 39, 55 (1977).
[18]F. David, Thése 3^ème^cycle, (to be published).

# An algorithm to construct evolution equations with a given set of conserved densities

A. Galindo

*Department of Theoretical Physics, Universidad Complutense de Madrid, Madrid-3, Spain*
(Received 4 August 1978)

This paper provides a simple formula to construct evolution equations $u_t = f(u,u_i,u_{ij},\cdots)$ having a prescribed finite set $\{\rho^{(1)}, \rho^{(2)},\cdots,\rho^{(N)}\}$ amongst its conserved densities. Besides its usefulness as a constructive algorithm, such a formula can also yield valuable information on the structure of the set $\mathscr{C}(f)$ of all conserved densities under a given evolution equation $u_t = f$.

## I. INTRODUCTION

In recent years a considerable effort in the study of nonlinear partial differential equations has produced most remarkable results both at the classical and quantum level, with the development of new and powerful techniques such as the inverse scattering method,[1] nonperturbative analysis of extended systems,[2] or the systematic application of Lie–Bäcklund invariance arguments,[3] amongst others. The physical interest of the subject is wide, ranging from nonlinear optics to hadron physics. Mathematically, this research area represents a real challenge where hard functional estimates usually combine with geometrical and physical insight to establish results which are not only rigorous but interesting as well.[4]

An important aspect in the theory and applications of nonlinear PDE's is the analysis of their conservation laws. Prominent equations, as the Korteweg–de Vries or the nonlinear cubic Schrödinger equations, are known to possess an infinite number of conserved currents.[5] That this may be exceptional should not surprise, since for instance such a simple evolution equation as $u_t = u_{xx}$ has essentially just *one* conserved density (namely $\rho = u$),[6] whereas the (nonlinear if $n \geqslant 2$) equation $u_t = u_{xx} + u^n$ has *none*.[7]

The aim of this paper is to provide a simple formula to construct evolution equations

$$u_t = f(u,u_i,u_{ij},\cdots) \tag{1}$$

having a prescribed finite set $\mathscr{R} = \{\rho^{(1)},\dots,\rho^{(N)}\}$ among its conserved densities. Apart from the clear interest it has as a constructive algorithm, such a formula can also yield fragmentary, yet valuable, information on the reciprocal, relevant and obviously more difficult question: Given (1), determine the set $\mathscr{C}(f)$ of all its conserved densities and properties thereof.

In Sec. II the notation is set up and our problem is precisely formulated. In Sec. III we derive and illustrate the algorithm for the simpler $(1+1)$-dimensional case; in addition, several applications are presented which partially bear on the aforementioned converse problem in some specific instances. Finally we generalize our algorithm to $(n+1)$-space–time in the last Sec. IV.

## II. FORMULATION OF THE PROBLEM

Let $u(\mathbf{x},t)$, $\mathbf{x} \equiv (x_1,\dots,x_n) \in \mathbb{R}^n$, $t \in \mathbb{R}$, stand for a (real valued) function which will be tacitly assumed sufficiently smooth ($C^\infty$ if necessary). We shall write $u_t \equiv \partial u/\partial t$, $u_{i_1\cdots i_r} \equiv \partial^r u/\partial x_{i_1}\cdots\partial x_{i_r}$ for its partial derivatives.

Let $\mathscr{F}_m$, $m \geqslant 0$, denote the quotient field of the ring of all (real valued) functions $g(u,u_i,\dots,u_{i_1\cdots i_m})$ depending (also smoothly) on the partial derivatives of $u$ up to order $m$. Any $f \in \overline{\mathscr{F}}_m \equiv \mathscr{F}_m - \mathscr{F}_{m-1}$ will be called of order $m$. We shall write $\mathscr{F} \equiv \cup_m \mathscr{F}_m$. The subsets of $\mathscr{F}_m, \overline{\mathscr{F}}_m, \mathscr{F}$ obtained upon restriction to functions $f(.,.,\dots)$ with polynomial dependence on their arguments will be denoted by $\mathscr{P}_m$, $\overline{\mathscr{P}}_m$, $\mathscr{P}$ respectively.

It is plain that the operator $\partial/\partial x_i (i = 1,\dots,n)$ extends naturally to $\mathscr{F}$ as the *total* derivative opertor $D_i$ in the $x_i$ direction

$$D_i \equiv \sum_{i_1\cdots i_k} u_{i_1\cdots i_k i} \frac{\partial}{\partial u_{i_1\cdots i_k}}. \tag{2}$$

Similarly, given an evolution equation (1), with $f \in \mathscr{F}$, the associated total derivative operator $D_t$ in the time direction

$$D_t \equiv \sum_{i_1\cdots i_k} \frac{\partial u_{i_1\cdots i_k}}{\partial t} \frac{\partial}{\partial u_{i_1\cdots i_k}}$$

$$= \sum_{i_1\cdots i_k} (D_{i_1\cdots i_k} f) \frac{\partial}{\partial u_{i_1\cdots i_k}} \equiv D_t^{(f)}, \tag{3}$$

where $D_{i_1\cdots i_k} \equiv D_{i_k}\cdots D_{i_1}$, will also map $\mathscr{F}$ into $\mathscr{F}$.

Clearly

$$[D_i, D_j] = [D_i, D_t] = 0, \quad \forall\, i,j \in \{1,\dots,n\}. \tag{4}$$

An element $\rho \in \mathscr{F}$ is called a *conserved density* for the evolution equation (1), where $f \in \mathscr{F}$, and we shall then write $\rho \in \mathscr{C}(f)$, whenever there exists $\mathbf{a} \equiv (a_1,\dots,a_n) \in \mathscr{F}$ (i.e., $a_j \in \mathscr{F}$) such that

$$D_t^{(f)} \rho = \mathbf{D}\cdot\mathbf{a} \left(\equiv \sum_j D_j a_j\right) \tag{5}$$

for every solution $u(\mathbf{x},t)$ of (1).

 0022-2488/79/061256-04$01.00  

It follows from (4) that $\mathbf{D}\cdot\mathbf{b}\in\mathscr{C}(f), \forall\mathbf{b}\in\mathscr{F}$, and hence from those $\rho\in\mathscr{F}$ which are total divergences represent conserved densities for all $f$. Such densities $\rho$ in the range of $\mathbf{D}$ are called *trivial*, and we will write $\rho\sim 0$. When sufficiently small at infinity, a trivial $\rho$ leads to vanishing total charge $\int\rho\, d^n x$.

Both $\mathscr{C}(f)$ and $\mathrm{Ran}\mathbf{D}$ are linear subspaces of $\mathscr{F}$. The linear dimension of the quotient $C(f)\equiv\mathscr{C}(f)/\mathrm{Ran}\mathbf{D}$ can be interpreted as the number of essentially linearly independent conserved densities for $u_t = f$. And so, $\dim C(u_{xx}) = 1$, $\dim C(u_{xx} + u^n) = 0$, whilst $\dim C(u_{xx} + uu_x)$ is infinite.

In terms of the variational derivative

$$\frac{\delta}{\delta u}\equiv\sum_{i_1\cdots i_k}(-1)^k D_{i_1\cdots i_k}\frac{\partial}{\partial u_{i_1\cdots i_k}} \tag{6}$$

it is well known[8] that $\rho\in\mathscr{C}(f)$ if and only if $f(\delta\rho/\delta u)\sim 0$. Therefore, since $(\delta/\delta u)D_j = 0$, if $\rho\in\mathscr{C}(f)$ then $(\delta/\delta u)[f(\delta\rho/\delta u)] = 0$. Conversely,[9] if $f(\delta\rho/\delta u)\in\mathrm{Ker}(\delta/\delta u)$ *and* $f(0,0,\cdots) = 0$, then $\rho\in\mathscr{C}(f)$.

Let now $\mathscr{R}\equiv\{\rho^{(1)},\dots,\rho^{(N)}\}$ be a finite subset of $\mathscr{F}$. The question arises which is the most general $f\in\mathscr{F}$ such that $\mathscr{R}\subset\mathscr{C}(f)$. This will constitute our *main* problem to be dealt with in the sequel. Clearly no generality will be lost by assuming $\rho^{(1)},\dots,\rho^{(N)}$ to be essentially linearly independent [$\equiv$ linearly independent modulo Ker $(\delta/\delta u)$].

For latter use the following notation will prove useful: Given $\mathscr{R}$, and taking first $n = 1$, let us recursively define

$$r_i\equiv\left(\frac{\delta\rho^{(i)}}{\delta u}\right)^{-1},\quad r_{ij}\equiv\left(D\frac{r_i}{r_j}\right)^{-1},$$

$\cdots$

$$r_{i_1\cdots i_m}\equiv\left(D\frac{r_{i_1\cdots i_{m-1}}}{r_{i_1\cdots i_{m-2}i_m}}\right)^{-1}, \tag{7}$$

where $D\equiv D_1$ (total derivative with respect to the single spatial variable $x_1\equiv x$), and $i, j(\neq i), i_j(\neq i_k$ if $j\neq k)$ are elements of $\{1,\dots,N\}$. Since $\rho^{(1)},\dots,\rho^{(N)}$ are supposed essentially linearly independent, no $r^{-1}_{i_1\cdots i_m}$ can identically vanish.

Similarly, for $n>1$, and $\boldsymbol{\alpha}_{i_1\cdots i_m}$ arbitrary elements of $\mathscr{F}$, define

$$r_i\equiv\left(\frac{\delta\rho^{(i)}}{\delta u}\right)^{-1},\quad \mathbf{r}_{ij}\equiv\left(\mathbf{D}\frac{r_i}{r_j}\cdot\boldsymbol{\alpha}_{ij}\right)^{-1}\boldsymbol{\alpha}_{ij},$$

$\cdots$

$$\mathbf{r}_{i_1\cdots i_m}\equiv\left(\mathbf{D}\frac{\mathbf{r}_{i_1\cdots i_{m-1}}\cdot\boldsymbol{\alpha}_{i_1\cdots i_{m-1}}}{\mathbf{r}_{i_1\cdots i_{m-2}i_m}\cdot\boldsymbol{\alpha}_{i_1\cdots i_{m-1}}}\cdot\boldsymbol{\alpha}_{i_1\cdots i_m}\right)^{-1}\boldsymbol{\alpha}_{i_1\cdots i_m}. \tag{8}$$

Again the expressions within parenthesis will not identically vanish for generically chosen $\boldsymbol{\alpha}$'s.

## III. THE (1 + 1)-DIMENSIONAL CASE

As announced in the Introduction, we shall now derive an algorithm to construct an evolution equation $u_t = f$, $f\in\mathscr{F}$, such that $\mathscr{C}(f)\supset\mathscr{R} = \{\rho^{(1)},\dots,\rho^{(N)}\}$, $u = u(x,t)$, $x\in\mathbb{R}$, and $\rho^{(1)},\dots,\rho^{(N)}$ are essentially linearly independent.

*Lemma* 1: $\rho^{(i)}\in\mathscr{C}(f)\Leftrightarrow f = r_i Dq_i,\quad q_i\in\mathscr{F}$. (Evident from our discussion in Sec. II.)

*Lemma* 2: $\rho^{(i)},\rho^{(j)}\in\mathscr{C}(f), i\neq j\Leftrightarrow f = r_i D[r_{ij}Dq_{ij}], q_{ij}\in\mathscr{F}$.

*Proof*: ($\Rightarrow$) Lemma 1 implies

$$f = r_i\, Dq_i = r_j\, Dq_j \tag{9}$$

and hence

$Dq_j = (r_i/r_j)Dq_i$.

Therefore, $q_j = (r_i/r_j)q_i - q_{ij}$, with $Dq_{ij} = [D(r_i/r_j)]q_i$,

whence $q_i = r_j Dq_{ij}$.

($\Leftarrow$) If $f = r_i D[r_{ij}\, Dq_{ij}]$, then (9) holds with $q_i\equiv r_{ij}\, Dq_{ij}$, $q_j\equiv(r_i/r_j)r_{ij}\, Dq_{ij} - q_{ij}$. It suffices now to apply Lemma 1.

*Lemma* 3: Let $(i_1,\dots,i_m)$, $m\geqslant 2$, be an arbitrarily ordered subset of $(1,\dots,N)$, and let $q_{i_1\cdots i_{m-1}}\in\mathscr{F}$. The following statements are equivalent:

$(a)_{i_1\cdots i_m}$: $\exists q_{i_m}\in\mathscr{F}$ such that

$$r_{i_1}D[r_{i_1i_2}D[r_{i_1i_2i_3}D[\cdots D[r_{i_1\cdots i_{m-1}}Dq_{i_1\cdots i_{m-1}}]\cdots]]] = r_{i_m}Dq_{i_m}$$

$(b)_{i_1\cdots i_m}$: $\exists q_{i_1\cdots i_m}\in\mathscr{F}$ such that

$q_{i_1\cdots i_{m-1}} = r_{i_1\cdots i_m}Dq_{i_1\cdots i_m}$.

*Proof*: For $m = 2$, it is a mere consequence of Lemma 2. Let us proceed inductively, and suppose thus that $(a)_{i_1\cdots i_{m-2}i_m}\Leftrightarrow(b)_{i_1\cdots i_{m-2}i_m}$.

$(a)_{i_1\cdots i_m}\Rightarrow(b)_{i_1\cdots i_m}$:

If $q_{i_1\cdots i_{m-2}}\equiv r_{i_1\cdots i_{m-1}}Dq_{i_1\cdots i_{m-1}}$, then $\exists q_{i_1\cdots i_{m-2}i_m}$, such that

$$q_{i_1\cdots i_{m-2}} = r_{i_1\cdots i_{m-2}i_m}Dq_{i_1\cdots i_{m-2}i_m}.$$

It suffices now to define

$$q_{i_1\cdots i_m}\equiv\frac{r_{i_1\cdots i_{m-1}}}{r_{i_1\cdots i_{m-2}i_m}}q_{i_1\cdots i_{m-1}} - q_{i_1\cdots i_{m-2}i_m}.$$

$(b)_{i_1\cdots i_m}\Rightarrow(a)_{i_1\cdots i_m}$:

If

$$q_{i_1\cdots i_{m-2}i_m}\equiv\frac{r_{i_1\cdots i_{m-1}}}{r_{i_1\cdots i_{m-2}i_m}}r_{i_1\cdots i_m}Dq_{i_1\cdots i_m} - q_{i_1\cdots i_m}$$

we get

$$q_{i_1\cdots i_{m-2}}\equiv r_{i_1\cdots i_{m-2}i_m}Dq_{i_1\cdots i_{m-2}i_m} = r_{i_1\cdots i_{m-1}}Dq_{i_1\cdots i_{m-1}}$$

and hence $\exists q_{i_m}$ such that

$$r_{i_1}D[r_{i_1i_2}D[\cdots D[r_{i_1\cdots i_{m-2}}Dq_{i_1\cdots i_{m-2}}]\cdots]] = r_{i_m}Dq_{i_m}.$$

This completes the proof.

*Theorem* 1: $\mathscr{R}\subset\mathscr{C}(f)$ if and only if

$$f = r_{i_1}D[r_{i_1i_2}D[r_{i_1i_2i_3}D[\cdots D[r_{i_1\cdots i_N}Dq_{i_1\cdots i_N}]\cdots]]],$$
$$q_{i_1\cdots i_N}\in\mathscr{F}, \tag{10}$$

where $(i_1,\dots,i_N)$ is an arbitrary permutation of $(1,\dots,N)$.

*Proof*: For $N = 1,2$ the claim follows from Lemmas 1 and 2. Let us argue by induction, assuming thus that our assertion holds when $\mathscr{R}$ has $N-1$ elements.

"*if*": Writing $q_{i_1\cdots i_{N-1}}\equiv r_{i_1\cdots i_N}Dq_{i_1\cdots i_N}$, the induction hypothesis yields immediately $\rho^{(i_1)}, \rho^{(i_2)},...,\rho^{(i_{N-1})}\in\mathscr{C}(f)$, and Lemma 3 implies $f = r_{i_N} Dq_{i_N}$ for some $q_{i_N}$, so that (Lemma 1) $\rho^{(i_N)}\in\mathscr{C}(f)$.

"*only if*": From $\rho^{(i_1)},...,\rho^{(i_{N-1})}\in\mathscr{C}(f)$ we have

$$f = r_{i_1}D\,[r_{i_1i_2}D\,[\cdots D\,[r_{i_1\cdots i_{N-1}}Dq_{i_1\cdots i_{N-1}}]\cdots]]$$

while $\rho^{(i_N)}\in\mathscr{C}(f)$ implies

$$f = r_{i_N}Dq_{i_N}.$$

Lemma 3 shows then that $q_{i_1\cdots i_{N-1}} = r_{i_1\cdots i_N}Dq_{i_1\cdots i_N}$ for some $q_{i_1\cdots i_N}\in\mathscr{F}$.

The proof is complete.

## A. Examples and applications

(1) Let $\mathscr{R} = \{(m!)^{-1}u^m, 1\leqslant m\leqslant N\}$. Then (7) leads to

$$r_1 = 1,\quad r_{12} = 1/Du,\cdots,\quad r_{12\cdots N} = 1/Du$$

and hence (10) yields

$$f = D\left[\frac{1}{u_1}D\left[\cdots D\left[\frac{1}{u_1}Dq_{12\cdots N}\right]\cdots\right]\right] \qquad (11)$$

as the most general expression for $f$ such that $\mathscr{R}\subset\mathscr{C}(f)$. (We write $u_n\equiv D^n u$.)

In particular, if

$$q_{12\cdots N} = \int_0^u dv_1\int_0^{v_1}dv_2\cdots\int_0^{v_{N-1}}dv_N\,g(v_N),$$

then $f = g(u)u_1$. Besides, $q_{1\cdots N}\in\overline{\mathscr{F}}_r, r>0\Rightarrow f\in\overline{\mathscr{F}}_{r+N}$.

Therefore,

*Corollary* 1: $\mathscr{P}_0\subset\mathscr{C}(f)\Leftrightarrow f\in D\mathscr{F}_0$.

For the purpose of illustration, we present the result of (11) for several simple choices of $q_{1\cdots N}$ ($\equiv q$ for short) and $N$:

$q = \frac{1}{2}u_1^2,\quad N = 2\rightarrow f = u_3,$

$q = \frac{1}{2}u_1^2 + \frac{1}{6}u^3,\quad N = 2\rightarrow f = u_3 + uu_1$(KdV),

$q = u_1^2u_2,\quad N = 2\rightarrow f = u_1u_4 + 5u_2u_3,$

$q = \frac{1}{4}u_1^4,\quad N = 3\rightarrow f = u_1u_4 + 5u_2u_3,$

$q = \frac{1}{6}u_1^6,\quad N = 4\rightarrow f = u_1^2u_5 + 13u_1u_2u_4 + 35u_2^2u_3 + 11u_1u_3^2,$

$q = u_1^6u_2,\quad N = 4\rightarrow f$
$= u_1^3u_6 + 24u_1^2u_2u_5 + 55u_1^2u_3u_4 + 165u_1u_2^2u_4 + 280u_1u_2u_3^2$
$+ 315u_2^3u_3.$

(2) Let $\mathscr{R} = \{-[(m+1)!]^{-1}u_1^{m+1},\ 1\leqslant m\leqslant N\}$. Then (10) leads to

$$f = \frac{1}{u_2}D\left[\frac{1}{u_2}D\left[\cdots D\left[\frac{1}{u_2}Dq_{12\cdots N}\right]\cdots\right]\right]. \qquad (12)$$

Thus, for example, if $N = 2$ and $q_{12} = \frac{1}{4}u_2^4$, then

$$f = u_2u_4 + 2u_3^2.$$

The choice

$$q_{1\cdots N} = \int_0^{u_1}dv_1\int_0^{v_1}dv_2\cdots\int_0^{v_{N-1}}dv_N\,g(v_N)$$

yields $f = g(u_1)$, whereas $q_{1\cdots N}\in\overline{\mathscr{F}}_r, r>1\Rightarrow f\in\overline{\mathscr{F}}_{r+N}$, and $q_{1\cdots N}\in\mathscr{F}_1, (\partial/\partial u)q_{1\cdots N}\not\equiv 0\Rightarrow f\in\overline{\mathscr{F}}_{1+N}$. So we have

*Corollary* 2: $\mathscr{C}(f)$ contains all polynomials in $u_1$ if and only if $f = f(u_1)$.

(3) Let $\mathscr{R} = \{u, -\frac{1}{2}u_1^2\}$. Now (10) becomes

$$f = D\,[(1/u_3)Dq_{12}]. \qquad (13)$$

In particular

$$q(\equiv q_{12}) = u\rightarrow f = \frac{u_2u_3 - u_1u_4}{u_3^2},$$

$$q = \tfrac{1}{2}u_2^2\rightarrow f = u_3,$$

$$q = u_3^2u_4\rightarrow f = u_3u_6 + 5u_4u_5.$$

Furthermore, it is easy to see from (13) that $\nexists f\in\overline{\mathscr{P}}_4$ such that $u_1, u_1^2\in\mathscr{C}(f)$.

(4) Let $f\in\overline{\mathscr{F}}_M$, $M\geqslant 2$, $\mathscr{R} = \{\rho^{(1)},\rho^{(2)}\}\subset\mathscr{F}_{\overline{M}}$, where $\overline{M}\equiv[M/2-1]$ ($[s]$ denoting the integer part of $s$). In order to have $\mathscr{R}\subset\mathscr{C}(f)$, Theorem 1 shows that *necessarily*

$$f = r_1D\,[r_{12}Dq_{12}]$$

with $q_{12}\in\mathscr{F}_{M-2}$, i.e.,:

$$f = \frac{au_M + bu_{M-1}^3 + cu_{M-1}^2 + du_{M-1} + e}{(a'u_{M-1} + b')^2}, \qquad (14)$$

where $a,...,e,a',b'\in\mathscr{F}_{M-2}$, and $a' = 0\Rightarrow b = 0$. In particular, if we impose in addition that $f$ be polynomic, then necessarily

$$f = au_M + bu_{M-1}^2 + cu_{M-1} + d \qquad (15)$$

with $a,b,c,d\in\mathscr{F}_{M-2}$. Therefore, we have

*Corollary* 3: $\mathscr{R} = \{\rho^{(1)},\rho^{(2)}\}\subset\mathscr{F}_{\overline{M}}\cap\mathscr{C}(f)$, $f\in\overline{\mathscr{F}}_M$ (resp. $\overline{\mathscr{P}}_M$), $M\geqslant 2$, only if $f$ is of the form (14) [resp. (15)].

*Remark*: For $M$ even $\geqslant 2$, it is known[7] that $\mathscr{F}_{\overline{M}}\cap\mathscr{C}(f) = \mathscr{C}(f)$ modulo Ran$D$, and thus (14) [(15)] gives the most general expression for $f\in\mathscr{F}_M$ ($\mathscr{P}_M$) such that $u_t = f$ might have at least two essentially linearly independent conserved densities.

(5) Let $\mathscr{R} = \{\rho^{(i)}, 1\leqslant i\leqslant N\}\subset\mathscr{F}_0$. It follows from (10) that

(a) $q_{1\cdots N}\in\mathscr{F}_0\Rightarrow f\in D\mathscr{F}_0$.

Conversely, it is clear that $\mathscr{F}_0\subset\mathscr{C}(f)$ for any $f\in D\mathscr{F}_0$.

(b) $q_{1\cdots N}\in\mathscr{F}_{M-N}$, $M>N\Rightarrow f = au_M + b$, $a\in\mathscr{F}_{M-N}$, $b\in\mathscr{F}_{M-1}$

Therefore, given $f\in\overline{\mathscr{F}}_M$, $M\geqslant 2$, the maximal number of essentially linearly independent conserved densities $\rho\in\mathscr{F}_0$ under $u_t = f$ is $M-1$. This bound is optimal, even for polynomials: if $\mathscr{R} = \{(m!)^{-1}u^m, 1\leqslant m\leqslant N\}$, $N\geqslant 2$, then the choice $q_{1\cdots N} = (2(N-1))^{-1}u_1^{2(N-1)}$ leads to a polynomial $f = u_1^{N-2}u_{N+1}$ + (lower order terms); besides, $u\in\mathscr{C}(u_2)$.

Similar arguments show finally that if $f\in\overline{\mathscr{F}}_M$, $M\geqslant 2$, the maximal number of ess. l.i. conserved densities $\rho(u_1)$ under $u_t = f$ is again $M-1$. However, for $f\in\overline{\mathscr{P}}_M$, such bound is $M-2$, which is also optimal.

## IV. GENERALIZATION TO THE $(n+1)$-DIMENSIONAL CASE

It is rather straightforward (so we omit the proof) to extend the previous theorem to the case where $\mathbf{x}\in\mathbb{R}^n, n > 1$, with the result

*Theorem* 2: $\mathscr{R}\subset\mathscr{C}(f)$ if and only if

$$f = r_{i_1}\mathbf{D}\cdot[\mathbf{r}_{i_1 i_2}\mathbf{D}\cdot[\mathbf{r}_{i_1 i_2 i_3}\mathbf{D}\cdot[\cdots\mathbf{D}\cdot[\mathbf{r}_{i_1\cdots i_N}\mathbf{D}\cdot q_{i_1\cdots i_N}]\cdots]]], \quad (16)$$

where $\mathbf{q}_{i_1\cdots i_N}\in\mathscr{F}$, $(i_1,\ldots,i_N)$ is any permutation of $(1,\ldots,N)$, and the $\mathbf{r}$'s are obtained through (8) for some set of $\boldsymbol{\alpha}$'s.

Let us illustrate (16) with a couple of examples:

(1) $\mathscr{R} = \{u, \frac{1}{2}u^2\}$. Then (16) becomes

$$f = \mathbf{D}\cdot\left[\frac{\boldsymbol{\alpha}_{12}}{\mathbf{D}u\cdot\boldsymbol{\alpha}_{12}}\mathbf{D}\cdot\mathbf{q}_{12}\right]. \quad (17)$$

So, if

$$\boldsymbol{\alpha}_{12} = (\Delta u)\mathbf{D}u + \mathbf{D}(\mathbf{D}u\cdot\mathbf{D}u), \quad \Delta \equiv \mathbf{D}\cdot\mathbf{D},$$

$$\mathbf{q}_{12} = (\mathbf{D}u\cdot\mathbf{D}u)\mathbf{D}u,$$

(17) yields

$$f = (\Delta u)^2 + 3\mathbf{D}u\cdot\mathbf{D}\Delta u + 2\sum_{ij}(D_iD_ju)^2 \quad (18)$$

which is invariant under SO($n$). Simpler $f$'s can be obtained if this rotational invariance is not required:

$$\boldsymbol{\alpha}_{12} = \mathbf{s}\in\mathbb{R}^n, \quad \mathbf{q}_{12} = \tfrac{1}{2}(\mathbf{s}\cdot\mathbf{D}u)^2\mathbf{s}\rightarrow$$

$$f = (\mathbf{s}\cdot\mathbf{D})^3u$$

$$\boldsymbol{\alpha}_{12} = \mathbf{s}\in\mathbb{R}^n, \quad \mathbf{q}_{12} = \frac{1}{3}(\mathbf{s}\cdot\mathbf{D}u)^3\mathbf{s}\rightarrow$$

$$f = [(\mathbf{s}\cdot\mathbf{D})^2u]^2 + (\mathbf{s}\cdot\mathbf{D}u)[(\mathbf{s}\cdot\mathbf{D})^3u]. \quad (19)$$

By averaging (19) over all directions of $\mathbf{s}$ the expression (18) results.

(2) $\mathscr{R} = \{u, -\frac{1}{2}\mathbf{D}u\cdot\mathbf{D}u\}$. Then (16) leads to

$$f = \mathbf{D}\cdot\left[\frac{\boldsymbol{\alpha}_{12}}{\mathbf{D}\Delta u\cdot\boldsymbol{\alpha}_{12}}\mathbf{D}\cdot\mathbf{q}_{12}\right] \quad (20)$$

which is the same as (17) with $u$ replaced by $\Delta u$. Therefore, just as for (18), we get

$$f = (\Delta^2u)^2 + 3\mathbf{D}\Delta u\cdot\mathbf{D}\Delta^2u + 2\sum_{ij}(D_iD_j\Delta u)^2 \quad (21)$$

and so on.

## ACKNOWLEDGMENTS

We are grateful to Professor L. Abellanas for many fruitful discussions. The partial financial support of the Instituto de Estudios Nucleares is also acknowledged.

[1]P.D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968); M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1462 (1973); **31**, 125 (1973), "The Inverse Scattering Transform-Fourier Analysis for Non-Linear Problems," Stud. Appl. Math. **3**, 249–315 (1974); V.E. Zakharov and A.B. Shabat, Sov. Phys.-JETP **34**, 62 (1972); **37**, 823 (1974); F. Calogero, "The Spectral Transform: A Tool to Solve and Investigate Nonlinear Evolution Equations," Proc. 1978 GIFT International Seminar (to be published).

[2]See "Extended Systems in Field Theory," edited by J.L. Gervais and A. Neveu, Phys. Rep. C **23**, 236 (1976).

[3]L. V. Ovsjannikov, *Group Properties of Differential Equations* Novosibirsk, Moscow (1962); G.W. Bluman and J.D. Cole, *Similarity Methods for Differential Equations* (Springer, Berlin, 1974); R.L. Anderson, S. Kumei, and C.E. Wulfman, Phys. Rev. Lett. **28**, 988 (1972); Rev. Mex. Fis. **21**, 1, 35 (1972); J. Math. Phys. **14**, 1527 (1973); N.H. Ibragimov, Lett. Math. Phys. **1**, 423 (1977); S. Kumei, J. Math. Phys. **18**, 256 (1977).

[4]W. Strauss, "Nonlinear Scattering Theory," in *Scattering Theory in Mathematical Physics*, edited by J.A. Lavita and J.P. Marchand (Reidel, New York, 1973); "Nonlinear Invariant Wave Equations" in 1977 *Erice Proceedings* edited by G. Velo and A.S. Wightman (Springer, Berlin 1978); M. Reed, *Abstract Nonlinear Wave Equations*, Springer Lecture Notes in Mathematics, No. 507 (Springer, Berlin 1976); see also the contributions by J. Fröhlich, K. Pohlmeyer, and M. Reed, in *Many Degrees of Freedom in Field Theory*, edited by L. Streit (Plenum, New York, 1978).

[5]R.M. Miura, C.S. Gardner, and M.D. Kruskal, J. Math. Phys. **9**, 1204 (1968); V.E. Zakharov and A.B. Shabat, Sov. Phys.-JETP **34**, 62 (1972).

[6]L. Abellanas and A. Galindo, Lett. Math. Phys. **2**, 399 (1978).

[7]L. Abellanas and A. Galindo, J. Math. Phys. **20**, 1239 (1979).

[8]I.M. Gel'fand and L.A. Dikii, Russ. Math. Surveys **30**, (5), 77 (1975).

[9]A. Galindo and L. Martínez, Lett. Math. Phys. **2**, 385 (1978).

# New expressions for the eigenvalues of the invariant operators of O(N) and Sp(2n)

C. O. Nwachuku[a)]

*International Centre for Theoretical Physics, Trieste, Italy*
(Received 22 August 1978)

In the spirit of the recent work of Popov for u(n), we derive a *direct* expansion of the eigenvalues of the invariant operators $C_p$ for the orthogonal and symplectic groups in terms of the power sums with completely specified coefficients $\beta_p(\nu)$, which are easy to compute. The resulting expression, which is a complete analog of the u(n) results, is closed, simple and manifests the general structure of the $C_p$. It is now possible to say for what value of $p$ a particular combination of the $S_k$'s begin to appear. Explicit applications of these results in computing the $C_p$ for $p \leq 8$ illustrate fully their simplicity. Thus our work simplifies and unifies the treatment of this aspect of the problem for the semisimple Lie groups.

## I. INTRODUCTION

The unitary, orthogonal, and symplectic groups which form the various series of the semisimple Lie groups, play an important role in the understanding of several branches of modern physics. The invariant operators of these groups (also called Casimir, Gel'fand operators) are useful in this respect. Their eigenvalues are used to label the irreducible representations of the groups, and can be identified with the quantum numbers of some physical observables. The problem of constructing generalized Casimir invariants for the semisimple Lie groups now seems completely solved. The colateral problem of finding simple methods of computing their eigenvalues continues to receive attention.[1–9]

A remarkable simplification was introduced into this problem by Perelomov and Popov[2] who working in tensor basis (rather than the complicated Cartan basis), were able to express the eigenvalues of the operators $C_p$ as sums of elements of a matrix raised to power $p$. By diagonalizing this matrix for u(n) they obtained a closed expression for the eigenvalues of the $C_p$ in terms of the integers $\lambda_i$ which characterize the irreducible (tensor) representations. A similar closed form [Eq. (2.2)] was obtained for the orthogonal and symplectic groups by Rashid and the present author[7] (also see Ref. 9). These forms though convenient for studying the structures of the $C_p$ do not manifest its polynomial dependence on the $\lambda_i$'s. Moreover, practical calculations using them (even for small values of $p$) appear tedious for any but very low values of $n$.

Further simplification to this problem was achieved[3,8] by introducing suitable generating functions, which enable a polynomial expansion of the $C_p$ in terms of power sums $S_k$ defined as functions of the $\lambda_i$'s [see Eq. (17) of Ref. 3 and (5.3) of Ref. 8]. However, these expansions are expressed indirectly through the quantities $B_q$, introduced in the text. In this form a great deal of work needs to be done to obtain, for instance, the coefficient of an arbitrary $S_k$ (or an arbitrary combination of the products of $S_k$'s) in the eigenvalue of $C_p$ for any $p$. Moreover this form of the expansion lacks coherence and conceals the general structure of the $C_p$. Realizing these defects, Popov, in a recent paper,[10] developed further the applications of the generating functions technique to the case of the unitary groups, obtaining a *direct* expansion of the $C_p$ in terms of the power sums $S_k$ (and various combinations of products of $S_k$) with easily obtainable coefficients $\beta_p(\nu)$ [Eq. (3.24)]. This results in considerable simplification of the calculations and clarifies the general structure of the $C_p$.

In this paper we aim at solving the corresponding problem for the cases of the orthogonal and symplectic groups. We obtain in Eq. (3.22) a direct expansion of the eigenvalues of the $C_p$ in terms of power sums with completely specified coefficients $\beta_p(\nu)$, which are relatively easy to determine. This equation is the complete analog of the expansion of Popov for u(n). The simplicity of our results is illustrated in the particular cases treated. Specific values of the $C_p$ for $p \leqslant 8$ computed using this equation are displayed in Table II. Thus our result simplifies and unifies the treatment of this aspect of the problem for the semisimple Lie groups.

A brief summary of previous results needed in what follows, together with some definitions, is provided in Sec. II of this paper. Section III contains the derivation of the main expansion, the applications of which are illustrated in Sec. IV. Section V gives the concluding remarks. There is an appendix which demonstrates the proof of a useful identity.

## II. SUMMARY

We define the invariant operators of order $p$ of the orthogonal group O(N) ($N$ stands for $2n$ or $2n+1$) and the symplectic group Sp(2n) by

$$C_p = \sum_{i_1, i_2, \ldots, i_p} X^{i_1}_{i_2} X^{i_2}_{i_3} \cdots X^{i_p}_{i_1}, \tag{2.1}$$

where $X^i_j$ ($i,j = -n,\ldots,n$) are the infinitesimal generators of the groups whose commutations relations are specified in, for instance, Ref. 7. Here the indices $i,j$ include zero only for the case of the O(2n + 1). There are $N(N-1)/2$, $n(2n+1)$ independent generators for the O(N) and the Sp(2n), respectively.

---

[a)]Permanent address: Department of Mathematics, University of Benin, Benin City, Nigeria.

 0022-2488/79/061260-07$01.00  

The irreducible (tensor) representations of these groups are characterized by $n$ integers $f_i$ (ordered such that $f_n \geqslant f_{n-1} \geqslant \cdots \geqslant f_1$), which correspond to the eigenvalues of the $n$-independent diagonal generators on the highest weight of the representation, $n$ being the rank of the group. It has been shown that the eigenvalues of $C_p$ for the O(2$n$), Sp(2$n$), and O(2$n$ + 1) in this order, can be written in the form[7-9]

$$C_p = \sum_{i=-n}^{n} \lambda_i^p \left\{ \begin{array}{c} \dfrac{\lambda_i - n + 1}{\lambda_i - n + \frac{1}{2}} \\ \dfrac{\lambda_i - n - 1}{\lambda_i - n - \frac{1}{2}} \\ \dfrac{\lambda_i - n - \frac{1}{2}}{\lambda_i - n - 1} \end{array} \right\}$$

$$\times \prod_{\substack{j=-n \\ \neq i}}^{n} \left(1 - \frac{1}{\lambda_i - \lambda_j}\right), \qquad (2.2)$$

where the summation and the product include zero only for the O(2$n$ + 1). In Eq. (2.2)

$$\lambda_i = \begin{cases} f_i + n + i - (1 + \epsilon_i), & \text{for the O}(2n) \\ f_i + n + i, & \text{for the Sp}(2n) \\ f_i + n + i - \theta_{0i}, & \text{for the O}(2n+1) \end{cases}, \qquad (2.3)$$

where

$$\epsilon_i = 1, \ -1, 0 \quad \text{for } i > 0, i < 0, \text{and } i = 0, \text{ respectively,}$$

and

$$\theta_{ji} = 1(0) \quad \text{for } j < i \ (j \geqslant i).$$

We define the power sums $S_k$, for nonnegative integers $k$, by

$$S_k = \sum_{i=-n}^{n} (\lambda_i^k - \rho_i^k) \quad (S_0 = S_1 = 0), \qquad (2.4)$$

where

$$\rho_i = \lambda_i - f_i = n + i - (1 + \epsilon_i), n + i, n + i - \theta_{0i}, \qquad (2.5)$$

for the O(2$n$), Sp(2$n$), and O(2$n$ + 1) groups, respectively. This definition is related to that of Ref. 2, p. 1127, Eq. (19), as follows,

$$\lambda_i = l_i + \alpha \quad \text{and} \quad \rho_i = r_i + \alpha, \qquad (2.6)$$

where $\alpha = n - 1, n, n - \frac{1}{2}$ for the O(2$n$), Sp(2$n$), and O(2$n$ + 1), respectively. The definitions of this reference has the advantage that the symmetry of the Weyl $S$ group (the group of reflections in hyperplanes perpendicular to the root vectors) can be most conveniently applied only when $C_p$ is expressed in terms of the variables $l_i$. However, our definition has the advantage that the $\rho_i$'s are precisely the values assumed by the $\lambda_i$'s for the identity representation. This property is used in the derivation of the results below.

## III. DERIVATION OF THE EXPANSION

From now on we concentrate on the case of the O(2$n$) and Sp(2$n$) groups. The final results for the O(2$n$ + 1) are obtained from those of the O(2$n$) by [11] replacing $n$ by $n + \frac{1}{2}$, according to the ansatz prescribed on p. 1389 of Ref. 8. The upper (lower) sign stands for the O(2$n$) [Sp(2$n$)] throughout.

Equation (2.2) can be transformed into a contour integral in the $\lambda$ plane,

$$C_p = -\frac{1}{2\pi i} \oint d\lambda\, \lambda^p \frac{\lambda - n \pm 1}{\lambda - n \pm \frac{1}{2}}$$

$$\times \prod_{i=-n}^{n} \left(1 - \frac{1}{\lambda - \lambda_i}\right) \pm \tfrac{1}{2}(n \mp \tfrac{1}{2})^p, \qquad (3.1)$$

the integration being taken in a positive sense along any large circle with origin as center and containing all the poles of the integrand. The additional term takes care of the poles at $\lambda = n \mp \frac{1}{2}$ that have no corresponding terms in Eq. (2.2).

On making the substitution $\lambda = 1/z$, Eq. (3.1) becomes

$$C_p = -\frac{1}{2\pi i} \oint \frac{F(z)dz}{z^{p+2}} \pm \tfrac{1}{2}(n \mp \tfrac{1}{2})^p, \qquad (3.2)$$

where

$$F(z) = \frac{1 - (n \mp 1)z}{1 - (n \mp \frac{1}{2})z} \prod_{i=-n}^{n} \left(1 - \frac{z}{1 - \lambda_i z}\right). \qquad (3.3)$$

The integration in Eq. (3.2) is in a positive sense along any small circle with center at the origin but excluding all the poles of $F(z)$.

From Eq. (3.3)

$$F(z) = \ln[1 - (n \mp 1)z] - \ln[1 - (n \mp \tfrac{1}{2})z] - \sum_{k=1}^{\infty} \frac{z^k}{k} \sum_{i=-n}^{n} [(\lambda_i + 1)^k - \lambda_i^k].$$

The sum $\Sigma_{i=-n}^{n}[(\lambda_i + 1)^k - \lambda_i^k]$. can be expressed[8] in the form

$$\sum_{l=0}^{k-1} \binom{k}{l} \sum_{i=-n}^{n} (\lambda_i^l - \rho_i^l) + \sum_{i=-n}^{n} [(\rho_i + 1)^k - \rho_i^k].$$

$$= \sum_{l=0}^{k-1} \binom{k}{l} S_l + (2n \mp 1)^k + n^k - (n \mp 1)^k,$$

so that

$$\ln F(z) = -\sum_{k=1}^{\infty} \frac{z^k}{k} \sum_{l=0}^{k-1} \binom{k}{l} S_l - \ln[1 - (n \mp \tfrac{1}{2})z]$$

$$+ \ln[1 - (2n \mp 1)z] + \ln[1 - nz]. \qquad (3.4)$$

Now let $f(z)$ be defined by

$$F(z) = F_0(z) f(z), \qquad (3.5)$$

where $F_0(z)$ is the function $F(z)$ for the identity representation given by $\lambda_i = \rho_i$. [For this representation $S_k = 0$ for all $k$.]

For the O(2$n$),

$$F_0(z) = \frac{1 - (n-1)z}{1 - (n - \frac{1}{2})z} \prod_{i=-n}^{n} \left(1 - \frac{z}{1 - (n + i - 1 - \epsilon_i)z}\right),$$

so that

$$\ln F_0(z)$$
$$= \ln[1-(n-1)z] - \ln[1-(n-\tfrac{1}{2})z] + \sum_{i=-n}^{n} \{\ln[1-(n+i-\epsilon_i)z] - \ln[1-(n+i-1-\epsilon_i)z]\}$$
$$= \ln[1-(2n-1)z] + \ln[1-nz] - \ln[1-(n-\tfrac{1}{2})z]. \quad (3.6)$$

Similarly for the Sp(2$n$)

$$\ln F_0(z) = \ln(1-(2n+1)z] + \ln[1-nz] - \ln[1-(n+\tfrac{1}{2})z]. \quad (3.7)$$

Equations (3.4)–(3.7), for both cases, result in

$$\ln f(z) = -\sum_{k=1}^{\infty} \frac{z^k}{k} \sum_{l=0}^{k-1} \binom{k}{l} S_l. \quad (3.8)$$

The right-hand side of Eq. (3.8) is the same as

$$-\sum_{k>1} \frac{z^k}{k} \sum_{l=0}^{\infty} \binom{k}{l} S_l$$
$$= -\sum_{l=0}^{\infty} z^l S_l \sum_{r=1}^{\infty} \binom{l+r}{l} \frac{z^r}{l+r}$$
$$= -\sum_{k=0}^{\infty} z^k S_k \sum_{l=1}^{\infty} \binom{k+l}{k} \frac{z^l}{k+l}$$
$$= -\sum_{k=2}^{\infty} S_k z^{k+1} \phi_k(z) \quad (S_0 = S_1 = 0),$$

where

$$\phi_k^{(z)} = \sum_{l=0}^{\infty} \frac{(k+l)! z^l}{k!(l+1)!} = \frac{1}{kz}[(1-z)^{-k}-1]. \quad (3.9)$$

Therefore,

$$f(z) = \exp\left\{-\sum_{k=2}^{\infty} S_k z^{k+1} \phi_k(z)\right\}$$
$$= \prod_{k=2}^{\infty} \sum_{\nu_k=0}^{\infty} z^{(k+1)\nu_k} \frac{(-)^{\nu_k}}{\nu_k!} [S_k \phi_k(z)]^{\nu_k}. \quad (3.10)$$

A typical product term of the form $S_2^{\nu_2} S_3^{\nu_3} \cdots S_k^{\nu_k}$ occurs in the expansion of Eq. (3.10) with a factor

$$\frac{(-)^{\bar{\nu}}}{[\nu!]} [\phi_2(z)]^{\nu_2} [\phi_3(z)]^{\nu_3} \cdots [\phi_k(z)]^{\nu_k} z^{K+1},$$

where $\nu_2, \nu_3, \cdots, \nu_k$ are nonnegative integers satisfying

$$K+1 = 3\nu_2 + 4\nu_3 + \cdots + (k+1)\nu_k \quad (3.11)$$

and

$$\bar{\nu} = \nu_2 + \nu_3 + \cdots + \nu_k, \quad [\nu!] = \nu_2! \nu_3! \cdots \nu_k!,$$

so we can write

$$f(z) = \sum_{l=0}^{\infty} \sum_{(\nu)} \frac{(-)^{\bar{\nu}}}{[\nu!]} Q_l(\nu) S_2^{\nu_2} S_3^{\nu_3} \cdots S_k^{\nu_k} z^{k+l+1}, \quad (3.12)$$

where $Q_l(\nu)$ is defined by

$$\sum_{l=0}^{\infty} Q_l(\nu) z^l = [\phi_2(z)]^{\nu_2} [\phi_3(z)]^{\nu_3} \cdots [\phi_k(z)]^{\nu_k} \quad (3.13)$$

and $(\nu)$ means the set of nonnegative integers satisfying the constraint (3.11).

We next introduce the Casimir operator $C_p$ by means of the generating function $G(z)$ defined by

$$G(z) = \sum_{p=0}^{\infty} C_p z^p$$

which, by virtue of Eq. (3.2), implies the relation

$$zG(z) = \sum_{p=0}^{\infty} C_p z^{p+1}$$
$$= 1 - F(z) \pm \tfrac{1}{2} \sum_{p=0}^{\infty} (n \mp \tfrac{1}{2})^p z^{p+1}.$$

Thus

$$1 - \sum_{p=0}^{\infty} [C_p \mp \tfrac{1}{2}(n \mp \tfrac{1}{2})^p] z^{p+1}$$
$$= F(z) = F_0(z) f(z)$$
$$= \frac{[1-(2n\mp 1)z][1-nz]}{[1-(n\mp\tfrac{1}{2})z]} f(z) \quad (3.14)$$

[the last step follows from Eqs. (3.6) and (3.7)].

Now from Eq. (3.12)

$$\frac{[1-(2n\mp 1)z](1-nz)}{[1-(n\mp\tfrac{1}{2})z]} f(z)$$
$$= \frac{1}{[1-(n\mp\tfrac{1}{2})z]} \Big( \sum_{l=0}^{\infty} \sum_{(\nu)} \frac{(-)^{\bar{\nu}}}{[\nu!]} \times [Q_l(\nu) - (3n\mp 1)Q_{l-1}(\nu) + n(2n\mp 1)Q_{l-2}(\nu)] \times S_2^{\nu_2} S_3^{\nu_3} \cdots S_k^{\nu_k} z^{k+l+1} \Big)$$
$$= \sum_{l=0}^{\infty} \Big\{ \sum_{(\nu)} \frac{(-)^{\bar{\nu}}}{[\nu!]} S_2^{\nu_2} S_3^{\nu_3} \cdots S_k^{\nu_k} \sum_{m=0}^{l} [Q_m(\nu) - (3n\mp 1)Q_{m-1}(\nu) + n(2n\mp 1)Q_{m-2}(\nu)] \times (n\mp\tfrac{1}{2})^{l-m} \Big\} z^{k+l+1} \quad (Q_{-1} = Q_{-2} = 0). \quad (3.15)$$

Thus comparing coefficients of $z^{p+1}$ in Eq. (3.14), using Eq. (3.15), we get

$$C_p = \pm \tfrac{1}{2}(n\mp\tfrac{1}{2})^p - \sum_{(\nu)} \beta_p(\nu) S_2^{\nu_2} S_3^{\nu_3} \cdots S_k^{\nu_k} \quad (p \geqslant 0), \quad (3.16)$$

where

$$\beta_p(\nu) = \frac{(-)^{\bar{\nu}}}{[\nu!]} \sum_{m=0}^{l} [Q_m(\nu) - (3n\mp 1)Q_{m-1}(\nu) + n(2n\mp 1)Q_{m-2}(\nu)](n\mp\tfrac{1}{2})^{l-m} \quad (l = p-K \geqslant 0). \quad (3.17)$$

Equation (3.16) is the preliminary form of the expansion we set out to derive. Before casting it into its final form, we need to consider first some particular cases.

First, from the definition (3.13)

$$Q_0(\nu) = 1 \quad \text{for all } (\nu), \quad Q_l(0) = 0 \quad \text{for all } l \geqslant 1. \tag{3.18}$$

For $p = 0$ only $(\nu) = (0)$, i.e., $\nu_k = 0$ for all $k$ is a solution to (3.11) corresponding to $K = -1$ or $l = 1$. Thus from (3.16)–(3.18)

$$C_0 = 2n. \tag{3.19}$$

It is not difficult to prove that for all $p \geqslant 1$, the contribution to $C_p$ in Eq. (3.16) corresponding to the zero set (0), i.e., to $K = -1$ vanishes.

Indeed, for this case $l = p + 1 \geqslant 2$. The contribution due to $K = -1$ is

$$\begin{aligned} C_p(K = -1) &= \pm \tfrac{1}{2}(n \mp \tfrac{1}{2})^p - [(n \mp \tfrac{1}{2})^{p+1} - (3n \mp 1)(n \mp 1)^p \\ &\quad + n(2n-1)(n \mp \tfrac{1}{2})^{p-1}] \\ &= (n \mp \tfrac{1}{2})^{p-1}[\pm \tfrac{1}{2}(n \mp \tfrac{1}{2}) - (n \mp \tfrac{1}{2})^2 \\ &\quad + (3n \mp 1)(n \mp \tfrac{1}{2}) - n(2n-1)] \\ &= 0. \end{aligned} \tag{3.20}$$

As a corollary, since only $K = -1$ contributes when $p = 1$, it follows that

$$C_1 = 0. \tag{3.21}$$

Incorporating Eq.s (3.20) and (3.21) into Eq. (3.16) we finally obtain

$$C_p = -\sum_{(\nu)} \beta_p(\nu) S_2^{\nu_1} S_3^{\nu_2} \cdots S_k^{\nu_k} \quad (p \geqslant K \geqslant 2), \tag{3.22}$$

where $\beta_p(\nu)$ is given by Eq. (3.17) and can be recast in the more covenient form:

$$\begin{aligned} &\frac{(-)^{\bar{\nu}}}{[\nu!]}\Bigg(\sum_{m=0}^{l} Q_m(n \mp \tfrac{1}{2})^{l-m} - (3n \mp 1)\sum_{m=0}^{l-1} Q_m(n \mp \tfrac{1}{2})^{l-m-1} \\ &\quad + n(2n \mp 1)\sum_{m=0}^{l-2} Q_m(n \mp \tfrac{1}{2})^{l-m-2}\Bigg) \\ &= \frac{(-)^{\bar{\nu}}}{[\nu!]}\{[Q_0(n \mp \tfrac{1}{2})^{l-2} + Q_1(n \mp \tfrac{1}{2})^{l-3} + \cdots + Q_{l-2}] \\ &\quad \times [(n \mp \tfrac{1}{2})^2 - (3n \mp 1)(n \mp \tfrac{1}{2}) + n(2n \mp 1)] \\ &\quad + Q_{l-1}[(n \mp \tfrac{1}{2}) - (3n \mp 1)] + Q_l\}, \end{aligned}$$

that is,

$$\beta_p(\nu) = \frac{(-)^{\bar{\nu}}}{[\nu!]}\Bigg[Q_l(\nu) - (2n \mp \tfrac{1}{2})Q_{l-1}(\nu) \pm \tfrac{1}{2}\sum_{m=0}^{l-2} Q_m(\nu)(n \mp \tfrac{1}{2})^{l-m-1}\Bigg],$$

$$l = p - K, \quad p \geqslant K \geqslant 2, \quad Q_{-1}(\nu) = 0. \tag{3.23}$$

Equation (3.22) together with (3.23) is the main expansion we seek. The problem of finding the eigenvalues of the $C_p$ for the O($2n$) and Sp($2n$) is reduced to the problem of finding the coefficients $\beta_p(\nu)$. The results for the O($2n+1$) are obtained by replacing $n$ in the O($2n$) expansion by $n + \frac{1}{2}$. These results may be compared with the corresponding results of Popov[10] for the u($n$),

$$C_p = -\sum_{(\nu)} \beta_p(\nu) S_1^{\nu_1} S_2^{\nu_2} \cdots S_k^{\nu_k} \quad (p \geqslant K \geqslant 1), \tag{3.24}$$

where

$$\beta_p(\nu) = \frac{(-)^{\bar{\nu}}}{[\nu!]}[Q_l(\nu) - nQ_{l-1}] \quad (l = p - K)$$

and

$$K + 1 = 2\nu_1 + 3\nu_2 + \cdots + (k+1)\nu_k.$$

Thus our result ties together the treatment of this problem for the semisimple Lie groups. The practical uses of these equations are illustrated in Sec. IV below.

## IV. COMPUTATION OF THE COEFFICIENTS $\beta_p(\nu)$

(a) *The terms of layer K*: The first step towards the computation of $\beta_p(\nu)$ is to determine the sets $(\nu)$ which satisfy Eq. (3.11) for any given $K$ satisfying $2 \leqslant K \leqslant p$. This is easily done by solving Eq. (3.11) directly. For example, for $K = 2$ the only solution set is $\nu_2 = 1, \nu_k = 0, k \neq 2$. While for $K = 7$, the solution sets are $\nu_7 = 1, \nu_k = 0, k \neq 7$; $\nu_4 = 1, \nu_2 = 1, \nu_k = 0, k \neq 4,2$ and $\nu_3 = 2, \nu_k = 0, k \neq 3$. Thus for $K = 2$ only the terms $S_2$ appears in the $C_p$, while for $K = 7$, there are three terms $S_7$, $S_2S_4$, and $S_3^2$. The terms appearing for values of $K$ up to 10 are shown in Table I, where use is made of the

TABLE I. All terms on and above the line $K = p$ appear in the coefficient $\beta_p$. The total number of such terms is denoted by $n_p$, while $n_k$ denotes the number of terms appearing for each value of $K$.

| | $K$ | Terms of larger $K$ | | | | | | $n_k$ | $n_p$ |
|---|---|---|---|---|---|---|---|---|---|
| | 2 | 2 | | | | | | 1 | 1 |
| | 3 | 3 | | | | | | 1 | 2 |
| | 4 | 4 | | | | | | 1 | 3 |
| $p$ | 5 | 5 | $2^2$ | | | | | 2 | 5 |
| | 6 | 6 | (3,2) | | | | | 2 | 7 |
| | 7 | 7 | (4,2) | $3^2$ | | | | 3 | 10 |
| | 8 | 8 | (5,2) | (4,3) | $2^3$ | | | 4 | 14 |
| | 9 | 9 | (6,2) | (5,3) | $4^2$ | $(3,2^3)$ | | 5 | 19 |
| | 10 | 10 | (7,2) | (6,3) | (5,4) | $(4,2^2)$ | $(3^2,2)$ | 6 | 25 |

abbreviation, 7 for $S_7$, (2,4) for $S_2S_4$, and $3^2$ for $S_3^2$, etc. [Note that this table[10] is the same as that for the SU($n$).]

(b) *The coefficients* $Q_l(\nu)$: The next step is to calculate the coefficients $Q_l(\nu)$. To do so we need to use the identity (established in the Appendix)

$$\phi_{k_1}(z)\phi_{k_2}(z)\cdots\phi_{k_r}(z)$$
$$= \frac{1}{k_1k_2\cdots k^r z^{r-1}}\{(k_1+k_2+\cdots+k_r)\phi_{k_1+k_2+\cdots+k_r}(z)$$
$$-[(k_1+k_2+\cdots+k_{r-1})\phi_{k_1+k_2+\cdots+k_{r-1}}(z)$$
$$-(k_1+k_2+\cdots+k_{r-2})\phi_{k_1+k_2+\cdots+k_{r-2}}$$
$$+\cdots+(-)^r k_1\phi_{k_1}(z)+\text{com.}k_1k_2\cdots k_r]\}, \tag{4.1}$$

where com. means all combinations of $k_1,k_2,\cdots,k_r$ for each term in the square bracket. Equations (A1) and (A4) are examples of (4.1) for $r=2,3$, respectively.

From Eqs. (3.9) and (3.13) the coefficient of $z^l$ in the Taylor series expansion of (4.1) is

$$Q_l(S_{k_1}S_{k_2}\cdots S_{K_r})$$
$$= \frac{1}{k_1k_2\cdots k_r(l+r)!}\Bigg[\frac{(k_1+k_2+\cdots+k_r+l+r-1)!}{(k_1+k_2+\cdots+k_r-1)!}$$
$$-\Bigg(\frac{(k_1+k_2+\cdots+k_{r-1}+l+r-1)!}{(k_1+k_2+\cdots+k_{r-1}-1)!}$$
$$-\frac{(k_1+k_2+\cdots+k_{r-2}+l+r-1)!}{(k_1+k_2+\cdots+k_{r-2}-1)!}$$
$$+\cdots+(-)^l\frac{(k_1+l+r-1)!}{(k_1-1)!}+\text{com.}k_1,k_2,\cdots,k_r\Bigg)\Bigg]. \tag{4.2}$$

Here, $Q_l(S_{k_1}\cdots S_{k_r})$ stands for $Q_l(\nu)$ for the particular set $(\nu)$ given by $\nu_{k_1}=\nu_{k_2}=\cdots=\nu_{k_r}=1$, all other $\nu$'s vanishing. This notation will be used whenever convenient. For example:

(i) For $r=1$, set $k_1=k$,

$$Q_l(S_k)=\frac{1}{(l+1)!}\frac{(k+l)!}{k!}. \tag{4.3}$$

(ii) For $r=2$

$$Q_l(S_{k_1}S_{k_2})$$
$$=\frac{1}{k_1k_2(l+2)!}\Bigg(\frac{(k_1+k_2+l+1)!}{(k+k_2-1)!}$$
$$-\frac{(k_1+l+1)!}{(k_1-1)!}-\frac{(k_2+l+1)!}{(k_2-1)!}\Bigg). \tag{4.4}$$

In particular if $k_1=k_2=k$,

$$Q_l(S_k^2)=\frac{1}{k^2(l+2)!}\Bigg(\frac{(2k+l+1)!}{(2k-1)!}-\frac{2(k+l+1)!}{(k-1)!}\Bigg). \tag{4.5}$$

(iii) For $r=3$

$$Q_l(S_{k_1}S_{k_2}S_{k_3})$$
$$=\frac{1}{k_1k_2k_3(l+3)!}\Bigg(\frac{(k_1+k_2+k_3+l+2)!}{(k_1+k_2+k_3-1)!}$$
$$-\frac{(k_1+k_2+l+2)!}{(k_1+k_2-1)!}-\frac{(k_1+k_3+l+2)!}{(k_1+k_3-1)!}$$
$$-\frac{(k_2+k_3+l+2)!}{(k_2+k_3-1)!}+\frac{(k_1+l+2)!}{(k_1-1)!}$$
$$+\frac{(k_2+l+2)!}{(k_2-1)!}+\frac{(k_3+l+2)!}{(k_3-1)!}\Bigg). \tag{4.6}$$

If in this case, $k_2=k_3=k$

$$Q_l(S_{k_1}S_k^2)$$
$$=\frac{1}{k^2k_1(l+3)!}\Bigg(\frac{(2k+k_1+l+2)!}{(2k+k_1-1)!}-\frac{(2k+l+2)!}{(2k-1)!}\Bigg)$$
$$-\frac{2(k+k_1+l+2)!}{(k+k_1-1)!}+\frac{2(k+l+2)!}{(k-1)!}+\frac{(k_1+l+2)!}{(k_1-1)!}\Bigg). \tag{4.7}$$

If further, $k_1=k_2=k_3=k$,

$$Q_l(S_k^3)=\frac{1}{k^3(l+3)!}\Bigg(\frac{(3k+l+2)!}{(3k-1)!}$$
$$-\frac{3(2k+l+2)!}{(2k-1)!}+\frac{3(k+l+2)!}{(k-1)!}\Bigg), \tag{4.8}$$

and so on.

Equations (4.3)–(4.8) give all the $Q_l(\nu)$ needed for computing all the $\beta_p(\nu)$ and hence the $C_p$ for $p\leqslant 10$. We illustrate this with a few examples.

(c) *The coefficient* $\beta_p(S_k)$: From the solution set to Eq. (3.11) (or from Table II), the term $S_k$ appears for the first time in $C_p$, when $p=k$. The maximum value of $K$, $K_{\max}=p=k$, so that $l=0$. Therefore, from Eqs. (4.3) and (3.23)

$$\beta_p(S_k)=-1,\quad p=k.$$

It is economical to determine immediately also the coefficient of $S_k$ for $p>k$. In this case $S_k$ appears only when $K=k$, so that $l=p-K\geqslant 1$, and $\beta_p(S_k)$ is given by Eq. (3.23) with $Q_l(S_k)$ determined by Eq. (4.3). In this way we get, for instance, $\beta_3(S_2)=2(n-1)$, $(2n-1)$ for O($2n$), Sp($2n$), respectively.

(d) *The coefficient* $\beta_p(S_{k_1}S_{k_2})$: Similarly, from Eq. (3.11) or from Table II the term $S_{k_1}S_{k_2}$ occurs for the first time in $C_p$, when $p=k_1+k_2+1$, and $l=p-k_1-k_2-1$. $\beta_p(S_{k_1}S_{k_2})$ is given by Eqs. (4.4) and (3.23). In particular for $k_1=2$, $k_2=3$, $p\geqslant 6$, and $\beta_6(S_2S_3)=1$, $\beta_7(S_2S_3)=4-2n$, $(3-2n)$ for the O($2n$) [Sp($2n$)]. For $k_1=k_2=2$, $\beta_p(S_2^2)$ occurs for $p\geqslant 5$ with $\beta_5(S_2^2)=\frac{1}{2}$. Just one more example follows.

(e) *The coefficient* $\beta_p(S_{k_1}S_{k_2}S_{k_3})$: This appears for $p\geqslant k_1+k_2+k_3+2$, $l=p-k_1-k_2-k_3-2$, with coeffi-

TABLE II. Upper (lower) entries and signs in [ ] apply to the O(2n) [Sp(2n)]. Common middle terms in [ ] apply to both. To obtain the results for the O(2n + 1) replace n by $(n + \frac{1}{2})$ in the corresponding results for the O(2n) according to the ansatz of Ref. 8.

| | $K$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ | $C_6$ | $C_7$ | $C_8$ |
|---|---|---|---|---|---|---|---|---|
| $S_2$ | 2 | 1 | $\left[\begin{matrix}2\\1\end{matrix}\ -2n\right]$. | $\frac{1}{2}\left[\begin{matrix}5-5n\\2-7n\end{matrix}\right]$ | $\frac{1}{4}\left[\begin{matrix}13-15n\\4-21n\end{matrix}\ \pm 2n^2\right]$ | $\frac{1}{8}\left[\begin{matrix}31-35n\\8-57n-12n^2\end{matrix}\ \pm 4n^3\right]$ | $\frac{1}{16}\left[\begin{matrix}73-87n+10n^2-4n^3\\16-145n-46n^2-28n^3\end{matrix}\ \pm 8n^4\right]$ | $\frac{1}{32}\left[\begin{matrix}167-199n+8n^2+24n^3-16n^4\\32-352n+144n^2-120n^3-64n^4\end{matrix}\ \pm 16n^5\right]$ |
| $S_3$ | 3 | | 1 | $\frac{1}{2}\left[\begin{matrix}5\\3\end{matrix}\ -4n\right]$ | $\frac{1}{12}\left[\begin{matrix}49-42n\\25-54n\end{matrix}\right]$ | $\frac{1}{24}\left[\begin{matrix}151-148n\\65-196n\end{matrix}\ \pm 12n^2\right]$ | $\frac{1}{48}\left[\begin{matrix}425-430n+12n^2\\161-624n-84n^2\end{matrix}\ \pm 24n^3\right]$ | $\frac{1}{96}\left[\begin{matrix}1143-1216n+88n^2\\385-1840n+376n^2-192n^3\end{matrix}\ \pm 48n^4\right]$ |
| $S_4$ | 4 | | | 1 | $\left[\begin{matrix}3\\2\end{matrix}\ -2n\right]$ | $\frac{1}{2}\left[\begin{matrix}12-9n\\7-11n\end{matrix}\right]$ | $\frac{1}{4}\left[\begin{matrix}43-37n\\22-47n\end{matrix}\ \pm 2n^2\right]$ | $\frac{1}{8}\left[\begin{matrix}139-127n+4n^2\\64-173n-16n^2\end{matrix}\ \pm 4n^3\right]$ |
| $S_5$ | 5 | | | | 1 | $\frac{1}{2}\left[\begin{matrix}7\\5\end{matrix}\ -4n\right]$ | $\frac{1}{4}\left[\begin{matrix}33-22n\\21-26n\end{matrix}\right]$ | $\frac{1}{8}\left[\begin{matrix}135-104n\\77-128n\end{matrix}\ \pm 4n^2\right]$ |
| $S_2^2$ | 5 | | | | $-\frac{1}{2}$ | $-\frac{1}{4}\left[\begin{matrix}7\\5\end{matrix}\ -4n\right]$ | $-\frac{1}{4}\left[\begin{matrix}15-11n\\9-13n\end{matrix}\right]$ | $-\frac{1}{4}\left[\begin{matrix}27-23n\\14-29n\end{matrix}\ \pm n^2\right]$ |
| $S_6$ | 6 | | | | | 1 | $\left[\begin{matrix}4\\3\end{matrix}\ -2n\right]$ | $\frac{1}{6}\left[\begin{matrix}65-39n\\44-45n\end{matrix}\right]$ |
| $S_2S_3$ | 6 | | | | | $-1$ | $-\left[\begin{matrix}4\\3\end{matrix}\ -2n\right]$ | $-\frac{1}{6}\left[\begin{matrix}59-39n\\38-45n\end{matrix}\right]$ |
| $S_7$ | 7 | | | | | | 1 | $\frac{1}{2}\left[\begin{matrix}9\\7\end{matrix}\ -4n\right]$ |
| $S_2S_4$ | 7 | | | | | | $-1$ | $-\frac{1}{2}\left[\begin{matrix}9\\7\end{matrix}\ -4n\right]$ |
| $S_3^2$ | 7 | | | | | | $-\frac{1}{2}$ | $-\frac{1}{4}\left[\begin{matrix}9\\7\end{matrix}\ -4n\right]$ |
| $S_8$ | 8 | | | | | | | 1 |
| $S_2S_5$ | 8 | | | | | | | $-1$ |
| $S_3S_4$ | 8 | | | | | | | $-1$ |
| $S_2^3$ | 8 | | | | | | | $\frac{1}{6}$ |

cients given by Eqs. (4.6) and (3.23). In particular, if $k_1 = k_2 = k_3 = 2$, $\beta_p(S_2^3)$ occurs for $p \geqslant 8$, with $\beta_8(S_2^3) = -\frac{1}{6}$, and so on.

## V. CONCLUDING REMARKS

(a) The examples (c), (d), and (e) of Sec. IV exhaust the possible types of $\beta_p$ occuring for $p \leqslant 10$, and sufficiently illustrate the ease with which these coefficients can be computed in a systematic way. It is now possible to say for what values of $p$ a particular combination of the $S_k$'s appears and with what coefficients. The eigenvalues of the Casimir operators $C_p$ for $p \leqslant 8$ are displayed in Table II for the orthogonal and symplectic groups. these results agree with those particular cases that were also treated in Ref. 8. The corresponding results for the Su($n$) due to Popov[10] are also reproduced in Table III for completeness and for purposes of comparison. The explicit values of $S_k$ in terms of the partition numbers $f_i$ are given in Ref. 8, for some particular representations and there is no need to repeat them here.

(b) We have considered here only the tensor representations. In addition to these, there exists two inequivalent spinor representations for the O($2n$). But the eigenvalues of the Casimir operators corresponding to these have already been elegantly worked out using the Weyl reflection symmetry in Ref. 2b with the result

$$C_n = (-)^{n/2(n-1)} 2^n n! l_1 l_2 \cdots l_n, \tag{5.1}$$

where $l_i$ is defined in Eq. (2.6) of this paper.

(c) Our definition of the Casimir operator $C_p$ in Eq. (2.1) is only one of the many possible ways of contracting the indices. However, owing to the commutation relations satisfied by the infinitesimal generators, any other contraction can be expressed as a linear combination of $C_p$ and Casimir operators of order less than $p$. In some cases this relation is quite simple. For instance, if the definition

TABLE III. $C_p$ for the SU($n$), for $p \leqslant 8$ (Ref. 10).

| | $K$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ | $C_6$ | $C_7$ | $C_8$ |
|---|---|---|---|---|---|---|---|---|
| $S_2$ | 2 | 1 | $-n+\frac{3}{2}$ | $-\frac{3}{2}n+2$ | $-2n+\frac{5}{2}$ | $-5n/2+3$ | $-3n+\frac{7}{2}$ | $-\frac{7}{2}n+4$ |
| $S_3$ | 3 | | 1 | $-n+2$ | $-2n+\frac{10}{3}$ | $-\frac{10}{3}n+5$ | $-5n+7$ | $-7n+\frac{28}{3}$ |
| $S_4$ | 4 | | | 1 | $-n+\frac{5}{2}$ | $-5n/2+5$ | $-5n+\frac{35}{4}$ | $-\frac{35}{4}n+14$ |
| $S_5$ | 5 | | | | 1 | $-n+3$ | $-3n+7$ | $-7n+14$ |
| $S_2^2$ | 5 | | | | $-\frac{1}{2}$ | $+\frac{1}{2}(n-3)$ | $+\frac{1}{2}(3n-\frac{35}{4})$ | $-\frac{1}{2}(\frac{25}{2}n-11)$ |
| $S_6$ | 6 | | | | | 1 | $-(n-\frac{7}{2})$ | $-\frac{7}{2}n+\frac{28}{3}$ |
| $S_2S_3$ | 6 | | | | | $-1$ | $+(n-\frac{7}{3})$ | $+(\frac{7}{2}n-\frac{25}{3})$ |
| $S_7$ | 7 | | | | | | 1 | $-(n-4)$ |
| $S_2S_4$ | 7 | | | | | | $-1$ | $+(n-4)$ |
| $S_3^2$ | 7 | | | | | | $-\frac{1}{2}$ | $+\frac{1}{2}(n-4)$ |
| $S_8$ | 8 | | | | | | | 1 |
| $S_2S_5$ | 8 | | | | | | | $-1$ |
| $S_3S_4$ | 8 | | | | | | | $-1$ |
| $S_2^3$ | 8 | | | | | | | $\frac{1}{6}$ |

$$C_p^1 = \sum_{i_1, i_2, \ldots i_p} X_{i_1}^{i_2} X_{i_2}^{i_3} \cdots X_{i_p}^{i_1}, \tag{5.2}$$

is adopted, a relation of the form

$$C_p^1 = (-)^p C_p \tag{5.3}$$

is obtained.

(d) Finally, we have not found it necessary to express the quantities $Q_l(\nu)$ in terms of the functions $\gamma_l(k,\nu)$ introduced in Eq. (3.18) of Ref. 10 for the case of the u($n$). This is because the use of these functions is distracting and makes computation cumbersome. For instance, for $l = 5$, and for a set $(\nu)$ containing three distinct nonvanishing terms, there are as many as 21 terms in the expansion of $Q_5(\nu)$ using this equation. Thus our method here is by far the simpler.

## ACKNOWLEDGMENT

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

## APPENDIX

The proof of the identity (4.1) can be done by induction. For $r = 1$, Eq. (4.1) is identically true. For $r = 2$, using the defining Eq. (3.9)

$$\begin{aligned}
&\phi_{k_1}(z)\phi_{k_2}(z) \\
&\quad = \frac{1}{k_1 k_2 z^2}[(1-z)^{-k_1} - 1][(1-z)^{-k_2} - 1] \\
&\quad = \frac{1}{k_1 k_2 z^2}[\{(1-z)^{-k_1-k_2} - 1\} - \{(1-z)^{-k_1} - 1\} \\
&\qquad - \{(1-z)^{-k_2} - 1\}] \\
&\quad = \frac{1}{k_1 k_2 z}[(k_1+k_2)\phi_{k_1+k_2}(z) - k_1\phi_{k_1}(z) - k_2\phi_{k_2}(z)].
\end{aligned} \tag{A1}$$

Assume the identity (4.1) true for $r$. To show it is true for $(r+1)$ consider the product

$$\begin{aligned}
&\phi_{k_1}(z)\phi_{k_2}(z)\cdots\phi_{k_{r+1}}(z) \\
&\quad = \frac{1}{k_1 k_2 \cdots k_r z^{r-1}} \\
&\qquad \times \{(k_1+k_2+\cdots+k_r)\phi_{k_1+k_2+\cdots+k_r}(z)\phi_{k_{r+1}}(z) \\
&\qquad - [(k_1+k_2+\cdots+k_{r-1})\phi_{k_1+k_2+\cdots+k_{r-1}}(z)\phi_{k_{r+1}} \\
&\qquad + \cdots(-)^r k_1\phi_{k_1}(z)\phi_{k_{r+1}}(z) + \text{com.}k_1k_2\ldots k_r]\}.
\end{aligned} \tag{A2}$$

Using Eq. (A1) on the first term of the right-hand side of (A2) we get

$$\begin{aligned}
&(k_1+k_2+\cdots+k_r)\phi_{k_1+k_2+\cdots+k_r}(z)\phi_{k_{r+1}}(z) \\
&\quad = \frac{1}{k_{r+1}z}[(k_1+k_2+\cdots+k_{r+1})\phi_{k_1+k_2+\cdots+k_{r+1}}(z) \\
&\qquad - (k_1+k_2+\cdots+k_r)\phi_{k_1+k_2+\cdots+k_r}(z) \\
&\qquad - k_{r+1}\phi_{k_{r+1}}(z)],
\end{aligned}$$

and the first term in the square bracket, for instance, gives

$$\begin{aligned}
&\frac{1}{k_{r+1}z}[(k_1+k_2+\cdots+k_{r-1}+k_{r+1}) \\
&\quad \times \phi_{k_1+k_2+\cdots+k_{r-1}+k_{r+1}}(z) \\
&\quad - (k_1+k_2+\cdots+k_{r-1})\phi_{k_1+k_2+\cdots+k_{r-1}}(z) \\
&\quad - k_{r+1}\phi_{k_{r+1}}(z)],
\end{aligned}$$

and so on. Thus collecting the terms in Eq. (A2)

$$\begin{aligned}
&\phi_{k_1}(z)\phi_{k_2}(z)\cdots\phi_{k_{r+1}}(z) \\
&\quad = \frac{1}{k_1 k_2 \cdots k_{r+1} z}\{(k_1+k_2+\cdots+k_{r+1}) \\
&\qquad \times \phi_{k_1+k_2+\cdots+k_{r+1}}(z)\} \\
&\qquad - [(k_1+k_2+\cdots+k_r)\phi_{k_1+k_2+\cdots+k_r}(z) \\
&\qquad + \cdots + (-)^{r+1}k_1\phi_{k_1}(z) + \text{com.}k_1,k_2,\ldots,k_{r+1}]\},
\end{aligned} \tag{A3}$$

which proves the identity.

For $r = 3$, for example,

$$\begin{aligned}
&\phi_{k_1}(z)\phi_{k_2}(z)\phi_{k_3}(z) \\
&\quad = \frac{1}{k_1 k_2 k_3 z^2}[(k_1+k_2+k_3)\phi_{k_1+k_2+k_3}(z) - (k_1+k_2) \\
&\qquad \times \phi_{k_1+k_2}(z) \\
&\qquad - (k_1+k_3)\phi_{k_1+k_3}(z) - (k_2+k_3)\phi_{k_2+k_3}(z) \\
&\qquad + k_1\phi_{k_1}(z) + k_2\phi_{k_2}(z) + k_3\phi_{k_3}(z)].
\end{aligned} \tag{A4}$$

[1]M. Micu, Nucl. Phys. **60**, 353 (1964); M. Umezawa, Nucl. Phys. **48**, 111 (1963); **53**, 54 (1964); **57**, 65 (1964).
[2]A.M. Perelomov and V.S. Popov, Yad. Fiz. **3**, 924, 1127 (1966) [Soviet J. Nucl. Phys. **3**, 676, 819 (1966)].
[3]A.M. Perelomov and V.S Popóv, Yad. Fiz. **5**, 693 (1967) [Soviet J. Nucl. Phys. **5**, 489 (1967)].
[4]J.D. Louck and L.C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).
[5]S. Okubo, J. Math. Phys. **16**, 528 (1975).
[6]M.X.F. Wong and H.Y. Yeh, J. Math. Phys. **16**, 1239 (1975). This reference contains more comprehensive references on the literature.
[7]C.O. Nwachuku and M.A. Rashid, J. Math. Phys. **17**, 1611 (1976).
[8]C.O. Nwachuku and M.A. Rashid, J. Math. Phys. **18**, 1387 (1977).
[9]S.A. Edwards, J. Math. Phys. **19**, 164 (1978).
[10]V.S. Popov, Theor. Mat. Fiz. **29**, 357 (1976) [Theor. Math. Phys. **29**, 1122 (1976), translated August 1977].
[11]It is important to note that this ansatz cannot be applied to equations of the form (2.2).

# Mayer equations for Bose field theoretical models[a)]

Roman Gielerak

*Institute of Theoretical Physics, University of Wrocław, Cybulskiego 36, 50-205 Wrocław, Poland*
(Received 28 February 1978)

We present a simple proof that volume and ultraviolet cutoff boson Hamiltonians with unbounded entire type self-interacting terms are essentially self-adjoint operators on the Fock space. We present the rigorous derivation of Mayer equations in the considered Bose field theoretical models. A new proof of the divergence of the Feynman perturbation expansion is presented.

## INTRODUCTION

Euclidean methods of the quantum field theory developed in the last years have caused considerable progress in our understanding of models constructed in super-renormalizable quantum field theory.

The present mathematical techniques seem to be insufficient for the study of the more singular models containing nontrivial ultraviolet divergencies.

The principal techniques used in the construction of two- and three-dimensional boson field models such as: $P(\varphi)$:$_2$, :cosh$\alpha\varphi$:$_2$,:cos$\alpha\varphi$:$_2$,:$\varphi^4$:$_3$ models are the cluster expansion of Gilmm and Jaffe and of Spencer[1-3] and the lattice approximation. The basic problem of these models is the thermodynamic limit. In the lattice approximation in higher dimensions, $d > 2$, the thermodynamic limit can be controlled in these models.[4] The aim of this paper is to show in new cases how the modern Euclidean methods simplified many of the problem of the constructive field theory. We will consider here a large class of bosons self-interactions with ultraviolet and volume cutoff. In Sec. 1 we show the simple proof of the essential self-adjointness of the doubly cut-off Hamiltonians including also a certain class of nonpolynomial and unbounded self-interactions; we just simplify and generalize the results of Refs. 5 and 6. In Sec. 2, following the ideas of Ref. 7, we rediscover after Symanzik[8] that the structure of the peturbation theory for bosons is very similar to that encountered in the theory of classical gases.[9] We present below a rigorous derivation of the system of linear, integral equations with almost the same structure as Mayer equations in the theory of classical gases which hold for so-called correlation functions. In particular, we consider the Kirkwood–Salsburg type equations and we conclude that due to the unstability of the potential, the fixed point method cannot be used in general. Only in the bound self-interaction case can it be used to control the thermodynamic limit. This fact was used in Ref. 7. From the instability of the potential we deduce the divergence of the doubly cut-off Feynman perturbation expansion for bosons interacting via an unbounded function of the field. As to the construction of the thermodynamic limit, the cluster expansion must be used and this will be the subject of a subsequent paper.

[a)]I acknowledge the assistance of the U.S. National Science under Grant GF-41959.

## 1. Doubly cut-off interactions

Let $\{S'_r(R^d),\Sigma_1 d\tilde{\mu}_0\}$ stand for the probabilistic space with the following identification: $S'_r(R^d)\equiv S'_{\rm real}(R^d)$ is the weak dual space of the real Schwartz space $S(R^d)$, $\Sigma$ is the $\sigma$-algebra of the Borel sets and $d\tilde{\mu}_0$ is the Gaussian, probabilistic, cylinder set measure given by mean zero and covariance

$$\tilde{S}(\mathbf{x}) = \int_{R^d} \frac{e^{i\mathbf{kx}}}{(\mathbf{k}^2+m^2)^{1/2}} d\mathbf{k}. \tag{1.1}$$

It is well known that this probabilistic space can be viewed as the spectral representation of an Abelian algebra generated by a time-zero, free, scalar boson field $\tilde{\varphi}_0$ with mass $m > 0$. The free scalar field is given by: $\tilde{\varphi}_0(\mathbf{x}) = \int e^{i\mathbf{kx}}[a^+(\mathbf{k}) + a(-\mathbf{k})](\mathbf{k}^2+m^2)^{-1/4} d\mathbf{k}$ where $a^+,a$ are ordinary creation and annihilation operators, respectively.

By Segal's isomorphism[1] its action may be represented on $L^2_{\tilde{\mu}_0}\equiv(S_1,\Sigma_1 d\mu_0)$ as a multiplication by coordinate function $\tilde{\varphi}_0(f)$. The free vacuum is the function $\Omega\equiv 1$. Let $j_s$ stand for the map (see Ref. 1)

$$j_s{:}H_{-1/2} \ni f \to (J_s f) = \delta(x_0 - s) f(\mathbf{x}) \tag{1.2}$$

and $J_s$ is its biquantization.[2] It is shown that $J_s$ are isometries from $\Gamma(H-\frac{1}{2})$ to $\Gamma(H_{-1})$.[1]

Later we use the following ultraviolet cutoff. Let $\chi_\epsilon$ be a net in $C_0^\infty(R^d)$ weakly converging to $\delta$ as $\epsilon\to 0$ and such that: (i) $0\leqslant|\hat{\chi}_\epsilon|\leqslant 1$ means the Fourier transform, (ii) for $x\in R^d$ such that $\Lambda_i|x_i| > 0,1{:}\chi_\epsilon(\mathbf{x})\equiv 0$, (iii) $\chi_\epsilon$ is rotational invariant. [The assumption (ii) is made in connection with the thermodynamic limit problem.]
Define

$$\varphi_\epsilon(\mathbf{x},0)\equiv\tilde{\varphi}_\epsilon(\mathbf{x})\equiv(\varphi * \chi_\epsilon)(\mathbf{x}) = \int_{R^d} \tilde{\varphi}(\mathbf{y})\chi_\epsilon(\mathbf{x}-\mathbf{y})d\mathbf{y}, \tag{1.3}$$

because the suitably smeared sharp-time free field is essentially a self-adjoint operator on the Fock space. Equation (13) defines the family of self-adjoint operators indexed by points $\mathbf{x}\in R^d$. In the diagonal representation described above the $\tilde{\varphi}_\epsilon(\mathbf{x})$ stands for the multiplication by the function $S'\ni\omega\to(\omega * \chi_\epsilon)(\mathbf{x})$. Since the convolution of the distribution with the test function is a smooth function, we observe that the field $\varphi_\epsilon(\mathbf{x})$ in the Schrödinger representation is multiplication by a well-defined function. Thanks to that, no difficulties appear in the definition of the local powers of such fields.[11]

 0022-2488/79/061267-07$01.00  

The mean of the random process $\tilde{\varphi}_\epsilon$ is zero and the covariance is

$$\tilde{S}_\epsilon(\mathbf{x}) \equiv \int_{S'_r} d\tilde{\mu}_0(\varphi)\varphi_\epsilon(\mathbf{x})\varphi_\epsilon(0)$$

$$= \int_{R^d} \frac{e^{i\mathbf{kx}}}{(\mathbf{k}^2+m^2)^{1/2}} |\hat{\chi}_\epsilon(\mathbf{k})|^2 d\,\mathbf{k}. \quad (1.4)$$

The local powers are defined in the following way:

$$S'_r \ni \omega \to (\tilde{\varphi}^r_\epsilon(\mathbf{x}))(\omega) \equiv [\omega_\epsilon(\mathbf{x})]^r \quad (1.5)$$

In connection with additive renormalization procedures we also define the Wick powers as the following random functions:

$$S'_r \ni \omega \to (:\tilde{\varphi}^r_\epsilon:(\mathbf{x}))(\omega) = [\tfrac{1}{2}\tilde{S}_\epsilon(0)]^{r/2} \mathscr{H}_r \left[ \frac{\omega_\epsilon(\mathbf{x})}{(2\tilde{S}_\epsilon(0))^{1/2}} \right] \quad (1.6)$$

Here $\mathscr{H}_n$ stands for the $n$th Hermitean polynomial. Now let

$$V(z) \sum_{n=0}^{\infty} (C_n/n!) z^r$$

be the entire function such that

(i) $\inf V(x) = B_V > -\infty$,

(ii) $V(\tilde{\varphi}_\epsilon)(\mathbf{x}) \equiv \sum_{n=0}^{\infty} \frac{C_n}{n!} \tilde{\varphi}^n_\epsilon(\mathbf{x}) \epsilon L^2_{\tilde{\mu}_0}$,

or

(i′) $\inf :V:(\varphi)(\mathbf{x}) > -\infty$

$$\Leftrightarrow \sum_{n=0}^{\infty} \frac{C_n}{n!} [\tfrac{1}{2}\tilde{S}_\epsilon(0)]^{n/2} \mathscr{H}_n \left[ \frac{x}{(2S_\epsilon(0))^{1/2}} \right] > -\infty,$$

(ii′) $:V:(\tilde{\varphi}_\epsilon)(\mathbf{x}) \equiv \sum_{n=0}^{\infty} \frac{C_n}{n!} :\varphi^r_\epsilon:(\mathbf{x}) \epsilon L^2_{\tilde{\mu}_0}$.

We define now the doubly cut-off interaction Hamiltonians $H^{\rm int}_{\Lambda,\epsilon}$

$$H^{\rm int}_{\Lambda,\epsilon} \equiv \lambda \int_\Lambda |V|(\varphi_\epsilon)(\mathbf{x}) d\mathbf{x}. \quad (1.7)$$

Here $\lambda$ is the major coupling constant, $\Lambda$ bounded region in $R^d$, $|\ |$ stands for $V$ or $:V:$ when (i), (ii) or (i′), (ii′), respectively, hold.

Our main result in this section is Theorem 1.1 which states that the full Hamiltonian $H_0 + H^{\rm int}_{\Lambda,\epsilon}$ is an essentially self-adjoint operator on the Fock space. In the case of the polynomial self-interaction terms such a result was obtained earlier in Ref. 5 by somewhat difficult functional analysis arguments.

*Theorem* 1.1: Let $\lambda \geqslant 0, |\Lambda| < \infty$ ($|\Lambda|$ means the volume of $\Lambda$) and $V$ be such that (i), (ii) or (i′), (ii′) hold.

Then the full Hamiltonian

$$H_{\epsilon,\Lambda} = H_0 + H^{\rm int}_{\epsilon,\Lambda}, \quad (1.8)$$

where $H_0$ is the free-field Hamiltonian,

(a) is essentially a self-adjoint operator on the Fock space [the domain $C^\infty(H_0)$ of analytic vectors for $H_0$ is a core for $H_{\epsilon,\Lambda}$],

(b) is bounded from below,

$$H_{\epsilon,\Lambda} \geqslant -\lambda |\Lambda| B_V, \quad (1.9)$$

and the following identification holds:

$$\exp[-t(H_0 + H^{\rm int}_{\epsilon,\Lambda})]$$
$$= J_t^\dagger \exp\left[ -\lambda \int_0^t dx_0 \int_\Lambda d\mathbf{x} J_{x_0} H^{\rm int}_{\Lambda,\epsilon} \right] J_0. \quad (1.10)$$

As a corollary of this theorem, by use of the Lie–Kato–Trotter formula or the arguments along the line presented in Ref. 2 in the $P(\varphi)_2$ theory context, we obtain

*Theorem* 1.2 (*Feynmann–Nelson–Kac formula*): Let $f_1,\ldots,f_n \in S(R^d)$ and $G_0,\ldots,G_r$ be cylindrical functions on $S'_r$ which are polynomially bounded. Then for $-\infty < S_0 < S_1 < \cdots < S_r \leqslant S < \infty$

$$\int_{S'_{d+1}} d\mu_0(\varphi) J_{S_0} G_0 \cdots J_{S_r} G_r \exp\left[ -\lambda \int_{S_0}^S dx_0 \int_\Lambda J_{x_0} H^{\rm int}_{\Lambda,\epsilon} \right]$$
$$= \langle \Omega_0, G_0 e^{-(S_1 - S_0)H_{\epsilon,\Lambda}} \cdots e^{-(S_r - S_{n-1})H_{\epsilon,\Lambda}} G_n \Omega_0 \rangle. \quad (1.11)$$

Note also that the cut-off Gell-Mann–Low formula holds

$$\lim_{|S_0 - S_n| \to \infty} \int d\mu_0(\varphi) J_{S_0} G_0 \cdots J_{S_r} G_r$$
$$\times \exp\left[ -\lambda \int_{S_0}^S dx_0 \int_\Lambda d\mathbf{x} J_{x_0} H^{\rm int}_{\Lambda,\epsilon} \right] Z^{-1}_{\Lambda,\epsilon}$$
$$= \langle \Omega(\epsilon,\lambda,\Lambda), G_0 e^{-(S_1 - S_0)H_{\Lambda,\epsilon}} \cdots e^{-(S_n - S_{n-1})H} \Omega(\epsilon,\lambda,\Lambda) \rangle \quad (1.12)$$

The existence of $\Omega(\epsilon,\lambda,\Lambda)$, its uniqueness and the finiteness of the ground state energy is a result of a general theory of semigroups preserving positively and the fact that $e^{-tH_{\epsilon,\Lambda}}$ forms a self-adjoint, exponentially bounded, positivity preserving semigroup.

A summary of the original bound from the $P(\varphi)_2$ theory in our simplified models is given by part (b) of Theorem 1.1.

*Sketch of the proof of Theorem* 1.1: One can prove this theorem by the same reasoning as in the $P(\varphi)_2$ case. The following trivial observations imply that proof of Theorem V.12 from Ref. 2 can be repeated in our case without essential changes.

First note that because the regularization is local in time the random elements

$$\int_a^b dx_0 \int d\mathbf{x}\, J_{x_0} |V(\varphi)|(\mathbf{x}) \quad (1.13)$$

are measurable with respect to the $\sigma$-algebra generated by the free Markov field $\varphi$ supported in the time direction on the interval $[a,b]$. This fact implies that one can construct the transfer matrix in the time direction and show the semigroup property by use of the Markov property that holds for the free field in the time direction. By the simple estimate

$$\inf_{\omega \in S'_r(R^d)} H_{\epsilon,\Lambda}(\omega) \geqslant -\lambda B_n |\Lambda| \quad (1.14)$$

we obtain the exponential bound

$$\exp\left[ -\lambda \int_\Lambda d\mathbf{x} |V|(\tilde{\varphi}_\epsilon)(\mathbf{x}) \right] \leqslant e^{\lambda B_n |\Lambda|}, \quad (1.15)$$

from which we conclude that

$$\exp\left[-\lambda\int_\Lambda d\mathbf{x}|V|(\tilde{\varphi_\epsilon})(\mathbf{x})\right]\in\bigcap_{p\geqslant 1}L^p_{d\tilde{\mu}_0}. \qquad (1.16)$$

Assumptions (ii), (ii′), respectively, are used for the proof of the last part via Duhamel's formula as in the original proof in Ref. 2.

## 2. MAYER EQUATIONS

In this section we will study the Schwinger functional of the theories described by self-interactions as in Theorem 1.1. We start with the formal considerations of the Feynman perturbation expansion in the Euclidean domain to obtain the structures similar to those in the theory of classical gases. Next we concentrate on the mathematical precision of the earlier obtained results. In the end we present simple arguments to prove the divergence of the ultraviolet and volume cut-off Feynman perturbation series when the self-interaction is a unbounded function of the field.

### 2.1 Correlation functions

We define the Schwinger generating functional for the self-interacting, Euclidean, boson, scalar field:

$$S_{\epsilon,\Lambda}(f)\equiv\left[\int_{S'_r(R^{d+1})}d\mu_0(\varphi)\exp\left(-\lambda\int_\Lambda P(\varphi_\epsilon)dx\right)\right]^{-1}$$
$$\times\int_{S'_r(R^{d+1})}\exp(i\varphi(f))\exp\left[-\lambda\int_\Lambda P(\varphi_\epsilon)(\mathbf{x})dx\right] \qquad (2.1)$$

In the following we restrict ourselves to the case $P(\varphi)=\varphi^4_\epsilon$. The general case with our without Wick ordering may be considered quite analogically.

Using the result of Sec. 1 and the Bochner–Minlos theorem, we have that $S_{\epsilon,\Lambda}$ is a Fourier transform of some probabilistic cylinder set measure on Borel $\sigma$-algebra of sets in $S'_r(R^{d+1})$. The quantity

$$Z_{\epsilon,\Lambda}=\int_{S'_r}d\mu_0(\varphi)\exp\left(-\lambda\int_\Lambda\varphi^4_\epsilon\,dx\right) \qquad (2.2)$$

will be called the partition function and its expansion in powers of $\lambda$, the grand partition function $\bar{Z}_{\epsilon,\Lambda}$:

$$\bar{Z}_{\epsilon,\Lambda}\equiv\sum_{n=0}^\infty\frac{(-\lambda)^n}{n!}\int_\Lambda\int_\Lambda dx_1\cdots dx_n$$
$$\times\int_{S'_r}d\mu_0(\varphi)\prod_{i=1}^n\varphi^4_\epsilon(x_i). \qquad (2.3)$$

It is well known that this power series is divergent for arbitrary values of $\Lambda\neq\phi$ and $\lambda\neq 0$. In the end of this section we present (it seems to us) new proof of this fact.

Now we use the formal power series (2.3) and the formal substitution

$$\varphi^4_\epsilon(x)\equiv D_t(e^{t\varphi_\epsilon(x)})=\frac{d^4}{dt^4}(e^{t\varphi_\epsilon(x)})\Big|_{t=0}. \qquad (2.4)$$

The meaning of this will be discussed in subsection 2.3. We express (2.1) in terms of the Feynman perturbation expansion and substitute (2.4) to obtain

$$S_{\Lambda,\epsilon}(f)Z_{\Lambda,\epsilon}=\sum_{n=0}^\infty\frac{(-\lambda)^n}{n!}\int_\Lambda\cdots\int_\Lambda\prod_{i=1}^n D_{t_i}\left\{\int d\mu_0(\varphi)\exp\left[\sum_{i=1}^n t_i\varphi_\epsilon(x_i)\right]\right\}\prod_{i=1}^n dx_i$$

$$=e^{(-1/2)\|f\|^2}\sum_{n=0}^\infty\frac{(-\lambda)^n}{n!}\int_\Lambda\cdots\int_\Lambda\prod_{i=1}^n dx_i\prod_{i=1}^n D_{t_i}\left\{\exp\left[i\sum_{i=1}^n t_i f^\epsilon(x_i)\right]\exp\left[\frac{1}{2}\sum_{1\leq i,j\leq n}t_it_jS_\epsilon(x_i-x_j)\right]\right\}. \qquad (2.5)$$

We used the following expression for the Laplace's transform of the Gaussian measure $d\mu_0$

$$\int_{S'_r}d\mu_0(\varphi)e^{z\varphi_0(f)}=e^{(z^2/2)\|f\|^2_1} \qquad (2.6)$$

and

$$f^\epsilon_{(x)}=(f^*S_\epsilon)(x)=\int f(y)S_\epsilon(x-y)dy. \qquad (2.7)$$

By the Mayer trick,

$$\prod_{i=1}^n e^{it_if^\epsilon(x_i)}=\prod_{i=1}^n[(e^{it_if^\epsilon(x_i)}-1)+1],$$

symmetry of the integrands in (2.5), and definition (2.3), we obtain

$$S_{\Lambda,\epsilon}(f)$$
$$=e^{-(1/2)\|f\|^2}\left\{\sum_{r=0}^\infty\frac{1}{r!}\int_\Lambda\cdots\int_\Lambda\prod_{i=1}^r dx_i\prod_{i=1}^n D_{ti}\right.$$
$$\left.\times\left(\prod_{i=1}^n[e^{it_if_\epsilon(x_i)}-1]\tilde{\rho}^r_{\epsilon,\Lambda}(x_1t_1,\ldots,x_rt_r)\right)\right\} \qquad (2.8)$$

where we defined the grand correlation functions by the following formal power series in $\lambda$:

$$\tilde{\rho}^r_{\Lambda,\epsilon}(x_1t_1,\dots,x_rt_r) = \bar{Z}^{-1}_{\epsilon,\Lambda} \times \sum_{n=0}^{\infty} \frac{(-\lambda)^{n+r}}{n!} \int_\Lambda \cdots \int_\Lambda \prod_{i=1}^{n} dx_{r+1} \prod_{j=1}^{r} D_{t_j} \times \left\{ \prod_{4j=1}^{n+r} e^{t_i t_j S_\epsilon(x_i - x_j)} \right\} \quad (2.9)$$

We observe on the formal perturbation level that the grand correlation functions may be written as

$$\rho^r_{\Lambda,\epsilon}(x_1t_1,\dots,x_rt_r) = \text{expansion in } \lambda \text{ of} (-\lambda)^r \int_{S'_r} \prod_{i=1}^{r} e^{t_i\varphi_\epsilon(x_i)} \left[ \exp - \lambda \int_\Lambda \varphi^4_\epsilon(y) dy d\mu_0(\varphi) \times Z^{-1}_{\epsilon,\Lambda} \right]. \quad (2.10)$$

This suggests the definition of the correlation function in canonical Gibb's ensamble:

$$\rho^r_{\Lambda,\epsilon}(x_1t_1,\dots,x_rt_r) = (-\lambda)^r \int_{S'_r} d\mu_0(\varphi) \prod_{i=1}^{r} e^{t_i\varphi_\epsilon(x_i)} \exp\left[ -\lambda \int_\Lambda \varphi^4_\epsilon(y) dy \right] \times Z^{-1}_{\Lambda,\epsilon}. \quad (2.11)$$

This will be studied in the next subsection.

## 2.2 Integral equations for the grand correlation functions

We see from (2.8) that the only nontrivial volume dependence of $S_{\Lambda,\epsilon}$ comes from that of $\tilde{\rho}^r_{\Lambda,\epsilon}$. We show that formula (2.8) exactly hold with $p_{\Lambda,\epsilon}$ replacing $\tilde{\rho}^r_{\Lambda,\epsilon}$. Now we concentrate on $\tilde{\rho}^r_{\Lambda,\epsilon}$. On the formal peturbation level we state:

*Proposition* 2.2.1: For any multi-index $s = (s_i)$, $i = 1,2,\dots,s_i$ integer, such that $1 \leqslant s_{-i} \leqslant i$, the grand correlation functions fulfill the following systems of linear integral equations:

$$\tilde{\rho}^1_{\Lambda,\epsilon}(x_1t_1) = (-\lambda)\left\{1 + \sum_{k=1}^{\infty} \frac{1}{k!} \int_\Lambda \cdots \int_\Lambda \prod_{i=1}^{k} dy_i \prod_{i=1}^{k} D_{s_i} K\begin{pmatrix} t_1 & s_1 \cdots s_k \\ x_1 & y_1 \cdots y_k \end{pmatrix} \tilde{\rho}^k_{\Lambda,\epsilon}(y_1s_1,\dots,y_ks_k)\right\}, \quad (2.12)$$

$$\tilde{\rho}^r_{\Lambda,\epsilon}(x_1t_1,\dots,x_rt_r) = (-\lambda)^{s_r} \exp\left[ \sum_{i=1}^{s_r} \sum_{j=1}^{r} t_it_jS_\epsilon(x_i - x_j) \right] \times \{ \tilde{\rho}^{r-s_r}_{\Lambda,\epsilon}(x_{s_{r+1}}t_{s_{r+1}},\dots,x_rt_r)$$

$$\times \sum_{k=0}^{\infty} \frac{1}{k!} \int_\Lambda \int_\Lambda \prod_{i=1}^{k} dy_i \prod_{i=1}^{k} D_{s_i} \left\{ K\begin{bmatrix} t_1 \cdots t_r & s_1 \cdots s_k \\ x_1 \cdots x_r & y_1 \cdots y_k \end{bmatrix} \right\} \tilde{\rho}^{k+r-s_r}_{\Lambda,\epsilon}(x_{s_{r+1}}t_{s_{r+1}}\cdots x_rt_r y_1s_1,\dots,y_ks_k)$$

$$K\begin{bmatrix} t_1 \cdots t_r & s_1 \cdots s_k \\ x_1 \cdots x_r & y_1 \cdots y_k \end{bmatrix} = \prod_{i=1}^{r} \left\{ \exp\left[ \sum_{j=1}^{k} t_is_jS_\epsilon(x_i - y_j) \right] - 1 \right\}. \quad (2.13)$$

*Proof*: It is abstracted from Ref. 13. From the $r$-point correlation function $\tilde{\rho}^r_{\Lambda,\epsilon}$ we first set apart the contribution from the $s_r \leqslant r$ particles:

$$\exp\left[ \sum_{i,j=1}^{r+k} t_it_jS_\epsilon(x_i - x_j) \right] = \exp\left[ \sum_{i=1}^{s_r} \sum_{j=1}^{n} S_\epsilon(x_i - x_j)t_it_j \right] \exp\left[ \sum_{i=1}^{s_r} \sum_{j=k+1}^{r+k} t_it_jS_\epsilon(x_i - x_j) \right]$$

$$\times \exp\left[ \sum_{i,j=S_r+1}^{r+k} t_it_jS_\epsilon(x_i - x_j) \right]. \quad (2.14)$$

Using this, we have

$$\tilde{\rho}^r_{\Lambda,\epsilon}(x_1t_1,\dots,x_rt_r) = \sum_{k=0}^{\infty} \frac{(-\lambda)^{k+r}}{k!} \exp\left[ \sum_{i=1}^{s_r} \sum_{j=1}^{r} t_it_jS_\epsilon(x_i - x_j) \right]$$

$$\times \int_\Lambda \cdots \int \prod_{i=1}^{k} dx_{r+i} \left( \prod_{i=1}^{k} D_{t_{r+i}} \left\{ \exp\left[ \sum_{i=s_{r+1}}^{r+k} \sum_{j=r+1}^{r+k} t_it_jS_\epsilon(x_i - x_j) \right] \right\} \right).$$

Now repeat the trick:

$$\exp\left[ \sum_{i=1}^{s_r} \sum_{j=s_r+1}^{r+k} t_it_jS_\epsilon(x_ix_j) \right] = \sum_{q=0}^{k} \prod_{\eta\in\pi_q} \left\{ \exp\left[ \sum_{j=1}^{s_r} t_\eta t_jS_\epsilon(x_j - x_\eta) \right] - 1 \right\}$$

Now observe that $\Sigma_q \Pi_\eta$ can be substituted in the integrals by $\Sigma_{q=0}^{k}\binom{k}{q}$ because of its symmetry. In this way we obtain

$$\rho^r_{\Lambda,\epsilon}(x_1,t_1,\ldots,x_r,t_r) = \sum_{k=0}^{\infty} \frac{(-\lambda)^{k+r}}{k!} \exp\Big[-\frac{1}{2}\sum_{j=1}^{r}\sum_{i=1}^{s_r} t_i t_j S_\epsilon(x_i - x_j)\Big] \prod_{j=1}^{k} D_{t_{r+j}} \Big\{ \sum_{q=0}^{\infty}\binom{k}{q} \int_\Lambda \cdots \int_\Lambda \prod_{j=1}^{k} dx_{r+j} \Big\}$$

$$\times \exp\Big[\frac{1}{2}\sum_{i=s_r+1}^{r+k}\sum_{j=r+1}^{r+k} t_i t_j S_\epsilon(x_i - x_j)\Big] \prod_{\eta=1}^{q} \exp\Big[\sum_{i=1}^{s_r} t_i t_\eta S_\epsilon(x_i - x_{r+\eta})\Big]$$

$$= \exp\Big[\frac{1}{2}\sum_{i=1}^{s_r}\sum_{j=1}^{r} t_i t_j S_\epsilon(x_i - x_j)\Big] \sum_{k=0}^{\infty} \frac{(-\lambda)^{k+r}}{k!} \sum_{q=0}^{k} \frac{k!}{q!(k-q)!} \prod_{i=1}^{k} D_{t_{r+i}}$$

$$\times \Big\{ \int_\Lambda \cdots \int_\Lambda \prod_{i=1}^{k} dx_{r+i} K\begin{bmatrix} t_1\cdots t_{s_r} & t_{r+1}\cdots t_{r+q} \\ x_1\cdots x_{s_r} & x_{r+1}\cdots x_{r+q}\end{bmatrix} \exp\Big[\frac{1}{2}\sum_{i=s_r+1}^{r+k}\sum_{j=r+1}^{r+k} t_i t_j S_\epsilon(x_i - x_j)\Big]\Big\}.$$

Now we change the variables

$x_{r+j} \to y_j$,

$t_{r+j} \to \eta_j$

and write

$$\rho^r_{\Lambda,\epsilon}(x_1,t_1,\ldots,x_r t_r) = \exp\Big[\frac{1}{2}\sum_{i=1}^{s_r}\sum_{j=1}^{r} t_i t_j S_\epsilon(x_i - x_j)\Big] \sum_{k=0}^{\infty}\sum_{q=1}^{k} \frac{(-\lambda)^{k+r}}{q!(k-q)!} \prod_{j=1}^{q} D_{\eta_j} \Big\{ \int_\Lambda \int_\Lambda \prod_{j=1}^{q} dy_j K\begin{bmatrix} t_1\cdots t_{s_r} & \eta_1\cdots\eta_q \\ x_1\cdots x_{s_r} & y_1\cdots y_q\end{bmatrix}$$

$$\times \int_\Lambda \cdots \int_\Lambda \prod_{i=1}^{k-q} dy_{q+i} \prod_{i=1}^{k-q} D_{\eta_{q+i}} \Big\} \exp\Big[\frac{1}{2}\sum_{i=s_r+1}^{r+k}\sum_{j=1}^{k} t_i \eta_j S_\epsilon(x_i - y_j)\Big]$$

$$= (-\lambda)^{s_r} \exp\Big\{\frac{1}{2}\sum_{i=1}^{s_r}\sum_{j=1}^{r} t_i t_j S_\epsilon(x_i - x_j)\Big\}$$

$$\times \Big\{ \tilde\rho^{\,r-s_r}_{\Lambda,\epsilon}(x_{s_r+1},t_{s_r+1},\ldots,x_r t_r) + \sum_{q=1}^{\infty} \frac{1}{q!} \int_\Lambda \cdots \int_\Lambda \prod_{i=1}^{q} dx_i \prod_{i=1}^{q} D_{\eta_i} K$$

$$\times \begin{bmatrix} t_1\cdots t_{s_r} & \eta_1\cdots\eta_q \\ x_1\cdots x_{s_r} & y_1\cdots y_q\end{bmatrix} \tilde\rho^{\,r+q-s_r}_{\Lambda,\epsilon}(x_{s_r+1},t_{s_r+1},\ldots,y_q,\eta_q)\Big\}.$$

In the last step we change the order of summation and used the definition (2.9) of the grand correlation functions. Q.E.D.

## 2.3 Mathematical motivations

This subsection is devoted to mathematical motivations of the formal manipulation made in subsections 2.1 and 2.2. We start from the following lemma which can be easily extracted from Ref. 14 but we present here its simple proof for the reader convenience.

*Lemma* 2.3.1. Let $\{M,\Sigma,dv\}$ be a measure space with finite total mass, i.e., $\int_M dv < \infty$. Let $G$ be an open region in the $n$-fold tensor product of the complex plane $C$. Let $\Gamma_\zeta$ be a map from $G$ to $L^\infty(M,dv)$ such that

(i) $\Gamma_\zeta$ is continuous in strong $L^2(dv)$ topology.

(ii) $\Gamma_\zeta$ has first complex derivatives in the strong $L^2(dv)$ sense, such that up to a set of measure zero the following hold:

$$G \to \mathop{\Lambda}_{\zeta=(z_1\cdots z_n)} \mathop{\Lambda}_{i\in\{1,\ldots,n\}} s - \Big[\frac{\partial}{\partial x_i} + i\frac{\partial}{\partial y_i}\Big]\Gamma_\zeta = 0. \quad (2.15)$$

(iii) For any compact $H \subset G$, $\Gamma^H \equiv \sup_{\zeta\in H}|\Gamma_\zeta| < \infty$.

The the following holds:

(1) All the strong $L_2$ derivatives of the map $\Gamma$ exist.

(2) The following Cauchy–Bochner formulas holds:

$$\mathop{\Lambda}_{\zeta\in G} \Gamma_\zeta = \frac{1}{(2\pi i)^n} \oint_{\Pi_\zeta} \prod_{i=1}^{n} d\xi_i (\zeta_i - \xi_i)^{-1} \Gamma_\xi, \quad (2.16)$$

$$\frac{\partial^{|R|}}{\partial \xi_r^{(R)}} = \frac{\partial^{r_1+\cdots+r_n}}{\partial \xi_1^{r_1}\cdots d\xi_N^{r_n}},$$

$$\Big(\frac{\partial^{|R|}}{\partial \xi_r^{(R)}}\Gamma_\xi\Big)(\zeta) = \frac{R!}{(2\pi i)^r} \oint_{\Pi_\zeta} \prod_{i=1}^{n} d\eta_i (\xi_i - \eta_i)^{-(r+1)} r_\eta,$$

$$R! = r_1!\cdots r_n!, \quad |R| = r_1 + \cdots + r_r. \quad (2.17)$$

Up to a set of measures zero this representations does not depend of the concrete polidisk $\Pi_\xi$ lying in $G$.

*Proof*: From assumptions and Schwartz inequality we easily deduce that the following function,

$$H(\xi,\xi')\equiv\int_M \Gamma_\xi\Gamma_{\xi}dv, \tag{2.18}$$

defined on $G\times G$ is holomorphic function in $(\xi,\xi')$ variables. We use the Cauchy integral fromula for it. Let $\Pi_\xi$ be any polidisk in $G$ containing the point $\xi$ in its interior.

Using the Fubini–Tonelli theorem [by assumption (iii)] and definition, we see that it is posssible that

$$\int_M d\mu\Gamma_\xi\Gamma_{\xi'}$$
$$=\frac{1}{(2\pi i)^{2n}}\oint_{\Pi_\zeta}d\zeta\oint_{\Pi_\zeta}d\zeta'$$
$$\times\int_M dv(m)(\xi'-\zeta)^{-1}(\xi-\zeta)^{-1}\Gamma_\xi\Gamma_{\xi'}. \tag{2.19}$$

On the other hand,

$$\int_M dv(m)\Big(\Gamma_\zeta-\frac{1}{(2\pi i)^n}\oint_{\Pi_\zeta}d\xi\,(\xi-\zeta)^{-1}\Gamma_\xi\Big)^2$$
$$=H(\zeta,\bar\xi)-\frac{2}{(2\pi i)^n}\oint_{\Pi_\zeta}(\xi-\zeta)^{-1}H(\zeta,\xi)d\xi$$
$$-\frac{1}{(2\pi i)^{2n}}\oint_{\Pi_\zeta}\oint_{\Pi_\zeta}(\xi-\zeta)^{-1}(\xi'-\zeta)^{-1}H(\xi,\xi').$$

This means that up to a set of measure zero

$$\Gamma_\zeta=\frac{1}{(2\pi i)^n}\oint_{\Pi_\zeta}d\xi\,(\xi-\zeta)^{-1}\Gamma_\xi.$$

To prove statement (1) we observe

$$\Big\{\Big[\int dv(m)\Big[\frac{\partial^{|R|}}{\partial\xi^{(R)}}\Gamma\Big]_\zeta-\frac{R\,!}{(2\pi i)^R}\oint_{\Pi_\zeta}d\xi\,(\xi-\zeta)^{-(R+1)}\Gamma_\xi\Big\}^2$$
$$=\Big[\frac{\partial^{2|R|}}{\partial\zeta^R\partial\bar\zeta^R}H\Big](\zeta,\zeta)-\frac{2R\,!}{(2\pi i)^r}\oint d\xi\,(\xi-\zeta)^{-(R+1)}$$
$$\times\frac{\partial^{|R|}}{\partial\xi^{(R)}}H(\xi,\zeta)-\frac{(R\,!)^2}{(2\pi i)^{2r}}$$
$$\times\oint_{\Pi_\zeta}\oint_{\Pi_\zeta}d\xi d\xi'(\xi-\zeta)^{-(R+1)}$$
$$\times(\xi'-\zeta)^{-(R+1)}H(\xi,\xi')$$
$$=0.$$

We used assumption (iii) to change the order of the integration with the differentation.

Q.E.D.

Let us now define the following cutoff on the Euclidean field $\varphi_\epsilon(x)$:

$$(\varphi_\epsilon)_N(x)=\begin{cases}\varphi_\epsilon(x) & \text{if } |\varphi_\epsilon(x)|\leqslant N,\\ 0 & \text{otherwise.}\end{cases} \tag{2.20}$$

This makes the field bounded function on $S'_\Gamma$. The following holds:

*Lemma* 2.3.2: The maps

$$\Gamma_{\mathbf z}=\Gamma_{(z_1,\dots,z_N)}=\prod_{i=1}^N e^{z_i(\varphi_\epsilon)_N(x_i)} \tag{2.21}$$

from $G\subset C^N$ ($G$ arbitrary region in $C^N$) to $L^\infty_{d\mu_0}$ fulfill the condition of Lemma 2.3.1 so that

$$\frac{\partial^{|R|}}{\partial\mathbf z^{(R)}}\Gamma_{\mathbf z}=\frac{R\,!}{(2\pi i)^r}\oint_{\Pi_\zeta}d\,\zeta(\zeta-\mathbf z)^{-(R+1)}\Gamma_\zeta. \tag{2.22}$$

Moreover

$$\prod_{i=1}^N(\varphi_\epsilon)^4_M(x_i)=\frac{(4!)^N}{(2\pi i)^N}\oint_{\Pi_0}d\,\zeta\,\zeta_i^{-5}e^{\zeta_i(\varphi_\epsilon)_M(x_i)} \tag{2.23}$$

with probability one.

*Proof*: We use the fact that for any fixed $x_1,\dots,x_n$ the maps are cylindric functions on $S'$. The assumptions (i)–(iii) of Lemma 2.3.1 can be easily extracted from the following formula for the cylindric functions integrated with respect to the Gaussian measure $d\mu_0$

$$\int_{S'_\Gamma}d\mu_0(\varphi)\Gamma^2_{\mathbf z}=\frac{1}{(2\pi)^{N/2}\det(s_\epsilon(x_1-x_j))}$$
$$\times\int_{-N}^{+N}\cdots\int_{-N}^{+N}\prod_{i=1}^N dq_i\exp\Big(\sum_{i=1}^N 2z_i^nq_i^n\Big)$$
$$\times\exp\Big[-\frac12\sum_{ij}q_iq_jS_\epsilon^{-1}(x_i-x_j)\Big],$$

$$\sum_{i,j}S_\epsilon^{-1}(x_i-x_j)S_\epsilon(x_i-x_j)=\delta_{ij}. \tag{2.24}$$

Q.E.D.

Now we use the results of Ref. 7 which stated that in the case of the bounded ultraviolet cutoff self-interaction the Feynman perturbation series are absolutely convergent for small values of coupling constant $\lambda$.

Because $(\varphi_\epsilon)^4_N$ is a bounded function it may be represented as uniform closure in $L^\infty(d\mu_0)$ of the functions of the following kind:

$$\int e^{i\alpha\varphi_\epsilon(x)}d\mu^N_R(\alpha), \tag{2.25}$$

where $d\mu^N_R$ is a net of measures with bounded variation and such that

$$d\mu^N_R(-\alpha)=\overline{d\mu^N_R}(\alpha) \tag{2.25a}$$

and

$$\lim_{R\to\infty}\int e^{i\alpha\varphi_\epsilon(x)}d\mu^N_R(\alpha)=(\varphi_\epsilon)^4_N(x) \tag{2.25b}$$

This approximation together with the diagonal procedure is used to show that Mayer given by Prop. 2.2.1 exactly hold for the canonical functions given by formula (2.11). The following easily to prove proposition is critical to show that Prop. 2.2.1 hold also on the rigorous, nonperturbative level.

*Proposition* 2.3.3: For any $p>1$, $N$ integer the following limits

$$\lim_{N\to\infty}\prod_{i=1}^M e^{z_i(\varphi_\epsilon)_N(x_i)}e^{-\lambda\int_\Lambda(\varphi_\epsilon)^4_N(x)\,dx} \tag{2.26}$$

exist in all $L^p(d\mu_0)$ spaces and equal

$$\prod_{i=1}^M e^{z_i'(\varphi_\epsilon)(x_i)}e^{-\lambda\int_\Lambda\varphi^4_\epsilon(x)dx}. \tag{2.27}$$

*Proof*: This is a simple consequence of the Hölder inequality and expressions for the cylindric integral with respect to the Guassian measure $d\mu_0$.

From the discussion above we finally obtain

*Theorem*: The equations given in Prop. 2.2.1 exactly hold for the canonical correlation functions in place of the $\tilde{\rho}^{\Gamma}$.

## 2.4 The Kirkwood–Salsburg type equations and the divergence of the perturbation expansion

The equations given by Prop. 2.2.1 in the case $s_i = 1$ for all $i$ have a structure identical to the Kirkwood–Salsburg equations in the theory of classical gases.[8] However because of

$$\exp\Big[\sum_{i,j=1} z_i z_j S_\epsilon(x_i - x_j)\Big] = \sum_{i,j=1}^{r} \Gamma_i \Gamma_j e^{i(\theta_i + \theta_j)} S_\epsilon(x_i - x_j) \tag{2.28}$$

with

$$z_i = \Gamma_i e^{i}\theta_i, \tag{2.28a}$$

we conclude that the following bound

$$\Big|\exp\Big[\sum_{i=1}^{n}\sum_{j=1}^{n} Z_i Z_j S_\epsilon(x_i - x_j)\Big]\Big| \leqslant e^{2n^2 \Gamma_*^2 S_\epsilon(0)} \tag{2.29}$$

is optimal. Here

$$\Gamma_* = \max_i\{|\Gamma_i|\}. \tag{2.29a}$$

But this means that the gas with interior degrees of freedom described by complex parameters $z_i$ and interaction of the form $z_i z_j S_\epsilon(x_i - x_j)$ is unstable so that the well-known contraction principle cannot be used in our case. The unstability of the interaction combined with Theorem 3.2.2 from Ref. 9 and absence of screening leads to

*Proposition* 2.4.1: In the case of unbounded self-interaction the volume and ultraviolet cutoff Feynman perturbation expansions for the partition functions are divergent.

*Remark*: The unstability itself does not necessarily imply the divergence of correlation functions in the big canonical ensamble. See Ref. 1 for this point.


## ACKNOWLEDGMENTS

The author is grateful to Dr. W. Karwowski for his fruitful comments and for a critical reading of the manuscript. He also thanks Dr. Sz. Rabsztyn and Dr. J. Hanckowiak for stimulating discussions.



[1] *Constructive Quantum Field Theory*, The 1973 "Ettore majorana International School of Mathematical Physics," edited by G. Velo and A. Wightman (Springer-Verlag, Berlin, 1973).

[2] *The $P(\phi)_2$ Euclidean Quantum Field Theory*, Princeton Series in Physics (Princeton U.P., Princeton, N.J., 1974).

[3] J. Fröhlich and J. Feldman, in *Renormalization Theory*, edited by G. Velo and A. Wightman (NATO Advanced Study Institutes Series) (Dordrecht, Holland, 1975).

[4] G. Baker, J. Math. Phys. **16**, 1324–46 (1975).

[5] A. Jaffe, O.E. Lanford, and A.S. Wightman, Commun. Math. Phys. **15**, 179–93 (1969).

[6] R. Hoegh-Krohn, Commun. Math. Phys. **17**, 179–93 (1970).

[7] S. Albeverio and R. Hoegh-Krohn, Commun. Math. Phys. **30**, 171–200 (1973).

[8] K. Symanzik, in *Euclidean Quantum Field Theory*, edited by R. Jost (Academic, New York, 1969).

[9] D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).

[10] K. Hepp, "Theorie de la renormalization," in *Lecture Notes in Physics*, edited by J. Ehlerrs (Steidelberg, 1969), Vol. 2.

[11] J.E. Segal, Ann. Math. **92**, 462–81 (1970).

[12] A. Jaffe, Commun. Math. Phys. **1**, 127–49 (1965).

[13] N. K. Bolotin, and Y. Judkin, Theoret. Math. Phys. **21**, 146–52 (1974).

[14] J. Diedounne, *Elements d'Analise* (Gauthier-Villars, Paris, 1969), new edition, Vol. 14.

[15] J. Fröhlich, Commun. Math. Phys. **47**, 233–68 (1976).

# Factored irreducible symmetry operators and space groups

N. O. Folland

*Physics Department, Kansas State University, Manhattan, Kansas 66506*

(Received 7 July 1978)

Induction techniques for the systematic construction of factorized irreducible symmetry operators (FISO) are applied to general space groups. Detailed results are given for the diamond structure space group and space double group. In the case of the diamond structure it is shown how the relations provided by time-reversal symmetry may be efficiently utilized to simplify calculations.

## I. INTRODUCTION

General techniques for the systematic construction of factored irreducible symmetry operators (FISO's) for finite groups are described in a previous paper[1] hereafter referred to as Ref. 1. In this paper these techniques are applied to and facilitate the symmetry analysis of space groups and space double groups. Explicit results are given for the case of the diamond structure. The diamond structure is chosen for purposes of illustration because of the extensive interest in crystals having this structure, and because FISO's for special points on the surface of the Brillouin zone are of interest in themselves and have not been reported previously. A further objective is to demonstrate how time-reversal symmetry may be treated simply and directly using FISO's.

The presentation of this paper reflects certain simplifications afforded by FISO's. In general an irreducible symmetry operator (ISO) is defined

$$P(A)_{ij} = \frac{n_A}{G^0}\sum_{g\in G} D^A(g^{-1})_{ij} g, \tag{1}$$

where the $n_A \times n_A$ matrices $A = \{D^A(g)\}$ are an irreducible unitary representation (IUR) $A$ of group $G$. ISO's display the following properties:

$$[P(A)_{ij}]^\dagger = P(A)_{ji}, \tag{2}$$

$$gP(A)_{ij} = \sum_m P(A)_{im} D^A(g)_{mj}, \tag{3}$$

$$P(A)_{ij} P(B)_{mn} = \delta_{AB}\delta_{in} P(A)_{mj}, \tag{4}$$

$$g = \sum_A \sum_m \sum_n P(A)_{mn} D^A(g)_{nm}, \tag{5}$$

where in Eq. (2) the superscript dagger means adjoint. It follows that the IOS's for a given IUR $A$ are completely described by the partner operators $P(A)_{1j}, j = 1,\ldots,n_A$, because

$$P(A)_{1j}[P(A)_{1i}]^\dagger = P(A)_{ij}. \tag{6}$$

In the case of the space group, FISO's are obtained by induction with respect to a subgroup $H$ having IUR $a$ and partner ISO's $P(a)_{1i}$. If the entire space group is $G = SH$ where $S = \{s_1 = e, s_2,\ldots,s_n\}$ is a set of left coset generators, then

$$P(A)_{(11)(pi)} = s_p P(a)_{1i} \tag{7}$$

are a set of partner operators for unitary representation $A$ of the space group. Such an induced representation is an IUR if the irreducibility conditions

$$P(a)_{ii}\, s_p\, P(a)_{jj}\, s_p^{-1} = 0, \quad i,j = 1,\ldots,n_a, \quad p = 2,\ldots,n \tag{8}$$

are satisfied. In Ref. 1 it was shown that Eqs. (8) are necessary and sufficient conditions that $A$ be an IUR. Equations (8) are a restatement of the group orthogonality conditions in the language of ISO's. Thus, a complete description of space group IUR is given by listing $P(a)_{1j}$ and the set $S$ of left coset generators for the partner operators. This does not constitute a sacrifice in practicality for the sake of brevity, since only the partner operators are needed in any practical application.

The construction of FISO's for space groups in general is treated in Sec. II. The FISO's for the diamond structure and the use of time-reversal symmetry are presented and discussed in Sec. III.

## II. IRREDUCIBLE SYMMETRY OPERATORS FOR SPACE GROUPS

A space group is the set of rotation-translation operators which leave a crystal structure invariant. A general space group operator $g$ has the form $g = [\bar{g}|\boldsymbol{\tau}_g]$ which acts on a position vector $\mathbf{x}$ as

$$g\mathbf{x} = \bar{g}\mathbf{x} + \boldsymbol{\tau}_g, \tag{9}$$

where $\bar{g}$ represents a rotation about a point and $\boldsymbol{\tau}_g$ is a translation vector. The translation group $\tau = \{t(\mathbf{R})\}$ consisting of a set of pure translation operators $t(\mathbf{R}) = [\bar{e}|\mathbf{R}]$ by a lattice vector $\mathbf{R}$ is a normal, Abelian subgroup of the space group. Any space group may be expressed as $S\tau$ where $S = \{s_1 e, s_2,\cdots\}$ is a set of $S^0$ rotation-translations whose translational parts are smaller in magnitude than any lattice vector. Any space group operator is representable in the form $g = s_i t(\mathbf{R})$ where $\bar{s}_i = \bar{g}$. In general $S$ is not a group, although the corresponding set $\bar{S} = \{\bar{e}, \bar{s}_2,\cdots\}$ of point rotation operators is a subgroup of the point group of the lattice.

The construction of FISO's for a general space group is approached as the systematic decomposition of the regular representation space of the space group, first into noninteracting subspaces and then into irreducible subspaces.[1]

It is convenient and usual to treat the translation subgroup as a finite group by imposing periodic boundary conditions (PBC's). This is done explicitly by defining a Born–Von Karman (BVK) lattice $\{\mathbf{R}_c\}$ in terms of primitive BVK lattice vectors $\mathbf{A}_i = N_i \mathbf{a}_i, i = 1,2,3$, which are integer $N_i$ mul-

 0022-2488/79/061274-07$01.00  

tiplies of the Bravais lattice vectors $\mathbf{a}_i$. PBC's allow only functions which are invariant to translations by a BVK lattice vector. In this sense lattice vectors which differ by a BVK lattice vector are equivalent. $\tau^0$ nonequivalent lattice vectors may be chosen in a BVK cell bounded by the primitive BVK lattice vectors. The ISO's for the translation group with PBC's are

$$T(\mathbf{k}) = \frac{1}{\tau^0}\sum_{\mathbf{R}} \exp(-\mathbf{k}\cdot i\mathbf{R})t(\mathbf{R}), \tag{10}$$

where the sum on $\mathbf{R}$ is restricted to lattice vectors in the BVK cell, and the vectors $\mathbf{k}$ ($\mathbf{K}$) are reciprocal to the BVK (space) lattice.

$$\mathbf{k}\cdot\mathbf{R}_c = 2\pi N, \quad (\mathbf{K}\cdot\mathbf{R} = 2\pi N'), \tag{11}$$

where $N(N')$ is an integer. Thus,

$$T(\mathbf{k}+\mathbf{K}) = T(\mathbf{k}), \tag{12}$$

and, in the sense that $\mathbf{k}' = \mathbf{k} + \mathbf{K}$ and $\mathbf{k}$ have identical translation ISO's, $\mathbf{k}$ and $\mathbf{k}'$ are equivalent. The set of $\tau^0$ nonequivalent $\mathbf{k}$ vectors may be chosen in a contiguous region of reciprocal space called the Brillouin zone. The properties of the translation group ISO's, paralleling Eqs. (2)–(5) are

$$T(\mathbf{k})^\dagger = T(\mathbf{k}), \tag{13}$$

$$t(\mathbf{R})T(\mathbf{k}) = \exp(i\mathbf{k}\cdot\mathbf{R})T(\mathbf{k}), \tag{14}$$

$$T(\mathbf{k})T(\mathbf{k}') = \Delta(\mathbf{k},\mathbf{k}')T(\mathbf{k}), \tag{15}$$

$$e = t(\mathbf{0}) = \sum_{\mathbf{k}} T(\mathbf{k}), \tag{16}$$

where $\Delta(\mathbf{k},\mathbf{k}') = 1(0)$ if $\mathbf{k}$ is (non)equivalent to $\mathbf{k}'$ and the sum on $\mathbf{k}$ includes the $\tau^0$ inequivalent $\mathbf{k}$ vectors from the Brillouin zone.

For each $T(\mathbf{k})$ the carrier space $ST(\mathbf{k})$ is stable and induces a unitary representation of the space group. This induced representation is irreducible if the conditions of Eq. (8) are fulfilled. It may be shown in general that, for space group operator $g$,

$$gT(\mathbf{k})g^{-1} = T(\bar{\mathbf{g}}\mathbf{k}). \tag{17}$$

Thus, the irreducibility conditions require that

$$T(\mathbf{k})T(\bar{s}_i\mathbf{k}) = 0, \quad i = 2,...,S^0 \tag{18}$$

or that all $\bar{s}_i\mathbf{k}$, $i = 2,...,S^0$ be inequivalent to $\mathbf{k}$. Points in the Brillouin zone which satisfy Eq. (18) are called general points.

Symmetry points, lines and planes in the Brillouin zone induce reducible representations from space $ST(\mathbf{k})$. An $n$-dimensional subspace $S_\mathbf{k}\,T(\mathbf{k}) = \{h_1 = e, h_2,...,h_n\}\,T(\mathbf{k})$ is defined consisting of $T(\mathbf{k})$ and all other components of $ST(\mathbf{k})$ which fail the irreducibility conditions Eq. (18). The set of rotational operators $H_\mathbf{k} = \{\bar{e},\bar{h}_2,...,\bar{h}_n\}$ of elements of $S_\mathbf{k}$ form a point group called the group of the $k$-vector. Let $\hat{h}_i = \exp(-i\mathbf{k}\cdot\boldsymbol{\tau}_i)h_iT(\mathbf{k})$. If $h_ih_j = h_mt(\mathbf{R}_{ijm})$ and $\bar{h}_i\mathbf{k} = \mathbf{k} + \mathbf{K}_i$, then $\bar{h}_i\bar{h}_j = \bar{h}_m$, $\mathbf{R}_{ijm} = \bar{h}_j^{-1}\boldsymbol{\tau}_j + \bar{h}_m^{-1}(\boldsymbol{\tau}_i - \boldsymbol{\tau}_m)$, and

$$\hat{h}_i\hat{h}_j = \exp[i\mathbf{K}_j\cdot\boldsymbol{\tau}_j + i\mathbf{K}_m\cdot(\boldsymbol{\tau}_i - \boldsymbol{\tau}_m)]\hat{h}_m. \tag{19}$$

For points inside the Brillouin zone $\bar{h}_i\mathbf{k} = \mathbf{k}(\mathbf{K}_i = \mathbf{0})$ the phase factor in Eq. (19) is unitary and components of the space $\hat{H}_\mathbf{k} = \{\hat{h}_1,\hat{h}_2,...,\hat{h}_n\}$ are isomorphic to the point group $H_\mathbf{k}$. In general, for points on the surface of the Brillouin zone the phase factor is not unitary and the components of $\hat{H}_k$ form a multiplier group.

The space $\hat{H}_k$ is a subspace of $ST(\mathbf{k})$ which is stable with respect to operations of the space subgroup $S_\mathbf{k}\tau$ and may be transformed into irreducible subspaces with respect to $S_\mathbf{k}\tau$ using the induction techniques as described in this paper and Ref. 1 or other methods. In general components of these irreducible subspaces may be chosen to be ISO's $P(\mathbf{k}a)_{ij}$ which are expressible in the factored form of $T(\mathbf{k})$ times a linear combination $\mathscr{P}(\mathbf{k}a)_{ij}$ of elements of $S_\mathbf{k}$ with complex coefficients,

$$P(\mathbf{k}a)_{ij} = T(\mathbf{k})\mathscr{P}(\mathbf{k}a)_{ij} = \mathscr{P}(\mathbf{k}a)_{ij}T(\mathbf{k}). \tag{20}$$

The ISO's for the space group $S\tau$ are induced with the "star" left coset generators $S_{\mathbf{k}\star} = \{s_{1\star} = e, s_{2\star},...,s_{m\star}\}$ of the space group with respect to $S_\mathbf{k}\tau$, $S\tau = S_{\mathbf{k}\star}(S_\mathbf{k}\tau)$. The partner ISO's for the space group have the form

$$P(\mathbf{k}A)_{(11)(pj)} = s_{p\star}P(\mathbf{k}a)_{ij}, \tag{21}$$

and a general induced ISO is

$$P(\mathbf{k}A)_{(qi)(pj)} = s_{p\star}P(\mathbf{k}a)_{ij}\,s_{q\star}^{-1}. \tag{22}$$

In this case the irreducibility conditions, Eq. (8),

$$P(\mathbf{k}a)_{ii}s_{p\star}P(\mathbf{k}a)_{jj}s_{p\star}^{-1}$$
$$= \mathscr{P}\mathbf{k}a)_{ii}T(\mathbf{k})T(\bar{s}_{p\star}\mathbf{k})s_{p\star}\mathscr{P}(\mathbf{k}a)_{jj}s_{p\star}^{-1} = 0,$$
$$i,j = 1,...,n_a, \quad p = 2,...,m,$$

are satisfied because $T(\mathbf{k})T(\bar{s}_{p\star}\mathbf{k}) = 0$. By the definition of $S_{\mathbf{k}\star}$ $\mathbf{k}$ is inequivalent to $\bar{s}_{p\star}\mathbf{k}$, $p = 2,...,m$. The set $\{\bar{s}_{p\star}\mathbf{k}\}$ of symmetry-related, but inequivalent $k$-vectors is called the star of the $k$-vector. Thus, the decompositon of the regular representation space of a general space group is completed. It may be helpful to verify this statement directly by reviewing the derivation and checking to see that all $S^0\tau^0$ operators are included. First, the $\tau^0$ inequivalent translation group operators were transformed to form $\tau^0$ stable subspaces $ST(\mathbf{k})$ each of dimension $S^0$. For a given $\mathbf{k}$, $m$ of these subspaces $ST(\bar{s}_{p\star}\mathbf{k})$, $p = 1,...,m$ are symmetry-related and there are $S^0m$ components associated with the star of $\mathbf{k}$. The decompostion of this $(S^0m)$-dimensional space began by reducing the $n$-dimensional subspace $S_\mathbf{k}\,T(\mathbf{k})$ into irreducible subspaces labeled $a$ of dimension $n_a$. For each IUR $a$ of $S_\mathbf{k}\tau$ the induction using $S_{k\star}$ produced an $(mn_a)$-dimensional IUR of $S\tau$. The total number of such ISO's is $(mn_a)^2$ for each IUR $a$. Thus, the total number of ISO's is $\Sigma_a(mn_a)^2 = m^2\Sigma_a n_a^2 = m^2 n = S^0 m$ because $\Sigma_a n_a^2 = n$, the dimension of $S_\mathbf{k}\,T(\mathbf{k})$, and by definition $S^0 = mn$.

In describing the FISO's for the diamond structure in Sec. III only the "angular" factors for the partner operators $\mathscr{P}(\mathbf{k}a)_{1j}$ defined by Eq. (20) and the members of the star coset $S_{\mathbf{k}\star}$ are given. The remaining operators may be constructed as described above.

## III. FISO's FOR THE DIAMOND STRUCTURE

The space lattice for the diamond structure is face-centered-cubic with primitive lattice vectors $\mathbf{a}_1 = (0,1,1)a/2$,

TABLE I. Selected space group operators for the diamond structure. The operators listed correspond to the coset expansion $S\tau$ of the space group with respect to the translation group $\tau$. The coset generators are $S = T_d(E,I)$ where $T_d = (E,T_3,T_3^2)(E,C_x)(E,C_y)(E,ID_x)$ is the tetrahedral group. The origin of coordinates is chosen to be a tetrahedral site with the equivalent tetrahedral site at $\mathbf{d} = (1,1,1)a/4$. Short and full symbols for the operators are listed in the first two columns of the table. The effect of the 3-space rotation operators on a point vector $(x,y,z)$ is shown in column 3. Spin-space operators listed in column 4 are represented in terms of 180° rotations ${}^{\sigma}R_a$ about selected symmetry axes $\hat{R}_a$: $\hat{T}_w = (1,1,1)/\sqrt{3}$, $\hat{C}_x = (1,0,0)$, $\hat{C}_y = (0,1,0)$, $\hat{C}_z = (0,0,1)$, $\hat{D}_x = (0,-1,1)/\sqrt{2}$. Then, ${}^{\sigma}R_a = -i\hat{R}_a\cdot\sigma$, where $\sigma_i$, $i = 1,2,3$ are matrices isomorphic to the Pauli spin matrices. A 360° rotation in spin space is denoted ${}^{\sigma}E_2$ and the identity by ${}^{\sigma}E$.

| Operator (short) | Operator (long) | Three-space rotations | spin–space rotations |
|---|---|---|---|
| $E$ | $[{}^sE\vert\mathbf{0}]{}^{\sigma}E$ | $(x,y,z)$ | ${}^{\sigma}E$ |
| $E_2$ | $[{}^sE\vert\mathbf{0}]{}^{\sigma}E_2$ | $(x,y,z)$ | $-{}^{\sigma}E$ |
| $T_3$ | $[{}^sT_3\vert\mathbf{0}]{}^{\sigma}T_3$ | $(y,z,x)$ | $[-{}^{\sigma}E + (3)^{1/2}\,{}^{\sigma}T_w]/2$ |
| $T_3^2$ | $[{}^sT_3^2\vert\mathbf{0}]{}^{\sigma}T_3^2$ | $(z,x,y)$ | $(-{}^{\sigma}E - (3)^{1/2}\,{}^{\sigma}T_w)/2$ |
| $C_x$ | $[{}^sC_x\vert\mathbf{0}]{}^{\sigma}C_x$ | $(x,-y,-z)$ | ${}^{\sigma}C_x$ |
| $C_y$ | $[{}^sC_y\vert\mathbf{0}]{}^{\sigma}C_y$ | $(-x,y,-z)$ | ${}^{\sigma}C_y$ |
| $C_z$ | $[{}^sC_z\vert\mathbf{0}]{}^{\sigma}C_z$ | $(-x,-y,z)$ | ${}^{\sigma}C_z$ |
| $ID_x$ | $[{}^sID_x\vert\mathbf{0}]{}^{\sigma}ID_x$ | $(x,z,y)$ | ${}^{\sigma}D_x$ |
| $I$ | $[{}^sI\vert\mathbf{d}]{}^{\sigma}I$ | $(-x,-y,-z)$ | ${}^{\sigma}E$ |
| $D_x$ | $[{}^sD_x\vert\mathbf{d}]{}^{\sigma}D_x$ | $(-x,-z,-y)$ | ${}^{\sigma}D_x$ |

$\mathbf{a}_2 = (1,0,1)a/2$, $\mathbf{a}_3 = (1,1,0)a/2$ expressed in terms of the conventional cube distance $a$. The corresponding reciprocal lattice is body-centered-cubic with primitive lattice vectors $\mathbf{b}_1 = (-1,1,1)2\pi/a$, $\mathbf{b}_2 = (1,-1,1)2\pi/a$, $\mathbf{b}_3 = (1,1,-1)\times 2\pi/a$. The point group is the full cubic group and the highest site symmetry group is the tetrahedral group $T_d$.

Both the relativistic (double) space group and the nonrelativistic space group will be treated. In the double space group each rotation in 3-space is accompanied by a rotation in spin-space. Selected space group operators are defined in Table I. Only those ten elements of the 96 cubic double group elements which appear explicitly in the FISO's are listed. The remaining operators may be obtained as products of those listed. A presuperscript $s(\sigma)$ distinguishes between 3(spin)-space operators in situations where an ambiguity might arise; otherwise the superscripts are suppressed.

As shown in Ref. 1 alternative spin representations may be obtained by associating the Pauli spin matrices

$$\pi_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \pi_2 = \begin{pmatrix} 0 & -1 \\ i & 0 \end{pmatrix}, \quad \pi_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

with an orthogonal set of unit vectors $\hat{n}_1$, $i = 1,2,3(\hat{n}_1 = \hat{n}_2\times\hat{n}_3)$. The spin matrices corresponding to the coordiante axes $\hat{x}_j$, $j = 1,2,3$ are then

$$\sigma_j = \hat{x}_j\cdot\boldsymbol{\sigma}, \tag{23}$$

where

$$\boldsymbol{\sigma} = \sum_i \hat{n}_i\pi_i. \tag{24}$$

The spin matrices $\sigma_j$ are isomorphic to the Pauli spin matrices $\pi_j$. The time-reversal operator is defined

$$K = K_0(i\hat{n}_2\cdot\boldsymbol{\sigma}) = K_0 i\pi_2, \tag{25}$$

where $K_0$ is the complex conjugation operator. The time-reversal operator commutes with each space group operator. In this paper only two-dimensional spinors will be treated. The results may be extended easily to four-dimensional (Dirac equation) spinors.

The flexibility obtained in choosing the spin representation is used to simplify the time-reversal properties of the FISO's and aids in their construction. For all the points and lines of the diamond structure two distinct spin representations are used. For the line $\Lambda$ and the points $\Gamma$ and $L$, $\hat{n}_2 = \hat{D}_x = (0,-1,1)/\sqrt{2}$, $\hat{n}_3 = \hat{T}_w = (1,1,1)/\sqrt{3}$, and $\hat{n}_1 = \hat{D}_x\times\hat{T}_w = (-2,1,1)/\sqrt{6}$. The corresponding spin-space rotations (presuperscript $\sigma$ is suppressed) are defined: $(C_a = -i\hat{n}_a\cdot\boldsymbol{\sigma})$,

$$C_{WD} = (-2C_x + C_y + C_z)/\sqrt{6}, \tag{26}$$

$$C_D = (-C_y + C_z)/\sqrt{2}, \tag{27}$$

$$C_W = (C_x + C_y + C_z)/\sqrt{3}. \tag{28}$$

For the lines $\Delta$, $\Sigma$, and $Z$ and the points $X$ and $W$ $\hat{n}_2 = \hat{D}_x$, $\hat{n}_3 = \hat{C}_x = (1,0,0)$ and $\hat{n}_1 = \hat{D}_x\times\hat{C}_x = (0,1,1)/\sqrt{2}$. The corresponding spin-space rotations (presuperscripts suppressed) are defined:

$$C_{XD} = (C_y + C_z)/\sqrt{2}, \tag{29}$$

$$C_D = (-C_y + C_z)/\sqrt{2}, \tag{30}$$

$$C_X = C_X. \tag{31}$$

The operator sets $(C_{WD},C_D,C_W)$, $(C_{XD},C_D,C_X)$, and $(C_X,C_Y,C_Z)$ are isomorphic to one another. In both representations ${}^{\sigma}D_x = {}^{\sigma}C_D$ and the time-reversal operator is $K = -K_0{}^{\sigma}D_x = -K_0{}^{\sigma}C_D$.

It may be helpful to the reader to be aware of the mnemonic significance inherent in the notation used to symbolize the group elements The notation for proper rotations makes specific reference to the axis of the rotation. The rotations $C_x,C_y,C_z$ are 180° rotations about the *cubic* coordinate axes $\hat{C}_x,\hat{C}_y,\hat{C}_z$. The rotations $T_3$ are 240° rotations about the principle *threefold* axis $\hat{T}_w$. The subscript $w$ here comes somewhat lamely from the observation that the remaining symmetry-related threefold axes are $\hat{T}_x = C_x\hat{T}_w$, $\hat{T}_y = C_y\hat{T}_w$,

TABLE II. Multiplication and commutation relations for selected operators. The relations are expressed in terms of double space group operators. They may be specialized to relations for point group operators by setting translations by lattice vectors $\mathbf{a}_i$, $i = 1,2,3$ to zero. Relations involving nonhybrid operators may be specialized to space (spin) operators by setting the spin (space) parts to unity. Relations between hybrid space-spin operators may be specialized to spin operators by setting the space parts to unity.

A. Relations between space double group operators.

$C_x^2 = C_y^2 = C_z^2 = ID_x^2 = E_2,\quad T_3T_3^2 = I^2 = E,\quad C_yC_z = E_2C_zC_y = C_x,\quad C_zC_x = E_2C_xC_z = C_y,$
$C_xC_y = E_2C_yC_x = C_z,\quad C_xT_3 = T_3C_y,\quad C_yT_3 = T_3C_z,\quad C_zT_3 = T_3C_x,\quad C_xT_3^2 = T_3^2C_z,\quad C_yT_3^2 = T_3^2C_x,$
$C_zT_3^2 = T_3^2C_y,\quad C_xID_x = E_2ID_xC_x,\quad C_yID_x = E_2ID_xC_z,\quad C_zID_x = E_2ID_xC_y,\quad T_3ID_x = ID_xT_3^2,\quad T_3^2ID_x = ID_xT_3,$
$I(ID_x) = ID_xI = D_x,\quad T_3I = IT_3,\quad IC_x = t(\mathbf{a}_1)C_xI,\quad IC_y = t(\mathbf{a}_2)C_yI,\quad IC_z = t(\mathbf{a}_3)C_zI.$

B. Relations involving hybrid operators Eqs. (26)–(31).

$C_{WD}^2 = C_D^2 = C_W^2 = C_{XD}^2 = E_2,\quad C_WT_3 = T_3C_W,\quad C_WT_3^2 = T_3^2C_W,\quad C_DC_W = E_2C_WC_D = C_{WD},\quad C_WC_{WD} = E_2C_{WD}C_W = C_D,$
$C_{WD}C_D = E_2C_DC_{WD} = C_W,\quad C_DC_X = E_2C_XC_D = C_{XD},\quad C_XC_{XD} = E_2C_{XD}C_X = C_D,\quad C_{XD}C_D = E_2C_DC_{XD} = C_X,$
$C_DT_3 = T_3C_D(-E - \sqrt{3}C_W)/2,\quad T_3C_D = C_D(-E + \sqrt{3}C_W)T_3/2,$
$C_DT_3^2 = T_3^2C_D(-E + \sqrt{3}C_W)/2,\quad T_3^2C_D = C_D(-E - \sqrt{3C_W})T_3^2/2,\quad ID_xC_W = E_2C_WID_x,\quad ID_XC_D = -E_2C_DID_X.$

$\widehat{T}_z = C_z\widehat{T}_w$. The rotation $D_x$ is a 180° rotation about the *dihedral* axis $\widehat{D}_x$. Similarly, the notation for the hybrid operators explicitly refers to the symmetry axis or related orthogonal axis with respect to which the hybrid operators were formed.

As shown in Ref. 1 the key to obtaining FISO's for the double cubic group is found in the construction of "hybrid" operators for the double space subgroup $DV = (E,C_X,C_Y,C_Z)(E,E_2)$. Since, the operators of $DV$ are isomorphic to the spin operators ${}^{\sigma}DV = (E^{\sigma},{}^{\sigma}C_X,{}^{\sigma}C_Y,{}^{\sigma}C_Z) \times ({}^{\sigma}E,{}^{\sigma}E_2)$, it follows that the hybrid operator sets defined by Eqs. (26)–(28) and Eqs. (29)–(30) are isomorphic to the operator set $(C_X,C_Y,C_z)$ of $DV$.

Selected multiplication and commutation properties of the space double group elements with themselves and with the hybrid operators are listed in Table II. Only those relations are listed which help one to directly verify the properties of the FISO. The FISO's for the diamond structure are listed in Tables III–V. These tables are intended to be essentially independent of the text. Each table includes nonrelativistic (Part A) and relativistic (Part B) FISO's for related symmetry points and lines. Each table caption specifies the relevant point groups of the $k$-vector and the associated cosets which generate the star of the $k$-vector. The FISO's allow a very compact notation. The connection between the present notation and the conventional notations of Bouckaert, Smoluchowski, and Wigner[2] (with an obvious modification to indicate inversion symmetry) and Koster[3] are given in each caption. An exception is the point $W$ where the IUR's found here have a similar form, but differ in detail from the results found by Koster. The expressions for the FISO's and the choice of degenerate IUR's used here are not unique. The choices made here reflect such considerations as the desirability of being able to determine compatibility between FISO's (IURs) for associated symmetry points and lines by inspection and the use of time-reversal to simplify the calculation of matrix elements with symmetry-adapted functions. A property common to all FISO's which are projection operators is that the individual factors may be placed in any order.

Also, in each of Tables III–V are listed relativistic symmetry-adapted functions (Parts C) in the form of linear combinations of products of nonrelativistic symmetry-adapted functions with appropriate spinors. The role of time-reversal symmetry here merits a more detailed discussion.

The time-reversal operator $K$ commutes with the Hamiltonian of the system and the operators of the space group. However, because $K$ involves a complex conjugation $K_0$ it has a distinctly different nature from the space group operators. An excellent summary of the properties of $K$ is given by Lax.[4] In applications $K$ performs one of two useful roles. Either it provides an additional relationship between functions which are symmetry related, or it leads to additional degeneracy. In the former case the additional relationship may be used to choose functions which give real Hamiltonian matrix elements.

Time-reversal symmetry has been approached here by first treating the nonrelativistic case where $K_0$ alone commutes with the nonrelativistic Hamiltonian $H_0$ and the diamond structure group elements. It is easy to show in general that all nonrelativistic eigenfunctions of $H_0\psi_E$ may be chosen to satisfy the relation

$$IK_0\psi_E = \psi_E. \tag{32}$$

Since a plane wave

$$\psi_{\mathbf{k}} = \exp\frac{[i\mathbf{k}\cdot(\mathbf{r} - \mathbf{d}/2)]}{\sqrt{V}} \tag{33}$$

satisfies Eq. (32), it follows that matrix elements of plane waves with an operator $H_0$, which commutes with $IK_0$,

$$(\psi_{\mathbf{k}},H_0\psi_{\mathbf{k}'})^* = (IK_0\psi_{\mathbf{k}}.H_0IK_0\psi_{\mathbf{k}'}) = (\psi_{\mathbf{k}},H_0\psi_{\mathbf{k}'}), \tag{34}$$

are real. Thus, a Hamiltonian matrix formed with respect to a complete set of plane waves will be real and the eigenfunctions $\psi_E = \Sigma_{\mathbf{k}}c_{\mathbf{k}}\psi_{\mathbf{k}}$ may be chosen to have real coefficients $c_{\mathbf{k}}$. Hence, $IK_0\psi_E = \psi_E$ and any other basis functions may be chosen to have the property Eq. (32). In particular symmetry-adapted basis functions,

$$f_{\mathbf{k}aj}(\mathbf{r})_I = P(\mathbf{k}a)_{1j}f(\mathbf{r})_I \tag{35}$$

generated with ISO $P(\mathbf{k}a)_{1j}$ from a basis function which satis-

TABLE III. FISO's and symmetry-adapted functions for symmetry points gamma, ($\mathbf{k}_\Gamma = \mathbf{0}$) and $L$ ($\mathbf{k}_L = (1,1,1)\pi/a$), and the associated symmetry line lambda ($\mathbf{k}_\Lambda = (p,p,p)\pi/a, 0<p<1$). The point groups of the $k$-vector are: $G_\Lambda = (E,T_3,T_3^2)(E,ID_x)$, $G_L = G_\Lambda(E,I)$, and $G_\Gamma = G_L(E,C_x)(E,C_z)$. The generators for the stars of the $k$-vector are $S_L = (E,C_x)(E,C_z)$, $S_\Lambda = S_L(E,I)$, and $S_\Gamma = E$. The connection between the conventional irreducible representation labels and the notation used here is $A++,\Gamma_1$ $A+-,\Gamma_{2'}$ $A-+,\Gamma_2$ $A--,\Gamma_{1'}$ $Ep,\Gamma_{12p}$ $T-+,\Gamma_{25'}$ $T--,\Gamma_{15},T++,\Gamma_{15'}$, $T+-,\Gamma_{25}$ $B-p,\Gamma_{6p}$ $B+p,\Gamma_{7p}$ $Fp,\Gamma_{8p}$ $\Lambda+,\Lambda_1$ $\Lambda-,\Lambda_2$ $\Lambda 3,\Lambda_3$ $\Lambda^r-,\Lambda_4$ $\Lambda^r+\Lambda_5$ $\Lambda^r,\Lambda_6$ $L+p,L_{1p}$ $L-p,L_{2p}$ $L3p,L_{3p}$ $L^r-p$, $L_{4p}$ $L^r+p,L_{5p}$ $L^rp,L_{6p}$. In the tables below only the angular dependent factors [Eq. (20)] of the FISO's are listed. The FISO's are to be completed by multiplying by an appropriate translation operator. The factors in projection operators may be arranged in any order.

A. Nonrelativistic angular factors for FISO's [$s,p = +/-, \omega = \exp(2\pi i/3)$].

$P(\Lambda s) = (E+T_3+T_3^2)(E+sID_x)/6$, $P(Lsp) = P(\Lambda s)(E+pI)/2$, $P(Asp) = P(Lsp)(E+C_x)(E+C_z)/4$,
$P(\Lambda 3)_{11} = (2E-T_3-T_3^2)(E+ID_x)/6$, $P(\Lambda 3)_{12} = (T_3-T_3^2)P(\Lambda 3)_{11}/\sqrt{3}$, $P(L3p)_{1j} = P(\Lambda 3)_{1j}(E+pI)/2$,
$P(Ep)_{1j} = P(L3p)_{1j}(E+C_x)(E+C_z)/4$, $j=1,2$, $P(Tsp)_{11} = (E+C_x)(E-C_z)(E-sID_x)(E+pI)/16$,
$P(Tsp)_{1j} = T_3^{j-1}P(Tsp)_{11}$, $j=1,2,3$.

All operators commute with $IK_0$.

B. Angular factors for relativistic FISO's.
($s,p = +/-, \omega = \exp(2\pi i/3)$)

$P(\Lambda^r s) = (E+T_3+T_3^2)(E+isID_x)(E-E_2)/12$, $P(L^r sp) = p(\Lambda^r s)(E+pI)/2$,
$P(Fp)_{11} = (E+T_3+T_3^2)(E+pI)(E-iC_w)(E-E_2)/24$,
$P(Fp)_{12} = ID_xP(Fp)_{11}$, $P(Fp)_{13} = C_DP(Fp)_{11}$, $P(Fp)_{14} = ID_xC_DP(Fp)_{11}$,
$P(\Lambda^r)_{11} = (E+\omega^*T_3+\omega T_3^2)(E-E_2)/6$, $P(\Lambda^r)_{12} = ID_xP(\Lambda^r)_{11}$,
$P(L^rp)_{1j} = P(\Lambda^r)_{1j}(E+pI)/2$, $j=1,2$,
$P(Bsp)_{1j} = P(L^rp)_{1j}(E-iC_W)(E-sID_xC_D)/4$, $j=1,2$

Time-reversal: $ID_xIKP(D)_{ij} = P(D)_{ij}ID_xIK$, $D = \Lambda^r, L^rp, Bsp, Fp$,
$IKP(\Lambda^r s) = P(\Lambda^r - s)IK$, $IKP(L^r sp) = P(L^r - sp)IK$.

C. Symmetrized relativistic functions. [The spin-space representation is defined by Eqs. (26)–(28).]

$\alpha = \begin{pmatrix}1\\0\end{pmatrix}, \beta = \begin{pmatrix}0\\1\end{pmatrix}, \gamma_n = (\omega^{2n}\sqrt{2}\alpha + \omega^n\beta)/\sqrt{3} = {}^\sigma T_3^n\gamma_0, \gamma^n = (-\omega^{2n}\alpha + \sqrt{2}\omega^n)/\sqrt{3} = {}^\sigma T_3^n\gamma^0, K\beta = \alpha, K\gamma^n = \gamma_n$,
${}^\sigma D_x\gamma_0 = \gamma^0, {}^\sigma D_x\gamma_1 = \gamma^2 \; {}^\sigma D_x\gamma_2 = \gamma^1$.

$\phi(\Lambda^r)_{11} = \sqrt{s}\psi(Asp)\beta$, $\phi(\Lambda^r)_{12} = -s\sqrt{s}\psi(\Lambda s)\alpha$,
$\phi(L^rp)_{11} = \sqrt{s}\psi(Lsp)\beta$, $\phi(L^rp)_{12} = -s\sqrt{s}\psi(Lsp)\alpha$,
$\phi(Bsp)_{11} = \sqrt{s}\psi(Asp)\beta$, $\phi(Bsp)_{12} = -s\sqrt{s}\psi(Bsp)\alpha$,

$\phi(\Lambda^r s) = \sqrt{s}[(\psi(\Lambda 3)_{11} - i\psi(\Lambda 3)_{12})\alpha + is(\psi(\Lambda 3)_{11} + i\psi(\Lambda 3)_{12})\beta]/2$,
$\phi(L^r sp) = \sqrt{s}[(\psi(L3p)_{11} - i\psi(L3p)_{12})\alpha + is(\psi(L_3p)_{11} + i\psi(L3p)_{12})\beta]/2$,
$\phi(\Lambda^r)_{11} = (\psi(\Lambda 3)_{11} + i\psi(\Lambda 3)_{12})\alpha/\sqrt{2}$, $\phi(\Lambda^r)_{12} = (\psi(\Lambda 3)_{11} - i\psi(\Lambda 3)_{12})\beta/\sqrt{2}$,
$\phi(L^rp)_{11} = (\psi(L3p)_{11} + i\psi(L3p)_{12})\alpha/\sqrt{2}$, $\phi(L^rp)_{12} = (\psi(L3p)_{11} - i\psi(L3p)_{12})\beta/\sqrt{2}$,
$\phi(Fp)_{11} = (\psi(Ep)_{11} + i\psi(Ep)_{12})\beta/\sqrt{2}$, $\phi(Fp)_{12} = -(\psi(Ep)_{11} - i\psi(Ep)_{12})\alpha/\sqrt{2}$,
$\phi(Fp)_{13} = -(\psi(Ep)_{11} + i\psi(Ep)_{12})\alpha/\sqrt{2}$, $\phi(Fp)_{14} = -(\psi(Ep)_{11} - i\psi(Ep)_{12})\beta/\sqrt{2}$.

$\phi(Bsp)_{11} = \sqrt{s}(\psi(Tsp)_{11}\gamma_0 + \omega^*\psi(Tsp)_{12}\gamma_1 + \omega\psi(Tsp)_{13}\gamma_2)/\sqrt{3}$,
$\phi(Bsp)_{12} = -s\sqrt{s}[\psi(Tsp)_{11}\gamma^0 + \omega\psi(Tsp)_{12}\gamma^1 + \omega^*\psi(Tsp)_{13}\gamma^2]/\sqrt{3}$,
$\phi(Fp)_{11} = \sqrt{s}[\psi(Tsp)_{11}\gamma_0 + \psi(Tsp)_{12}\gamma_1 + \psi(Tsp)_{13}\gamma_2]/\sqrt{3}$,
$\phi(Fp)_{12} = -s\sqrt{s}[\psi(Tsp)_{11}\gamma^0 + \psi(Tsp)_{12}\gamma^1 + \psi(Tsp)_{13}\gamma^2]/\sqrt{3}$,
$\phi(Fp)_{13} = -\sqrt{s}[\psi(Tsp)_{11}\gamma^0 + \omega^*\psi(Tsp)_{12}\gamma^1 + \omega\psi(Tsp)_{13}\gamma^2]/\sqrt{3}$,
$\phi(Fp)_{14} = -s\sqrt{s}[\psi(Tsp)_{11}\gamma_0 + \omega\psi(Tsp)_{12}\gamma_1 + \omega^*\psi(Tsp)_{13}\gamma_2]/\sqrt{3}$.

Time reversal: $ID_xIK\Phi(D)_{1j} = \Phi(D)_{1j}$, $D = \Lambda^r, L^rP, Bsp, Fp, IK\phi(\Lambda^r s) = -\phi(\Lambda^r - s)$, $IK\phi(L^r sp) = -\phi(L^r - sp)$. All nonrelativistic functions $\psi$ satisfy $IK_0\psi = \psi$.

fies Eq. (32) will itself satisfy Eq. (32) provided $P(\mathbf{k}a)_{1j}$ commutes with $IK_0$. All nonrelativistic FISO's listed in Tables III–V are chosen to commute with $IK_0$, and matrix elements

$$(f_{\mathbf{k}aj}(\mathbf{r})_I, H_0 f(\mathbf{k}'a'j'(\mathbf{r})_J) = \delta_{\mathbf{k}\mathbf{k}'}\delta_{aa'}\delta_{jj'}(f(\mathbf{r})_I, H_0 P(\mathbf{k}a)_{11} f(\mathbf{r})_J) \qquad (36)$$

with any other operator $H_0$ that commutes with the relevant group and $IK_0$ are real.

The symmetry-adapted relativistic functions $\phi(A^r)_{1j}$ of Tables III–V are obtained by applying relativistic ISO's $P(A^r)_{1j}$ to the product of a symmetry-adapted nonrelativistic function $\psi(\mathbf{k}a)_{1i}$ and a spinor $\gamma$,

$$\phi(\mathbf{k}A^r)_{1j} = P(\mathbf{k}A^r)_{1j}N_a\psi(\mathbf{k}a)_{1i}\gamma, \qquad (37)$$

where $N_a$ is a normalization factor. Having specified all relevant transformation properties of $\psi(ka)_{1i}$, all that remains

TABLE IV. FISO's and symmetry-adapted functions for the symmetry point $X(\mathbf{k}_X=(2,0,0)\pi/a)$, and the associated symmetry lines $(\mathbf{k}_\Delta=(p,0,0)\pi/a, 0<p<2)$ and $(\mathbf{k}_\Sigma=(0,p,p,)\pi/a, 0<p<2)$. The point groups of the $k$-vector are $G_\Delta=(E,C_x)(E,ID_x)(E,IC_z), G_\Sigma=(E,ID_x)(E,IC_x)$, and $G_X=G_\Delta(E,I)$. The stars of the $k$-vector are generated by $S_X=(E,T_3,T_3^2)$, $S_\Delta=S_X(E,I)$, and $S_\Sigma=S_X(E,C_x)(E,C_z)$. The connection between the conventional irreducible representation labels and the notation used here is $\Delta++,\Delta_1$ $\Delta+-,\Delta_{2'}$, $\Delta-+,\Delta_2$ $\Delta--,\Delta_{1'}$ $\Delta_5,\Delta_5$ $\Delta'+,\Delta_6$ $\Delta'-,\Delta_7$ $\Sigma++,\Sigma_1$ $\Sigma+-,\Sigma_3$ $\Sigma-+,\Sigma_4$ $\Sigma--,\Sigma_2$ $\Sigma',\Sigma_5$ $X+,X_1$ $X-,X_2$ $X'+,X_3$ $X'-,X_4$ $X',X_5$. In the tables below only the angular dependent factors of the FISO's are listed. The complete FISO's are obtained by multiplying by an appropriate translation operator. The factors in projection operators may be arranged in any order.

A. Nonrelativistic angular factors for FISO's. $(s,t=+/-.\ e=\exp(-i\mathbf{k}\cdot\mathbf{d})$. (Where $\mathbf{k}$ is the $k$-vector for the FISO.)
$P(\Delta st)=(E+C_x)(E+sID_x)(E+teIC_y)/8,\quad P(Xs)_{11}=P(\Delta s+),\quad P(Xs)_{12}=IP(Xs)_{11},$
$P(\Delta_5)_{11}=(E-C_x)(E+eIC_y)/4,\quad P(\Delta_5)_{12}=ID_xP(\Delta_5)_{11},$
$P(X's)_{11}=(E-C_x)(E+sID_x)(E+I)/8,\quad P(X's)_{12}=eIC_yP(X's)_{11}$
$P(\Sigma st)=(E+sID_x)(E+teIC_x)/4.$
All operators commute with $IK_0$.
B. Angular factors for relativistic operators. $(s=+/-.\ e=\exp(-i\mathbf{k}\cdot\mathbf{d}))$.
$P(\Delta's)_{11}=(E+iC_x)(E-seID_xIC_D)(E-E_2)/8,\quad P(\Delta's)_{12}=ID_xP(\Delta's)_{11},$
$P(X')_{11}=P(\Delta'-)_{11},\quad P(X')_{12}=ID_xP(X')_{11},\quad P(X')_{13}=IP(X')_{11},$
$P(X')_{14}=D_xP(X')_{11},\quad P(\Sigma')_{11}=(E+ieIC_x)(E-E_2)/4,\quad P(\Sigma')_{12}=ID_xP(\Sigma')_{11},$

Time-reversal: All FISO's commute with $ID_xIK$.

C. Symmetrized relativistic functions. [The spin-space representation is defined by Eqs. (26)–(28). $\alpha=\binom{1}{0}, \beta=\binom{0}{1}$.]
$\phi(\Delta's)_{11}=\sqrt{s}\,\psi(\Delta st)\alpha,\quad \phi(\Delta'st)_{12}=s\sqrt{s}\,\psi(\Delta st)\beta,$
$\phi(\Delta's)_{11}=\sqrt{is}\,(\psi(\Delta_5)_{11}-is\psi(\Delta_5)_{12})\beta/\sqrt{2},$
$\phi(\Delta's)_{12}=is\sqrt{is}\,(\psi(\Delta_5)_{11}+is\psi(\Delta_5)_{12})\alpha/\sqrt{2},$
$\phi(\Sigma')_{11}=\sqrt{s}\,\psi(\Sigma s+)\alpha,\quad \phi(\Sigma')_{12}=s\sqrt{s}\,\psi(\Sigma s+)\beta,$
$\phi(\Sigma')_{11}=\sqrt{s}\,\psi(\Sigma s-)\beta,\quad \phi(\Sigma')_{12}=-s\sqrt{s}\,\psi(\Sigma s-)\alpha,$
$\phi(X')_{11}=i\psi(X-)_{11}\alpha,\quad \phi(X')_{12}=-i\psi(X-)_{11}\beta,\quad \phi(X')_{13}=i\psi(X-)_{12}\alpha,\quad \phi(X')_{14}=-i\psi(X-)_{12}\beta,$
$\phi(X')_{11}=\psi(X+)_{12}\alpha,\quad \phi(X')_{12}=\psi(X+)_{12}\beta,\quad \phi(X')_{13}=\psi(X+)_{11}\alpha,\quad \phi(X')_{14}=\psi(X+)_{11}\beta,$
$\phi(X')_{11}=\sqrt{s}\,[\psi(X's)_{11}+is\psi(X's)_{12}]\beta/\sqrt{2},\quad \phi(X')_{12}=-s\sqrt{s}\,[\psi(X's)_{11}-is\psi(X's)_{12}]\alpha/\sqrt{2},$
$\phi(X')_{13}=\sqrt{s}\,[\psi(X's)_{11}-is\psi(X's)_{12}]\beta/\sqrt{2},\quad \phi(X')_{14}=-s\sqrt{s}\,[\psi(X's)_{11}+is(X's)_{12}]\alpha/\sqrt{2},$

Time-reversal: All relativistic functions $\phi$ satisfy $ID_x IK\phi=\phi$. All nonrelativistic funtions $\psi$ satisfy $IK_0\phi=\phi$.

TABLE V. FISO's and symmetry-adapted functions for symmetry point $W(\mathbf{k}_W=(1/2,0,1)2\pi/a)$ and the associated symmetry line $Z(\mathbf{k}_Z=(p,0,1)2\pi/a, 0<p<1/2)$. The point groups of the $k$-vector are $G_Z=(E,C_x)(E,IC_z)$ and $G_W=G_Z(E,D_x)$. The generators of the stars of the $k$-vector are $S_W=(E,T_3,T_3^2)(E,I)$ and $S_Z=S_W(E,ID_x)$. In the tables below only the angular dependent factors of the FISO's are listed. The FISO's are completed by multiplying by an appropriate translation operator. The factors in projection operators may be arranged in any order.

A. Nonrelativistic angular factors for FISO's. [$s=+/-$, Hybrid 3-space operators are defined: $\bar{C}_D=e(-IC_y+iIC_z)/\sqrt{2}$, $\bar{C}_{XD}=e(IC_y+iIC_z)/\sqrt{2}$ where $e=\exp(-i\mathbf{k}\cdot\mathbf{d})$.]
$P(Z)_{11}=(E+\bar{C}_D)/2,\quad P(Z)_{12}=iC_xP(Z)_{11},$
$P(Ws)_{11}=(E+\bar{C}_D)(E+sD_x)/4,\quad P(Ws)_{12}=iC_xP(Ws)_{11}$

All FISO's commute with $IK_0$.

B. Angular factors for relativistic FISO's. [$s,t,u=+/-$, $e=\exp(-i\mathbf{k}\cdot\mathbf{d})$]
$P(Z'st)=(E+isC_x)(E-teIC_z)(E-E_2)/8,$
$P(W'tu)=(E-iC_x)(E-teIC_z)(E+iuD_x)(E-E_2)/16$
$P(W't)_{11}=(E+iC_x)(E-teIC_z)(E-E_2)/8,\quad P(W't)_{12}=D_xP(W't)_{11}$

Time-reversal: $IKP(Z'st)=P(Z's-t)IK,\quad IKP(W'tu)=P(W'-t-u)IK,$
$IKD_xP(W't)_{ij}=P(W't)_{ij}IKD_x.$

C. Symmetrized relativistic functions. [The spin-space representation is defined by Eqs. (26)–(28),$\alpha=\binom{1}{0}$, $\beta=\binom{0}{1}$.]

$\phi(Z'st)=s^{-1/4}[\psi(Z)_{11}(\alpha+\sqrt{st}\,\beta)-is\psi(Z)_{12}(\alpha-\sqrt{st}\,\beta)]/2,$
$\phi(W'tu)=\exp(-i\pi/4)[\psi(W(tu))_{11}(\alpha+it\beta)+i\psi(W(tu))_{12}(\alpha-it\beta)]/2$
$\phi(W't)_{11}=\sqrt{s}\,[\psi(Ws)_{11}(\alpha+t\beta)-i\psi(Ws)_{12}(\alpha-t\beta)]/2$
$\phi(W't)_{12}=-ts\sqrt{s}\,[\psi(Ws)_{11}(\alpha-t\beta)-i\psi(Ws)_{12}(\alpha+t\beta)]/2$
Time-reversal: $IK\phi(Z_r st)=t\phi(Z's-t),\quad IK(W'tu)=t\phi(W'-t-u),$
$IKD_x\phi(W't)_{ij}=\phi(W't)_{ij}$. All nonrelativistic functions $\psi$ satisfy $IK_0\psi=\psi$.

free to choose are the spin representation, the IUR for the degenerate FISO's and an overall phase factor. The relativistic ISO's cannot be chosen to commute with $IK$ and $IK\psi(\mathbf{k}a)_{1i}\gamma$ is orthogonal to $\psi(\mathbf{k}a)_{1j}\gamma$, so an exact parallel to the nonrelativistic treatment of time-reversal symmetry is not possible. Degenerate relativistic FISO's were chosen so that either $ID_x$ or $D_x$ produces partner ISO's and they all commute with either $ID_xIK$ or $D_xIK$. The spin representations used in Eqs. (26)–(31) were chosen to simplify these operators,

$$ID_xIK = {}^sD_xK_0 \quad \text{or} \quad D_xIK = {}^sD_xIK_0. \tag{38}$$

The phase factor for degenerate symmetry-adapted relativistic functions $\phi$ is chosen to satisfy

$$ID_xIK\phi = \phi \quad \text{or} \quad D_xIK\phi = \phi. \tag{39}$$

In the case of nondegenerate symmetry-adapted relativistic functions, $IK$ provides an additional degeneracy and the phase factor was chosen to give a simple relation between the time-reversal related functions.


[1]N.O. Folland, J. Math. Phys. **18**, 31 (1977).
[2]L.P. Bouckaert, R. Smoluchowski, and E. Wigner, Phys. Rev. **50**, 58 (1936).
[3]G.F. Koster, *Solid State Physics* (Academic, New York, 1957), Vol. V, 174.
[4]M. Lax, *Symmetry Principles in Solid State and Molecular Physics* (Wiley, New York, 1974).

# Inhomogeneous local gauge transformations in spacetimes with torsion

Francis J. Flaherty[a)] and G. David Kerlick[b)]

*Department of Mathematics, Oregon State University, Corvallis, Oregon 97331*
(Received 10 July 1978)

The formalism of bundle maps is developed, then applied to several situations of physical interest. The bundle isomorphisms are defined to be inhomogeneous gauge transformations. First, we examine the compatibility between local gauge invariance with respect to an internal symmetry and the minimal coupling of the free gauge field Lagrangian to spacetime. In both homogeneous and inhomogeneous cases, restrictions arise which limit the torsion of the spacetime covariant derivative when applied to the gauge potential. Second, we examine the group of diffeomorphisms as a spacetime symmetry in the local gauge theoretic approaches to general relativity and the Einstein–Cartan theory of gravity.

## 1. INTRODUCTION

Symmetry principles are fundamental to most physical theories. The framework of local gauge invariance[1–4] fixes this idea in a concrete form, yielding not only conservation laws corresponding to each observed symmetry but also new physical fields and field equations. For example, invariance of a matter field with respect to local phase transformations [circle group SO(2) or U(1)] yields Maxwell's field $F^{\mu\nu}$.

In general relativity, it is often stated that the entire group of diffeomorphisms of spacetime are the symmetries. Yet this group fits imperfectly into the local gauge framework, since the diffeomorphisms act on spacetime itself rather than internally (on the fibers). Our purpose here is to provide a better formalism for dealing with diffeomorphisms as a gauge symmetry, for which purpose we shall introduce bundle mappings and "inhomogeneous gauge transformations."

In what follows, we shall consider only classical field theory, thus reserving questions of quantization or symmetry breaking for later investigation.

Let $P$ be a principal bundle over a manifold $M$ (spacetime) with group $G$. The group $G$ acts naturally on $P$ by left or right multiplication. The right action is the global gauge group. Denote the projection of $P \to M$ by $\pi$. Since $P$ is a local product of $M$ with $G$, local coordinates $x$ on $M$ and $s$ in $G$ provide local coordinates $(x,s)$ on $P$. A *bundle map* of $P$ is a smooth function $\eta$: $P \to P$ that maps fibers into fibers and commutes with the *left* action of $G$. The map $\eta$ then induces a map $h$: $M \to M$ and the following diagram commutes:

$$\begin{array}{ccc} P & \xrightarrow{\ \eta\ } & P \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ M & \xrightarrow{\ h\ } & M. \end{array}$$

In terms of local coordinates,

$$\eta(x,s) = (h(x), s\tilde{\eta}(x)). \tag{1}$$

[The $\tilde{\eta}(x)$ is written on the right because $\eta$ commutes with left multiplication.] This generalizes the usual notion of local gauge transformation. Indeed, if $\eta$ induces the identity on $M$ or

$$\eta(x,s) = (x, s\tilde{\eta}(x)), \tag{2}$$

then $\eta$ is an ordinary homogeneous local gauge transformation (compare: Atiyah, Hitchen, Singer[5]; Souriau[6]; Utiyama[3]). The most interesting specialization of bundle maps occurs when $\eta$ is an isomorphism, in which case $\eta$ has an inverse which is also a bundle map. Clearly, this is the case when $\eta$ induces a diffeomorphism on $M$. Such an isomorphism will be called an inhomogeneous (local) gauge transformation. Denoting the group of gauge transformations (those inducing the identity on $M$) by $G$ and the group of diffeomorphisms of $M$ by $\mathscr{D}$, it is easy to see that $G$ is a normal subgroup of $I$, the group of inhomogeneous gauge transformations and that the quotient group $I/G$ is isomorphic to $\mathscr{D}$. Hence $I$ is isomorphic to a *direct* product of $\mathscr{D}$ with $G$ (in contrast to the semidirect product which occurs in the Poincare group). So if $\eta$ is in $I$, then

$$\eta = (h, \bar{\eta}), \quad h \text{ in } \mathscr{D}, \text{ and } \bar{\eta} \text{ in } G.$$

Further $\eta = (h, \mathrm{id}_P)\cdot(\mathrm{id}_{M'}, \bar{\eta})$, and we shall identify $h$ with $(h, \mathrm{id}_P)$, $\bar{\eta}$ with $(\mathrm{id}_{M'}, \bar{\eta})$, and write $\eta = h\bar{\eta}$.

For simplicity let us assume tht $G$ is a subgroup of the general linear group in $k$-real variables. A gauge potential $A$ in $P$ then has a local representation

$$A = s^{-1}ds + s^{-1}\theta s, \tag{3}$$

with $\theta$ a $k \times k$ matrix of 1-forms locally defined on $M$. One can think of $\theta$ as coming from a covariant derivative on the associated vector bundle to $P$ with fiber dimension $k$. The effect of the gauge transformation $\eta = h\bar{\eta}$ on $A$, denoted by $\eta\cdot A$ is

$$\eta\cdot A = \bar{\eta}^{-1}h^*(A)\bar{\eta} + \bar{\eta}^{-1}d\bar{\eta}. \tag{4}$$

Note that $h^*(A)$ makes sense, since $h$ is being identified with $(h, \mathrm{id}_P)$. The effect of an inhomogeneous gauge transformation on the gauge field or curvature form is readily computed to be

$$\eta\cdot F = \bar{\eta}^{-1}h^*(F)\bar{\eta},$$

with $F = dA + A \wedge A$.

[a)]Research partially supported by NSF Grant MCS-7704868.
[b)]The authors wish to thank the referee for his helpful comments.

 0022-2488/79/061281-04$01.00  

Thus, we see that gauge invariance extends to inhomogeneous gauge transformations if we act not only by the homogeneous part but also pullback via the diffeomorphism.

Any form that behaves as the gauge field under an inhomogeneous gauge transformation is said to satisfy the *naturality condition*. Moreover, any gauge theory whose gauge field satisfies the naturality condition will be called a *natural gauge theory*.

## 2. EXAMPLES OF INHOMOGENEOUS LOCAL GAUGE TRANSFORATIONS

### A. Maxwell's Theory

In the ordinary (homogeneous) gauge theory of electromagnetism, if the matter field $\psi$ undergoes the global gauge transformation $\psi \to \bar{\psi} = e^{i\Lambda}\psi$, the action remains invariant. Requiring invariance of the action under the local transformation $\psi \to \bar{\psi} = e^{i\Lambda(x)}\psi$ means one must modify the derivative $d\psi$; thus,

$$d\psi \to d\psi + ieA\psi, \tag{5}$$

where the gauge potential $A$ transforms inhomogeneously under the gauge transformation

$$A \to \bar{A} = A + d\Lambda. \tag{6}$$

The gauge field,

$$F = dA, \tag{7}$$

and the action for the free gauge field,

$$S = \int F \wedge {}^*F, \tag{8}$$

remain invariant under a gauge transformation.

In a relativistic theory, the equivalence principle suggests the principle of minimal coupling, namely $d \to D$ everywhere, where $D$ is the covariant exterior derivative, which may have torsion. We shall assume that the metric of the base space is preserved under covariant differentiation, so $Dg = 0$.

This requirement does not effect the gauge potential (6) since $D\Lambda = d\Lambda$, but it does effect the gauge field $F$, since $F{:} = DA$ implies

$$\bar{F} = D\bar{A} = DA + D(d\Lambda).$$

Thus, $\bar{F} = F$ implies that $D$ must be torsionless.

Hojman, Rosenbaum, Ryan, and Shepley[7] (HRRS) attempt to circumvent this difficulty by modifying (6) to read

$$A \to \bar{A} = A + bd\Lambda, \tag{9}$$

where $b$ is a field of (1,1) tensors. We can interpret this in our picture as follows: The gauge group is pulled back under diffeomorphisms but the connection is left unchanged. In this case $b$ corresponds to $h^*$. Using (9) and assuming minimal coupling $F = DA$, one is led to

$$F \to \bar{F} = DA + D(bd\Lambda). \tag{10}$$

Requiring local gauge invariance, $\bar{F} = F$, then implies

$$D(bd\Lambda) = 0 \tag{11}$$

(compare HRRS[7], Eq. 15). The simple choice of $b = e^{\lambda}\cdot 1$, 1 the unit tensor, implies a torsion proportional to the gradient of $\lambda$, which can propagate.

The assertion (9) violates the naturality condition because the gauge potential $A$ is introduced on the pullback bundle. Rather, an inhomogeneous local gauge theory would require

$$A \to \bar{A} = h^*(A + d\Lambda), \tag{12}$$

that is, the connection is also pulled back. Minimal coupling now implies,

$$\bar{F} = Dh^*(A + d\Lambda) + h^*(DA + Dd\Lambda). \tag{13}$$

The condition of naturality, which generalizes local gauge invariance, is simply

$$\bar{F} = h^*F, \tag{14}$$

which in turn implies

$$Dh^*(A + d\Lambda) + h^*(Dd\Lambda) = 0. \tag{15}$$

Calculation then shows that the torsion of $D$ must vanish as in the ordinary case.

### B. Yang–Mills Theory[2]

Here the gauge group is SU(2), non-Abelian. A homogeneous local gauge transformation $\phi \epsilon G$ takes $\psi$ into $\phi\psi$ and takes a gauge potential $A$ into $A' = \phi^{-1}A\phi + \phi^{-1}d\phi$. An inhomogeneous transformation $\eta = h\bar{\eta}$ takes $A$ into

$$\bar{A} = \bar{\eta}^{-1}h^*(A)\bar{\eta} + \bar{\eta}^{-1}d\bar{\eta}, \tag{16}$$

where $h^*(A) = s^{-1}ds + s^{-1}h^*s$ as in Sec. 1. If $F = DA + A\Lambda A$, the naturality condition states

$$\eta \cdot F = \bar{F} = \bar{\eta}^{-1}h^*(F)\bar{\eta}. \tag{17}$$

Hence $A$ satisfies the equation

$$Dh^*(A)\cdot\bar{\eta} + D(d\bar{\eta}) = 0, \tag{18}$$

which after a lengthy computation shows that $D$ must be torsionless.

### C. Inhomogeneous local gauge theory of the Lorentz group (comments on the Einstein–Cartan theory)

Here, we are dealing with a spacetime symmetry, but since the Lorentz group acts on the fibers of the bundle of orthonormal frames over $M$, we may regard it as the homogeneous part of the gauge transformation just as in Yang–Mills. Note that the theory which results will not be a gauge theory of the Poincarè group, but rather one of the direct product of the Lorentz group by the diffeomorphisms.

The results look very much like the Yang–Mills field in the previous example. The rotational gauge potential (i.e., the gauge potential for the Lorentz group) $A$ must transform as

$$A \to \bar{A} = \bar{\eta}^{-1}h^*(A)\bar{\eta} + \bar{\eta}^{-1}d\bar{\eta}, \tag{19}$$

where $\bar{\eta}$ is a Lorentz transformation. Minimal coupling and naturality again produce a constraint analogous to (18) which requires the connection to be torsionless.

How are we to interpret this? What we must remember is that, in gauge theories of spacetime symmetries another, additional assumption is made after the gauge potentials and gauge fields are derived. Namely, an identification is made between the gauge potentials and geometrical objects in a manifold different from the base space of the frame bundle. Thus, it is conventional to identify the Lorentz gauge potential with the Ricci rotation coefficients in a Riemann–Cartan space.[4,8] Note, however, that this new space, which may have torsion, is not the same one as the base space of the frame bundle where we first constructed the gauge theory. A second assumption must now be made when we choose a Lagrangian for the free gauge field, since the standard choice, quadratic in the curvature,[9] yields equations of higher order than those of the Einstein–Cartan (E–C) theory. Thus we have reinforced the argument of Trautman[10] who claims that general relativity and the E–C theory are not local gauge theories.

## 3. LAGRANGIANS AND FIELD EQUATIONS

In this section we discuss inhomogeneous gauge transformations and their effect on Lagrangians and matter fields. To begin with, matter fields will be sections of a vector bundle of fiber dimension $k$, $V\xrightarrow{\pi}M$ in which $G$ is the group of the bundle. So if $s_1,\cdots,s_k$ is a local field of sections trivializing $V$, then any cross section $\psi$ can be written locally as

$$\psi = \sum s_\alpha \psi^\alpha \tag{20}$$

where $\psi^\alpha$ are smooth functions over $M$. Supposing that $\nabla$ is a covariant derivative on $V$,

$$\nabla s_\alpha = \sum s_\beta \theta^\beta_\alpha, \tag{21}$$

where $(\theta^\beta_\alpha)$ is a matrix of 1-forms (same notation as Sec. 1). Working locally with $\theta^\alpha_\beta = \Gamma^\alpha_{\beta i}\, dy^i$

and

$$v = \sum \frac{\partial}{\partial y^i} v^i = \frac{\partial}{\partial y^i} v^i,$$

$$(\nabla_v \psi)^\alpha = \frac{\partial \psi^\alpha}{\partial y^i} v^i + \Gamma^\alpha_{\beta i} \psi^\beta v^i. \tag{22}$$

A diffeomorphism $h$ acts on a field $\psi$ by composition $\psi \rightarrow \bar{\psi} = \psi \circ h$, thus,

$$\bar{\psi} = \sum \bar{s}_\alpha \bar{\psi}^\alpha. \tag{23}$$

The problem now is tht $\psi \circ h = \bar{\psi}$ is a field over $h$, meaning $\pi\bar{\psi} = h$. Such fields can be covariantly differentiated with respect to $\bar{\nabla}$, locally represented by $(h^* \theta^\alpha_\beta)$. Using $x^i = y^i \circ h$ as local coordinates,

$$(\bar{\nabla}_u \bar{\psi})^\alpha = \frac{\partial \bar{\psi}^\alpha}{\partial x^j} \frac{\partial h^j}{\partial x^i} u^i$$

$$+ \Gamma^\alpha_{\beta j} \circ h \bar{\psi}^\beta \frac{\partial h^j}{\partial x^i} u^i. \tag{24}$$

Viewing a Lagrangian $L(\psi, d\psi)$ as a scalar valued 4-form

$$L(\psi, d\psi) \rightarrow L(\bar{\psi}, \bar{\nabla} \bar{\psi}) \det h^*, \tag{25}$$

and by the change of variable theorem if $h$ preserves orientation

$$S_M = \int_M L(\psi, d\psi) = \int_M L(\bar{\psi}, \bar{\nabla} \bar{\psi}) \det h^*. \tag{26}$$

Invariance of the action $S_M$ under diffeomorphisms leads to a conservation law for energy–momentum. Suppose that a one-parameter subgroup of the diffeomorphisms $\phi(t)$ is generated by the vector field $\xi$ i.e., $\xi$ is tangent to $\phi(t)$ at $t = 0$). Then Eq. (26) implies

$$\int [L(\psi, d\psi) - L(\bar{\psi}, \bar{\nabla} \bar{\psi}) \det h^*] = 0, \tag{27}$$

which further implies

$$\textstyle\int_M \mathscr{L}_\xi L = 0. \tag{28}$$

Here, $\mathscr{L}_\xi$ is the Lie derivative along $\xi$. Substitution of the Euler–Lagrange equation, $\delta L / \delta\psi = 0$, into (28) and the fact that $\xi$ is arbitrary together yield a conservation law of the form $\nabla_\nu T^{\mu\nu} = 0$. For details of this calculation in the case of Einstein's theory see Hawking and Ellis (Ref. 11, page 67).

Several remarks are in order about action integrals and field equations. If $G$ is semisimple, then using the Killing form[12] $\|\ \|^2$ and the volume element $\omega$ of $M$,

$$\|\eta \cdot \mathrm{F}\|^2 = \|\mathrm{h}^*(\mathrm{F})\|^2, \tag{29}$$

for all inhomogeneous gauge transformations $\eta = h\bar{\eta}$.

But

$$\int_M \|h^*(F)\|^2 \omega \tag{30}$$

can surely be invariant if $h$ is an isometry.

Now if $V$ has an inner product structure in the fibers whose $\nabla$-covariant derivative is zero, then the elementary symmetric functions of $(F^{\alpha\beta})$ lead to the Pontryagin classes of $V$.[13] By the naturality of Pontryagin classes (see Milnor and Stasheff[14]), the elementary symmetric functions of $h^*(F^{\alpha\beta})$ represent the same classes in the de Rham cohomology). Hence the inhomogeneous gauge transformations preserve the Pontryagin classes and Pontrygain index, without any assumptions (semisimplicity or isometry).

## 4. CONCLUSIONS

General relativity (Einstein–Cartan) is not a gauge theory in the usual sense of Maxwell's theory or Yang–Mills theory. The field equations of general relativity can be obtained from a gaugelike approach, but only after a geometrical reidentification of the gauge potentials and nonstandard choice of freefield Lagrangian. The usual choice according to Yang–Mills is not possible simply because $\mathscr{D}$ is such a large Lie group.

In trying to append invariance under $\mathscr{D}$ to a local gauge theory of internal symmetry, one inevitably meets two types of obstacles. First, invariance of the action integral allows only those diffeomorphisms which are isometries. Secondly, the naturality condition forces the torsion of the covariant derivative to vanish. Thus, ignoring the difficulties encountered with the invariance of the action integral, the Einstein–Cartan theory cannot be considered a natural gauge theory.


[1]H. Weyl, Sitzber. Preuss. Akad. Wiss. Berlin, 465–80 (1918).
[2]C.N. Yang and R.L. Mills, Phys. Rev. **96**, 191 (1954).
[3]R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[4]T.W.B. Kibble, J. Math. Phys. **2**, 212 (1961).
[5]M. Atiyah, N. Hitchen, and I.M. Singer, preprint (U.C. Berkeley, 1978).
[6]J.-M. Souriau, *Géometrie et Relativité* (Hermann, Paris, 1964).
[7]S. Hojman, M. Rosenbaum, M.P. Ryan, and L.C. Shepley, Phys. Rev. D **19**, 430 (1979).
[8]F.W. Hehl, P. von der Heyde, G.D. Kerlick, and J.M. Nester, Rev. Mod. Phys. **48**, 393 (1976).
[9]C.N. Yang, Phys. Rev. Lett. **33**, 445 (1974).
[10]A. Trautman, Rep. Math. Phys. **1**, 29, (1970).
[11]S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space–Time* Cambridge U.P. Cambridge, 1973).
[12]J. Wolf, *Spaces of Constant Curvature* (Publish or Perish, Boston, Massachusetts, 1974).
[13]S. Chern, Topics in differential geometry (IAS, 1951).
[14]J. Milnor and J. Stasheff, *Characteristic Classes* (Princeton U.P., Princeton, New Jersey, 1974).

# Two-body gravidynamics[a)]

J. Ibañez and J. L. Sanz[b)]

*Departmento de Física Teórica, Universidad Autónoma de Madrid, Canto Blanco, Madrid–34 Spain*
(Received 6 February 1978; revised manuscript received 25 July 1978)

We develope a theory of gravitational interaction between two massive punctual structureless particles, from the point of view of predictive relativistic mechanics, whose main characteristics are: (i) the geometric framework is Minkowski flat space–time, $M_4$; (ii) it is a predictive theory as opposed to the heredity which characterizes the dynamics derived from the classical field theory; (iii) the interaction occurs through light cones with a second order covariant symmetrical tensorial function on $(TM_4)^2$ which we call "field"; (iv) we do not impose any "field equation" on this field, but only certain symmetries which lead to a certain indeterminacy which appears in the dynamic; (v) the dynamic is invariant with respect to the Poincaré group. From the outset, we adopt a fast motion approximation which consists of supposing that the dynamics can be developed in power series of the particle masses, and we evaluate this dynamic up to the second order of the masses. The comparison of the dynamics obtained from the Einstein, Infeld, and Hoffmann Lagrangian (which describes the gravitational interaction up to $c^{-2}$ order in general relativity) helps us to fix five of the twelve arbitrary parameters which appear in our dynamics. It appears possible to explain all the gravitational experiments carried out up to now, in which neither light nor internal particle structure appear.

## I. INTRODUCTION

The predictive relativistic mechanic was developed in two completely independent formalisms: the manifestly predictive[1] and the manifestly invariant,[2] although they are equivalent.[3]

In the latter formalism, the evolution of an $N$ pointlike structureless charge system is given by a second order differential equation on the Minkowsky flat space–time $M_4$.[4]

$$\frac{dx_a^\alpha}{d\lambda} = \pi_a^\alpha, \quad \frac{d\pi_a^\alpha}{d\lambda} = \theta_a^\alpha(x_b^\beta,\pi_c^\gamma), \tag{I.1}$$

where the $\theta_a^\alpha$ functions must satisfy:

$$(\pi_a\theta_a) = 0, \tag{I.2}$$

$$\pi_{a'}^\rho \frac{\partial\theta_a^\alpha}{\partial x^{a'\rho}} + \theta_{a'}^\rho \frac{\partial\theta_a^\alpha}{\partial\pi^{a'\rho}} = 0. \tag{I.3}$$

Equation (I.3) is usually known as the "predictive" equations or the Droz-Vincent equations.[2] The invariance of the system under the Poincaré group gives us a set of linear equations for the $\theta_a^\alpha$ functions.[5]

The predictive relativistic mechanic has been used, with success, in a perturbation scheme, with the scalar interaction (short and long range) and vectorial interaction (particularly the electromagnetic interaction).

The electrodynamic theory has the Minkowsky flat space-time $M_4$ as the geometric framework. The electromagnetic field is represented by a second order antisymmetric covariant tensor $F_{\alpha\beta}(x^\lambda)$ in this space, which satisfies a linear partial differential equation system, the Maxwell equations[6]:

$$\partial_\alpha F^{\alpha\beta} = -\frac{4\pi}{c}j^\beta,$$
$$\partial_\alpha F_{\beta\gamma} + \partial_\beta F_{\gamma\alpha} + \partial_\gamma F_{\alpha\beta} = 0, \tag{I.4}$$

where $j^\beta$ is the current density which depends only on the source variables. For an $N$ pointlike structureless charge system, the charge $a$ is affected by a field obtained by means of Lienard–Wiechert retarded fields whose sources are the other particles.

We then postulate that the movement equations are the Lorentz equations

$$\ddot{x}_a^\alpha(\tau) = e_a m_a^{-1}\dot{x}_{\alpha\beta}\sum_{a'}\hat{F}_{a'}^{\alpha\beta}[x_a^\rho(\tau)], \tag{I.5}$$

where $x_a^\alpha(\tau)$ is the $a$ trajectory, $\dot{x}_a^\alpha \equiv dx_a^\alpha/d\tau$, $\dot{x}_a^2 \equiv -(\dot{x}_a\dot{x}_a) = 1$, the $a$ velocity in the $x_a^\alpha$ point in $M_4$ $\hat{F}_{a'\alpha\beta}(x^\rho) \equiv F_{a'\alpha\beta}[x_{a'}^\gamma(\hat{\tau}_{a'}),\dot{x}_{a'}^\lambda(\hat{\tau}_{a'}),\ddot{x}_{a'}^\mu(\hat{\tau}_{a'});x^\rho]$, $\hat{\tau}_{a'}$ being such that $(x^\rho - x_{a'}^\rho(\hat{\tau}_{a'}))(x_\rho - x'_{a\rho}(\hat{\tau}_{a'})) = 0$, $x^0 > x_{a'}^0(\hat{\tau}_{a'})$ and $F_{a'\alpha\beta}(x^\lambda)$ must satisfy (I.4) with current[6]

$$j_{a'}^\alpha = e_{a'}\int_{-\infty}^{+\infty}\dot{x}_{a'}^\alpha(\tau)\delta^4[x^\rho - x_{a'}^\rho(\tau)]d\tau.$$

Equation (I.5) can be put

$$\ddot{x}_a^\alpha(\tau) = e_a m_a^{-1}\sum_{a'} e_{a'}W_{aa'}^\alpha \times [x_a^\beta(\tau),\dot{x}_{a'}^\gamma(\tau),x_{a'}^\rho(\hat{\tau}_{a'}),\dot{x}_{a'}^\sigma(\hat{\tau}_{a'}),\ddot{x}_{a'}^\delta(\hat{\tau}_{a'})], \tag{I.6}$$

where the functionals $W_{aa'}^\alpha$ are known. The (I.6) equation is not an ordinary differential equation, but a second order differential-difference equation system in which the difference is not a constant and about which Chern and Havas[7] have demonstrated that there is not uniqueness. That is to say, different solutions can apear for the same initial conditions.

The fact that the interaction propagates itself with a finite velocity, through light cones, leads us to *hereditary* dynamics, i.e., the positions and velocities of the particles

[a)]Research supported by the Instituto de Estudios Nucleares, Madrid, Spain.
[b)]Present address: Physics Department, Queen Mary College, London.

 0022-2488/79/061285-11$01.00  

given at fixed times are not sufficient in predicting the later evolution of the system.

In the predictive relativistic mechanic (I.6) is not considered a motion equation but a complementary condition contributing to calculate the dynamics of the system, which is the solution of (I.1).

Up to now, the predictive relativistic mechanic has not been applied to gravitational interaction whose geometric frame work is a curved space–time $V_4$. The gravitational field is represented by a symmetrical second order covariant tensor $g_{\alpha\beta}(x^\lambda)$ on this manifold, which must verify the Einstein equations

$$R_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}R = \frac{8\pi G}{c^2} T_{\alpha\beta}, \qquad (I.7)$$

where $R_{\alpha\beta}$ is the Ricci tensor, $R$ the scalar curvature, and $T_{\alpha\beta}$ the stress–energy tensor, which depends on the $g_{\alpha\beta}$ and on the source variables. The no–linearity of the (I.7) equation is the origin of many difficulties. Particularly for $N$ pointlike structureless particle systems the interaction scheme is not as the electromagnetic interaction scheme. We could evaluate the gravitational field created by all the particles of the system in any point of the space–time and then, the field which acts on $a$ would be obtained from the preceding gravitational fields, thus eliminating the divergences.

From the Bianchi identities and (I.7) we obtain the motion equations

$$T^{\alpha\beta}{}_{;\beta} = 0. \qquad (I.8)$$

From a pointlike particle system the stress–energy tensor is[8]:

$$T^{\alpha\beta}(x^\lambda) = \frac{1}{\sqrt{-g}} \sum_a m_a \times \int_{-\infty}^{+\infty} \frac{\dot{x}_a^\alpha \dot{x}_a^\beta}{(-g_{\mu\nu}\dot{x}_a^\mu \dot{x}_a^\nu)^{1/2}} \delta^4 \times [x^\rho - x_a^\rho(\tau_a)] d\tau_a, \qquad (I.9)$$

$$g \equiv \det(g_{\alpha\beta}).$$

By writing (I.8) in terms of the particle variables and taking (I.9) into account, we obtain

$$\ddot{x}_a^\alpha = -\Gamma^\alpha_{\rho\sigma}(x_a^\lambda)\dot{x}_a^\rho \dot{x}_a^\sigma, \qquad \Gamma^\alpha_{\rho\sigma} \equiv \tfrac{1}{2}g^{\alpha\lambda}(\partial_\rho g_{\sigma\lambda} + \partial_\sigma g_{\rho\lambda} - \partial_\lambda g_{\rho\sigma}). \qquad (I.10)$$

Obviously $\Gamma^\alpha_{\rho\sigma}(x_a^\lambda)$ is obtained from $\Gamma^\alpha_{\rho\sigma}(x^\lambda)$, by substituting $x^\lambda$ for $x_a^\lambda$ and eliminating the divergences.

Two different approximation methods have been developed at this time: The slow motion approximation which consists in developing $g_{\alpha\beta}$ in series of $c^{-1}$, and the fast-motion approximation which consists in developing in series of $G$ in the following way

$$g_{\alpha\beta} = \eta_{\alpha\beta} + \sum_{n=1}^{\infty} G^n \overset{(n)}{g}_{\alpha\beta}.$$

The introduction of this development up to the first order in (I.7) and (I.9), leads us to[8]:

$$\Box\gamma_{\alpha\beta} = -\frac{16\pi}{c^2}\sum_a m_a \times \int_{-\infty}^{+\infty} \dot{x}_{a\alpha}\dot{x}_{a\beta}\delta^4[x^\rho - x_a^\rho(\tau_a)]d\tau_a, \qquad (I.11)$$

$$\overset{(1)}{g}_{\alpha\beta} \equiv \gamma_{\alpha\beta} - \tfrac{1}{2}\eta_{\alpha\beta}\gamma^\rho_\rho, \quad \dot{x}_a^2 \equiv -(\dot{x}_a\dot{x}_a) = 1,$$

where we have supposed that $\partial^\beta\gamma_{\alpha\beta} = 0$, which is equivalent to imposing the harmonic coordinate condition to the first order: $\partial_\beta(\sqrt{-g}g^{\alpha\beta}) = 0$. Up to this order we dispose of a Lorentz covariant wave equation and the interaction propagates itself through light cones. If we consider the retarded solution as the physical solution and we take into account the geodesic equation (I.10) up to the first order in $G$, it is clear that (I.10) will lead us to a differential-difference system

$$\ddot{x}_a^\alpha(\tau) = G\sum_{a'} m_{a'} Z^\alpha_{aa'}[x_a^\beta(\tau),\dot{x}_a^\gamma(\tau), x_{a'}^\rho(\hat{\tau}_{a'}),\dot{x}_{a'}^\sigma(\hat{\tau}_{a'});\ddot{x}_{a'}^\delta(\hat{\tau}_{a'})] + O(G^2), \qquad (I.12)$$

where the $Z^\alpha_{aa'}$ functionals are known from (I.11) and (I.10).

This equation is very similar to (I.6). That is to say, up to first order in $G$, there is a complete analogy between gravitational and electromagnetic interactions. Therefore, we try to develop the gravitational interaction in the Predictive relativistic mechanic framework.

In this paper we consider a *first approximation* to the gravitational interaction problem, considering a two particle system. In order to use (I.10) as motion equations, we must change the condition (I.2). Equation (I.9) expresses a homogeneity relation between accelerations and velocities and (I.2) is a orthogonality condition. Therefore, we replace (I.2) by a homogeneity condition,[9]

$$\pi_b^\rho \frac{\partial\theta_a^\alpha}{\partial\pi^{b\rho}} = 2\delta_{ba}\theta_a^\alpha. \qquad (I.13)$$

Equations (I.3) and (I.13) constitute which we shall call "Künzle Predictive system.[10]"

We begin with a very simple hypothesis: We consider the Minkowsky flat space–time as the geometric framework and we do not impose any "field eqution" on the gravitational field but only certain symmetries which do not univocally determine it. We take mass series developments of each of the particles for the fundamental magnitudes which appear (i.e., field and dynamic magnitudes). We obtain the approximate dynamic in masses for the two-body problem in a gravitational interaction which up to the $C^{-2}$ order is the same as those obtained with the Einstein, Infeld, and Hoffmann Lagrangian.[11]

## II. KÜNZLE PREDICTIVE SYSTEM

### A. Definition and properties

In the predictive relativistic mechanics framework, in its covariant formalism, certain isolated and nonisolated $N$ pointlike structureless particle systems can be described by Künzle predictive systems.[12] Such a system is, by definition, a second order autonomous ordinary differential system in $M_4$:

$$\frac{dx_a^\alpha}{d\lambda} = \pi_a^\alpha, \quad \frac{d\pi_a^\alpha}{d\lambda} = \theta_a^\alpha(x_b^\beta\ \pi_c^\gamma), \tag{II.1}$$

which satisfies the following properties:

$$\text{(i) } \pi_a^2 \equiv -(\pi_a\ \pi_a) > 0, \quad \pi_a^0 > 0 \tag{II.2a}$$

$$\text{(ii) } \pi_{a'}^\rho \frac{\partial\theta_a^\alpha}{\partial x^{a'\rho}} + \theta_{a'}^\rho \frac{\partial\theta_a^\alpha}{\partial\pi^{a'\rho}} = 0 \tag{II.2b}$$

$$\text{(ii) } \pi_b^\rho \frac{\partial\theta_a^\alpha}{\partial\pi^{b\rho}} = 2\delta_{ab}\theta_a^\alpha. \tag{II.2c}$$

The (II.2b) equations which we will henceforth call "Droz-Vincent equations," were introduced by Droz-Vincent.[7] The (II.2c) equations which clearly represent a homogeneity condition for the $\theta_a^\alpha$ functions, were introduced, in a more general way, by Künzle.[12] In order to study the geometric significance of the (II.2) conditions, we consider the general integral associated with the (II.1) system:

$$x_a^\alpha = \varphi_a^\alpha({}_0x_b^\beta, {}_0\pi_c^\gamma; \lambda),$$

$$ {}_0x_a^\alpha = \varphi_a^\alpha({}_0x_b^\beta, {}_0\pi_c^\gamma; 0), \quad {}_0\pi_a^\alpha = \dot\varphi_a^\alpha({}_0x_b^\beta, {}_0\pi_c^\gamma; 0). \tag{II.3}$$

It is easy to see that such conditions give the following properties:

$$\text{(i) } \dot\varphi_a^\alpha \dot\varphi_{a\alpha}({}_0x_b, {}_0\pi_c; \lambda) > 0, \quad \dot\varphi_a^0({}_0x_b, {}_0\pi_c; \lambda) > 0, \tag{II.4a}$$

$$\text{(ii) } \varphi_a^\alpha[\varphi_b^\beta({}_0x, {}_0\pi, \lambda_b), \dot\varphi_c^\gamma({}_0x, {}_0\pi; \lambda_c); \lambda] = \varphi_a^\alpha({}_0x, {}_0\pi; \lambda_a + \lambda) \tag{II.4b}$$

$$\text{(iii) } \varphi_a^\alpha({}_0x_b^\beta, \rho_c\, {}_0\pi_c^\gamma; \lambda) = \varphi_a^\alpha({}_0x_b^\beta, {}_0\pi_c^\gamma; \rho_a\lambda). \tag{II.4c}$$

(II.4a) expresses that the trajectories are oriented in time and towards the future. (II.4b) indicates that if $x_a^\alpha$ is a set of $N$ points in the trajectories obtained from certain initial conditions $({}_0x_a^\alpha, {}_0\pi_c^\beta)$ and if $\pi_b^\beta$ are the corresponding tangent vectors, then the trajectories which correspond to the initial conditions $(x_a^\alpha, \pi_b^\beta)$ are the same as those corresponding to $({}_0x_a^\alpha, {}_0\pi_b^\beta)$. This reduces the number of essential parameters, on which the general integral depends, from $8N$ to $7N$.

The (II.4c) condition expresses that the general integral does not depend on the modulus of the $\pi_a^\alpha$ vectors, for we can take $\rho_c \equiv {}_0\pi_c^{-1}$. This again reduces the number of essential parameters, of which the general integral depends, from $7N$ to $6N$.

Let us study the compatibility of the (II.2) conditions. It is clear that the general solution of (II.2c) can be written

$$\theta_a^\alpha(x_b^\beta, \pi_c^\gamma) = \pi_a^2 \xi_a^\alpha(x_b^\beta, \pi_c^{-1}\pi_c^\gamma), \tag{II.5}$$

where $\xi_a^\alpha$ is an arbitrary function of its arguments. By substituting (II.5) in the (II.2b) equation, we obtain

$$u_{a'}^\rho \frac{\partial\xi_a^\alpha}{\partial x^{a'\rho}} + [\xi_{a'}^\rho + (u_{a'}\xi_{a'})u_{a'}^\rho] \frac{\partial\xi_a^\alpha}{\partial u^{a'\rho}} = 0, \tag{II.6}$$

with $u_a^\rho \equiv \pi_a^{-1}\pi_a^\rho$. The differential conditions that the $\theta_a^\alpha$ functions, which characterize a Künzle predictive system, must satisfy have been reduced to the (II.6) equation, which constitutes a nonlinear equation system in first order partial derivatives in the $\xi_a^\alpha$ functions. Some nontrivial exact particular solutions are known but they do not have a physical meaning and they are only interesting because they demonstrate that the concept of the Künzle predictive system is not trivial.

## B. Approximate solutions

Equation (II.6) can obviously be rewritten in the form

$$D_{a'}\xi_a^\alpha = -\zeta_{a'}^\rho \frac{\partial\xi_a^\alpha}{\partial u^{a'\rho}}, \tag{II.7}$$

where

$$D_{a'} \equiv u_{a'}^\rho \frac{\partial}{\partial x^{a'\rho}}, \quad \zeta_{a'}^\rho \equiv \xi_{a'}^\rho + (u_{a'}\xi_{a'})u_{a'}^\rho. \tag{II.8}$$

We are going to try and find approximate solutions of (II.7) for the two particle case, $N = 2$. In order to do this, we shall assume that the function $\xi_a^\alpha$ can be developed in series of two parameters, characteristic of the particles, $m_a$ (which we will call the *masses* of the particles).

$$\xi_a^\alpha = \sum_{r,s=0}^{\infty} m_a^r m_{a'}^s\ {}^{(r,s)}\xi_a{}^\alpha, \tag{II.9}$$

where the ${}^{(r,s)}\xi_a{}^\alpha$ function verifies

$$ {}^{(r,0)}\xi_a{}^\alpha \equiv 0, \tag{II.10}$$

in such a way that (II.9) can be written up to $r + s = 2$ order in the following way

$$\xi_a^\alpha = m_{a'}\ {}^{(1)}\xi_a{}^\alpha + m_{a'}^2\ {}^{(2)}\xi_a{}^\alpha + m_a m_{a'}\ {}^{(2)}\xi_{aa'}{}^\alpha + \cdots. \tag{II.11}$$

Hypothesis (II.10) is based on the following reasons: First of all, we shall deal with *massive* particles (i.e., $m_b \neq 0$) and secondly, we shall identify $m_a$ as a "strength" parameter in the sense that the bigger (or smaller) the $m_a$ parameter is in comparison with $m_{a'}$, the bigger (or smaller) the influence of the particle $a$ on the particle $a'$. If, for example, $m_{a'}$ can be neglected in comparison with $m_a$ the influence of $a'$ on $a$ can be neglected, that is to say, the particle $a$ is approximately free (i.e., $\theta_a^\alpha = 0$). This justifies condition (II.10). On the other hand, for the lower order we obtain a term $m_{a'}\ {}^{(1)}\xi_a{}^\alpha$ for $a$, which is independent of the $a$ mass. This is, up to this order, an "equivalence principle."

By introducing (II.1) in (II.7) we obtain up to second order:

$$D_{a'}\ {}^{(1)}\xi_a{}^\alpha = 0, \tag{II.12a}$$

$$D_{a'}\ {}^{(2)}\xi_a{}^\alpha = 0, \tag{II.12b}$$

$$D_{a'}\ {}^{(2)}\xi_{aa'}^\alpha = -\ {}^{(1)}\zeta_{a'}^\rho \frac{\partial\ {}^{(1)}\xi_a^\alpha}{\partial u^{a'\rho}}. \tag{II.12c}$$

For each order, the equations which appear are of the form

$$D_{a'}F_a = -A_{aa'}(x_b^\beta, u_c^\gamma), \tag{II.13}$$

whose general solution we will write in the form

$F_a = F^*_a + F^P_a$, where $F^*_a$ is the general solution of the homogeneous equation ($D_{a'}F^*_a = 0$) and $F^P_a$ is a particular solution of the complete equation ($D_{a'}F^P_a = -A_{aa'}$). We can then see that the general solution of (II.13) can be written in two different ways ($\epsilon = \pm 1$):

$$F_a = F^*_a(x^\rho_a, x^\sigma_{a'} - \tau_a u^\sigma_{a'}, u^\lambda_b) - \int_0^{\tau_a} dy A_{aa'}(x^\alpha_a, x^\beta_{a'} - yu^\beta_{a'}, u^\gamma_b), \tag{II.14}$$

where $F^*_a$ is an arbitrary function of its arguments,

$$x^\alpha \equiv x^\alpha_1 - x^\alpha_2, \quad r_a \equiv +[x^2 + (xu'_a)^2]^{1/2}$$
$$\tau_a \equiv \eta_a(xu_a) - \epsilon r_a, \quad \eta_a : \eta_1 = +1, \quad \eta_2 = -1, \tag{II.15}$$

and where the quantity under the integral sign, as the notation indicates, can be obtained from the function $A_{aa'}(x^\beta_b, u^\gamma_c)$, by making the change $x^\alpha_{a'} \to x^\alpha_{a'} - yu^\alpha_{a'}$.[13]

For further applications we will write in an explicit manner the development of the $\xi^\alpha_a$ functions up to the second order

$$\xi^\alpha_a = m_{a'}\,{}^{(1)}\xi^{*\alpha}_a + m^2_{a'}\,{}^{(2)}\xi^{*\alpha}_a + m_a m_{a'} \times \left\{ {}^{(2)}\xi^{*\alpha}_{aa'} - \int_0^{\tau_a} dy \left( {}^{(1)}\xi^\rho_{a'} \frac{\partial\,{}^{(1)}\xi^\alpha_a}{\partial u^{a'\rho}} \right) (x^\beta_a, x^\gamma_{a'} - yu^\gamma_{a'}, u^\lambda_b) \right\} + \cdots, \tag{II.16}$$

where ${}^{(2)}\xi^{*\alpha}_a$, ${}^{(2)}\xi^{*\alpha}_{aa'}$ are arbitrary functions of the arguments $x^\beta_a$, $x^\gamma_{a'} - \tau_a u^\gamma_{a'}, u^\lambda_b)$.

## III. TWO PARTICLE GRAVIDYNAMICS

In this section we study the dynamics of two particles which interact gravitationally. First, we develop the most general expression of the gravitational field between two particles, by imposing certain symmetries. We then postulate the motion equations of the system and introduce a perturbative scheme in the masses. All this is used as a complementary principle of the formalism which gives us the evolution of the system which we will suppose is given by a Künzle predictive system. This concept has been developed in the preceding section.

### A. Gravitational field created by particles

Let us consider two particles in space–time, each characterized by a parameter which we can identify with the mass. We will suppose that the interaction scheme, as is usual in the classical field theory, has the following form

particle $a' \to$ field due to $a' \to$ particle $a$,

where the field created by $a'$ will be described in this case by a symmetrical second order tensorial covariant function on $(TM_4)^2$,[14] which we will indicate by $g_{a'\alpha\beta}(x^\rho_a, x^\sigma_{a'}, u^\lambda_a, u^\mu_{a'})$, $u^2_b \equiv -(u_b u_b) = 1$. Now we will impose the following conditions on the field:

(i) *Principle of constants economy*: The only parameters which appear in the $g_{a'\alpha\beta}$ fields are the $m_b$ masses of the two particles, and the only universal constants which can appear are $G$ and $c$, i.e., the gravitational universal constant and the velocity of light in vacuum. To show this dependence we will write

$$g_{a'\alpha\beta}(x^\rho_a, x^\sigma_{a'}, u^\lambda_a, u^\mu_{a'}; m_a, m_{a'}; G, c). \tag{III.1}$$

Generally, the field created by $a'$ can depend on the characteristics of $a$ (i.e., its $m_a$ mass) and its $u^\rho_a$ velocity. Guided by the electrodynamics analogy we shall not consider the dependence on higher order derivatives possible.

(ii) *Symmetries*: let us consider the instantaneous reference system in which $a'$ is at rest in $x^\alpha_{a'}$ (i.e., $u^\alpha_{a'} \equiv \delta^\alpha_0$). We shall assume that the field created by $a'$ is

(a) invariant under the realization derived from the Aristoteles group[15] on $(TM_4)^2$. This realization is defined by:

$$x^{0'}_b = x^0_b - A^0, \quad u^{0'}_b = u^0_b,$$
$$x^{i'}_b = R^i_j(x^j_b - A^j), \quad u^{i'}_b = R^i_j u^j_b, \tag{III.2}$$

where $A^\alpha$ are four real numbers and $R^i_j$ is a matrix of the rotation group. When we define $g_{a'\alpha\beta}$ as an invariant field we mean that it behaves as a tensorial function invariant under the (III.2) realization, i.e.,

$$g_{a'\alpha\beta}[L^\rho_\nu(x^\nu_b - A^\nu), L^\sigma_\tau u^\tau_b] = L_\alpha^{-1\lambda} L_\beta^{-1\mu} g_{a'\lambda\mu}(x^\rho_b, u^\sigma_c),$$
$$u^\alpha_{a'} = \delta^\alpha_0, \tag{III.3}$$

where $L^\alpha_\beta$ is a matrix of the Lorentz group having the structure:

$$L^\alpha_\beta : L^0_0 \equiv 1, \quad L^0_i = L^i_0 \equiv 0, \quad L^i_j \equiv R^i_j, \quad L^\alpha_\rho L_\beta^{-1\rho} = \delta^\alpha_\beta.$$

This property shows the space–time homogeneity and the isotropy of the field created by each particle in the instantaneous reference system in which it is at rest.

(b) static, that is to say, it is independent of $x^0_{a'}$. Invariance under space–time translations of $g_{a'\alpha\beta}$ automatically implies that it depends on the difference of the $x^\rho \equiv x^\rho_1 - x^\rho_2$ positions and the staticity clearly implies that it does not depend on $x^0$, so we will simply write

$$g_{a'\alpha\beta}(x^i, u^0_a, u^j_a), \quad u^\alpha_{a'} \equiv \delta^\alpha_0, \tag{III.4}$$

and the condition (III.3) can then be written

$$g_{a'\alpha\beta}(R^i_k x^k, u^0_a, R^j_l u^l_a) = L_\alpha^{-1\lambda} L_\beta^{-1\mu} g_{a'\lambda\mu}(x^i, u^0_a, u^j_a), \tag{III.5}$$
$$u^\alpha_{a'} \equiv \delta^\alpha_0,$$

which shows that $g_{a'\alpha\beta}$ is a tensorial function of the vectorial variables, $x^i$, $u^i_a$, with respect to the rotation group.

(iii) *Regularity*: we will suppose that $g_{a'\alpha\beta}$ can be developed in a power series of the $m_b$ masses in the following way:

$$g_{a'\alpha\beta} = \sum_{r,s=0}^{\infty} m^r_{a'} m^s_a\, {}^{(r,s)}g_{a'\alpha\beta} \tag{III.6}$$

where the ${}^{(r,s)}g_{a'\alpha\beta}$ are independent of $m_b$ and verify:

$$ {}^{(0,0)}g_{a'\alpha\beta} \equiv \eta_{\alpha\beta}, \quad {}^{(0,s)}g_{a'\alpha\beta} \equiv 0, \quad s>0, $$
$$ {}^{(r,0)}g_{a'\alpha\beta} \equiv {}^{(r)}g_{a'\alpha\beta}(x^\rho, u^\sigma_{a'}). \tag{III.7} $$

This hypothesis is motivated by the following considerations. We believe that $g_{a'\alpha\beta}$ are sufficiently regular functions of the masses [that is why we use (III.6)] and that when $m_{a'} \ll m_a$ the field $g_{a'\alpha\beta} \simeq \eta_{\alpha\beta}$. We also believe that if $m_a \ll m_{a'}$, $\theta^\alpha_{a'} \simeq 0$ and it is then logical to assume that $g_{a'\alpha\beta}$ does not depend on $u^\rho_{a'}$.

We can write (III.6) by introducing (III.7) up to the $r+s=2$ order, in the following way:

$$ g_{a'\alpha\beta} = \eta_{\alpha\beta} + m_{a'}\ {}^{(1)}g_{a'\alpha\beta} + m^2_{a'}\ {}^{(2)}g_{a'\alpha\beta} $$
$$ + m_{a'}m_a\ {}^{(2)}g_{a'a\alpha\beta} + \cdots, \tag{III.8} $$

where ${}^{(1)}g_{a'\alpha\beta}(x^\rho, u^\sigma_{a'})$ and ${}^{(2)}g_{a'a\alpha\beta}(x^\rho, u^\sigma_c)$.

We can conclude as a corollary of this hypothesis that the field is dimensionless, for $\eta_{\alpha\beta}$ is by hypothesis dimensionless.

$$ [g_{a'\alpha\beta}] = [\eta_{\alpha\beta}] = 1. \tag{III.9} $$

(iv) $S_2$ *invariance*: As we believe we are dealing with two particles of the same type, i.e, each characterized only by its mass, we shall suppose that the field is invariant under the two element permutation group $S_2$, i.e.,

$$ g_{a'\alpha\beta}(x^\rho_{a'}, x^\sigma_a, u^\lambda_{a'}, u^\mu_a; m_{a'}, m_a) $$
$$ = g_{a\alpha\beta}(x^\rho_{a'}, x^\sigma_a, u^\lambda_{a'}, u^\mu_a; m_{a'}, m_a). \tag{III.10} $$

Let us now see what restriction the four preceding conditions impose on $g_{a'\alpha\beta}$. Let us first consider the term in $m_{a'}\cdot m_a$ in Eq. (III.8) which is the only one which can depend on $u^\rho_a$ up to that order. If we consider the instantaneous reference system in which $a'$ is at rest (i.e., $u^\alpha_{a'} \equiv \delta^\alpha_0$) and we take into account (III.4), we can write this term as follows

$$ m_{a'}m_a\ {}^{(2)}g_{a'a\alpha\beta}(x^i, u^0_a, u^j_a; G, c). \tag{III.11} $$

Taking (III.9) into account, which demonstrates that the field is dimensionless, we can conclude that this term must necessarily be of the form

$$ \frac{G^2 m_a m_{a'}}{c^4 r^2} h_{a'\alpha\beta}\left(\frac{x^i}{r}, u^0_a, u^i_a\right) r \equiv +(x^i x_i)^{1/2}, \tag{III.12} $$

where $h_{a'\alpha\beta}$ are arbitrary functions of the specified dimensionless arguments.

When we identify, in the usual manner, $u^0_a \equiv (1 - v^2_a/c^2)^{-1/2}$, $u^i_a \equiv u^0_a v^i_a/c$, $v^i_a$ being the 3-velocity of the $a$ particle, we can rewrite (III.12) as

$$ \frac{G^2 m_a m_{a'}}{c^4 r^2} h_{a'\alpha\beta}\left[\frac{x^i}{r}, \left(1-\frac{v^2_a}{c^2}\right)^{-1/2}, \left(1-\frac{v^2_a}{c^2}\right)^{-1/2}\frac{v^i_a}{c}\right] \tag{III.13} $$

and supposing the differentiality of $h_{a'\alpha\beta}$ in a neighborhood of $c^{-1}=0$,

$$ \frac{G^2 m_a m_{a'}}{c^4 r^2} \bar{h}_{a'\alpha\beta}\left(\frac{x_i}{r}\right) + O(c^{-5}), $$

we can then conclude that the term in $m_{a'}m_a$ is of the form

$$ m_{a'}m_a\ {}^{(2)}h_{a'\alpha\beta}(x^i; G, c) + O(c^{-5}), \tag{III.14} $$

meaning that it does not depend on $u^\rho_a$ up to the $c^{-4}$ order. Summing up, in the instantaneous reference system in which $a'$ is at rest, the expression of the field which this particle creates up to the second order is

$$ g_{a'\alpha\beta} = m_{a'}\ {}^{(1)}g_{a'\alpha\beta}(x^i; G, c) + m^2_{a'}\ {}^{(2)}g_{a'\alpha\beta}(x^i; G, c) $$
$$ + m_a m_{a'}[{}^{(2)}h_{a'\alpha\beta}(x^i; G, c) + O(c^{-5})] + \cdots. \tag{III.15} $$

Condition (III.5) breaks up, trivially, up to that order in the following way:

$$ g_{a'00}(R^i_k x^k) = g_{a'00}(x^i), $$
$$ g_{a'0i}(R^i_k x^k) = R_i^{-1l} g_{a'0l}(x^m), \quad R^i_k R_j^{-1k} = \delta^i_j, \tag{III.16} $$
$$ g_{a'ij}(R^k_l x^l) = R_i^{-1k} R_j^{-1l} g_{a'kl}(x^m) $$

where $g_{a'\alpha\beta}$ must not include powers higher than $c^{-4}$ [16] in $m_{a'}m_{a'}$. It can be easily seen that the general form of $g_{a'\alpha\beta}$, verifying (III.16), is:

$$ g_{a'00} = -a_{a'}(r), $$
$$ g_{a'0i} = b_{a'}(r)\frac{x_i}{r}\eta_a, \quad \eta_a: \eta_1 = +1, \eta_2 = -1, \tag{III.17} $$
$$ g_{a'ij} = c_{a'}(r)\delta_{ij} + d_{a'}(r)\frac{x_i x_j}{r^2}, $$

where $a_{a'}$, $b_{a'}$, $c_{a'}$, and $d_{a'}$ which we will generically denote by $f_{a'}(r, m_{a'}, m_a, G, c)$ are arbitrary functions of the specified arguments. The dimensionless character of $f_{a'}$, for $g_{a'\alpha\beta}$ is also dimensionless, and the (III.15) developments, together with the $S_2$ invariance shown in (III.10), lead to the following structure on $f_{a'}$:

$$ f_{a'} = {}^{(0)}f + \frac{Gm_{a'}}{c^2 r}\ {}^{(1)}f + \frac{G^2 m^2_{a'}}{c^4 r^2}\ {}^{(2)}f $$
$$ + \frac{G^2 m_a m_{a'}}{c^4 r^2}\ {}^{(2)}\underline{f} + \cdots, \tag{III.18a} $$

where

$$ {}^{(1)}f: \ {}^{(0)}a \equiv 1, \quad {}^{(0)}b \equiv 0, \quad {}^{(0)}c \equiv 1, \quad {}^{(0)}d \equiv 0, \tag{III.18b} $$

and ${}^{(2)}f$, ${}^{(2)}f$, ${}^{(2)}\underline{f}, \cdots$ are arbitrary constants (real numbers), in such a way that up to that order they appear as a total of twelve arbitrary constants. This arbitrariness is the price we pay for not imposing any "field equation" on the field.

Once the general form of $g_{a'\alpha\beta}$ has been found, given by (III.17) and (III.18) in the instantaneous reference system in which $a'$ is at rest [as we have supposed that $g_{a'\alpha\beta}$ is a tensorial covariant second order function on $(TM_4)^2$], obviously when the field created by $a'$ moves with velocity, $u^\rho_{a'} \neq \delta^\rho_0$ will be obtained taking into account the transformation law:

$$ g_{a'\alpha\beta}(x^\rho_b, u^\lambda_{a'}) = L_\alpha^{-1\rho} L_\beta^{-1\sigma} g_{a'\rho\sigma}[x^i \rightarrow L_\rho^{-1i} x^\rho], $$
$$ L_\beta^{-1\alpha}: L_\alpha^{-10} = -u_\alpha L_0^{-1i} = -u^i L_j^{-1i} \tag{III.19} $$
$$ = \delta^i_j + \frac{u^i u_j}{u^0 + 1}, $$

which leads to:

$$g_{a'\alpha\beta}(x^\rho,u^\sigma_{a'};m_b,G,c) = c_{a'}(r_{a'})\eta_{\alpha\beta} + \frac{d_{a'}(r_{a'})}{r^2_{a'}} x_\alpha x_\beta + \frac{1}{r_{a'}}\left(\eta_{a'}b_{a'}(r_{a'}) + \frac{(xu_{a'})}{r_{a'}} d_{a'}(r_{a'})\right)(x_\alpha u_{a'\beta} + x_\beta u_{a'\alpha}) + \left( - a_{a'}(r_{a'}) + 2\eta_{a'}\frac{(xu_{a'})}{r_{a'}} b_{a'}(r_{a'}) + c_{a'}(r_{a'}) + \frac{(xu_{a'})^2}{r^2_{a'}} d_{a'}(r_{a'})\right)u_{a'\alpha}u_{a'\beta}, \tag{III.20}$$

where $f_{a'}(r_{a'})$ is obtained from (III.18) substituting the $r$ argument by $r_{a'} \equiv + [x^2 + (xu_{a'})^2]^{1/2}$.

## B. Motion equations

Let us consider the fields ($\epsilon = \pm 1$):

$$\hat{g}_{a'\alpha\beta}(\hat{x}^\rho_{aa'},\hat{u}^\sigma_{a'};m_b;G,c) = \hat{c}_{a'}\eta_{\alpha\beta} + (\hat{x}_{aa'}\hat{u}_{a'})^{-2}\hat{d}_{a'}\hat{x}_{aa'\alpha}\hat{x}_{aa'\beta} + (\hat{x}_{aa'}\hat{u}_{a'})^{-1} \times(-\epsilon\hat{b}_{a'} + \hat{d}_{a'})(\hat{x}_{aa'\alpha}\hat{u}_{a'\beta} + \hat{x}_{aa'\beta}\hat{u}_{a'\alpha}) + (-\hat{a}_{a'} - 2\epsilon\hat{b}_{a'} + \hat{c}_{a'} + \hat{d}_{a'})\hat{u}_{a'\alpha}\hat{u}_{a'\beta} \tag{III.21}$$

obtained from (III.20) considering light cone conditions, i.e.,

$$\hat{x}^\alpha_{aa'} \equiv x^\alpha_a - x^\alpha_{a'}, \quad \hat{x}^\alpha_{aa'}\hat{x}_{aa'\alpha} = 0, \quad \epsilon x^o_{aa'} < 0,$$

and thus obtaining $f_{a'}[\epsilon(\hat{x}_{aa'}\hat{u}_{a'})]$ from (III.18) with the substitution $r \to \epsilon(\hat{x}_{aa'},u_{a'})$. This field hereinafter will be called the advanced field ($\epsilon = +1$) or retarded field ($\epsilon = -1$) created by $a'$. It is interesting to note its analogy with the Linenard–Wiechert potentials in electrodynamics.

Let us now consider the problem of the gravitational interaction of two pointlike structureless particles characterized by their masses, $m_b$. We postulate that the motion $x^\alpha_a = \varphi^\alpha_a(\lambda)$ of each of the particles must be a solution of the "geodesic equation" corresponding to the retarded field[17] $\hat{g}_{a'\alpha\beta}$ whose source is the $m_{a'}$ mass (the causality must be understood in this sense):

$$\ddot{\varphi}^\alpha_a(\lambda) = -\hat{\Gamma}^\alpha_{a'\rho\sigma}\dot{\varphi}^\rho_a\dot{\varphi}^\sigma_a, \hat{\Gamma}^\alpha_{a'\rho\sigma} \equiv \tfrac{1}{2}\hat{g}^{\alpha\lambda}_{a'}\left(\frac{\partial\hat{g}_{a'\rho\lambda}}{\partial x^{a\sigma}} + \frac{\partial\hat{g}_{a'\sigma\lambda}}{\partial x^{a\rho}} - \frac{\partial\hat{g}_{a'\rho\sigma}}{\partial x^{a\lambda}}\right), \hat{g}^{\alpha\lambda}_{a'}\hat{g}_{a'\lambda\beta} = \delta^\alpha_\beta, \tag{III.22}$$

with $g_{a'\alpha\beta}$ given by (III.21).

Note that the following is accepted: (a) the no-existence of self-interaction, (b) that the interaction propagates through light cones. All this has a formal analogy with electrodynamics, except for the nonlinearity of the theory. This would mean that for $N > 2$ we would have to choose between a forces superposition principle and a field superposition principle.

(III.22) can be written

$$\ddot{\varphi}^\alpha_a(\lambda) = \dot{\varphi}^2_a(\lambda)W^\alpha_a[\varphi^\beta_a(\lambda),\varphi^\gamma_{a'}(\hat{\lambda}),\dot{\varphi}_a^{-1}(\lambda)\dot{\varphi}^\rho_a(\lambda),\dot{\varphi}_{a'}^{-1}(\hat{\lambda})\dot{\varphi}^\sigma_{a'}(\hat{\lambda});\dot{\varphi}_{a'}^{-2}(\hat{\lambda})\ddot{\varphi}^\delta_{a'}(\hat{\lambda})] \tag{III.23}$$

with

$$\hat{\lambda}: \hat{\varphi}_{aa'\alpha}\hat{\varphi}^\alpha_{aa'} = 0, \quad \hat{\varphi}^0_{aa'} > 0, \quad \hat{\varphi}^\alpha_{aa'} \equiv \varphi^\alpha_a(\lambda) - \varphi^\alpha_{a'}(\hat{\lambda}),$$

and where the functional $W^\alpha_a$ is too lengthy to write here. We shall nevertheless say that (III.23) is a differential-difference system of the same style as those obtained in the problem of two charges interacting electromagnetically using the retarded Lienard–Wiechert potential. There are no existence and uniqueness theorems for these systems.

## C. Predictivity and causality

We see now that the (III.23) equations can be considered, in the framework of predictive relativistic mechanics, as complementary conditions which will contribute to determining univocally the dynamic of the system.[18] The causality, understood in the sense of the preceding paragraph, is then a complementary principle to the predictivity principle and not something which contradicts it.

Supposing that the evolution of the system is given by a Künzle predictivity system, characterized by functions $\xi^\alpha_a(x^\beta_b,u^\gamma_c)$ [see (II.5)], it is easy to see that such functions must satisfy the functional relations which follow[19]:

$$\xi^\alpha_a(x^\beta_a,\hat{x}^\gamma_{a'},u^\rho_a,\hat{u}^\sigma_{a'}) = W^\alpha_a[x^\beta_a,\hat{x}^\gamma_{a'},u^\rho_a,\hat{u}^\sigma_{a'};\xi^\delta_{a'}(\hat{x}^\lambda_{a'},x^\mu_a,\hat{u}^\nu_{a'},u^\tau_a)], \tag{III.24}$$

$\hat{x}^\alpha_{a'}$ being an arbitrary point of the future cone ($\epsilon = +1$) or past cone ($\epsilon = -1$), whose vertex is at $x^\alpha_a$ and $\hat{u}^\alpha_{a'}$ a unitary vector in the $\hat{x}^\alpha_{a'}$ point.

We now intend to use the (III.24) condition to calculate the developments of the $\xi^\alpha_a$ functions considered in Sec. II A.

Taking into account the structure of the $W^{\alpha}_{a}$ functions in (III.23) we can write the (III.24) conditions up to the second order in the following way:

$$
\begin{aligned}
&{}^{(1)}\xi^{\alpha}_{a}(x^{\beta}_{a},\hat{x}^{\gamma}_{a'},u^{\rho}_{a},\hat{u}^{\sigma}_{a'}) \\
&\quad = -\epsilon G l_{a'}^{-3}\Big\langle\{-2l_{a'}(l-\hat{k})\beta\}u^{\alpha}_{a} + \{-\hat{k}^2\alpha + \hat{\Lambda}^2\beta + [\hat{\Lambda}^2 - \tfrac{3}{2}(l-\hat{k})^2]\,{}^{(1)}d\}\hat{x}^{\alpha}_{aa'} \\
&\qquad + l_{a'}\{-\hat{k}(2l-\hat{k})\alpha + (2\hat{k}l - 1 - \hat{k}^2)\beta + [\hat{\Lambda}^2 - 2(l-\hat{k})^2](-\epsilon\,{}^{(1)}b + {}^{(1)}d) + \tfrac{1}{2}(l-\hat{k})^2\,{}^{(1)}d\}\hat{u}^{\alpha}_{a'}\Big\rangle,
\end{aligned} \tag{III.25a}
$$

$$
\begin{aligned}
&c^2\,{}^{(2)}\xi^{\alpha}_{a}(x^{\beta}_{a},\hat{x}^{\gamma}_{a'},u^{\rho}_{a},\hat{u}^{\sigma}_{a'}) \\
&\quad = \{2G^2 l_{a'}^{-3}(l-\hat{k})({}^{(2)}c - 2\beta^2)\}u^{\alpha}_{a} - G^2 l_{a'}^{-4}\{-\hat{k}^2\,{}^{(2)}a + \hat{\Lambda}^2\,{}^{(2)}c + [\hat{\Lambda}^2 - 2(l-\hat{k})^2\,{}^{(2)}d] + 2\hat{k}^2\alpha\beta \\
&\qquad + 2\epsilon\hat{k}(l-\hat{k})\alpha\,{}^{(1)}b - [-\hat{k}^2\alpha + \tfrac{1}{2}(l-\hat{k})^2\,{}^{(1)}d]\,{}^{(1)}d + [-3\hat{\Lambda}^2 + 5(\hat{k}-l)^2]\beta\,{}^{(1)}d \\
&\qquad + [-\hat{\Lambda}^2 + 2(l-\hat{k})^2]({}^{(1)}d^2 - {}^{(1)}b^2) - 2\hat{\Lambda}^2\beta^2\}\hat{x}^{\alpha}_{aa'} - G^2 l_{a'}^{-3}\{-\hat{k}(2l-\hat{k})\,{}^{(2)}a + (2\hat{k}l - 1 - \hat{k}^2)\,{}^{(2)}c \\
&\qquad + [\hat{\Lambda}^2 - 3(l-\hat{k})^2](-\epsilon\,{}^{(2)}b + {}^{(2)}d) + (l-\hat{k})^2\,{}^{(2)}d + \hat{k}(2l-3\hat{k})\epsilon\,{}^{(1)}b\alpha + [-\hat{\Lambda}^2 - 2(\hat{k}-l)^2] \\
&\qquad (-\epsilon\,{}^{(1)}b\beta - {}^{(1)}b^2 - 3\,{}^{(1)}d\beta - \epsilon\,{}^{(1)}b\,{}^{(1)}d + {}^{(1)}d^2 - 2\epsilon\,{}^{(1)}b\alpha) + 4\hat{k}(l-\hat{k})\alpha^2 + \hat{k}^2(2\beta + {}^{(1)}d)\alpha \\
&\qquad + \tfrac{1}{2}(l-\hat{k})^2[\epsilon\,{}^{(1)}b\,{}^{(1)}d - {}^{(1)}d^2 - 2\,{}^{(1)}d\beta] - 2(2\hat{k}l - 1 - \hat{k}^2)\beta^2\}\hat{u}^{\alpha}_{a'},
\end{aligned} \tag{III.25b}
$$

$$
\begin{aligned}
&c^2\,{}^{(2)}\xi^{\alpha}_{aa'}(x^{\beta}_{a},\hat{x}^{\gamma}_{a'},u^{\rho}_{a},\hat{u}^{\sigma}_{a'}) \\
&\quad = \{2\epsilon l l_{a'}^{-2} M_{a'} G\beta + 2l_{a'}^{-3}(l-\hat{k})G^2\,{}^{(2)}\underline{c}\}u^{\alpha}_{a} - \langle G^2 l_{a'}^{-4}\{-\hat{k}^2\,{}^{(2)}\underline{a} + \hat{\Lambda}^2\,{}^{(2)}\underline{c} + [\hat{\Lambda}^2 - 2(l-\hat{k})^2]\,{}^{(2)}\underline{d}\} \\
&\qquad + \epsilon G l_{a'}^{-3}\{M_{a'}[-\hat{k}^2\alpha + \hat{\Lambda}^2\beta - \epsilon\hat{k}^2\,{}^{(1)}b + \tfrac{1}{2}(\hat{k}^2 - 3l^2)\,{}^{(1)}d] - 2\hat{k}N_{a'}l_{a'}(-\alpha - \epsilon\,{}^{(1)}b + \beta + \tfrac{1}{2}\,{}^{(1)}d)\}\rangle\hat{x}^{\alpha}_{aa'} \\
&\qquad - \langle\epsilon G l_{a'}^{-2} l\{M_{a'}[2\hat{k}(-\alpha + \beta) + \epsilon(l - 2\hat{k})\,{}^{(1)}b - (l-\hat{k})\,{}^{(1)}d] - l(\hat{x}_{aa'}\hat{\xi}^{(1)}_{a'})(-\epsilon\,{}^{(1)}b + {}^{(1)}d) \\
&\qquad - l_{a'}[N_{a'} + \hat{k}(\hat{u}_{a'}\hat{\xi}^{(1)}_{a'})](-2\alpha - 2\epsilon\,{}^{(1)}b + 2\beta + {}^{(1)}d)\} + G^2 l_{a'}^{-3}\{-\hat{k}(2l-\hat{k})\,{}^{(2)}\underline{a} + (2\hat{k}l - 1 - \hat{k}^2)\,{}^{(2)}\underline{c} \\
&\qquad + [\hat{\Lambda}^2 - 3(l-\hat{k})^2](-\epsilon\,{}^{(2)}\underline{b} + {}^{(2)}\underline{d}) + (l-\hat{k})^2\,{}^{(2)}d\}\rangle\hat{u}^{\alpha}_{a'} + 2\epsilon l l_{a'}^{-1} G\hat{\xi}^{(1)\alpha}_{a'} \\
&\qquad [-\tfrac{1}{2}l(-\epsilon\,{}^{(1)}b + {}^{(1)}d) + \hat{k}(-\alpha - \epsilon\,{}^{(1)}b + \beta + {}^{(1)}d)],
\end{aligned} \tag{III.25c}
$$

where

$$
l_{a'} \equiv (\hat{x}_{aa'}\hat{u}_{a'}), \quad l \equiv (\hat{x}_{aa'}u_{a})\,l_{a'}^{-1}, \quad \hat{k} \equiv -(u_{a}\hat{u}_{a'}), \quad \hat{\Lambda}^2 \equiv \hat{k}^2 - 1,
$$

$$
\alpha \equiv {}^{(1)}\frac{a}{2}, \quad \beta \equiv {}^{(1)}\frac{c}{2}, \quad \hat{\xi}^{(1)\alpha}_{a'} \equiv {}^{(1)}\xi^{\alpha}_{a'}(\hat{x}^{\beta}_{a'},x^{\gamma}_{a},\hat{u}^{\lambda}_{a'},u^{\mu}_{a}),
$$

$$
M_{a'} \equiv (\hat{x}_{aa'}\hat{\xi}^{(1)}_{a'}) + l_{a'}(\hat{u}_{a'}\hat{\xi}^{(1)}_{a'}), \quad N_{a'} \equiv \hat{k}(\hat{u}_{a'}\hat{\xi}^{(1)}_{a'}) - (u_{a}\hat{\xi}^{(1)}_{a'}).
$$

(III.25) is a system of recurrent relations adapted to the perturbation technique which we had previously adopted. According to (II.15), $\tau_a$ is a function of the arguments $(x^{\rho}_{a},x^{\sigma}_{a'},u^{\lambda}_{a'})$ which is null for $x^{\rho}_{a'} = \hat{x}^{\rho}_{a'}$, so taking into account (II.16) the (III.25) conditions impose the following conditions on ${}^{(1)}\xi^{*\alpha}_{a}$, ${}^{(2)}\xi^{*\alpha}_{a}$, and ${}^{(2)}\xi^{*\alpha}_{aa'}$:

$$
{}^{(1)}\xi^{*\alpha}_{a}(x^{\rho}_{a},\hat{x}^{\sigma}_{a'},u^{\lambda}_{a},\hat{u}^{\mu}_{a'}) = {}^{(1)}W^{\alpha}_{a}(x^{\rho}_{a},\hat{x}^{\rho}_{a'},u^{\lambda}_{a},\hat{u}^{\mu}_{a'}), \tag{III.26a}
$$

$$
{}^{(2)}\xi^{*\alpha}_{a}(x^{\rho}_{a},\hat{x}^{\sigma}_{a'},u^{\lambda}_{a},\hat{u}^{\mu}_{a'}) = {}^{(2)}W^{\alpha}_{a}(x^{\rho}_{a},\hat{x}^{\rho}_{a'},u^{\lambda}_{a},\hat{u}^{\mu}_{a'}), \tag{III.26b}
$$

$$
{}^{(2)}\xi^{*\alpha}_{aa'}(x^{\rho}_{a},\hat{x}^{\sigma}_{a'},u^{\lambda}_{a},\hat{u}^{\mu}_{a'}) = {}^{(2)}W^{\alpha}_{aa'}[x^{\rho}_{a},\hat{x}^{\rho}_{a'},u^{\lambda}_{a},\hat{u}^{\mu}_{a'};\xi^{\delta}_{a'}(\hat{x}^{\beta}_{a'},x^{\gamma}_{a},\hat{u}^{\nu}_{a'},u^{\tau}_{a})], \tag{III.26c}
$$

where ${}^{(1)}W^{\alpha}_{a}$, ${}^{(2)}W^{\alpha}_{a}$, and ${}^{(2)}W^{\alpha}_{aa'}$ are the right-hand side of (III.25), respectively.

In this way we see that the problem of building developments is reduced to determining, to each order, ${}^{(\ )}\xi^{*\alpha}$ which depend on the arguments $(x_a^\rho, x_{a'}^\sigma - \tau_a u_{a'}^\sigma, u_b^\lambda)$ and which verify (III.26). It can be seen that said functions are unique and have the structure[19]:

$$^{(r)}\xi_a^{*\alpha}(x_a^\rho,\bar{x}_{a'}^\sigma,u_b^\lambda) = {}^{(r)}W_a^\alpha(x_a^\rho,\bar{x}_{a'}^\sigma,u_a^\lambda,u_{a'}^\mu), \tag{III.27a}$$

$$r = 1,2 \quad \bar{x}_{a'}^\alpha \equiv x_{a'}^\alpha - \tau_a u_{a'}^\alpha, \tag{III.27b}$$

$$^{(2)}\xi_{aa'}^{*\alpha}(x_a^\rho,\bar{x}_{a'}^\sigma,u_b^\lambda) = {}^{(2)}W_{aa'}^\alpha[x_a^\rho,\bar{x}_{a'}^\sigma,u_a^\lambda,u_{a'}^\mu;{}^{(1)}\xi_{a'}^\delta(\bar{x}_{a'}^\beta,x_a^\gamma,u_{a'}^\nu,u_a^\tau)], \tag{III.28}$$

After a rather lengthy calculation, the expressions (II.16), (III.25), and (III.26) lead to:

$$\begin{aligned}{}^{(1)}\xi_a^\alpha = {} & -Gr_a^{-3}\langle\{-2\beta s_a\}u_a^\alpha + \{-\alpha k^2 + \beta\Lambda^2 + {}^{(1)}d(\Lambda^2 - \tfrac{3}{2}t_a^2)\}x_{aa'}^\alpha \\ & + \{2(-\alpha+\beta)k(x_{aa'}u_a) - [(-\alpha+\beta)k^2 + \beta] \\ & \times(x_{aa'}u_{a'}) + \tfrac{1}{2}r_a^{-1}s_a^2\,{}^{(1)}b + (\Lambda^2 - \tfrac{3}{2}t_a^2)[-r_a\,{}^{(1)}b + (x_{aa'}u_{a'})\,{}^{(1)}d\,]\}u_{a'}^\alpha\rangle\end{aligned} \tag{III.29a}$$

$$\begin{aligned}{}^{(2)}\xi_a^\alpha & \\ = {} & -G^2c^{-2}r_a^{-4}\langle\{-2s_a({}^{(2)}c - 2\beta^2)\}u_a^\alpha + \{k^2(-{}^{(2)}a + 2\alpha\beta) + \Lambda^2({}^{(2)}c - 2\beta^2) \\ & + (\Lambda^2 - 2t_a^2)\,{}^{(2)}d - [-2\alpha k t_a - {}^{(1)}b(\Lambda^2 - 2t_a^2)]\,{}^{(1)}b - {}^{(1)}d[-\alpha k^2 + \beta(3\Lambda^2 - 5t_a^2) + {}^{(1)}d(\Lambda^2 - \tfrac{3}{2}t_a^2)]\}x_{aa'}^\alpha \\ & + \{2ks_a(-{}^{(2)}a + 2\alpha^2) + k^2(x_{aa'}u_{a'})(-{}^{(2)}a + 2\alpha\beta) + (x_{aa'}u_{a'})(\Lambda^2 - 2t_a^2)\,{}^{(2)}d \\ & + [2ks_a + \Lambda^2(x_{aa'}u_{a'})]({}^{(2)}c - 2\beta^2) - r_a(\Lambda^2 - 3t_a^2)\,{}^{(2)}b + {}^{(1)}b\,[r_a(\Lambda^2 - 2t_a^2)(2\alpha - \beta) + k\,(2t_a(x_{aa'}u_{a'}) - kr_a)\alpha \\ & - (x_{aa'}u_{a'})(-\Lambda^2 + 2t_a^2)\,{}^{(1)}b\,] + [(-3\Lambda^2 + 5t_a^2)\beta + (-\Lambda^2 + \tfrac{3}{2}t_a^2)\,{}^{(1)}d + k^2\alpha] \\ & \times(x_{aa'}u_{a'})\,{}^{(1)}d + \Gamma_a(\Lambda^2 - t_a^2)\,{}^{(1)}b\,{}^{(1)}d\,\}u_{a'}^\alpha\rangle,\end{aligned} \tag{III.29b}$$

where

$$k \equiv -(u_1u_2), \quad s_a \equiv (x_{aa'}u_a) - k\,(x_{aa'}u_{a'}), \quad \Lambda^2 \equiv k^2 - 1, \quad t_a \equiv r_a^{-1}s_a. \tag{III.30}$$

The term ${}^{(2)}\xi_{aa'}$ has not been exactly calculated because of its difficulty (in fact, we have supposed from the start that with the term in $m_a m_{a'}$ in $g_{a'\alpha\beta}$ we operated up to the $c^{-4}$ order). We shall only calculate this term for its future use, up to the $c^{-2}$ order

$${}^{(2)}\xi_{aa'}^i\left[x_a^0 = x_{a'}^0, x_a^i, u_b^0 \equiv \left(1 - \frac{v_b^2}{c^2}\right)^{-1/2}, u_d^j \equiv u_d^0\,\frac{v_d^j}{c}\right] = G^2c^{-2}\eta_a r^{-4}(2\alpha^2 - 2\alpha\beta - \alpha\,{}^{(1)}d + {}^{(2)}\underline{a})x^i + O(c^{-4}). \tag{III.31}$$

We note that in the expression (III.29a) four arbitrary parameters appear: $\alpha, \beta, {}^{(1)}b, {}^{(1)}d$; in (III.29b) the same four parameters and four more appear: ${}^{(2)}a$, ${}^{(2)}b$, ${}^{(2)}c$ and ${}^{(2)}d$ and finally in (III.31), up to the $c^{-4}$ order only ${}^{(2)}\underline{a}$ appears, besides $\alpha, \beta$, and ${}^{(1)}d$.

It can be verified that the $\theta_a^\alpha$ dynamic built in agreement with (II.5) satisfies:

$$\epsilon_b\frac{\partial\theta_a^\alpha}{\partial x_b^\beta} = 0, \quad \epsilon_a = +1, \tag{III.32a}$$

$$(\delta_\lambda^\alpha\eta_{\mu\beta} - \delta_\mu^\alpha\eta_{\lambda\beta})\left(x_b^\beta\frac{\partial\theta_a^\alpha}{\partial x_b^\alpha} + \Pi_b^\beta\frac{\partial\theta_a^\alpha}{\partial\Pi_b^\alpha}\right) = (\delta_\lambda^\rho\eta_{\mu\gamma} - \delta_\mu^\rho\eta_{\lambda\gamma})\theta_a^\gamma. \tag{III.32b}$$

If we call $\varphi_a^\alpha({}_0x_b^\beta,{}_0\pi_c^\gamma;\lambda)$ the general integral associated to the $\theta_a^\alpha$ dynamic, we can easily conclude that it satisfies

$$\varphi_a^\alpha[L_\rho^\beta({}_0x_b^\rho - A^\rho), L_\sigma^\gamma\,{}_0\pi_c^\sigma;\lambda] = L_\mu^\alpha[\varphi_a^\mu({}_0x_b^\beta,{}_0\pi_c^\gamma;\lambda) - A^\mu], \tag{III.33}$$

where $(L_\beta^\alpha, A^\gamma)$ is an element of the Poincaré group. (III.32) and (III.33) express the invariance of the dynamic system (II.1) under the Poincaré group. That is to say, in the way it has been built, the theory is Poincaré invariant.

## IV. COMPARISON WITH THE DYNAMIC OBTAINED FROM THE EIH LAGRANGIAN

In this section we compare the accelerations which are derived from the dynamics calculated in the preceding section with the accelerations obtained from the EIH Lagrangian[11] (Einstein, Infeld, Hoffmann) for the two-particle system. As we know, the accelerations derived from the said

Lagrangian are valid up to $c^{-2}$ order, so that in order to carry out the comparison we need to make a development in $c^{-1}$ of our approximate dynamic up to $c^{-2}$ order. It can be clearly seen that the terms higher than the second order in masses, contain powers superior to $c^{-2}$ order in such a way that there is no contradiction when we carry out this double development and operate in the indicated way. In this manner, we manage to fix five of the 12 parameters which appear in (III.18).

## A. EIH Lagrangian

The EIH Lagrangian for a two-particle system is given by[11]

$$L = \epsilon^a \frac{m_a v_a^2}{2}\left(1 + \frac{v_a^2}{4c^2}\right) + G\frac{m_1 m_2}{r} + G\frac{m_1 m_2}{2c^2 r} \times\left(3(v_1^2 + v_2^2) - 7(\bar{v}_1\bar{v}_2) - \frac{(\bar{x}\bar{v}_1)(\bar{x}\bar{v}_2)}{r^2}\right) - G^2\frac{m_1 m_2(m_1 + m_2)}{2c^2 r^2}. \tag{IV.1}$$

This Lagrangian is obtained in general relativity for the approximate description of the two-body problem up to the $c^{-2}$ order and it is well known that it provides relativistic invariant trajectories.[20]

It can be easily shown that the accelerations obtained from this Lagrangian are:

$$\mu_a^i = -\frac{Gm_{a'}}{r^3}x_{aa'}^i + \frac{1}{c^2}\left\langle \frac{Gm_{a'}}{r^3}\left[\left(v_a^2 - 2v^2 + \frac{3}{2r^2}(\bar{x}\bar{v}_{a'})^2\right)x_{aa'}^i + [4(\bar{x}\bar{v}) + \eta_a(\bar{x}\bar{v}_{a'})]v_{aa'}^i\right] + \frac{G^2 m_{a'}(5m_a + 4m_{a'})}{r^4}x_{aa'}^i\right\rangle, \tag{IV.2}$$

where

$$v^i \equiv v_1^i - v_2^i, \quad v_{aa'}^i \equiv v_a^i - v_{a'}^i, \quad v^2 \equiv v^i v_i.$$

The above $c^{-2}$ development can be rewritten as the following mass development:

$$\mu_a^i = m_{a'}\,{}^{(1)}\mu_a^i + m_{a'}^2\,{}^{(2)}\mu_a^i + m_a m_{a'}\,{}^{(2)}\mu_{aa'}^i + \cdots, \tag{IV.3}$$

$${}^{(1)}\mu_a^i = \frac{G}{r^3}\langle -x_{aa'}^i + c^{-2}\{[v_a^2 - 2v^2 + (3/2r^2)(\bar{x}\bar{v}_{a'})^2]x_{aa'}^i + [4(\bar{x}\bar{v}) + \eta_a(\bar{x}\bar{v}_{a'})]v_{aa'}^i \tag{IV.4a}$$

$${}^{(2)}\mu_a^i = 4\frac{G^2}{c^2 r^4}x_{aa'}^i, \tag{IV.4b}$$

$${}^{(2)}\mu_{aa'}^i = 5\frac{G^2}{c^2 r^4}x_{aa'}^i. \tag{IV.4c}$$

## B. Dynamic development up to $c^{-2}$

In the preceding section we obtained the dynamic in a power series development of the masses of both particles. We have used a manifestly covariant formalism, imposing that the evolution of a system of 2-particles is determined by a Künzle predictive system, i.e., a dynamic system on $M_4$ of the (II.1) type, fulfilling the (II.2) conditions. Let us now consider the ordinary differential system of second order on $\mathbb{R}^{3N}$,

$$\frac{dx_a^i}{dt} = v_a^i, \quad \frac{dv_a^i}{dt} = \mu_a^i(t, x_b^j, v_c^k), \tag{IV.5}$$

given by

$$\mu_a^i(t, x_b^j, v_c^k) \equiv \left(1 - \frac{v_a^2}{c^2}\right)\left(\xi_a^i - \xi_a^0\frac{v_a^i}{c}\right) \tag{IV.6a}$$

with

$$\xi_a^\alpha(t, x_b^j, v_c^k) \equiv \xi_a^\alpha\left[x_b^0 = ct, x_{c'}^j, u_d^0 = \left(1 - \frac{v_d^2}{c^2}\right)^{-1/2}, u_e^k = u_e^0\frac{v_e^k}{c}\right], \tag{IV.6b}$$

where $\xi_a^\alpha(x_b^\beta, u_c^\gamma)$ is obtained from the $\theta_a^\alpha(x_b^\beta, \pi_c^\gamma)$ dynamic by means of (II.5).

The equivalence of both descriptions can be easily demonstrated in the sense that any set of trajectories of the (II.1) system is a set of integral curves of the (IV.5) system and vice versa. This last formalism is usually known as the manifestly predicitve formalism.

From (IV.6), (III.29), and (III.31) we obtain, only up to the $c^{-2}$ order, a development similar to (IV.3), with:

$${}^{(1)}\mu_a^i = \frac{G}{r^3}\left\langle \alpha x_{aa'}^i + \frac{1}{c^2}\left\{-\alpha v_a^2 - (-\alpha + \beta + {}^{(1)}d)v^2 + \frac{3}{2r^2}[-\alpha(\bar{x}\bar{v}_{a'})^2 + {}^{(1)}d(\bar{x}\bar{v})^2]\right\}x_{aa'}^i + \frac{1}{c^2}[2(-\alpha + \beta)(\bar{x}\bar{v}) - \alpha\eta_a(\bar{x}\bar{v}_{a'})]v_{aa'}^i\right\rangle, \tag{IV.7a}$$

$${}^{(2)}\mu_a^i = ({}^{(2)}a + 2\alpha\beta - \alpha\,{}^{(1)}d)\frac{G^2}{c^2 r^4}x_{aa'}^i, \tag{IV.7b}$$

$${}^{(2)}\mu_{aa'}^i = (2\alpha^2 - 2\alpha\beta + \alpha\,{}^{(1)}d + {}^{(2)}\underline{a})\frac{G^2}{c^2 r^4}x_{aa'}^i \tag{IV.7c}$$

At the beginning of Sec. IV, we pointed out that all the terms which contribute up to $c^{-2}$, appear in the terms up to second order in a development in the masses. Accordingly, (IV.4) and (IV.7) can be directly compared. We thus deduce that to obtain the same accelerations with our dynamic, as those we obtain from the EIH Lagrangian:

$$\alpha = -1, \quad \beta = 1, \quad {}^{(1)}d = 0, \quad {}^{(2)}a = 2, \quad {}^{(2)}\underline{a} = 1. \tag{IV.8}$$

We then conclude that of the 12 arbitrary parameters which appear up to second order in a mass development, we must fix the following parameters in the (III.18) development:

$$\text{order } m_{a'}: {}^{(1)}a = -2, \quad {}^{(1)}c = 2, \quad {}^{(1)}d = 0,$$
$$\text{order } m_{a'}^2: {}^{(2)}a = 2, \qquad \text{(IV.9)}$$
$$\text{order } m_a m_{a'}: {}^{(2)}\underline{a} = 1,$$

leaving the rest arbitrary: ${}^{(1)}b$, ${}^{(2)}b$, ${}^{(2)}\underline{b}$, ${}^{(2)}c$, ${}^{(2)}\underline{c}$, ${}^{(2)}d$ and ${}^{(2)}\underline{d}$ if we want to obtain the EIH dynamic. The $g_{a'\alpha\beta}$ field given by (III.17) is then written:

$$g_{a'00} = -1 + 2\frac{{}_0r_{a'}}{r} - 2\frac{{}_0r_{a'}^2}{r^2} - \frac{{}_0r_a\,{}_0r_{a'}}{r^2} + \cdots,$$
$$g_{a'0i} = \left({}^{(1)}b\frac{{}_0r_{a'}}{r} + {}^{(2)}b\frac{{}_0r_{a'}^2}{r^2} + {}^{(2)}\underline{b}\frac{{}_0r_a\,{}_0r_{a'}}{r^2}\right)\eta_a\frac{x_i}{r} + \cdots, \qquad \text{(IV.10)}$$
$$g_{a'ij} = \left(1 + 2\frac{{}_0r_{a'}}{r} + {}^{(2)}c\frac{{}_0r_{a'}^2}{r^2} + {}^{(2)}\underline{c}\frac{{}_0r_a\,{}_0r_{a'}}{r^2}\right)\delta_{ij}$$
$$+ \left({}^{(2)}d\frac{{}_0r_{a'}^2}{r^2} + {}^{(2)}\underline{d}\frac{{}_0r_a\,{}_0r_{a'}}{r^2}\right)\frac{x_ix_j}{r^2} + \cdots,$$

where ${}_0r_a \equiv Gm_a/c^2$ is the Schwarzschild radius of the $m_a$ mass.

Regarding the (IV.10) metric, let us note that in the case $m_a/m_{a'} \ll 1$, if we suppose ${}^{(1)}b \equiv {}^{(2)}b \equiv {}^{(2)}\underline{b} \equiv 0$ and that ${}^{(2)}c$ (respectively ${}^{(2)}d$) is not negligible in comparison with ${}^{(2)}\underline{c}$ (respectively ${}^{(2)}\underline{d}$), we obtain:

$$g_{a'00} = -1 + 2\frac{{}_0r_{a'}}{r} - 2\frac{{}_0r_{a'}^2}{r^2} + \cdots, \quad g_{a'0i} = 0, \qquad \text{(IV.11)}$$
$$g_{a'ij} = \left(1 + 2\frac{{}_0r_{a'}}{r}\right)\delta_{ij} + \frac{{}_0r_{a'}^2}{r^2}\left({}^{(2)}c\delta_{ij} + {}^{(2)}d\frac{x_ix_j}{r^2}\right) + \cdots$$

which is the approximate metric[21] which leads to the explanation of three classical effects: advance of the perihelion of Mercury, bending of light, and red shift.

## V. CONCLUSIONS AND COMMENTARIES

We have introduced a theory on gravitational interaction between two punctual structureless particles in the framework of predictive relativistic mechanics. We have adopted a variation of the formalism which is used for the description of other interactions[22] (scalar and vectorial, and in particular, electrodynamic) which we have called the Künzle predictive formalism and which is compatible with the motion equations postulated for the particles. We have also supposed that the interaction propagates through light cones and that it can be expressed by a symmetric second order covariant tensorial function $g_{\alpha\beta}$ on $(TM_4)^2$ which we have called "field," thus committing certain language abuse as this field is not required to satisfy and "field equation," but only certain symmetries. The result of this very weak imposition on the field implies the appearance of a number of arbitrary parameters which become manifest in the particle accelerations. The comparison with the dynamic obtained from the EIH Lagrangian allows us to fix five of these parameters, the remainder being arbitrary. On the other hand, our dynamics is valid for any velocity (for we have adopted fast motion approximation).

Let us note that because of the way the theory is built, it is Poincaré invariant. We also believe that it will not be difficult to build a Hamiltonization which is compatible with this group of symmetries,[23] and this will thus lead us to clearly define the ten conservative quantities: energy, momentum, angular momentum, and center of mass of the system.

Regarding the possible explanation of all the gravitational experiments done up to now, we can say that those in which light does not play a role (i.e., advance of the perihelion of Mercury) and do not require the interior structure of the particles, could be explained without any difficulty. In our opinion, light should be the object of a new interpretation.[24]

Taking into consideration the preceding, the arbitrariness and generality of the calculations which we have carried out are clear, but we consider it important to show that from very general principles and in the framework of predicitve relativistic mechanics a theory can be developed for the gravitational interaction, which agrees with certain theoretical results obtained with another theory (i.e, general relativity).

Some remarkable aspects are that the geometric framework which we have used is Minkowsky flat space–time and that the theory, as it has been built, is invariant under the Poincaré group. Finally, we are studying the possibility of introducing Lorentz invariant "field equations" in order to determine (univocally) the field.


## ACKNOWLEDGMENTS

We are deeply indebted to Dr. J. Martin for many helpful discussions and suggestions during the writing of this paper.


[1]D.G. Currie, Phys. Rev. **142**, 817 (1966). R.N. Hill, J. Math. Phys. **8**, 201 (1967). L. Bel, Ann. Inst. H. Poincaré **12**, 307 (1970). R. Arens, Arch. Rat. Mech. **47**, 255 (1972).

[2]Ph. Droz-Vincent, Lett. Nuovo Cimento **1**, 839 (1969); Physica Scripta **2**, 129 (1970); L. Bel, Ann. Inst. H. Poincaré **3**, 189 (1971).

[3]L. Bel, Ann. Inst. H. Poincaré **3**, 189 (1971); J.G. Wray, Phys. Rev. D **1**, 2212 (1970).

[4]$\alpha,\beta,\gamma,\cdots = 0,1,2,3$; $i,j,k,\cdots = 1,2,3$; $a,b,c,a' = ,\cdots = 1,\cdots,N$. $a'$ is always different from the index $a$, $a' \neq a$; we use these indices indistinctly in an up or down position and we use the summation convention for both the Greek and Latin indices, $\epsilon_a = \epsilon^a = +1$ for all values of the index $\underline{a}$; when $N = 2$ we use the symbol $\eta_a$ defined by $\eta_1 = +1$, $\eta_2 = -1$; we use the signature $+2$ of $M_4$ and $\eta_{\alpha\beta}$ denotes the coefficients of the Minkowski metric $\eta_{00} = -1$, $\eta_{0i} = \eta_{i0} = 0$, $\eta_{ij} = \delta_{ij}$, where $\delta_{ij}$ is the symmetric Kronecker tensor; $(xy) \equiv x^\rho y_\rho$, $(\bar{x}\bar{y}) \equiv x^i y_i$; $c$ is the velocity of light in vacuum and $G$ the gravitational constant.

[5]Concretely, they are Eqs. (III.32) of this article.


[6]J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962).

[7]D.C. Chern and Peter Havas, J. Math. Phys. **14**, 470 (1973).

[8]J.L. Andersson, *Principles of Relativity Physics* (Academic, New York, 1967).


[9]It is clear, that we may use (I.2) condition, if we change the world lines parameter.

[10]A more general homogeneity condition than (I.16) has been introduced by H.P. Künzle, Symposia Mathematica **14**, 53 (1974).


[11]A. Einstein, L. Infeld, and B. Hoffmann, Ann. Math. **39**, 65 (1938); L.D. Landau and E.M. Lifshitz, *Teoría Cleásica de Campos* (*Reverté*, Barcelona, 1966), p. 423.


[12]Actually, the spirit in which H.P. Künzle introduces his predictive system is somewhat different. See Ref. 10.

[13]This shift technique has also been used by: D. Hirondel, J. Math. Phys. **15**, 1689 (1974); J.L. Sanz and J. Martín, Ann. Inst. H. Poincaré **24**, 347 (1976).

[14]$TM_4$ denotes the set of pairs $(x^\alpha, u^\beta)$ where $x^\alpha$ is a point in $M_4$ and $u^\beta$ is a unitary temporal vector oriented towards the future in that point.

[15]The Aristoteles group is a group of seven parameters which acts on $M_4$ in the following way: $x^{0'} = x^0 - A^0, x^{i'} = R^i_j(x^j - A^j)$, where $A^\alpha$ are four real numbers which define a space–time translation and $R^i_j$ is a matrix of the rotation group, i.e., $\delta_{ij} R^i_k R^j_l = \delta_{kl}$.

[16]Hereinafter, the term in $m_a m_{a'}$ does not include posers higher than $c^{-4}$.

[17]We still keep $\epsilon$ so that we will be able to observe afterwards the differences which exist between the advanced and retarded cases.

[18]L. Bel, A. Salas, and J.M. Sánchez, Phys. Rev. D **7**, 1099 (1973); L. Bel, Journées Relativistes de Toulouse, Université de Toulouse, Department de Mathematiques; L. Bel and X. Fustero, Ann. Inst. H. Poincaré **25**, 411 (1976).

[19]J.L. Sanz and J. Martín, Ann. Inst. H. Poincaré **14**, 347 (1976).

[20]L. Mas, C.R. Acad. Sci. Paris A **271**, 206 (1970); F.J. Kennedy, J. Phys. **40**, 63 (1971).

[21]A.S. Eddington, *The Mathematical Theory of Relativity* (Cambridge U. P., Cambridge, 1930), Sec. 47.

[22]The two versions of the short and long range scalar theory are known as $\gamma = 0$ and $\gamma = 1$, which are connected to the motion equations which are postulated for the evolution of the system. See Ref. 5.

[23]We believe that the Hamiltonization of Künzle predictive systems must follow a different line although similar in idea to the one satisfactorily developed by L. Bel and J. Martín, Ann. Inst. H. Poincaré **22**, 173 (1975).

[24]In the framework of general relativity, the bending of light and the gravitational red shift are interpreted as being due to the effect of the curvature of space–time in the presence of mass. For a different point of view, see A.B. Volkov, Canadian J. Phys. **49**, 201 (1971), where the author interprets these effects as being due to the consequences of the energy–momentum conservation of photons in a gravitational field in a Minkowski flat space–time.

# Addendum to: Invariance and conservation laws for Lagrangian systems with one degree of freedom [J. Math. Phys. 19, 1049 (1978)]

Willy Sarlet

*Instituut voor Theoretische Mechanica, Rijksuniversiteit Gent, Krijgslaan 271-S9, B—9000 Gent, Belgium*

(Received 8 November 1978)

An invariance requirement on the Lagrangian of type (7) has a doubtful meaning when transformations of the independent variable $t$ are involved. A symmetry for $L$ is more commonly defined by

$$L(q,\dot{q},t) \equiv L(Q,Q',T)\frac{dT}{dt}, \qquad (7')$$

where we ignore the addition of a total time-derivative $dG(Q,T)/dT$ to $L$ on the right-hand side, because it is not needed for this paper. When (7′) is taken as the starting point, the equation of motion always inherits the symmetry of the Lagrangian. In that case the paper should be regarded to deal from the beginning only with transformations of type (14). All statements and conclusions then remain valid if the following modifications are taken into account:

Equation (19) should read: $\Lambda(q,\dot{q},t) = L(Q,Q',T)\,d$,

Equation (38) should read: $L(q,\dot{q},t) = L_1(t\dot{q},q)/t$,

Equation (53) should read:

$$\mathscr{L}(q,\dot{q},t) \equiv \mathscr{L}(q,d^{-1}\dot{q},dt)\,d \text{ and } \Phi(q,\dot{q},t) \equiv \Phi(q,d^{-1}\dot{q},dt),$$

Equation (56) should read:

$$\mathscr{L} = \sum_{k=0}^{n-1} L_k(q,\dot{q},t)\,d^k \quad \text{and} \quad \Phi = \sum_{k=0}^{n-1} \phi_k(q,\dot{q},t).$$

I am indebted to Dr. R.L. Schafir for suggesting the modifications of this addendum.



0022-2488/79/061295-01$00.00

[13]This shift technique has also been used by: D. Hirondel, J. Math. Phys. **15**, 1689 (1974); J.L. Sanz and J. Martín, Ann. Inst. H. Poincaré **24**, 347 (1976).

[14]$TM_4$ denotes the set of pairs $(x^\alpha,u^\beta)$ where $x^\alpha$ is a point in $M_4$ and $u^\beta$ is a unitary temporal vector oriented towards the future in that point.

[15]The Aristoteles group is a group of seven parameters which acts on $M_4$ in the following way: $x^{0'} = x^0 - A^0, x^{i'} = R^i_j(x^j - A^j)$, where $A^\alpha$ are four real numbers which define a space–time translation and $R^i_j$ is a matrix of the rotation group, i.e., $\delta_{ij}R^i_kR^j_l = \delta_{kl}$.

[16]Hereinafter, the term in $m_a m_{a'}$ does not include posers higher than $c^{-4}$.

[17]We still keep $\epsilon$ so that we will be able to observe afterwards the differences which exist between the advanced and retarded cases.

[18]L. Bel, A. Salas, and J.M. Sánchez, Phys. Rev. D **7**, 1099 (1973); L. Bel, Journées Relativistes de Toulouse, Université de Toulouse, Department de Mathematiques; L. Bel and X. Fustero, Ann. Inst. H. Poincaré **25**, 411 (1976).

[19]J.L. Sanz and J. Martín, Ann. Inst. H. Poincaré **14**, 347 (1976).

[20]L. Mas, C.R. Acad. Sci. Paris A **271**, 206 (1970); F.J. Kennedy, J. Phys. **40**, 63 (1971).

[21]A.S. Eddington, *The Mathematical Theory of Relativity* (Cambridge U. P., Cambridge, 1930), Sec. 47.

[22]The two versions of the short and long range scalar theory are known as $\gamma = 0$ and $\gamma = 1$, which are connected to the motion equations which are postulated for the evolution of the system. See Ref. 5.

[23]We believe that the Hamiltonization of Künzle predictive systems must follow a different line although similar in idea to the one satisfactorily developed by L. Bel and J. Martín, Ann. Inst. H. Poincaré **22**, 173 (1975).

[24]In the framework of general relativity, the bending of light and the gravitational red shift are interpreted as being due to the effect of the curvature of space–time in the presence of mass. For a different point of view, see A.B. Volkov, Canadian J. Phys. **49**, 201 (1971), where the author interprets these effects as being due to the consequences of the energy–momentum conservation of photons in a gravitational field in a Minkowski flat space–time.

# Addendum to: Invariance and conservation laws for Lagrangian systems with one degree of freedom [J. Math. Phys. 19, 1049 (1978)]

Willy Sarlet

*Instituut voor Theoretische Mechanica, Rijksuniversiteit Gent, Krijgslaan 271-S9, B—9000 Gent, Belgium*

(Received 8 November 1978)

An invariance requirement on the Lagrangian of type (7) has a doubtful meaning when transformations of the independent variable $t$ are involved. A symmetry for $L$ is more commonly defined by

$$L(q,\dot q,t)\equiv L(Q,Q',T)\,\frac{dT}{dt}, \qquad (7')$$

where we ignore the addition of a total time-derivative $dG(Q,T)/dT$ to $L$ on the right-hand side, because it is not needed for this paper. When (7′) is taken as the starting point, the equation of motion always inherits the symmetry of the Lagrangian. In that case the paper should be regarded to deal from the beginning only with transformations of type (14). All statements and conclusions then remain valid if the following modifications are taken into account:

Equation (19) should read: $\Lambda(q,\dot q,t) = L(Q,Q',T)\,d$,

Equation (38) should read: $L(q,\dot q,t) = L_1(t\dot q,q)/t$,

Equation (53) should read:

$$\mathscr{L}(q,\dot q,t)\equiv\mathscr{L}(q,d^{-1}\dot q,dt)\,d \text{ and } \Phi(q,\dot q,t)\equiv\Phi(q,d^{-1}\dot q,dt),$$

Equation (56) should read:

$$\mathscr{L} = \sum_{k=0}^{n-1} L_k(q,\dot q,t)\,d^k \quad\text{and}\quad \Phi = \sum_{k=0}^{n-1}\phi_k(q,\dot q,t).$$

I am indebted to Dr. R.L. Schafir for suggesting the modifications of this addendum.



0022-2488/79/061295-01$00.00