# Prolongation structures of nonlinear evolution equations*

H. D. Wahlquist and F. B. Estabrook

*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103*
(Received 12 August 1974)

A technique is developed for systematically deriving a "prolongation structure"—a set of interrelated potentials and pseudopotentials—for nonlinear partial differential equations in two independent variables. When this is applied to the Korteweg–de Vries equation, a new infinite set of conserved quantities is obtained. Known solution techniques are shown to result from the discovery of such a structure: related partial differential equations for the potential functions, linear "inverse scattering" equations for auxiliary functions, Bäcklund transformations. Generalizations of these techniques will result from the use of irreducible matrix representations of the prolongation structure.

## I. INTRODUCTION

The simplest nonlinear evolution equations having solitary wave solutions (solitons) are known to possess an infinite set of conservation laws.[1–3] It is a natural expectation that the existence of infinitely many conserved quantities is probably intimately connected with the soliton phenomenon,[3–5] but the precise relationship is still unclear. The differential and integral conservation laws (as distinct from "constants of the motion") which have been considered heretofore are constructed from the independent and dependent variables of the evolution equation plus successively higher partial derivatives of the dependent variables.[1–3]

An entirely different set of conservation laws can be developed from the hierarchy of potentials and pseudopotentials connected with the original equation. In a sense these laws proceed in the opposite direction, depending on successively higher *integrals* of the original, or primitive, variables. We use the term "potential" to denote an integral variable which can be defined by a quadrature over lower variables. The term "pseudopotential" refers to an integral variable which is defined by an integrable set of first-order differential equations (more precisely, by a Pfaffian 1-form), the solution to which cannot be written in terms of a quadrature.

When pseudopotentials are included, this set of conservation laws, too, appears to be infinite. It also appears to be much more closely related to useful solution techniques for the evolution equations. As we show for the Korteweg–de Vries (KdV) equation in Sec. VI (and for the nonlinear Schrödinger equation in a separate paper[6]), these conservation laws lead directly to the soliton solutions,[3–5] to the Bäcklund transformation[3,5,7] between solutions, and to the linear equations used in the inverse scattering approach to the initial value problem.[2–4,8]

The equations which ultimately define these conservation laws can be expressed in terms of an algebraic Lie structure which we have called the *prolongation structure* of the equation. Subsets of this structure form finite Lie algebras, whereas the entire structure appears to be open-ended, leading to an infinity of higher conservation laws. If the structure be closed arbitrarily, it then becomes a finite Lie algebra (as shown here, Sec. V, for the KdV equation). The structure constants of the resulting algebra can be identified with the eigenvalue parameter of the well known linear operator associated with the KdV equation.[4]

We have not yet had time to investigate these prolongation structures as abstract algebraic entities, nor to determine their general properties: finiteness, degeneracy, invariant substructures, etc. We believe, however, that these general properties must contain important clues toward answering the fundamental questions about the existence of solitons and Bäcklund transformations, the applicability of inverse scattering approaches to the initial value problem, and the possibility of linearization transformations. From the point of view of the prolongation structure, many of these techniques seem to depend on the existence of a *pseudopotential* for the given evolution equation. As will be brought out in Sec. IV, pseudopotentials are connected with the non-Abelian character of the prolongation structure, although the relationship is not entirely clear. We speculate that this basic property, among others, will be highly significant for a general analysis of nonlinear evolution equations.

As the best known equation exhibiting all these phenomena, the KdV equation provides an excellent prototype upon which to exercise and illustrate any new development. Accordingly, the present paper is concerned with obtaining the prolongation structure of the KdV equation, and illustrating its relation to the many known techniques for treating this equation. Since the analysis is performed in the perhaps unfamiliar language of Cartan's exterior differential forms,[9,10] Secs. II and III provide a brief introduction, defining the notation and setting up the KdV equation in terms of differential forms. While we do not emphasize the geometrical interpretation of our analysis (which is so well expressed by the differential form language), even analytically this notation is unquestionably superior for any treatment of conservation laws and integrability conditions. Finally, it is clear that there are many places in the analysis of this paper where further work needs to be done—sometimes just to settle a minor difficulty, but also to carry out major extensions.

## II. THE KORTEWEG-DE VRIES EQUATION

Using subscripts to denote partial derivatives, we may write the KdV equation as

$$u_t + u_{xxx} + 12uu_x = 0. \tag{1}$$

The constant multiplying the nonlinear term can be adjusted by scaling $u$, and the value 12 has been chosen here for convenience. In order to express this equation in differential forms, we define the variables

$$z \equiv u_x, \quad p \equiv z_x = u_{xx}, \tag{2}$$

whereupon Eq. (1) may be written as the first-order equation

$$u_t + p_x + 12uz = 0. \tag{3}$$

In the five-dimensional space of all these dependent and independent variables $\{x, t, u, z, p\}$, we adopt the basis forms $\{dx, dt, du, dz, dp\}$. The set of first-order equations in Eqs. (2) and (3) may then be expressed by the following set of second-rank differential forms (2-forms):

$$\begin{aligned} \alpha_1 &= du \wedge dt - z dx \wedge dt, \\ \alpha_2 &= dz \wedge dt - p dx \wedge dt, \\ \alpha_3 &= -du \wedge dx + dp \wedge dt + 12 uz dx \wedge dt, \end{aligned} \tag{4}$$

where $d$ denotes the exterior derivative and $\wedge$ denotes the exterior product (antisymmetric tensor product).[9] Any regular two-dimensional solution manifold in the five-dimensional space, $S_2 = \{u(x,t), u_x = z(x,t), z_x = p(x,t)\}$, satisfying Eq. (1) will *annul* this set of forms, as may be verified by *sectioning* the forms into the solution manifold.[10] On $S_2$ we will have

$$du = u_x dx + u_t dt, \tag{5}$$

and similarly for $z$ and $p$, so that by virtue of the antisymmetry of exterior multiplication, for example, $du \wedge dt = u_x dx \wedge dt$. Thus, the forms become

$$\begin{aligned} \tilde{\alpha}_1 &= (u_x - z) dx \wedge dt = 0, \\ \tilde{\alpha}_2 &= (z_x - p) dx \wedge dt = 0, \\ \tilde{\alpha}_3 &= (u_t + p_x + 12uz) dx \wedge dt = 0, \end{aligned} \tag{6}$$

where the sectioned forms are denoted by a tilda.

In order to assert complete equivalence between the forms and the differential equations, the set of forms must be "closed"; i.e., the exterior derivatives of all the forms must be contained in the *ring* of forms generated by the set,

$$d\alpha_i = \sum_{j=1}^{3} \eta_{ji} \wedge \alpha_j, \tag{7}$$

where $\eta_{ji}$ is some set of 1-forms. This is equivalent to ensuring that all integrability conditions of the set of first-order equations in Eq. (2) and Eq. (3) are satisfied. In the present case, we find

$$\begin{aligned} d\alpha_1 &= -dz \wedge dx \wedge dt = dx \wedge \alpha_2, \\ d\alpha_2 &= dx \wedge \alpha_3, \\ d\alpha_3 &= -12\, dx \wedge (z\alpha_1 + u\alpha_2). \end{aligned} \tag{8}$$

Thus, the set of forms, Eq. (4), constitutes a *closed ideal* of differential forms and Cartan's theory[10] of such systems may be applied. In a closed ideal any local surface element which annuls the $\alpha_i$ also annuls their exterior derivatives $d\alpha_i$. Cartan's theorem guarantees that these surface elements will "fit" together to produce a global 2-surface which constitutes a solution manifold for the forms.

## III. CONSERVATION LAWS AND POTENTIALS

Conservation laws associated with the KdV equation correspond to the existence of *exact* 2-forms contained in the ring of the $\alpha_i$. Let us suppose that we can find a set of functions, $f_i(x, t, u, z, p)$, such that the 2-form

$$\beta = f_1\alpha_1 + f_2\alpha_2 + f_3\alpha_3 \tag{9}$$

satisfies $d\beta = 0$, the condition for exactness. This is the integrability condition for the existence of a 1-form, say $\omega$, such that

$$\beta = d\omega, \tag{10}$$

which conversely implies, $d\beta = 0$, by the usual identity for the double exterior derivative of any differential form, $d(d\omega) = 0$.[9]

For example, we consider

$$\beta = -\alpha_3 - 12u\alpha_1, \tag{11}$$

and calculate its exterior derivative

$$d\beta = -d\alpha_3 - 12\, du \wedge \alpha_1 - 12u d\alpha_1 \equiv 0. \tag{12}$$

Substituting the forms from Eq. (4) and Eq. (8) verifies that this vanishes identically, as shown. We find then in accordance with Eq. (10) that $\beta$ can be derived from the 1-form

$$\omega = u dx - (p + 6u^2) dt. \tag{13}$$

The associated conservation law results from an application of Stokes' theorem[9,10]

$$\oint_{M_1} \omega = \int_{M_2} d\omega, \tag{14}$$

written for any simply-connected two-dimensional manifold $M_2$ with closed one-dimensional boundary $M_1$, and the notation implies that $\omega$ and $d\omega$ are to be evaluated on their respective manifolds. If for $M_2$ we choose a solution manifold $S_2$ which annuls the $\alpha_i$, we will have from Eq. (11)

$$d\tilde{\omega} = \tilde{\beta} = 0, \tag{15}$$

giving

$$\oint_{S_1} \tilde{\omega} = \oint_{S_1} [u(x,t)\, dx - (p + 6u^2)\, dt] = 0, \tag{16}$$

where $S_1$ is any closed curve in $S_2$. For appropriate asymptotic boundary conditions ($u, z, p \to 0$, $|x| \to \infty$), $S_1$ can be chosen in the usual way to exhibit the conserved quantity

$$w_0 = \int_{-\infty}^{\infty} u(x,t)\, dx. \tag{17}$$

Returning to Eq. (13), we can of course add to this $\omega$ any *exact* 1-form (say $dw$, where $w$ is an arbitrary scalar function), and so take instead

$$\omega = dw + u dx - (p + 6u^2) dt, \tag{18}$$

still having $\beta = d\omega$. It is now convenient to regard $w$ simply as a coordinate in an extended six-dimensional space of variables $\{x, t, u, z, p, w\}$ and to add the 1-form $\omega$ to our original set of forms. Since $d\omega$ is known to be in the ring of the original set, the new set of forms remains a closed ideal. We shall refer to this process of inventing new variables and larger closed ideals (existing in higher dimensional spaces), as "prolongation" of the original set.[10]

A two-dimensional solution manifold $S_2$ in this larger space will be required to annul all the forms of the prolonged ideal. Upon sectioning into $S_2$, where $dw = w_x dx + w_t dt$, we have

$$\tilde{\omega} = (w_x + u)\, dx + (w_t - p - 6u^2)\, dt = 0. \tag{19}$$

Thus, $w(x, t)$ now appears as a potential function for $u$, defined by the first-order equations

$$u = -w_x, \tag{20}$$

$$p + 6u^2 = w_t, \tag{21}$$

or, equivalently, by the quadrature

$$w(x, t) = -\int^x u(x', t)\, dx'. \tag{22}$$

The cross-derivative integrability condition $w_{xt} = w_{tx}$ (which requires $u$ to satisfy the KdV equation) simply restates the content of Eq. (11) showing $d\omega$ in the ring of the original set of forms $\alpha_i$ which represent the KdV equation. Eliminating $u$ and $p$ between Eq. (20) and Eq. (21), we find that $w$ itself satisfies the equation

$$w_t + w_{xxx} - 6w_x^2 = 0, \tag{23}$$

and so have discovered another nonlinear partial differential equation closely related to the KdV equation.

## IV. MULTIPLE PROLONGATION AND PSEUDOPOTENTIALS

It may be possible to find several different 1-forms (Pfaffians) having a structure similar to $\omega$; i.e.,

$$\omega_k = dy^k + F^k(x, t, u, z, p)\, dx + G^k(x, t, u, z, p)\, dt, \tag{24}$$

the exterior derivatives of which are in the ring of the initial set of forms

$$d\omega_k = \sum_{i=1}^{3} f^k{}_i \alpha_i, \tag{25}$$

where $f^k{}_i$ is some set of scalar functions. This last equation in fact provides the most convenient method to search for such Pfaffians. Expanding it, we have

$$dF^k \wedge dx + dG^k \wedge dt = \sum_{i=1}^{3} f_i^k \alpha_i, \tag{26}$$

or

$$\sum_u \left( \frac{\partial F^k}{\partial z^u} dz^u \wedge dx + \frac{\partial G^k}{\partial z^u} dz^u \wedge dt \right) - \sum_i f_i^k \alpha_i = 0, \tag{27}$$

where $z^u = \{x, t, u, z, p\}$ is the set of arguments of $F^k$ and $G^k$. Setting to zero the coefficients of the various independent 2-forms in this equation, we obtain a highly overdetermined set of coupled linear first-order equations for $F^k$ and $G^k$. Each independent solution determines a Pfaffian form $\omega_k$, and each such form leads to an associated conservation law, defines a new potential function $y^k$, and permits a corresponding prolongation.

We may consider two immediate generalizations of the process. Firstly, prolongation may be thought of as a *sequential* process. In the last section, for example, we achieved a first prolongation of the set of forms using the potential $w$. At the next step we might consider Pfaffians dependent on all six variables including $w$. Thus, we would search for 1-forms as in Eq. (24), but with $F^k$ and $G^k$ depending on $w$ as well, and require closure in the first-prolonged ideal of forms by

$$d\omega_k = \sum_{i=1}^{3} f^k{}_i \alpha_i + \eta^k \wedge \omega, \tag{28}$$

where $\eta^k$ is some 1-form. If new Pfaffians result, the ideal could be further prolonged and then the entire process repeated. This generalization is not empty, and in fact does lead to new and interesting Pfaffians.

But this suggests a further generalization in which we allow $F^k(z^u, y^i)$ and $G^k(z^u, y^i)$ from the start to depend on *all* variables, the primitive set $z^u = \{x, t, u, z, p\}$, and *any* set of prolongation variables $y^i$. The closure equation will read

$$d\omega_k - \sum_{i=1}^{3} f^k{}_i \alpha_i - \sum_{i=1}^{n} \eta_i^k \wedge \omega_i = 0, \tag{29}$$

where $n$ is the number of prolongation variables to be included and $\eta_i^k$ is some set of 1-forms. This equation can be treated in exactly the same manner as Eq. (25), and again results in an overdetermined set of partial differential equations for $F^k$ and $G^k$, which, however, are no longer strictly linear, since terms of the form

$$\sum_i \left( G^i \frac{\partial F^k}{\partial y^i} - F^i \frac{\partial G^k}{\partial y^i} \right) dx \wedge dt \tag{30}$$

will now occur. This nonlinearity is the price we must pay to avoid a tedious *sequential* process in solving for many of the interesting Pfaffians (those which themselves depend on prolongation variables). Fortunately, the nonlinear terms always have the simple "commutator" form, as shown by Eq. (30), and in fact lead to an elegant algebraic structure which is always "solvable" in principle.

One further consequence of this last generalization should be noted. In the previously considered sequential case, the functions $F^k$ and $G^k$ do not depend on the newest prolongation variable $y^k$ itself. It is then appropriate to call the new variable ($w$, for instance) a "potential." Now, however, we may find Pfaffians in which $F^k$ and $G^k$ do depend on $y^k$, and we shall refer to this type of prolongation variable as a "pseudopotential." The nonlinear terms of the closure equations are clearly essential for these variables, and pseudopotentials cannot be found by the sequential process using linear equations. As stated in the Introduction, the existence of pseudopotentials appears to be the key to Bäcklund transformations, inverse scattering equations, and solution generation techniques.

## V. PROLONGATION STRUCTURE OF THE KdV EQUATION

The following treatment of the KdV equation is restricted in that we do not allow $F^k$ and $G^k$ to be explicit functions of the independent variables $x$ and $t$. The primary reason for this is simplicity, but it seems plausible that it is not overly restrictive since the KdV equation itself has no explicit $(x, t)$ dependence. Nevertheless, this is one of the loose ends mentioned in the Introduction, and it remains to be verified that nothing essential is missed by this simplification.

When Eq. (29) is written out in detail, the following set of partial differential equations for $F^k(u, z, p, y^i)$ and $G^k(u, z, p, y^i)$ is obtained:

$$F^k_{,z} = 0, \quad F^k_{,p} = 0, \quad F^k_{,u} + G^k_{,p} = 0,$$

$$zG^k_{,u}+pG^k_{,z}-12uzG^k_{,p}+G^iF^k_{,y^i}-F^iG^k_{,y^i}=0, \tag{31}$$

where the comma notation for partial derivatives has been used and the summation convention for repeated indices. The integrability conditions found by taking further partial derivatives of these equations with respect to the primitive variables $(u, z, p)$ can be integrated to show that $F^k$ and $G^k$ may be expressed as

$$\begin{aligned} F^k &= 2X_1^k + 2uX_2^k + 3u^2X_3^k, \\ G^k &= -2(p+6u^2)X_2^k + 3(z^2 - 8u^3 - 2up)X_3^k \\ &\quad + 8X_4^k + 8uX_5^k + 4u^2X_6^k + 4zX_7^k, \end{aligned} \tag{32}$$

where the integration functions $X_m^k(y^i)$ $(m=1,\ldots,7)$ depend only on prolongation variables. All dependence of $F^k$, $G^k$ on the primitive variables is thereby explicitly determined, and the expressions do not depend on the number $n$ of prolongation variables assumed.

With this result Eq. (31) splits up into a set of equations for the $X_m^k(y^i)$. For example, one of these equations is

$$X_1^iX^k_{2,y^i} - X_2^iX^k_{1,y^i} + X_7^k = 0. \tag{33}$$

All the remaining equations have this commutator structure of the derivative operators and can be concisely expressed by using the Lie derivative, or Lie product, notation[9]

$$\underset{\vec{X}_m}{\pounds}\vec{X}_l = [\vec{X}_m, \vec{X}_l] = -[\vec{X}_l, \vec{X}_m], \tag{34}$$

where in components

$$[\vec{X}_m, \vec{X}_l]^k \equiv X_m^iX^k_{l,y^i} - X_lX^k_{m,y^i}. \tag{35}$$

Thus, Eq. (33) can be written

$$[\vec{X}_1, \vec{X}_2] = -\vec{X}_7. \tag{36}$$

Henceforth we shall omit the arrow over vectors. In this notation the entire set of equations which result from Eq. (31) for the $X_m^k$ is

$$\begin{aligned} &[X_1, X_3] = [X_2, X_3] = [X_1, X_4] = [X_2, X_6] = 0, \\ &[X_1, X_2] = -X_7, \quad [X_1, X_7] = X_5, \quad [X_2, X_7] = X_6, \\ &[X_1, X_5] + [X_2, X_4] = 0, \quad [X_3, X_4] + [X_1, X_6] + X_7 = 0. \end{aligned} \tag{37}$$

A number of further relations can be derived by operating on these equations with the $X_m$ to form new Lie products and then using the Jacobi identity. For instance, it is quickly found that $X_3$ must commute with all vectors except $X_4$; and also

$$[X_2, X_5] = [X_1, X_8], \quad [X_6, X_7] = X_6. \tag{38}$$

Algebraically, this structure clearly comes close to defining a Lie algebra. In fact, several subsets of these generators do constitute finite Lie algebras: for example, the set $\{X_2, X_3, X_6, X_7\}$.

Defining new generators $X_8$ and $X_9$ by

$$[X_3, X_4] \equiv -X_8, \quad [X_1, X_5] \equiv X_9 \tag{39}$$

allows us to split up the last two relations between commutators in Eq. (37). Further operations will give new algebraic relations between the unknown commutators, and solve for some of them, but do not permit the explicit determination of all commutators. In fact, the process is apparently open-ended; the whole structure does not appear to close itself off into a unique Lie algebra with any finite number of generators, although we have no proof of this either way. An open structure would imply the existence of an infinite number of possible prolongation variables and associated conservation laws. For the KdV equation, such a result is perhaps not surprising. It must be noted though that these conservation laws will not involve explicit $(x, t)$ dependence, nor dependence on derivatives of $u$ higher than $p = u_{xx}$. This is not the same infinite collection of conservation laws obtained in Ref. 1.

While a detailed investigation of this algebraic prolongation structure remains to be carried out, many interesting results can be immediately obtained from it. One approach is to *force* closure by arbitrarily imposing linear dependence among the generators at some level. Specifically, for instance, if we postulate

$$X_9 = \sum_{m=1}^{8} C_m X_m, \tag{40}$$

where the $C_m$ are constants, and demand that the first eight generators be linearly independent, we find that the equations require

$$C_m = 0 \quad (m \neq 7, 8), \quad C_7 = -C_8 \equiv \lambda,$$

where $\lambda$ is any constant. When all the equations are considered, we finally obtain the eight-parameter Lie algebra described by the relations

$$\begin{aligned} &[X_1, X_2] = -X_7, && [X_2, X_5] = -X_9/\lambda, && [X_4, X_7] = -\lambda X_5, \\ &[X_1, X_5] = X_9, && [X_2, X_7] = X_6, && [X_5, X_6] = X_9/\lambda, \\ &[X_1, X_6] = -X_9/\lambda, && [X_3, X_4] = -X_8, && [X_5, X_7] = -X_5 - \lambda X_6, \\ &[X_1, X_7] = X_5, && [X_4, X_5] = -\lambda X_9, && [X_6, X_7] = X_6, \\ &[X_2, X_4] = -X_9, && [X_4, X_6] = X_9, && X_9 \equiv \lambda(X_7 - X_8), \end{aligned} \tag{41}$$

and all other commutators vanish. The element $X_9$ is used simply as an abbreviation, and is not a generator of the algebra. Note that the generator $X_8$ commutes with all others.

It is not difficult to obtain an eight-dimensional realization of this algebra. Let basis vectors be defined by

$$b_k \equiv \frac{\partial}{\partial y^k} \quad (k = 1, \ldots, 8), \tag{42}$$

where $y^k$ is a set of coordinates in the space of prolongation variables. A nondegenerate representation of the generators then is

$$\begin{aligned} X_1 &= \tfrac{1}{2}[b_1 + \exp(2y_3)b_2 + y_8b_3 + y_7b_5 + (y_8^2 - \lambda)b_8], \\ X_2 &= \tfrac{1}{2}[b_7 + 2b_8], \\ X_3 &= \tfrac{1}{3}b_6, \\ X_4 &= -\tfrac{1}{2}\lambda[b_1 + \exp(2y_3)b_2 + y_8b_3 - b_4 \\ &\quad + (3/2\lambda)y_6b_5 + (y_8^2 - \lambda)b_8], \\ X_5 &= -\tfrac{1}{2}[\exp(2y_3)b_2 + y_8b_3 + (y_8^2 + \lambda)b_8], \\ X_6 &= b_8, \\ X_7 &= \tfrac{1}{2}[b_3 + \tfrac{1}{2}b_5 + 2y_8b_8], \end{aligned} \tag{43}$$

$X_8 = \frac{1}{4}b_5$.

Using these results in Eq. (32) to write out explicitly the eight Pfaffian forms

$$\omega_k = dy^k + F^k\,dx + G^k\,dt, \tag{44}$$

which correspond to the components of these vectors, we have

$$\begin{aligned}
\omega_1 &= dy_1 + dx - 4\lambda\,dt,\\
\omega_2 &= dy_2 + \exp(2y_3)\,dx - 4\exp(2y_3)(u+\lambda)\,dt,\\
\omega_3 &= dy_3 + y_8\,dx + [2z - 4y_8(u+\lambda)]\,dt,\\
\omega_4 &= dy_4 + 4\lambda\,dt,\\
\omega_5 &= dy_5 + y_7\,dx + (z - 6y_6)\,dt,\\
\omega_6 &= dy_6 + u^2\,dx + (z^2 - 8u^3 - 2up)\,dt,\\
\omega_7 &= dy_7 + u\,dx - (p + 6u^2)\,dt,\\
\omega_8 &= dy_8 + (2u + y_8{}^2 - \lambda)\,dx - 4[(u+\lambda)(2u + y_8{}^2 - \lambda)\\
&\quad + \tfrac{1}{2}p - zy_8]\,dt.
\end{aligned} \tag{45}$$

Of these, $\omega_1$ and $\omega_4$ are trivial; $\omega_6$ and $\omega_7$ define known potentials and conservation laws for the KdV equation which involve only the primitive variables,[1] and we see that $y_7 = w$, the potential discussed in Sec. III. The potential $y_5$ could have been obtained by the linear sequential process, whereas $\omega_2$ and $\omega_3$ give potentials which require discovery of the single pseudopotential $y_8$. As an example of closure for these Pfaffians, we calculate

$$\begin{aligned}
d\omega_8 &= -4(4u + y_8{}^2 + \lambda)\alpha_1 + 4y_8\alpha_2 - 2\alpha_3\\
&\quad - 2\{y_8\,dx + [2z - 4(u+\lambda)y_8]\,dt\} \wedge \omega_8,
\end{aligned} \tag{46}$$

verifying that $\omega_8$ is closed in the prolonged ideal of forms, $\{\alpha_i, \omega_k\}$.

## VI. SOLUTION TECHNIQUES

The prolongation structure can be shown to lead quite directly to the equations which have been used in a variety of methods for obtaining analytical solutions to the KdV equation. It also serves to relate the KdV equation to a number of other nonlinear equations, such as that satisfied by $y_7 = w$, Eq. (23). Additional related equations derivable from the Pfaffians of Eq. (45) are

$$\begin{aligned}
&v_t + v_{xxx} - 2v_x{}^3 + 6\lambda v_x = 0 \quad (v = y_3),\\
&y_t + y_{xxx} - 6y^2 y_x + 6\lambda y_x = 0 \quad (y = y_8).
\end{aligned} \tag{47}$$

The second of these is essentially the *modified* KdV equation. Miura's discovery that this equation is transformable to KdV was one of the earliest results used in the analytical treatment of solitons.[11]

We shall now briefly review the relationship of the prolongation structure to the known solution methods for the KdV equation. The most significant Pfaffian of the set in Eq. (45) is $\omega_8$, defining the pseudopotential $y_8$ for which we henceforth use the symbol $y$. On a solution manifold of the prolonged ideal, we will have from $\tilde{\omega}_8 = 0$

$$\begin{aligned}
y_x &= -(2u + y^2 - \lambda),\\
y_t &= 4[(u+\lambda)(2u + y^2 - \lambda) + \tfrac{1}{2}p - zy].
\end{aligned} \tag{48}$$

The first of these is a Riccati equation linearizable by the substitution

$$y = \psi_x/\psi, \tag{49}$$

giving

$$\psi_{xx} + (2u - \lambda)\psi = 0. \tag{50}$$

Analysis of this Schrödinger equation has uncovered the beauty of the solitary wave solutions of the KdV equation.[2-4] Note that the arbitrary parameter $\lambda$, which appears in Eq. (50) as an eigenvalue when the usual asymptotic boundary conditions are imposed, first appeared in the prolongation structure as a structure constant of the Lie algebra. This demonstrates the direct relationship between these eigenvalue "constants of the motion" and the set of differential conservation laws.

The equations of another method can be derived by combining $\omega_8$ and $\omega_3$. From the Pfaffian form $\tilde{\omega}_3$ we have

$$y = -y_{3,x}. \tag{51}$$

Considering Eq. (49) and the Pfaffian $\omega_2$, this suggests making the logarithmic variable substitution

$$y_3 = -\ln\psi, \tag{52}$$

together with $\phi = \psi_x$, so that Eq. (49) becomes

$$y = \phi/\psi. \tag{53}$$

If we define a new pair of Pfaffians in the ring of $\omega_3$ and $\omega_8$ by

$$\begin{aligned}
\omega_9 &\equiv \psi\omega_8 - \phi\omega_3\\
\omega_{10} &\equiv -\psi\omega_3,
\end{aligned} \tag{54}$$

and express them in the variables $\phi$ and $\psi$, we find

$$\begin{aligned}
\omega_9 &= d\phi + (2u - \lambda)\psi\,dx\\
&\quad + \{2z\phi - [4(u+\lambda)(2u-\lambda) + 2p]\psi\}\,dt,\\
\omega_{10} &= d\psi - \phi\,dx - [2z\psi - 4(u+\lambda)\phi]\,dt.
\end{aligned} \tag{55}$$

These are linear in $\phi$ and $\psi$ and constitute the Pfaffian differential form representation of the first-order inverse scattering equations. Both Eqs. (50) and (55) have been used to develop linear techniques for solving the initial value problem for the KdV equation.[3,8]

Another technique for generating analytic solutions can be deduced from the prolongation structure. Suppose that one particular solution of the prolonged ideal $\{\alpha_i, \omega_k\}$ is known. We may inquire whether another solution, say $u'$, of the KdV equation can be written as an algebraic function of all the variables in the space of the prolonged ideal; i.e., $u' = u'(u, z, p, y^i)$. The answer can be found by substituting this *ansatz* into the set of forms

$$\begin{aligned}
\alpha_1' &= du' \wedge dt - z'\,dx \wedge dt,\\
\alpha_2' &= dz' \wedge dt - p'\,dx \wedge dt,\\
\alpha_3' &= -du' \wedge dx + dp' \wedge dt + 12u'z'\,dx \wedge dt,
\end{aligned} \tag{56}$$

and, as usual, demanding that these be in the ring of the prolonged ideal.

After a tedious but straightforward calculation, the result is that

$$u' = -u - y^2 + \lambda \tag{57}$$

is always another solution.

Since $u = 0$ satisfies KdV, $u_0' = -y^2 + \lambda$ must also be a solution. From Eq. (48) for this case

$$y_x = -(y^2-\lambda),$$
$$y_t = 4\lambda(y^2-\lambda) = -4\lambda y_x, \tag{58}$$

the regular integral of these being

$$y = \lambda^{1/2}\tanh[\lambda^{1/2}(x-x_0-4\lambda t)]. \tag{59}$$

The result for $u_0'$ is the regular 1-soliton solution.

An equation like Eq. (57) is clearly equivalent in general to a Bäcklund transformation. Simply solving for $y$ and substituting into Eq. (48) will produce the usual form of the Bäcklund transformation. In fact, for the KdV equation,[12] it is simpler to use the potential function $w$. To see this, we use the Pfaffian $\tilde{\omega}_7$ to get

$$u = -y_{7,x} = -w_x, \tag{60}$$

so that Eq. (57) can be written

$$-w_x' = w_x - y^2 + \lambda = w_x + y_x - 2w_x, \tag{61}$$

where we have used Eq. (48) to write the second equality. Integrating and absorbing the integration constant in the potentials, we have

$$y = w - w', \tag{62}$$

so that Eq. (57) can finally be written as

$$-w_x' - w_x = u' + u = \lambda - (w'-w)^2. \tag{63}$$

With $\lambda = k^2$, this is precisely the space part of the Bäcklund transformation presented in Ref. 12. By using Eq. (57) and Eq. (62), the $y_t$ equation in Eq. (48) can also be rewritten to give

$$w_t' + w_t = 4(u'^2 + u'u + u^2) + 2(w'-w)(z'-z), \tag{64}$$

which expresses the other half of the Bäcklund transformation of Ref. 12 in a symmetric form.

An extension of the solution generating technique leading to Eq. (57) can be used to derive directly the hierarchy of solutions which are known to result from recursive application of the Bäcklund transformation. So far we have treated the Pfaffian $\omega_8$ as a *single* 1-form, but we can also consider it to represent a 1-parameter infinity of independent Pfaffians, parametrized by $\lambda$. That is, for any given solution $\{u, z, p\}$ of the KdV equation, $\omega_8$ defines a 1-parameter family of pseudopotentials, $y(x,t,\lambda)$, which are linearly independent functions. This suggests that we attempt to find more general solutions than Eq. (57) by entering the forms of Eq. (56) with the *ansatz*

$$u' = u'(u, y(\lambda_1), y(\lambda_2)), \tag{65}$$

for example. Another tedious calculation (we certainly suspect there must be a neater way to obtain these results) shows that

$$u' = u + \frac{(\lambda_2-\lambda_1)[y^2(\lambda_2)-\lambda_2-y^2(\lambda_1)+\lambda_1]}{[y(\lambda_2)-y(\lambda_1)]^2} \tag{66}$$

is indeed always another solution. Since we know from Eq. (57) that

$$u_1 = -u - y^2(\lambda_1) + \lambda_1, \quad u_2 = -u - y^2(\lambda_2) + \lambda_2 \tag{67}$$

are solutions, we can eliminate the $y$'s from Eq. (66) to obtain the superposition principle, or recursion relation,

$$u' = u + [(\lambda_2-\lambda_1)(u_1-u_2)]/[(\lambda_2-u_2-u)^{1/2} - (\lambda_1-u_1-u)^{1/2}]^2, \tag{68}$$

which can then be used to generate the Bäcklund hierarchy. Again, one has much simpler expressions if this analysis is carried out for the potential $w$ as in Ref. 12, and we shall not pursue it further here.

## VII. CONCLUSION

The formulation of nonlinear evolution equations in terms of ideals of differential forms leads in a very clear fashion to the derivation of potential functions and conservation theorems. The natural and important generalization to pseudopotential functions results in discovery of new conservation theorems; and even more importantly, pseudopotentials appear to be the unifying concept for understanding the relations between diverse known solution techniques (Bäcklund transformation, associated inverse scattering problems). The discussion here has been made concrete by reference throughout to the treatment of the Korteweg–de Vries equation.

The systematic search for the pseudopotentials of a closed set of forms leads to consideration of an associated overdetermined set of first-order nonlinear partial differential equations which we denote a *prolongation structure*. The prolongation structure is integrable precisely because it has the form of (a subset of) the commutation relations of a Lie group. We have not in the present paper been able to exploit some of the deeper known results for Lie groups systematically, but it seems clear that extremely powerful mathematical techniques are now at hand. For example, the search for linear, or matrix, representations of the group (or structure) can be undertaken in a completely algorithmic way. This results in representations of the vector generators of the form $X_i = a_{ijk}y^j\partial/\partial y^k$, where the $a_{ijk}$ are constants; and the *pseudopotentials* $y^k$ *thus found will enter the prolongation forms* $\omega$ *linearly*. The variables $\phi$ and $\psi$ and forms $\omega_9$ and $\omega_{10}$ of Eq. (55) thus are seen to belong to a two-dimensional matrix representation of the prolongation structure for the KdV equation, Eq. (37).

The inverse scattering technique has been shown to provide a linear method of solving the initial value problem for many nonlinear evolution equations. One of the primary obstacles to extending the method is the discovery of the appropriate linear equations or operators. The search for linear representations of the prolongation structure would appear to provide a straightforward approach to this problem which does not require ad hoc restrictions. It also suggests generalizations; for instance, an intriguing possible generalization of the known method of inverse scattering may result from higher-dimensional matrix representations. At present we can only speculate that for linear representations of sufficiently high dimension, the pseudopotentials generated will provide some ultimate linearization of the original problem similar to that achieved for KdV. A final comment is to remark on the close connection of Lie groups and generalized Fourier analysis; the natural expression of the superposition rules intuitively felt to underlie the soliton phenomenon may well be found in the use of the invariant functions dual to the vector generators of the prolongation structure.

*This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. NAS7-100, sponsored by the National Aeronautics and Space Administration.

[1]R.M. Miura, C.S. Gardner, and M.D. Kruskal, J. Math. Phys. **9**, 1204 (1968).
[2]P. Lax, Comm. on Pure and Appl. Math. **21**, 467 (1968).
[3]A.C. Scott, F.Y.F. Chu, and D.W. McLaughlin, Proc. IEEE **61**, 1443 (1973).
[4]C.S. Gardner, J.M. Green, M.D. Kruskal, and R.M. Miura, Phys. Rev. Lett. **19**, 1095 (1967); N.J. Zabusky, Phys. Rev. **168**, 124 (1968).
[5]G.L. Lamb, Jr., Rev. Mod. Phys. **43**, 99 (1971).
[6]F.B. Estabrook and H.D. Wahlquist, to be published.
[7]L.P. Eisenhart, *A Treatise on the Differential Geometry of Curves and Surfaces* (Dover, New York, 1960), pp. 280—90; A.R. Forsythe, *Theory of Differential Equations* (Dover, New York, 1959), Vol. IV, Chap. XXI, pp.. 425—55.
[8]M.J. Ablowitz, D.J. Kaup, A.C. Newell, and H. Segur, Phys. Rev. Lett. **31**, 125 (1973).
[9]H. Flanders, *Differential Forms* (Academic, New York, 1963), pp. 1—73; Y. Choquet-Bruhat, *Géométrie différentielle et systèmes extérieurs* (Dunod, Paris, 1968), pp. 54—84.
[10]E. Cartan, *Les systèmes différentielles extérieurs et leurs applications géométrique* (Hermann, Paris, 1945); W. Slebodzinski, *Exterior Forms and Applications* (Polish Scientific Publishers, Warsaw, 1970); B.K. Harrison and F.B. Estabrook, J. Math. Phys. **12**, 653 (1971).
[11]R.M. Miura, J. Math. Phys. **9**, 1202 (1968).
[12]H.D. Wahlquist and F.B. Estabrook, Phys. Rev. Lett. **31**, 1386 (1973).

# Singularities for fluids with $p = \omega$ equation of state

P. Letelier

*Instituto de Física, Universidad Católica de Chile, Santiago, Chile*

R. Tabensky

*Departamento de Física, Universidad de Chile, Santiago, Chile*
(Received 15 July 1974)

The structure of singularities is discussed for some exact solutions of Einstein equations for irrotational perfect fluids with equation of state pressure equal to energy density, $p = \omega$. It is found that all singularities studied are velocity-dominated of the semi-Kasner class. It is also found that data on the singularity are not enough to generate space–time for all times.

## 1. INTRODUCTION

The concept of a velocity-dominated singularity in irrotational hydrodynamics models[1] is used to study the singularities of three nonhomogeneous exact solutions of Einstein equations for irrotational perfect fluids with pressure $p$ equal to rest energy $\omega$.

The solutions studied have singularities with more complicated structure than other known exact solutions. The $P$ symbols defined by Liang[1] depend on one or two arbitrary functions of one variable.

The first solution studied (Sec. 2) is the general solution for plane symmetric fluids.[2] It is found that the solution has a velocity-dominated singularity of the semi-Kasner class. In Secs. 3 and 4 the singularities of two special classes of cylindrical symmetric solutions[3,4] are studied. It is found that the singularities are of the semi-Kasner class, too. In Sec. 5 some remarks on the spherical case are made.

The fact that the solutions have arbitrary functions that are completely wiped out near the "big bang" is indicated.

## 2. THE PLANE SYMMETRIC SINGULARITY

In this section we study the structure of the singularities of irrotational plane symmetric perfect fluids with $p = \omega$ equation of state. The general solution of Einstein equations for these fluids is known.[2] The result is summarized in the following set of formulas: Solve the linear equation

$$\sigma_{zz} = \sigma_{tt} + t^{-1}\sigma_t \tag{1}$$

to get

$$ds^2 = t^{-1/2}e^{\Omega}(dt^2 - dz^2) - t(dx^2 + dy^2) \tag{2a}$$

from

$$\Omega = \int t[(\sigma_t^2 + \sigma_z^2)\,dt + 2\sigma_t\sigma_z\,dz] \tag{2b}$$

and

$$p = \omega = \tfrac{1}{2}t^{1/2}(\sigma_t^2 - \sigma_z^2)e^{-\Omega}; \tag{3}$$

any solution of (1) generates the solution (2) of Einstein equations, with pressure given by (3).

Comoving coordinates are $\sigma$, $Z$, $x$, and $y$. $\sigma$ is the comoving time, $Z$ is a comoving spatial coordinate defined by

$$dZ = t(\sigma_z dt + \sigma_t dz). \tag{4}$$

In these coordinates the metric (2) reads

$$ds^2 = t^{-1/2}e^{\Omega}(\sigma_t^2 - \sigma_z^2)^{-1}(d\sigma^2 - t^{-2}dZ^2) - t(dx^2 + dy^2). \tag{5}$$

We shall study the behavior of the general solution near the singularity $t = 0$.

The formal Fourier transform of equation (1), in the variable $z$, leads to the solution

$$\sigma(t, z) = \int_0^{\pi} F(z + t\cos u)\,du + \int_0^{\infty} E(z + t\cosh u)\,du. \tag{6}$$

It is easily checked that (6) solves (1) for arbitrary functions $F$ and $E$ whenever the integrals exist and differentiation is allowed under the integral sign.

Unfortunately, there are some solutions that cannot be expressed as (6). An example is $\sigma = \ln t$. It is likely that all solutions can be obtained from (6) by some limiting procedure. An example is

$$\ln t = \lim_{\lambda \to 0}\sigma_\lambda(t, z)$$

where

$$\sigma_\lambda(t, z) = \pi^{-1}\int_0^{\pi}\ln\lambda/2\,du - \int_0^{\infty}\exp[-\lambda(z + t\cosh u)]\,du.$$

The behavior of $\sigma$ near the singularity can be found from (6) noting that[5]

$$\lim_{t\to 0} t\sigma_t = -E(z), \tag{7}$$

and therefore

$$\sigma \simeq -E(z)\ln t \quad \text{as } t \to 0 \tag{8}$$

The arbitrary $F(z)$ is completely dominated by $E(z)$ near $t = 0$. $F(z)$ plays a role at $t = 0$ only if $E(z) = 0$. If that happens the singularity is no longer spacelike. An example can be found in Ref. (6). Now the complete solution near $t = 0$ in comoving coordinates can be obtained. Equations (8) and (4) give

$$dZ \simeq -E(z)\,dz. \tag{9}$$

Thus $z$ is also a comoving coordinate near $t = 0$. Equations (8) and (2.b) give us

$$\Omega \simeq -E(z)\sigma. \tag{10}$$

From (10), (9), and (3) the metric and the pressure near the singularity are

$$ds^2 \simeq E^{-2}\exp[-(E + \tfrac{3}{2}E^{-1})\sigma]\,d\sigma^2 - \exp[-(E - \tfrac{1}{2}E^{-1})\sigma]\,dz^2 - \exp(-\sigma/E)(dx^2 + dy^2), \tag{11}$$

$$p = \omega \simeq \tfrac{1}{2}E^2 \exp[(E + \tfrac{3}{2}E^{-1})\sigma]. \tag{12}$$

The metric (11) shows that the singularity is velocity dominated, the $P$ symbol given by

$$P = \left(\frac{2E^2 - 1}{2E^2 + 3};\ \frac{2}{2E^2 + 3};\ \frac{2}{2E^2 + 3}\right) \tag{13}$$

as defined by Liang.[1] This class of singularities where $\Sigma P_i = 1$ and $\Sigma P_i^2 \neq 1$ are called semi-Kasner singularities.

## 3. CYLINDRICAL SYMMETRIC SINGULARITIES. I

In this section we study the singularities of a class of irrotational cylindrical symmetric perfect fluids with $p = \omega$ equation of state. A general solution of Einstein equation for fluids in this condition is unknown, but there are two particular classes of known solutions.[3,4] In this section we shall study the first class[3] and in the next section the second.[4] The main relations for the first class are

$$\sigma_{\rho\rho} + \rho^{-1}\sigma_\rho = \sigma_{tt} + t^{-1}\sigma_t, \tag{14}$$

$$ds^2 = \pm t^{-1/2}(t^2 - \rho^2)^{3/4} e^{\Omega}(dt^2 - d\rho^2) - t(\rho^2 d\theta^2 + dz^2), \tag{15a}$$

$$\Omega = \int \frac{\rho t}{t^2 - \rho^2}\{[t(\sigma_\rho^2 + \sigma_t^2) - 2\rho\sigma_\rho\sigma_t]d\rho + [2t\,\sigma_\rho\sigma_t - \rho(\sigma_\rho^2 + \sigma_t^2)]dt\} + A, \tag{15b}$$

$$p = \omega = \pm t^{1/2}(t^2 - \rho^2)^{-3/2} e^{-\Omega}(\sigma_t^2 - \sigma_\rho^2) \tag{16}$$

where $\sigma$, a solution of (14), generates the solution (15) with pressure (16). $A$ is an arbitrary constant. Comoving coordinates are $\sigma$, $R$, $\theta$, and $z$. $\sigma$ is the comoving time, $R$ is the comoving radial coordinate defined by

$$dR = t\rho(\sigma_\rho dt + \sigma_t d\rho). \tag{17}$$

In comoving coordinates the metric (15) is given by

$$ds^2 = \pm t^{-1/2}(t^2 - \rho^2)^{3/4}(\sigma_t^2 - \sigma_\rho^2)^{-1} e^{\Omega}[d\sigma^2 - (\rho t)^{-2} dR^2] - t(\rho^2 d\theta^2 + dz^2). \tag{18}$$

The solution to (14) is now

$$\sigma(\rho, t) = \int_0^\pi \int_0^\pi F(t\cos u + \rho\cos v)\,du\,dv + \int_0^\infty \int_0^\infty \times G(t\cosh u + \rho\cosh v)\,du\,dv. \tag{19}$$

[The remarks made on (6) are also valid for (19)]

Space–time (15) has two singularities, $t = 0$ and $\rho = 0$, the first is a spacelike singularity in the plus sign solution and the second is spacelike too in the minus sign solution.

*The plus sign solution*: Confronting (6) and (19), it follows that

$$\lim_{t\to 0} t\sigma_t = -\int_0^\infty G(\rho\cosh v)\,dv = -E(\rho). \tag{20}$$

Thus

$$\sigma \simeq -E(\rho)\ln t \quad \text{as } t \to 0. \tag{21}$$

Equation (17) near the singularity reads

$$dR \simeq -\rho E(\rho)\,d\rho. \tag{22}$$

Therefore $\rho$ is comoving too near $t = 0$.

Now the metric and the pressure in comoving coordinates can be found as above. They are

$$ds^2 \simeq \rho^{3/2}E^{-2}\exp[-(E + \tfrac{3}{2}E^{-1})\sigma]\,d\sigma^2 - \rho^{3/2}\exp[-(E - \tfrac{1}{2}E^{-1})\sigma] \times d\rho^2 - \exp(-\sigma/E)(\rho^2 d\theta^2 + dz^2), \tag{23}$$

$$p = \omega \simeq \tfrac{1}{2}E^2\rho^{-3/2}\exp[(E + \tfrac{3}{2}E^{-1})\sigma]. \tag{24}$$

In this solution $A$ has been chosen to make the metric and the pressure real. The metric (23) tells us that the singularity is velocity-dominated of the same class as the former solution with the same $P$ symbol.

*The minus sign solution*: Now $\rho$ is a temporal coordinate and $t$ a radial coordinate. Near the spacelike singularity $\rho = 0$ we have

$$\sigma = -E(t)\ln\rho, \tag{25}$$

$$dR \simeq -tE(t)\,dt, \tag{26}$$

$$\Omega \simeq E^2 \ln\rho \; (A = 0 \text{ without lost of generality}^3). \tag{27}$$

The metric and the pressure near $\rho = 0$ are

$$ds^2 \simeq tE^{-2}\exp[-(E + 2E^{-1})\sigma]\,d\sigma^2 - \exp[-E\sigma]\,dt^2 - t\exp(-2\sigma/E)\,d\theta^2 - t\,dz^2, \tag{28}$$

$$p = \omega \simeq \tfrac{1}{2}E^2 t^{-1}\exp[(E + 2E^{-1})\sigma]. \tag{29}$$

The **P** symbol in this case is

$$\mathbf{P} = \left[\frac{E^2}{E^2 + 2};\ \frac{2}{E^2 + 2};\ 0\right]. \tag{30}$$

Hence we have that the singularity is velocity-dominated of the semi-Kasner class too.

## 4. CYLINDRICAL SYMMETRIC SINGULARITIES. II

The other class of known solutions is given by

$$\sigma_{tt} = \sigma_{\rho\rho} + \rho^{-1}\sigma_\rho, \quad \lambda_{tt} = \lambda_{\rho\rho} + \rho^{-1}\lambda_\rho, \tag{31}$$

$$ds^2 = \pm\exp[2(\nu - \lambda)](dt^2 - d\rho^2) - \rho^2\exp(-2\lambda)\,d\theta^2 - \exp(2\lambda)\,dz^2, \tag{32a}$$

$$\nu = \int \rho[\sigma_t^2 + \sigma_\rho^2 + \lambda_t^2 + \lambda_\rho^2)d\rho + 2(\sigma_t\sigma_\rho + \lambda_t\lambda_\rho)\,dt], \tag{32b}$$

$$p = \omega = \pm(\sigma_t^2 - \sigma_\rho^2)\exp[-2(\nu - \lambda)]. \tag{33}$$

Where a pair of solutions of (31) generates the solutions (32) with pressure (33). Comoving coordinates are $\sigma$, $R$, $\theta$, and $z$. $R$ is given by

$$dR = \rho(\sigma_t d\rho + \sigma_\rho dt). \tag{34}$$

Near the singularity $\rho = 0$ we have from (6) and (7) that

$$\sigma \simeq -E(t)\ln\rho. \quad \lambda \simeq -E(t)\ln\rho. \tag{35}$$

Therefore,

$$dR \simeq -E(t)\,dt, \tag{36}$$

$$\nu \simeq (E^2 + \epsilon^2)\ln\rho. \tag{37}$$

To keep the pressure positive we must choose the minus sign in (32). Thus $\rho = 0$ is a spacelike singularity. The metric and the pressure in comoving coordinates are

$$ds^2 \simeq E^{-2}\exp[-2(E^2+\epsilon^2+\epsilon+1)\sigma/E]\,d\sigma^2 - \exp[-2(E^2+\epsilon^2+\epsilon)\sigma/E]\,dt^2 - \exp[-2(\epsilon+1)\sigma/E]\,d\theta^2 - \exp(2\epsilon\sigma/E)\,dz^2, \tag{38}$$

$$p=\omega \simeq E^2\exp[2(E^2+\epsilon^2+\epsilon+1)\sigma/E]. \tag{39}$$

And the **P** symbol is

$$\mathbf{P}=\left[\frac{E^2+\epsilon^2+\epsilon}{E^2+\epsilon^2+\epsilon+1};\ \frac{\epsilon+1}{E^2+\epsilon^2+\epsilon+1};\ \frac{-\epsilon}{E^2+\epsilon^2+\epsilon+1}\right]. \tag{40}$$

Hence the singularity is of the semi-Kasner class again.

## 5. SPHERICAL SYMMETRIC SINGULARITIES

In this section we make only some remarks on the spherical case. The general solution of Einstein equations for irrotational perfect spherical symmetric fluids with $p=\omega$ equation of state is unknown. However, two classes of solutions are known,[6] each class depending only on one parameter. Depending on the value of the parameter we can have solutions without a big bang, with "naked singularities" or velocity-dominated semi-Kasner singularities.


[1]E. P. T. Liang, J. Math. Phys. 13, 386 (1972), and references therein.
[2]R. Tabensky and A. H. Taub, Commun. Math. Phys. 29, 61 (1973).
[3]P. Letelier and R. Tabensky (unpublished).
[4]P. Letelier (unpublished).
[5]H. Lamb, *Hydrodynamics* (Dover, New York, 1932), p. 298.
[6]R. Tabensky, Ph.D. thesis, University of California (Berkeley) (1972).

# Analytic vector harmonic expansions on $SU(2)$ and $S^2$

Brian Lee Beers*

*Department of Physics and Astronomy, Southern Illinois University at Carbondale, Carbondale, Illinois 62901*

Richard S. Millman

*Department of Mathematics, Southern Illinois Univesity at Carbondale, Carbondale, Illinois 62901*

(Received 25 February 1974)

Irreducible vector fields (vector harmonics) are introduced on $SU(2)$. It is shown that an arbitrary vector field can be expanded in terms of these vector harmonics, and tight convergence conditions are derived for analytic vector fields. These expansions are related to well-known vector harmonic expansions on the two-sphere. The generalization to arbitrary tensor fields is discussed. A connection between the Lie algebra of vector fields on $SU(2)$ and the Virasoro algebra is noted.

## 1. INTRODUCTION

A topic of considerable interest in mathematical physics is the expansion of functions on groups and homogeneous spaces in terms of a complete set of basis functions on the manifold. These basis functions normally arise as basis vectors for irreducible representations of the group.[1] In general, these expansions are useful from a practical point of view only if the convergence of the expansion is very rapid, i.e., only if a few terms in the series need be retained. Rapid convergence is not a common property for arbitrary classes of functions. However, if the family of functions is confined to the analytic class, then, as is the case for functions of a real variable, it may be possible to show that rapid convergence prevails.

This program has been carried out for $SU(2)$,[2] $S^n$,[3] and $SU(3)$,[4] and the theorems obtained can be generalized to arbitrary compact semisimple Lie groups and their homogeneous spaces. These theorems are quite strong and greatly restrict the possible behavior of the expansion coefficients. These theorems also indicate the kinds of difficulties that may be encountered if one attempts to expand a nonanalytic function. The poor convergence of certain Dalitz plot expansions,[5] e.g., may be understood in terms of singularities on the phase-space manifold.[6]

A related tool has been the expansion of vector-valued functions in terms of vector harmonics.[7] Most important here is the multipole expansion of the radiation field,[8] the rapid convergence in this case being the reason for its ubiquitous nature. Since in the radiation zone, it is primarily the angular distribution which is being expanded, we have in a natural way the expansion of a vector field defined over the sphere $S^2$. In a similar fashion, the partial wave expansions of helicity amplitudes[9] are tensor expansions on $SU(2)$.

It is the purpose of this paper to put the convergence properties of such expansions on a firm foundation. The emphasis in this paper is on the strong theorems encountered when the fields are analytic. In order to carry out this program it is convenient to use the abstract language of manifold theory. This serves two purposes. First, it makes certain that all definitions and theorems are natural, i.e., independent of the coordinate patch chosen, a condition that is essential. One need only refer to the example cited in I where it is shown that the function $\sin\theta$ is not analytic on the two-sphere $S^2$, whereas $\cos\theta$ is, a fact that would not have been ascertained by remaining in the $(\theta, \varphi)$ coordinate patch. The second reason for working with the abstract language is that the generalization to other Lie groups and homogeneous spaces is most apparent in this language. Thus, although the answers to questions may vary as we change the category of group (simply connected to non-simply connected; compact to noncompact; etc.), the questions which we wish to ask remain relatively clear.

For two reasons, we will give (whenever possible) the description of the objects discussed in the conventional coordinate patches [i.e., the Euler angles on $SU(2)$ and polar coordinates on $S^2$] as well as the description of the objects in the language of manifolds. First, for a large number of physical applications it is ultimately necessary to choose some coordinates (e.g., the angular spectrum of the radiation field). Second, this article will have served no useful purpose if it does not make contact with series most commonly used in physical applications. This necessitates the display of the objects in their customary notational form.

Vector fields and their general properties are discussed in Sec. 2A. Irreducible vector fields, analytic continuation, and an invariant inner product are introduced in the remainder of Sec. 2. Sections 3 and 4 discuss the convergence properties of harmonic expansions of vector fields on $SU(2)$. Section 5 gives a construction which relates the customary vector spherical harmonics[7] to the above irreducible vector fields. Section 6 indicates the same type of results for arbitrary tensor fields. Finally, in Sec. 7, the full Lie algebra of vector fields on $SU(2)$ is discussed in connection with the Virasoro[10] algebra.

## 2. VECTOR FIELDS ON $SU(2)$

### A. Background

In order to avoid repetition, we will refer the reader to I for background material on real and complex analytic manifolds. The notation used here will conform to that in I, and the reader will be referred to specific formulas given there. The geometric material discussed here can be found in any modern text on differential geometry (e.g., Kobayashi and Nomizu[11]) or to a recent expository article in the *American Mathematical Monthly*.[12] We will mainly follow Helgason[13] because of the close relation between the topics discussed there and the theory of Lie groups. To avoid confusion and repetition, we will use the word "analytic" to mean real analytic, and "holomorphic" to mean complex analytic in the sense discussed in I.

The manifold under discussion will be the group $G$ of $2\times 2$ unitary unimodular matrices, $SU(2)$. Endowed with the unique analytic structure in which the group action is analytic, $SU(2)$ is analytically diffeomorphic to the sphere in four dimensions $S^3$ endowed with the usual analytic structure [Eqs. (2.8)–(2.10) of I]. Let $CF(G)$ be the family of complex-valued analytic functions on $G$. It is the space of real-valued analytic functions on $G$, which is natural in the sense that it is uniquely determined by the analytic structure on $G$, and in turn uniquely determines the analytic structure on $G$.[14] We may consider complexifying these real functions to pairs of real functions $f_1+if_2$ without complication. It is necessary to use this complex space for decomposing the left regular representation (defined below) into unitary irreducible representations. The introduction of unitary representations immediately necessitates the use of complex vector spaces. That unitary representations are used is related to the well-known fact that finite-dimensional representations of compact Lie groups are completely reducible over the complex field to a direct sum of unitary irreducible representations. This does not hold over the real field. We give first the abstract definition of a vector field on $G$, as this is the form we shall use.

*Definition* 2.1: A (complex) vector field $X$ on $G$ is a derivation on $CF(G)$, i.e., $X$ is a map $X: CF(G)\rightarrow CF(G)$ such that

$$X(\alpha f+\beta g)=\alpha X(f)+\beta X(g) \tag{2.1}$$

and

$$X(fg)=fX(g)+X(f)g, \tag{2.2}$$

where $\alpha$ and $\beta$ are complex and $f, g\in CF(G)$ (Helgason,[13] p. 9).

This definition merely formalizes our ideas about first order differential operators. Indeed, in a coordinate patch $\mathbf{y}=(y_1,\ldots,y_n)$,[15] $X$ has the form

$$X(f)=\left(\sum_i X^i\frac{\partial}{\partial y_i}\right)(f), \tag{2.3}$$

where the $X^i$'s are complex functions on the patch called the components of the (covariant) vector field, with respect to the patch $\mathbf{y}=(y_1,\ldots,y_n)$. Note also that we are allowing complex vector fields $X_1+iX_2$, where $X_{1,2}$ are real, because of our use of the complex function space in the definition. It, of course, easily follows that if $X$ is a vector field, then so is $fX$, for all $f\in CF(G)$.

Let us relate the above definition to the idea of the tangent space at a point $u$ of $G$. Let $X_u$ denote the linear mapping of the space of analytic functions at a point $u\in G$, defined by

$$X_u: f\rightarrow (Xf)(u). \tag{2.4}$$

From the form of (2.3), it is clear that this set of objects is spanned by the linearly independent operators $\hat{\mathbf{e}}_i$ given by

$$\hat{\mathbf{e}}_i: f\rightarrow \frac{\partial f}{\partial y_i}(\mathbf{y}^\circ) \tag{2.5}$$

where $y^\circ$ is the image of $u$. Denote the collection of all $X_u$ by $CT_u(G)$. From (2.5) we see that a general $X_u$ has the form

$$X_u=\sum_i X^i(u)\,\hat{\mathbf{e}}_i, \tag{2.6}$$

which manifestly displays $CT_u$ as a finite-dimensional vector space of dimension equal to the dimension of the manifold. The $X^i(u)$ are the components of $X_u$ with respect to the basis $\hat{\mathbf{e}}_i$. Intuitively, then, we may think of a vector field as a vector whose components vary as we change the point $u$. The only way we will get into trouble in this thinking is if we fail to remember that (2.5) and (2.6) are only local statements, since the patch does not in general cover all of $G$.

Recall that a function $f$ is said to be of class $C^n$ if derivatives of all orders up to and including $n$ exist and are continuous. We now define this concept for vector fields.

*Definition* 2.2: A vector field $X$ on $G$ is said to be $C^n$ at a point $u\in G$ if $Xf$ is of class $C^n$ at $u$, for all $f\in CF(G)$. $X$ is said to be of class $C^n$ if it is $C^n$ at all points $u\in G$. The concepts of $X$ being analytic at a point and analytic are similarly defined. From Eq. (2.3) it should be clear that the $C^n$ definitions are equivalent to the fact that the $X^i$ are of class $C^n$ in the usual Euclidean sense. Also, if $X$ is of class $C^n$, then so is $(fX)$ for all $f\in CF(G)$. Because we will be working primarily with analytic vector fields, we shall use $CT(G)$ to denote the set of all analytic vector fields on $G$.

$CT(G)$ has more structure. Clearly $CT(G)$ is a vector space over the complex numbers, with the obvious definition

$$(\alpha X+\beta Y)f=\alpha(Xf)+\beta(Yf), \tag{2.7}$$

where $\alpha,\beta$ are complex and $X, Y\in CT(G)$. In a similar fashion $CT(G)$ is a module[16] over the ring $CF(G)$. In an intuitive fashion, we also expect the commutator of two first order differential operators to be a first order differential operator. A simple application of definition (2.1) shows this to be true, so that we define a bracket operation [ , ]

$$[X,Y]\equiv XY-YX \tag{2.8}$$

for all $X, Y\in CT(G)$. It is easily checked that the usual Jacobi identity is satisfied, so that $CT(G)$ endowed with the bracket operation of Eq. (2.8) becomes an infinite-dimensional Lie algebra.

For a Lie group, a preferred position is accorded to the Lie subalgebra of $CT(G)$ which is invariant under the left group action. This subalgebra is normally called *the* Lie algebra, $\mathcal{A}$, of the Lie group.

To see how $\mathcal{A}$ arises, consider the process of left translation, $L_{u_1}: G\rightarrow G$, by an element $u_1$ given by

$$L_{u_1}: u\rightarrow u'=u_1u. \tag{2.9}$$

$L_{u_1}$ is an analytic diffeomorphism of $G$ onto itself for each $u_1 \in G$. It induces in a natural way a mapping $U_{u_1}: CF(G) \to CF(G)$ which is defined by

$U_{u_1}: f \to f^{u_1}$:

$$f^{u_1}(u_1 u) = f(u) \quad \text{for all } u \in G \tag{2.10}$$

or

$$f^{u_1}(u) = f(u_1^{-1} u) \tag{2.11}$$

or

$$f^{u_1} = f \circ L_{u_1^{-1}}, \tag{2.12}$$

where $\circ$ denotes composition of mappings. The set of operators $U_{u_1}$ clearly acts in a linear fashion on $CF(G)$, and in fact obeys the group rule

$$U_{u_1} U_{u_2} = U_{u_1 u_2} \tag{2.13}$$

so that it in fact forms a representation of $G$ called the *left regular* representation on $G$.

In a similar fashion $L_{u_1}$ induces in a natural way a mapping $dL_{u_1}$ of $CT(G)$ onto itself called the differential of $L_{u_1}$. We define it by the rule (Helgason, p. 22) $dL_{u_1}: X \to X^{u_1}$, where

$$(X^{u_1} f^{u_1})(u_1 u) = (Xf)(u) \tag{2.14}$$

for all $f \in CF(G)$ and $u \in G$ or in one of the following forms:

$$X^{u_1} f^{u_1} = (Xf)^{u_1}, \tag{2.15}$$

$$X^{u_1} f = (X f^{u_1^{-1}})^{u_1}. \tag{2.16}$$

The last equation is the most convenient form. From its definition (2.14), it follows that $dL_{u_1}$ acts linearly on $CT(G)$, while a repeated application of (2.16) shows that the family of maps $dL_u$ where $u \in G$ (which we denote $dL_G$) obeys the group rule

$$dL_{u_1} \circ dL_{u_2} = dL_{u_1 u_2}. \tag{2.17}$$

$dL_G$ is a representation of $G$ which we will call the *left differential representation* of $G$. Because the maps $L_u$ are analytic, $dL_u$ preserves the analytic character of $X$. $dL_u$ also preserves the structure of $CT(G)$ as a module over $CF(G)$ and as a Lie algebra (Helgason, p. 24), i.e.,

$$(fX)^{u_1} = f^{u_1} X^{u_1} \tag{2.18}$$

and

$$[X^{u_1}, Y^{u_1}] = [X, Y]^{u_1}. \tag{2.19}$$

### B. Irreducible vector field on $SU(2)$

Because we have a representation $dL_G$ of $G$ on the linear space $CT(G)$, we may ask if there are any finite-dimensional subspaces of $CT(G)$ in which $dL_G$ acts irreducibly. In particular, are there any one-dimensional subspaces which are invariant? We therefore seek vector fields $X$ such that

$$X^u = X \quad \text{for all } u \in G. \tag{2.20}$$

That this set is nonempty is seen by the following construction. Let $X_e$ be a vector in the tangent space at the identity $CT_e(G)$. Define the vector field $X$ by left translation on $X_e$, i.e.,

$$(Xf)(u) = [X_e f^{u^{-1}}](e). \tag{2.21}$$

With this definition, it is evident that $X$ is in fact left invariant. Further, $X$ is an analytic vector field (Helgason, p. 89). Also, because left translation preserves the vector space structure and the algebraic structure, all $X$ so defined form a finite-dimensional subspace of $CT(G)$ which is closed under [ , ]. In fact, these are all of the invariant vector fields.

The unique association of an element in $CT_e(G)$ with a left invariant vector field permits a bracket operation to be defined on $CT_e(G)$ in the obvious manner. Under this operation $CT_e(G)$ becomes a Lie algebra normally called *the* Lie algebra of the group $\mathcal{A}$. It is apparent that the structure of $G$ uniquely determines $\mathcal{A}$ and the invariant vector fields. The converse is only true locally.

Let us make contact with the customary physical notation. Let $\gamma(t)$ be any path through the identity. This defines an element of $CT_e(G)$ by the rule

$$X_e f = \frac{d}{dt}\{f(\gamma(t))\}_{t=0} \tag{2.22}$$

so that we may speak of $\gamma(t)$ having tangent vector $X_e$. For matrix groups, the functions we consider are naturally regarded as functions of a matrix, so that $CT_e(G)$ may naturally be identified with the matrices obtained by the rule (Helgason, p. 100)

$$\frac{d}{dt}\{u(t)\}_{t=0} \quad \text{where } u(t) \in G, \tag{2.23}$$

where $u(t)$ is a path in the matrix group $G$. With this identification, the brakcet operation on $CT_e(G)$ becomes the ordinary matrix commutator. Because we are working with the matrix group $SU(2)$, we will make free use of this equivalence.

For practical applications, it is convenient to choose a basis for $\mathcal{A}$. We choose the basis set $\frac{1}{2}\sigma_j$, where the $\sigma_j$'s are the standard Pauli matrices [I, Eq. (2.3)]. With this choice, the structure constants for the Lie algebra are pure imaginary. The corresponding invariant vector fields (denoted $\mathcal{J}_i$) are defined by the rule

$$(\mathcal{J}_i f)(u) = (-i)\,\frac{d}{d\lambda}\{f(u\exp(i\lambda\sigma_i/2)\}_{\lambda=0}. \tag{2.24}$$

In terms of the Euler angle parameterization of $SU(2)$ [I, Eq. (2.25)], the $\mathcal{J}_i$'s are given by

$$\mathcal{J}_1 = i\left(-\frac{\cos\psi}{\sin\theta}\frac{\partial}{\partial\varphi} + \sin\psi\frac{\partial}{\partial\theta} + \cos\psi\cot\theta\frac{\partial}{\partial\psi}\right) \tag{2.25a}$$

$$\mathcal{J}_2 = i\left(\frac{\sin\psi}{\cos\theta}\frac{\partial}{\partial\varphi} + \cos\psi\frac{\partial}{\partial\theta} - \sin\psi\cot\theta\frac{\partial}{\partial\psi}\right) \tag{2.25b}$$

$$\mathcal{J}_3 = i\frac{\partial}{\partial\psi}. \tag{2.25c}$$

The $\mathcal{J}_i$'s are a basis for the invariant vector fields on $SU(2)$. Each $\mathcal{J}_i$ spans a one-dimensional vector space which is invariant under left translation, being the carrier space for an identity representation of $SU(2)$.

Other irreducible subspaces of $CT(G)$ are easy to find. Let $D^j_{mm'}(u)$ be the matrix elements of the irreducible representations of $SU(2)$ [I, Eq. (3.8)]. If $m'$ is fixed between $-j$ and $j$, then the set of functions $\{D^j_{mm'}, -j \le m \le j\}$ is a basis for a $(2j+1)$-dimensional unitary irreducible representation of $SU(2)$ under left translation, i.e.,

$$[D^j_{mm'}]^{u_1}[u] = D^j_{mm'}[u_1^{-1}u] = \sum_{m''} D^j_{mm''}(u_1^{-1})\, D^j_{m''m'}[u]. \tag{2.26}$$

By using Eq. (2.18) and the invariance of $\mathcal{G}_i$, it is clear how to define irreducible vector fields. Define $\mathcal{G}^{Ji}_{mm'}$ by the rule

$$\mathcal{G}^{Ji}_{mm'} = D^J_{mm'}\mathcal{G}^i. \tag{2.27}$$

If $m'$ and $i$ are fixed then the set of objects $\{\mathcal{G}^{Ji}_{mm'}, -J \le m \le J\}$ is a basis for a $(2J+1)$-dimensional irreducible representation of $SU(2)$ under left translation. Each representation occurs $3\circ(2J+1)$ times in this construction (the general formula is $N\cdot M$, where $N$ is the dimension of the group, and $M$ is the dimension of the representation).

The objects $\mathcal{G}^{Ji}_{mm'}$ are the irreducible vector fields on $SU(2)$ or, in customary language, *vector harmonics* on $SU(2)$. Each $\mathcal{G}^{Ji}_{mm'}$ is analytic on $SU(2)$ because both $\mathcal{G}^i$ and $D^J_{mm'}$ are (I, Sec. 3 B). It will be shown in Sec. 4 that this set of objects provides a canonical decomposition of the left differential representation $dL_G$ into irreducible subspaces, i.e., that any analytic vector field on $SU(2)$ can be written as a convergent series of vector harmonics.

### C. Holomorphic vector fields on $SL(2,C)$

As discussed in I, $SL(2,C)$ is a holomorphic manifold which may be viewed as the complexification of $SU(2)$. In particular, each analytic function on $SU(2)$ has a unique extension to a holomorphic function on an open subset of $SL(2,C)$ containining $SU(2)$.

Holomorphic vector fields on open subsets are defined in the obvious fashion [see def. (2.2)]. Because each real analytic function has a unique extension, we can locally extend a real analytic vector field on $SU(2)$ to a holomorphic vector field on $SL(2,C)$.

*Theorem* 1: Each of the irreducible vector fields $\mathcal{G}^J_{mm'}$ is an entire holomorphic vector field on $SL(2,C)$.

*Proof*: Because we have shown in I that the functions $D^J_{mm'}$ are entire holomorphic functions on $SL(2,C)$, it suffices to prove the theorem for $\mathcal{G}_i$ only. By referring to the definition given in Eq. (2.24) it is clear that the extension of $\mathcal{G}_i$ is given by

$$(\mathcal{G}_i f)(g) = -i\frac{d}{d\lambda}\{f[g\exp(i\lambda\sigma_i/2)]\}\Big|_{\lambda=0} \tag{2.28}$$

for all $g \in SL(2,C)$. By choosing any holomorphic patch $\mathbf{Z} = \{z_i\}$ around $g$, $\mathcal{G}_i$ takes the form

$$\mathcal{G}_i = \sum_j (-i)\frac{dz_j}{d\lambda}\Big|_{\lambda=0}\frac{\partial}{\partial z_j}. \tag{2.29}$$

But the group operation is holomorphic in $\lambda$, so $z_j$ is holomorphic in $\lambda$, and hence its derivative. This proves the theorem for $\mathcal{G}_i$.

### D. The invariant inner product on $CT(G)$

In order to exhibit the vector harmonics as an orthogonal basis of $CT(G)$, it is necessary to convert $CT(G)$ into an inner product space. We know that when we consider only real vector fields, the introduction of an inner product [as a map from $RT(G)\times RT(G) \to RF(G)$] is equivalent to giving a (Riemannian) metric tensor field on $G$. There is a unique bi-invariant Riemannian metric on a compact semisimple Lie group (which is a negative multiple of the Killing form) (Helgason, p. 191). Integrating the resulting inner product over $G$ gives an invariant inner product on $RT(G)$ [i.e., $RT(G)\times RT(G) \to R$]. We give below the complex analog of this construction.

Recall the definition of the adjoint representation Ad, of a real Lie algebra $\mathcal{A}$. With an element $X\in\mathcal{A}$, we associate the matrix Ad$X$ acting in the linear space $\mathcal{A}$ which is defined by

$$\mathrm{Ad}X(Y) = [X, Y]. \tag{2.30}$$

The Killing form is a symmetric bilinear form $B(X,Y)$ on $\mathcal{A}$ defined by

$$B(X,Y) = \mathrm{Tr}\{\mathrm{Ad}X \circ \mathrm{Ad}Y\}. \tag{2.31}$$

Because the tangent space at any point of $G$ is isomorphic (as a vector space and a Lie algebra) to $\mathcal{A}$, this form may be pointwise defined for any pair of vector fields on $G$. The negative of this yields a Riemannian structure $g$ on $G$, which is given by

$$g(X,Y)(u) = -\tfrac{1}{2}\mathrm{Tr}\{\mathrm{Ad}X(u)\,\mathrm{Ad}Y(u)\} \tag{2.32}$$

for all $X, Y \in RT(G)$ and $u \in G$. The factor of $\frac{1}{2}$ is for later convenience.

In order to handle complex vector fields, we make a simple extension of $g$ given by

$$g(X_1,X_2)(u) = \tfrac{1}{2}\mathrm{Tr}\{(\mathrm{Ad}X_1)^*(u)\,\mathrm{Ad}X_2(u)\}, \tag{2.33}$$

where $^*$ denotes the Hermitian conjugate matrix. (Note that the transposition disposes of the minus sign.) Integrating (2.33) puts an inner product ( , ) on $CT(G)$. Thus,

$$(X,Y) = \tfrac{1}{2}\int \mathrm{Tr}\{[\mathrm{Ad}X(u)]^*[\mathrm{Ad}Y(u)]\}\, d\Omega(u) \quad \text{for all } X, Y \in CT(G), \tag{2.34}$$

where $d\Omega(u)$ is the bi-invariant measure on $G$ (I, Sec. 2 C). One readily verifies that ( , ) obeys all the properties for a proper inner product.

An elementary calculation shows that

$$(\mathcal{G}_i, \mathcal{G}_j) = \delta_{ij} \tag{2.35}$$

while an application of the rule

$$\mathrm{Ad}(f(u)\,X(u)) = f(u)\,\mathrm{Ad}X(u) \tag{2.36}$$

together with the orthogonality of the $D^j$ functions (I, Eq. (5.2)] gives

$$(\mathcal{G}^{Ji}_{mm'}, \mathcal{G}^{J'j}_{nn'}) = (2J+1)^{-1}\delta_{JJ'}\delta_{ij}\delta_{mn}\delta_{m'n'}. \tag{2.37}$$

## 3. INFINITE SERIES OF VECTOR HARMONICS

In this section we consider the convergence properties of series of the form

$$X(g)=\sum_{i=1}^{3}\sum_{J=0}^{\infty}\sum_{m=-J}^{J}\sum_{m'=-J}^{J}(2J+1)\,a_{mm'}^{Ji}\mathcal{J}_{mm'}^{Ji}(u), \tag{3.1}$$

which, of course, may be rewritten

$$X(g)=\sum_{i=1}^{3}X^{i}(u)\mathcal{J}^{i}(u), \tag{3.2}$$

where

$$X^{i}(u)=\sum_{J=0}^{\infty}\sum_{m=-J}^{J}\sum_{m'=-J}^{J}a_{mm'}^{Ji}D_{mm'}^{J}(u). \tag{3.3}$$

We must first define in the obvious manner the meaning of convergence of a sequence of vector fields.

*Definition* 3.1: A sequence of vector fields $\{X_n\}$ is said to converge at a point $u\in G$ if the sequence $\{(X_n f)\}$ converges at $u$ for all $f\in CF(u)$.

Other convergence concepts are similarly defined.

If we now consider the convergence of partial sums formed from (3.1), it is clear from (3.2)–(3.3) that the factor $(\mathcal{J}_i f)$ is common to each partial sum, so that we are in fact really considering the convergence of partial sums for $X^i(u)$ defined in Eq. (3.3). But the convergence properties of such series are already known from I, so that the convergence of series of type (3.1) is an immediate consequence of I.

We summarize this succinctly.

*Definition* 3.2: For a series of the form (3.1), the *exponent of convergence* $\alpha_0$ is defined by the formula

$$\alpha_0=-\log[\limsup|a_{mm'}^{Ji}|^{1/J}] \tag{3.4}$$

where the limit superior is taken over all $i,J,m,m'$ indicated in Eq. (3.1).

*Theorem* 2: Let $S(g)$ be a series of the form (3.1) with exponent of convergence $\alpha_0$. Then the series converges absolutely and uniformly to a holomorphic vector field everywhere in the interior of the superball (I, Sec. 4), $B^6(\alpha_0)$. This domain $B^6(\alpha_0)$ is analytically complete, and is maximal, i.e., there is no larger superball inside of which $S(g)$ converges absolutely and uniformly everywhere.

As with functions on $SU(2)$, the divergence properties of $S(g)$ are more complicated than for Taylor's series.

## 4. EXPANSIONS OF ANALYTIC VECTOR FIELDS

Let $X$ be a vector field on $SU(2)$. Let $S(g)$ be a series of the form (3.1) with coefficients obtained by the rule

$$a_{mm'}^{Ji}=(\mathcal{J}_{mm'}^{Ji},X). \tag{4.1}$$

We ask the question, "In what sense and under what conditions does $S(g)$ represent $X$?"

The power of the present formulation of the problem becomes apparent here, for the answer to this question, as in the last section, is an immediate consequence of the results given in I. Consider a coordinate patch $\mathbf{Z}=(z_1,\ldots,z_n)$. In this patch, $\{\mathcal{J}_i\}$ provides a basis for the tangent space at each point in the patch. Hence, in the patch the field may be written

$$X=\sum_i X^i(\{z_j\})\mathcal{J}_i(z_j). \tag{4.2}$$

But this is true in every patch, so that $X$ may be written

$$X=\sum_i X^i\mathcal{J}_i, \tag{4.3}$$

where the $X^i$'s are elements of $CF(G)$ and are the components of $X$ with respect to the invariant basis $\mathcal{J}_i$. Thus, our question really amounts to "In what sense and under what circumstances does a Wigner expansion of $X^i$ represent $X^i$?" We formulate the answer for analytic vector fields only, and refer the reader to I for discussion of expansion of nonanalytic vector fields.

*Theorem* 3: Let $X$ be an analytic vector field on $SU(2)$ and $B^6(\alpha_0)$ the maximal superball of holomorphy of its unique holomorphic extension. Then the harmonic expansion for $X$ given by Eqs. (4.1) and (3.1) converges absolutely and uniformly and is holomorphic in $B^6(\alpha_0)$. The harmonic series converges to $X$. Conversely, if the harmonic expansion for $X$ has a superball of convergence $B^6(\alpha_0)$, then $X$ can be continued to a holomorphic vector field in $B^6(\alpha_0)$ and its continuation agrees with its harmonic expansion. The exponent of convergence (Def. 3.1) for the harmonic expansion is $\alpha_0$.

## 5. VECTOR HARMONIC EXPANSIONS ON $S^2$

Once the elaborate machinery was set up, the theory of vector harmonics on $SU(2)$ was quite simple. In going from $SU(2)$ to the ordinary sphere in three dimensions, $S^2$, the situation becomes somewhat more complicated. These complications occur for several reasons. The action of $SU(2)$ on $S^2$ is, of course, that of rotation in the ordinary sense, and $S^2$ is identified with the coset space $SU(2)/U(1)$ (I, Sec. 6B). One wishes to translate results on $SU(2)$ to $S^2$ via the natural mapping $\pi$ of $SU(2)$ onto $S^2$. Unfortunately, the image of an irreducible vector field on $SU(2)$ is not necessarily even a vector field on $S^2$, never mind irreducible on $S^2$. Further, the sphere is only two dimensional, so that the image of the operators $\mathcal{J}_i$ can no longer be linearly independent in the tangent space at a point. Finally, in most physical applications, it is also desirable to consider vectors which are normal to the sphere.

To introduce normal vectors to a manifold requires in general the theory of immersions.[11] For the present case, since $S^2$ is the image of $\pi$, the theory of submersions[17] may be used. However, the general exposition of this construction seems a bit far afield for the application at hand. Instead, we present an explicit mapping which obviates the necessity for the general theory. This mapping is from the three-dimensional tangent space at a point of $SU(2)$ to the three-dimensional space attached to a point of $S^2$ formed by the direct sum of the tangent space and the normal space. The construction depends explicitly on the metric properties of the manifolds and of ordinary Euclidean 3-space. In making this construction, we automatically solve the other problems as well and bring our notation into line with normal physical usage.

The manifold $M=S^2$ under consideration may be identified with the set of left cosets $SU(2)/U(1)$, where $U(1)$ is the subgroup of transformations $\exp(-\frac{1}{2}i\psi\sigma_3)$, where $\psi$ is the third Euler angle. The identification may be made concrete with the following definition (I, Sec. 6B):

$$x_j(u) = \tfrac{1}{2}\,\mathrm{Tr}(\sigma_j u \sigma_3 u^{-1}), \tag{5.1}$$

where the $x_j$'s are the Cartesian coordinates of a point in $E^3$, and $u \in G$. It is easily verified that both $u$ and $u\exp(-\frac{1}{2}i\psi\sigma_3)$ get mapped to the same point $\mathbf{x}$, and that $\sum x_j^2 = 1$, i.e., $\pi : SU(2) \to S^2$ such that $\pi(u) = \mathbf{x}$ as in (5.1).

At any point $p \in R^3$, the tangent vector space $CT_p(R^3)$ is three dimensional. In the Cartesian coordinate patch (which covers all of $R^3$) $CT_p(R^3)$ is spanned by $\mathbf{e}_i = \partial/\partial x_i$. This is a global choice of basis. $R^3$ is converted to Euclidean space $E^3$ by the introduction of the ordinary scalar product

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij} \quad \text{(globally)}. \tag{5.2}$$

$S^2$ as given by (5.1) is a submanifold of $E^3$ (Helgason, p. 23). The differential[18] of the inclusion maps $\iota$ of $M$ into $E^3$, $d\iota$, maps the tangent space $T_p(M)$ at a point $p \in M$ into the tangent space $CT_p(E^3)$. If we identify the image of $CT_p(M)$ with $CT_p(M)$, we may then speak of $CT_p(M)$ as being a subspace of $CT_p(E^3)$. A vector $X$ in $CT_p(E^3)$ is *normal* to $S^2$ if

$$X^* \circ Y = 0 \quad \text{for all } Y \in CT_p(M). \tag{5.3}$$

The vector space of all normal vectors at $p$ will be written $CN_p(M)$. $CN_p(M)$ is clearly one dimensional. A globally defined unit normal vector at $p = (x_1, x_2, x_3)$ is given by $\mathbf{e}_r = \sum_{i=1}^3 x_i \mathbf{e}_i$. It is now clear that $CT_p(R^3) = CT_p(M) \oplus CN_p(M)$ if $p \in S^2$. A vector field of $E^3$ restricted to $S^2$ is a differentiable assignment to each $p \in S^2$ a vector $X_p \in CT_p(E^3)$. Clearly every such vector field can be written as the sum of a tangential and normal vector field. We shall write a "vector field on $S^2$" to mean a vector field of $E^3$ restricted to $S^2$ and a vector field "tangent (normal) to $S^2$" to mean $X_p \in CT_p(M)$ $[X_p \in CN_p(M)]$ for all $p \in M$.

If $A : E^3 \to E^3$ is a linear transformation, then $dA = A$ in the sense that $dA(\mathbf{e}_i) = A\mathbf{e}_i$. If $A \in SO(3)$ and $\mathbf{x} \in S^2$, then $A(x)^* \circ A(x) = (A^*A)(x)^* \circ x = x^* \circ x = 1$ so that $A : S^2 \to S^2$. This means that if $Y$ is a vector field tangent to $S^2$, then $dA(Y)$ is also tangent to $S^2$. If $X_p \in CN_p(M)$ then for any $Y \in T_p(M)$,

$$0 = X_p^* \cdot Y_p = AX_p^* \cdot AY_p \tag{5.4}$$

so that $dA(X_p)$ is also normal to $S^2$. Thus the action of $SO(3)$ on $E^3$ induces an action of $SO(3)$ on $CT(M)$ and an action of $SO(3)$ on $CN(M)$. If $X \in CT(M)$ [or $X \in CN(M)$] and $A \in SO(3)$, we denote by $X^A$ the vector field $X^A = dA(X)$. Note that $(X^A)^* \cdot Y^A = dA(X)^* \circ dA(Y) = AX^* \circ AY = X^* \circ Y$ so that $(X^A)^* \cdot Y^A = X^* \cdot Y$. This says that the usual metric tensor of $E^3$ is invariant under the action of $SO(3)$ and so $A \to X^A$ is a unitary representation of $SO(3)$ on $CT(E^3)$. Let $\Phi$ be the two-to-one mapping of $SU(2)$ onto $SO(3)$ given by Eq. (2.6b) of I. $\Phi$ induces a representation $d\Phi : u \to d\Phi_u$ of both tangential and normal vector fields to $S^2$ given by $d\Phi_u X = X^{\Phi(u)}$ (which we abbreviate $X^u$). From the above it is clear that $X^u \circ Y^u = X \circ Y$ so that $d\Phi$ is a unitary representation. We will now search for subspaces of both $CT(M)$ and $CN(M)$ which are irreducible under $d\Phi$. In order to utilize the results of the previous sections we must find a way of pulling back vector fields from $S^2 = SU(2)/U(1)$ to vector fields on $SU(2)$. This does present a problem because vector fields behave covariantly, not contravariantly, so that (unlike differential forms) we cannot just pull them back. It is possible, however, to pull back vector fields if the map is a Riemanian submersion.[17] To circumvent this theory, we will present an ad hoc construction which assigns to a vector field $X$ on $S^2$ a vector field $\tilde{X}$ on $SU(2)$ in a natural way.

Let $\pi : G \to M$ and $V_u = \{X \in T_u G \mid d\pi(X) = 0\}$.[18] We say that $H_u = \{X \in T_u G \mid g(X, Y) = 0 \text{ for all } Y \in V_u\}$ is the horizontal subspace at $u$. If $P \in V_u$ (resp. $H_u$) then $P$ is called a *vertical* (resp. *horizontal*) vector. If $X \in T(G)$, then $X$ is called *vertical* (resp. *horizontal*) if $X_u$ is vertical (resp. horizontal) for all $u \in G$. We note explicitly that: (1) $\pi$ is onto; (2) $d\pi$ is onto at every point; (3) $d\pi$ is an isometry on horizontal vectors [i.e., $g(X, Y) = d\pi(X)^* \circ d\pi(Y)$ if $X$ and $Y$ are horizontal]. The third assertion follows from the general theory of submersions (Ref. 17, p. 446). It will also follow from some computations which will appear later. These are the three axioms for a Riemannian submersion. Note that $d\pi : H_u \to CT_{\pi(u)}(M)$ is an isomorphism. If $X$ is a vector field on $S^2$, we define $\tilde{X}$ by the equation

$$g(\tilde{X}, P) = X^* \cdot d\pi(P) \quad \text{for all } P \in CT(G). \tag{5.5}$$

*Theorem* 4: Let $\tilde{X}$ be the vector field defined as in Eq. (5.5).

(a) If $X$ is real analytic then $\tilde{X}$ is real analytic.

(b) If $X, Y$ are tangent to $S^2$ then $\tilde{X}$ is horizontal, $d\pi(\tilde{X}) = X$ and $g(\tilde{X}, \tilde{Y}) = X^* \cdot Y$.

(c) If $X$ is normal, then $\tilde{X} = 0$.

*Proof*: (a) It suffices to write $\tilde{X}$ in local coordinates. For convenience, we choose the Euler angles on $SU(2)$ [which we denote temporarily by $(\Phi', \theta', \Psi)$] and polar coordinates on $S^2$ (denoted $\theta, \Phi$). From the explicit realization of $\pi$ given by Eq. (5.1) [see also Eq. (6.29) of I] we have that $\pi(\Phi', \theta', \Psi) = (\theta, \Phi)$ such that $\theta = \theta', \Phi = \Phi'$. Thus, we have the result

$$d\pi\left(\frac{\partial}{\partial \Phi'}\right) = \frac{\partial}{\partial \Phi}, \tag{5.6a}$$

$$d\pi\left(\frac{\partial}{\partial \theta'}\right) = \frac{\partial}{\partial \theta}, \tag{5.6b}$$

$$d\pi\left(\frac{\partial}{\partial \Psi}\right) = 0. \tag{5.6c}$$

Decomposing $X$ into tangential and normal components, we may write $X = X^\Phi \partial/\partial\Phi + X^\theta \partial/\partial\theta + X^r \mathbf{e}_r$, while $\tilde{X}$ may be written $\tilde{X} = \tilde{X}^\Phi \partial/\partial\Phi + \tilde{X}^\theta \partial/\partial\theta + \tilde{X}^\Psi \partial/\partial\Psi$, where primes have now been dropped. Choosing $P$ in Eq. (5.5) to be $\partial/\partial\Phi$, $\partial/\partial\theta$, $\partial/\partial\Psi$, respectively, using the metric components $g(\partial/\partial\Phi, \partial/\partial\Psi) = \cos\theta$, $(\partial/\partial\Phi) \cdot (\partial\Phi) = \sin^2\theta$, etc., applying Eqs. (5.6), and complex conjugating, we obtain the three equations

$$\tilde{X}^\Phi + \tilde{X}^\Psi \cos\theta = \sin^2\theta\, X^\Phi, \tag{5.7a}$$

$$\tilde{X}^\theta = X^\theta, \tag{5.7b}$$

$$\tilde{X}^\Phi \cos\theta + \tilde{X}^\Psi = 0. \tag{5.7c}$$

These have the solution

$$\tilde{X}^\Phi = X^\Phi, \tag{5.8a}$$

$$\tilde{X}^\theta = X^\theta, \tag{5.8b}$$

$$\tilde{X}^\Psi = -\cos\theta X^\Phi. \tag{5.8c}$$

Thus, $\tilde{X}$ is manifestly analytic in this patch if $X$ is analytic. Similar results prevail in other patches, proving the assertion.

(b) If $P$ is any vertical vector $[d\pi(P)=0]$, then $g(\tilde{X}, P) = X^* \cdot d\pi(P) = 0$ so that $\tilde{X}$ is horizontal. In the Euler angle patch we have, using the metric components again,

$$g(\tilde{X}, \tilde{Y}) = \bar{\tilde{X}}^\Phi \tilde{Y}^\Phi + \bar{\tilde{X}}^\theta \tilde{Y}^\theta + \bar{\tilde{X}}^\Psi \tilde{Y}^\Psi + (\bar{\tilde{X}}^\Phi \tilde{Y}^\Psi + \bar{\tilde{X}}^\Psi \tilde{Y}^\Phi)\cos\theta. \tag{5.9}$$

By using Eqs. (5.8) this becomes

$$g(\tilde{X}, \tilde{Y}) = \bar{X}^\Phi Y^\Phi \sin^2\theta + \bar{X}^\theta Y^\theta \tag{5.10}$$

which is just $X^* \cdot Y$ when $X^r = Y^r = 0$. This demonstrates the second assertion.

(c) Finally, it follows immediately from Eqs. (5.8) that $\tilde{X} = 0$ if $X^\theta = X^\Phi = 0$, which is the condition that $X$ is vertical.

We would like to extend the map $X \to \tilde{X}$ to be a vector space isomorphism $CT_{\pi(u)}(E^3) \to CT_u(G)$. We do this by defining $X \to \tilde{X}$ if $X$ is tangential, and $\mathbf{e}_r \to \tilde{\mathbf{e}}_r$ where we have picked some global unit vector field $\tilde{\mathbf{e}}_r$ on $G$ for once and for all such that $\tilde{\mathbf{e}}_r$ is vertical. Inspection of Eqs. (5.6c), (2.25c), and (2.35) tells us that $\tilde{\mathbf{e}}_r$ may be chosen to be $\mathcal{J}_3$. Thus, we let $\mathbf{e}_r \to \mathcal{J}_3$, and extend this map by linearity to other normal fields. If $\tilde{X}$ is any vector field on $SU(2)$, we may pointwise decompose $\tilde{X}$ into its horizontal $\tilde{X}^h$ and vertical parts $\tilde{X}^v$. If both $\tilde{X}^v$ and $\tilde{X}^h$ are projectable onto a vector field $X^v$ and $X^h$ which are normal to $S^2$ and tangent to $S^2$ respectvely, then we define $\widetilde{d\pi}(\tilde{X})$ to be $X^v + X^h$. Note that $\widetilde{d\pi}$ is not defined on all vector fields of $SU(2)$. The following is immediate from the construction of $\widetilde{d\pi}$ and Theorem 4.

*Proposition* 1: $g(\tilde{X}, \tilde{Y}) = X^* \cdot Y$ for all $X$ and $Y$ which are vector fields on $S^2$ .

We now explicitly give the irreducible vector fields $X$ on $S^2$ and their pullbacks $\tilde{X}$. Clearly the action of $u \in G$ on vector fields on $S^2$ gives rise to the same action on their pullbacks. This in turn must be the same as the differential action $dL_G$ on $CT(G)$. Thus, irreducibles on $S^2$ must be related to the irreducibles on $SU(2)$, which we already know. At this point, let us introduce some convenient linear combinations of irreducibles on $S^2$:

$$N_{lm} = \left(\frac{l}{2l+1}\right)^{1/2} W_{lm} - \left(\frac{(l+1)}{(2l+1)}\right)^{1/2} V_{lm}, \tag{5.11a}$$

$$A^\pm_{lm} = \frac{1}{\sqrt{2}}\left[X_{lm} \mp \left(\frac{l}{2l+1}\right)^{1/2} V_{lm} \mp \left(\frac{l+1}{2l+1}\right)^{1/2} W_{lm}\right]. \tag{5.11b}$$

$X_{lm}$, $V_{lm}$, and $W_{lm}$ are the irreducible vector fields on $S^2$ given in Ref. 19 which are in relatively common usage. $N_{lm}$ is normal, while $A^\pm_{lm}$ are tangential.

The pullback of normal fields is the easiest, as we have the relation

$$\widetilde{f\mathbf{e}_r} = (f \circ \pi)\mathcal{J}_3. \tag{5.12}$$

Using the fact that

$$N_{lm} = Y_l^m \mathbf{e}_r \tag{5.13}$$

and the relation

$$D^j_{m,0}(\Phi, \theta, \Psi) = \left(\frac{4\pi}{2j+1}\right)^{1/2} (-1)^m Y_j^{-m}(\theta, \Phi) \tag{5.14}$$

yields the relation

$$\tilde{N}_{lm} = (-1)^m (2l + 1/4\pi)^{1/2} \mathcal{J}^{13}_{-m0}. \tag{5.15}$$

The set of all $\mathcal{J}^{13}_{m,0}$, $l$ integer, $-m \leq l \leq m$, is all of the irreducible vertical vector fields on $SU(2)$.

Introducing the ladder operators $\mathcal{J}_\pm$ by the rule

$$\mathcal{J}_\pm = \mathcal{J}_1 \pm i\mathcal{J}_2 \tag{5.16}$$

and using Eqs. (2.25) and (5.8), we may write the pullback of a tangent vector field as

$$\tilde{X} = \frac{1}{2}\left\{\begin{matrix} (X^\theta + iX^\Phi \sin\theta)\exp(-i\Psi)\mathcal{J}_- \\ -(X^\theta - iX^\Phi\sin\theta)\exp(i\Psi)\mathcal{J}_+ \end{matrix}\right\}. \tag{5.17}$$

Associated with $\mathcal{J}_\pm$, we introduce the vector fields

$$\mathcal{J}^{J\pm}_{mm'} = \frac{1}{\sqrt{2}}\{\mathcal{J}^{J1}_{mm'} \pm i\mathcal{J}^{J2}_{mm'}\}. \tag{5.18}$$

From (5.17), and (5.18) and the form of $D^j_{mm'}$ it is clear that, for an irreducible to be a pullback, it must be of the form $\mathcal{J}^{J-}_{m1}$ or $\mathcal{J}^{J+}_{m-1}$. The fields $A^\pm_{lm}$ have the explicit form

$$A^\pm_{lm} = [2l(l+1)]^{-1/2}\left(\mp\frac{\partial Y_l^m}{\partial\theta} - \frac{m}{\sin\theta}Y_l^m\right)\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta} \pm i\frac{\partial}{\partial\Phi}\right). \tag{5.19}$$

Again using Eq. (5.14), as well as Eqs. (5.17) and (5.19), we find

$$\tilde{A}^\pm_{lm} = (-1)^m\left(\frac{2l+1}{8\pi l(l+1)}\right)^{1/2}\left[\exp(\mp i\Psi) \times \left(\frac{\partial}{\partial\theta} \mp \frac{m}{\sin\theta}\right)\left(D^l_{-m,0}\right)\right]\mathcal{J}_\pm. \tag{5.20}$$

Using the explicit form of the ladder operators $\mathcal{J}_\pm$, we find that

$$\mathcal{J}_\pm(D^l_{-m,0}) = \mp\exp(\mp i\Psi)\left(\frac{\partial}{\partial\theta} \mp \frac{m}{\sin\theta}\right)(D^l_{-m,0}). \tag{5.21}$$

It may further be verified that the operators $\mathcal{J}_\pm$ act as ladder operators on the second index so that

$$\mathcal{J}_\pm(D^l_{-m,\lambda}) = [(l\mp\lambda)(l\pm\lambda+1)]^{1/2} D^l_{-m,\lambda\pm1}, \tag{5.22}$$

where the sign must be determined by looking at the explicit expressions. Combining this information finally yields

$$\tilde{A}^\pm_{lm} = \mp(-1)^m\left(\frac{2l+1}{4\pi}\right)^{1/2}\mathcal{J}^{l\pm}_{-m,\mp1}. \tag{5.23}$$

The set of all $\mathcal{J}^{l\pm}_{m,\mp1}$ for $l$ integer, $-m < l < m$, is all of the irreducible horizontal fields on $SU(2)$.

An inner product $\langle\ ,\ \rangle$ is normally defined for vector fields on $S^2$ by the rule

$$\langle X, Y\rangle = \int d\Omega\, X^* \circ Y, \tag{5.24}$$

where $d\Omega$ is the usual measure on $S^2$ (with total area $4\pi$). An elementary calculation using Prop. 1 shows that

$$\langle X, Y\rangle = 4\pi(\tilde{X}, \tilde{Y}). \tag{5.25}$$

Note also that

$$\langle N_{lm}, N_{lm}\rangle = \langle A^{\pm}_{lm}, A^{\pm}_{lm}\rangle = 1, \tag{5.26}$$

with other scalar products vanishing. It follows immediately from the preceeding discussion that the convergence of vector harmonic series on $S^2$ and the representation of vector fields on $S^2$ in harmonic series are but special cases of the theorems given in Secs. 3 and 4.

## 6. TENSOR HARMONICS

In this section we shall carry out the program of Secs. 4 and 5 for general tensor fields. We shall first investigate the special case of differential 1-forms. Recall (Helgason, p. 8) that a complex covector at $m\in M$ is a linear map $\omega_m : CT_m(M)\to C$. A 1-form is a choice for each $m$ of a covector $\omega_m$ such that $\omega(X)$ is a differentiable complex valued function on $M$ for each $X\in CT(M)$. We denote the vector space of all 1-forms on $M$ by $CT^{0,1}(M)$. Suppose now that $M$ has a Riemannian metric $g$. If $X\in CT(M)$, then we define $\hat{X}$ to be the dual vector field to $X$ (with respect to $g$) by

$$\hat{X}(Y) = g(X, Y) \quad \text{for all } Y\in CT(M). \tag{6.1}$$

Classically, the assignment $X\to\hat{X}$ is called "raise an index."

If $G$ is a Lie group and $u\in G$, then we define $\omega^u\in CT^{0,1}(G)$ by

$$\omega^u(X) = (\omega(dL_{u^{-1}}X))^u = (\omega(X^{u^{-1}}))^u \tag{6.2}$$

where we have used (2.16) for the last equality. The representation $u\to\omega^u$ is called the left adjoint (contragredient) differential representation of $G$. We use (6.1) to define $\hat{\mathcal{J}}^i$ [see (2.24)] and $\hat{\mathcal{J}}^{Ji}_{mm'}$ [see (2.27)]. We may now consider a series expansion similar to (3.1):

$$\omega(u) = \sum_{i=1}^{3}\sum_{J=1}^{\infty}\sum_{m=-J}^{J}\sum_{m'=-J}^{J}(2J+1)\,a^{Ji}_{mm'}\hat{\mathcal{J}}^{Ji}_{mm'}(u). \tag{6.3}$$

*Theorem* 5: (a) $\hat{\mathcal{J}}^i$ generates a one-dimensional invariant subspace of $CT^{0,1}(G)$ on which the left adjoint differential representation is unity.

(b) If $m'$ and $i$ are fixed, then $\{\hat{\mathcal{J}}^{Ji}_{mm'} \mid -J\le m\le J\}$ is a basis for a $(2J+1)$-dimensional unitary irreducible representation.

(c) The expansion (6.3) has the same convergence properties as expansion (3.1).

*Proof* (a) From (6.2) $(\hat{\mathcal{J}}^i)^u(Y) = \hat{\mathcal{J}}^i(Y^{u^{-1}})$, thus $(\hat{\mathcal{J}}^i)^u(Y) = g(\mathcal{J}_i, Y^{u^{-1}}) = g(\mathcal{J}_i, Y) = \hat{\mathcal{J}}^i(Y)$ holds for all $Y\in CT(G)$ and $u\in G$. We therefore see that $\hat{\mathcal{J}}^i$ forms an invariant one dimensional subspace. (b) follows directly from the following computation:

$$\begin{aligned}(\hat{\mathcal{J}}^{Ji}_{mm'})^{u_1}(X) &= \hat{\mathcal{J}}^{Ji}_{mm'}(X^{u_1^{-1}})\\ &= g(\mathcal{J}^{Ji}_{mm'}, X^{u_1^{-1}})\\ &= g((\mathcal{J}^{Ji}_{mm'})^{u_1}, X).\end{aligned}$$

Invoking Eq. (2.26) with $u$ equal to the identity gives

$$g(D^J_{m\lambda}(u_1^{-1})\mathcal{J}^{Ji}_{\lambda m'}, X)$$

or

$$(\hat{\mathcal{J}}_{mm'})^{u_1}(X) = \bar{D}^J_{m\lambda}(u_1^{-1})\hat{\mathcal{J}}^{Ji}_{\lambda m'},$$

(c) follows because a sequence of vector fields $X_i$ converges in $CT(G)$ if and only if $g(X_i, Y)$ converges for all $Y\in CT(G)$ which is precisely the statement that the $\hat{X}_i$ converge in $T^{0,1}(G)$.

We now make some remarks concerning expansions of arbitrary tensors in terms of harmonic tensors. Let $CT^{r,s}(M)$ be the complex vector space of tensors of type $(r,s)$ on $M$ (Helgason, p. 9). The left differential representation of $G$ induces a representation called the left tensor representation on $CT^{r,s}(M)$ by using the left differential representation on the $s$ contravariant factors and the left adjoint differential representation on the $r$-covariant factors, to wit: If $\omega^{l_1},\dots,\omega^{l_s}\in CT^{0,1}(M)$ and $X_{i_1},\dots,X_{i_r}\in CT(M)$, then

$$\begin{aligned}&(X_{i_1}\otimes\cdots\otimes X_{i_r}\otimes\omega^{l_1}\otimes\cdots\otimes\omega^{l_s})^u\\ &\quad= X^u_{i_1}\otimes\cdots\otimes X^u_{i_r}\otimes(\omega^{l_1})^u\otimes\cdots\otimes(\omega^{l_s})^u.\end{aligned} \tag{6.4}$$

We define

$$\mathcal{J}^{l_1,\dots,l_s}_{i_1,\dots,i_r} = \mathcal{J}_{i_1}\otimes\cdots\otimes\mathcal{J}_{i_s}\otimes\hat{\mathcal{J}}^{l_1}\otimes\cdots\otimes\hat{\mathcal{J}}^{l_s}. \tag{6.5}$$

Let $J$, $m$, and $m'$ be integer or half-integer, and $\vec{i} = (i_1,\dots,i_r,l_1,\dots,l_s)$ be an $(r+s)$-tuple of integers each of which is either 1, 2, or 3. Define the tensor harmonics $\mathcal{J}^{J\vec{i}}_{mm'}$ by $\mathcal{J}^{J\vec{i}}_{mm'} = D^J_{mm'}\mathcal{J}^{l_1,\dots,l_s}_{i_1,\dots,i_r}$.

We may also consider series expansions of tensors in terms of tensor harmonics analogous to (3.1) and (6.3):

$$T(u) = \sum_{\vec{i}}\sum_{J=1}^{\infty}\sum_{m=-J}^{J}\sum_{m'=-J}^{J}(2J+1)\,a^{J\vec{i}}_{mm'}\mathcal{J}^{J\vec{i}}_{mm'} \tag{6.6}$$

and investigate the convergence properties of (6.6).

In analogy to definition 3.2 we define the exponent of convergence $\alpha_0$ of the series (6.6) to be $\alpha_0 = -\log(\lim\sup|a^{Ji}_{mm'}|^{1/J})$ where the limit superior is taken over all $J, M, M'$ and multi-indices $\vec{i}$. A tensor $T$ of type $(r,s)$ converges if $T(X_1,\dots,X_s)$ is a convergent $r$-tuple of vector fields for all vector fields $X_1,\dots,X_s$. It is also clear that the scalar product on $CT(G)$ can be extended first to $CT^{0,1}(G)$ and then to all of $T^{(r,s)}(G)$. We now let

$$a^{J\vec{i}}_{mm'} = (\mathcal{J}^{J\vec{i}}_{mm'}, X). \tag{6.7}$$

*Theorem* 6: (a) $\mathcal{J}^{l_1,\dots,l_s}_{i_1,\dots,i_r}$ is invariant under the left tensor representation.

(b) If $J$, $m'$, and $\vec{i}$ are fixed then $\{\mathcal{J}^{J\vec{i}}_{mm'} \mid -J\le m\le J\}$ generate a vector subspace of $T^{(r,s)}(G)$ on which the left tensor representation is unitary and irreducible.

(c) If $T$ is an analytic vector field on $SU(2)$ and $B^6(\alpha_0)$ is the superball of radius $\alpha_0$ [where $\alpha_0$ is the exponent of convergence of the series (6.6) with coefficients given by (6.7)], then the series (6.6) converges absolutely and unformly in $B^6(\alpha_0)$ to $T$.

We shall now construct the horizontal lift of tensors on $S^2$. If $\omega \in CT^{0,1}(S^2)$ and $\pi : SU(2) \to S^2$ (see Sec. 5), then $\pi^*\omega \in CT^{0,1}(SU(2))$ is defined by $\pi^*\omega(X) = \omega(d\pi(X))$. If we have a tensor, $T$, of type $(r,s)$ on $S^2$, then we may (at least locally) write

$$T = X_1 \otimes \cdots \otimes X_r \otimes \omega^1 \otimes \cdots \otimes \omega^s, \tag{6.8}$$

where $X_i \in CT(S^2)$ and $\omega^j \in CT^{0,1}(S^2)$. The horizontal lift of the tensor $T$ is then defined to be

$$\tilde{T} = \tilde{X}_1 \otimes \cdots \otimes \tilde{X}_r \otimes \pi^*(\omega^1) \otimes \cdots \otimes \pi^*(\omega^s). \tag{6.9}$$

If we have an arbitrary tensor as defined in (6.8), we may pull it back to $SU(2)$ as in (6.9), express it as a series [Eq. (6.6)] and then push it back down to $S^2$ via the map $\widetilde{d\pi}$ of Sec. 5. To determine which of the irreducible tensor fields on $SU(2)$ are pullbacks, one must use the Clebsch—Gordon series for the tensor products of irreducible vectors and forms. We may handle in a similar manner tensors which are normal to $S^2$, i.e., tensors $T$ of the form

$$T = f\mathbf{e}_r \otimes \cdots \otimes \mathbf{e}_r \otimes \hat{\mathbf{e}}_r \otimes \cdots \otimes \hat{\mathbf{e}}_r,$$

where $\mathbf{e}_r$ is the normal vector field on $S^2$ given as in Sec. 5 and $\hat{\mathbf{e}}_r$ is the dual of $\hat{\mathbf{e}}_r$ as in Eq. (5.5).

## 7. LIE ALGEBRA OF VECTOR FIELDS

We have shown that the set of all $\mathcal{J}^{Ji}_{mm'}$ form a basis for the space of vector fields on $SU(2)$, $CT(G)$. As mentioned $CT(G)$ is also an infinite-dimensional Lie algebra. This algebra is specified by giving the commutation relations of the basis. We have, for example,

$$\begin{aligned}[][\mathcal{J}^{J_1 3}_{m_1\lambda_1}, \mathcal{J}^{J_2 3}_{m_2\lambda_2}] &= (\lambda_2 - \lambda_1)\sum_J C(J_1 J_2 J; m_1 m_2)\, C(J_1 J_2 J; \lambda_1\lambda_2) \\ &\quad \times \mathcal{J}^{J3}_{m_1+m_2, \lambda_1+\lambda_2},\end{aligned} \tag{7.1}$$

where the C's are the usual Clebsch—Gordon coefficients. Similar relations may easily be displayed for the other commutators. The main point is that a canonical procedure has been specified whereby the Lie algebra of vector fields on the group may be explicitly displayed (all other relations being linear combinations of the bases relations). This canonical procedure for uncovering the structure of the algebra should be of use in any studies of the algebra and its representations.

We mention this because the Virasoro algebra[10] can arise in a similar construction.[20] On the simpler group $U(1) \approx S^1$, the irreducible vector fields are given by

$$L_m = (-i)\exp(im\Phi)\,\partial/\partial\Phi, \tag{7.2}$$

where $\Phi$ is the usual polar angle. The commutator is

$$[L_m, L_n] = (n-m)\, L_{m+n} \tag{7.3}$$

which is just the Virasoro algebra without the $C$-number term. It is interesting to speculate whether the generalization given by Eq. (7.1) and associated relations might permit more realistic dual-resonance models.

## ACKNOWLEDGMENTS

One of us (BLB) would like to thank Dr. J.H. Eberly for his kind hospitality at the Quantum Optics group of the University of Rochester during the summer of 1973, during which time part of this work was begun. He would also like to thank Dr. E. Wolf of the University of Rochester and Dr. P. Winternitz of the Université de Montréal for helpful conversations which provided motivation for this paper.

*Current address: Department of Physics, The Johns Hopkins University, Baltimore, Md. 21218.

[1]See, for example W. Ruhl, *The Lorentz Group and Harmonic Analysis* (Benjamin, New York, 1970); N.J. Vilenkin, *Special Functions and the Theory of Group Representations* (American Mathematical Society, Providence, R.I. 1968); G. Warner, *Harmonic Analysis on Semi-Simple Lie Groups, I and II* (Springer-Verlag, Berlin, 1972); P. Winternitz, "Expansions of Relativistic Scattering Amplitudes and Harmonic Analysis on the Lorentz Group, Parts I, II, and III" The University of Arizona Department of Mathematics 1970–71 Lecture Series).

[2]B. L. Beers and A.J. Dragt, J. Math. Phys. **11**, 2313 (1970), hereafter referred to as I.

[3]B. L. Beers, J. Math. Phys. **13**, 950 (1972).

[4]A.J. Dragt, private communication.

[5]H.R. Hicks, C. Shukre, and P. Winternitz, Phys.Rev. D **7**, 2659 (1973).

[6]B. L. Beers, thesis (University of Maryland, 1970) (unpublished).

[7]See, for example, J.M. Blatt and V.F. Weisskopf, *Theoretical Nuclear Physics* (Wiley, New York, 1952).

[8]A review and some recent results may be found in A.J. Devaney and E. Wolf, Phys. Rev. D **8**, 1044 (1973); University of Rochester preprint, 1973; A.J. Devaney, thesis (University of Rochester, 1971).

[9]C. Shukre and P. Winternitz, Phys. Rev. D **6**, 3592, 3607 (1972).

[10]M.A. Virasoro, Phys. Rev. D **1**, 2933 (1970).

[11]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry, I and II* (Interscience, New York, 1963, 1969).

[12]R.S. Millman and A.K. Stehney, Amer. Math. Monthly **80**, 475 (1973).

[13]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962), hereafter referred to as Helgason.

[14]S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, N.J., 1964).

[15]Throughout the course of this article, we will often adhere to the usual (rather vague) practice of physicists of indicating a patch simply by a set of variables $y = (y_1, \ldots, y_n)$ rather than a map $\Phi : U \to R^n$, where $U$ is an open subset of $G$. This leads us to say that $y_0$ is a point of $G$. In the same spirit, we often use the same symbol for the representative of the function in the patch and write $f(y)$ rather than introducing a separate symbol [say $f^*(y)$] for the representative $f \circ \varphi^{-1}$.

[16]O. Zariski and P. Samuel, *Commutative Algebra* (Van Nostrand, Princeton, N.J., 1958), Vol. I, p. 134.

[17]B. O'Neill, Michigan Math. J. **13**, 459 (1966).

[18]The general definition of the differential $d\Phi$ of a map $\Phi$ may be found on p. 22 of Helgason.

[19]E. L. Hill, Amer. J. Phys. **22**, 211 (1954).

[20]P. Ramond, Lett. Nuovo Cimento **4**, 422 (1972).

# Dual trees and resummation theorems*

R. Friedberg

*Department of Physics, Columbia University and Barnard College, New York, New York 10027*
(Received 24 May 1974)

Various resummation theorems known for graphical expansions in statistical physics and field theory are exhibited as special cases of a single symmetrical, easily remembered theorem on a generalized structure we call a dual tree.

## I. INTRODUCTION

Graphical summation is now a common technique in physics. Graphs are constructed according to some simple recursive rule, and to each graph is associated a value determined algebraically according to the topology of the graph. The physical quantity of interest is expressed as the (infinite) sum of the values of all graphs. In a resummation theorem, the sum is proved identical to a sum derived in some other way, typically from a smaller set of graphs with modified rules for assigning values.

Many resummation theorems have been proved for various graphical systems. These theorems have a family resemblance, although the detailed statement and proof varies from system to system according to the kind of graph considered and the way its value is determined. In this paper we shall bring a whole family of such theorems into a single theorem, stated and proved once for all. To this end we define a particular kind of graph, which we call a dual tree, composed of two elements designated arbitrarily as squares and circles. By suitable interpretation of these elements, the system of dual trees can be transformed into any of a remarkably wide variety of known graphical systems.

In Sec. II we shall define the dual tree system and state its resummation theorem. We have two proofs, both derived from arguments already in the literature. The proofs are given in Appendices A and B. In the following four sections we shall derive a number of well-known results from this theorem, thereby exhibiting the flexibility of the dual tree concept. We conclude in Sec. VII.

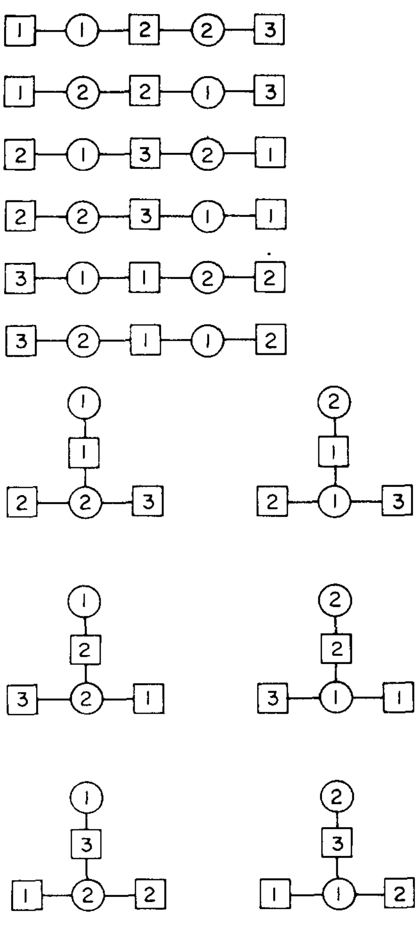


FIG. 1. All 12 inequivalent dual trees with $n_C=2$, $n_S=3$.

## II. CENTRAL THEOREM

By a "dual tree" we mean an assemblage of $n_S$ squares and $n_C$ circles $(n_S+n_C \geq 1)$ joined by $n_L$ lines under the following restrictions:

1. Each line joins a circle to a square—never a square to a square or a circle to a circle.

2. The whole graph is simply connected—it cannot be decomposed without severing a line, but the removal of any single line would render it decomposable.

The use of squares and circles has no geometrical significance; it is an aid to visual representation.

The $n_S$ squares are distinguishable and to that end are labeled with integers from 1 to $n_S$. Similarly the $n_C$ circles.

Two dual trees (from now on we shall drop the word "dual") are identical if they have the same topology, with the labeling taken into account. Fig. 1 shows the twelve *distinct* trees having $n_S=3$, $n_C=2$.

We assume that for each nonnegative integer $n$, two symmetric functions on $n$ arguments are given, which we call $S_n$ and $C_n$. The arguments of all these functions vary over the same domain $X$. For $n=0$ there are no arguments, so that $S_0$ and $C_0$ are constants.

Let $S(\xi)$, $C(\xi)$ represent arbitrary functions on an argument $\xi \in X$. We define the functionals

$$\mathcal{D}(S) \equiv \sum_{n=0}^{\infty} n!^{-1} \int_{\xi_1 \ldots \xi_n} C_n(\xi_1 \cdots \xi_n) \prod_1^n S(\xi_i), \tag{1}$$

$$\mathcal{J}(C) \equiv \sum_{n=0}^{\infty} n!^{-1} \int_{\xi_1 \ldots \xi_n} S_n(\xi_1 \cdots \xi_n) \prod_1^n C(\xi_1) \tag{2}$$

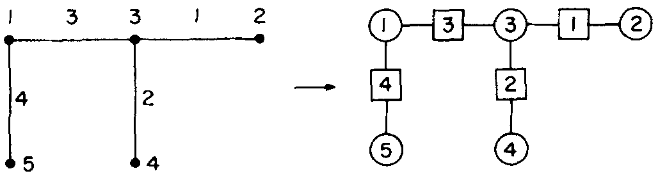



FIG. 2. Transformation of Cayley tree to dual tree.

where $\int_{\xi_1 \cdots \xi_n}$ represents integration or summation with respect to a measure defined on $X$. The term in $n=0$ has the value $C_0$ in (1), or $S_0$ in (2).

To each tree $T$ we assign a value $w(T)$ as follows. The $n_L$ lines are given an arbitrary order. To the $i$th line we associate a variable $\xi_i \in X$. To each square (attached to lines $i_1 \cdots i_m$) we associate the quantity $S_m(\xi_{i_1} \cdots \xi_{i_m})$. To each circle (attached to lines $i_1 \cdots i_q$) we associate $C_q(\xi_{i_1} \cdots \xi_{i_q})$. The product of all these $S$- and $C$-functions we call $\Pi(\xi_1 \cdots \xi_{n_L})$. We obtain $w(T)$ by summing $\Pi$ over all the $\xi$'s and dividing by $n_S! n_C!$. Thus the first six trees shown in Fig. 1 each have the value

$$w=\frac{1}{3!\,2!}\int_{\xi_1\xi_2\xi_3\xi_4} S_1(\xi_1)S_2(\xi_2,\xi_3)S_1(\xi_4)C_2(\xi_1,\xi_2)C_2(\xi_3,\xi_4)$$

and the last six have each the value

$$w=\frac{1}{3!\,2!}\int_{\xi_1\xi_2\xi_3\xi_4} S_1(\xi_1)S_2(\xi_2,\xi_3)S_1(\xi_4)C_3(\xi_1,\xi_3,\xi_4)C_1(\xi_2).$$

We now define

$$\tau \equiv \sum_T w(T) \tag{3}$$

where the sum goes over *all* distinct trees.[1] Regarding $\tau$ as a functional on $S_1$, $C_1$, $S_2$, $C_2$, $\cdots$, we define

$$\bar{S}(\xi)\equiv \delta\tau/\delta C_1(\xi), \quad \bar{C}(\xi)\equiv\delta\tau/\delta S_1(\xi). \tag{4}$$

Our central result is the following theorem.

*Theorem* 1: Given the above definitions, the following relations hold:

$$\mathcal{D}(\bar{S})=\sum_T n_C w(T), \quad \mathcal{J}(\bar{C})=\sum_T n_S w(T), \tag{5}$$

$$\bar{S}(\xi)=\frac{\delta\mathcal{J}}{\delta C(\xi)}\bigg|_{C=\bar{C}}, \quad \bar{C}(\xi)=\frac{\delta\mathcal{D}}{\delta S(\xi)}\bigg|_{S=\bar{S}}, \tag{6}$$

$$\tau=\mathcal{D}(\bar{S})+\mathcal{J}(\bar{C})-\int_\xi \bar{S}(\xi)\bar{C}(\xi). \tag{7}$$

We shall defer the proof (two versions) to appendices.

*Corollary*: Let the above definitions be modified as follows. Let $S_1$, $S_2$, $S_3$, $\cdots$ all be functions of *one* argument. For each tree $T$, let $\Pi(\xi_1 \cdots \xi_{n_S})$ be defined by attaching $\xi_i$ to the $i$th square, associating $S_m(\xi_i)$ to the $i$th square where $m$ is the number of attached lines, and associating $C_q(\xi_{i_1} \cdots \xi_{i_q})$ to each circle (attached to squares $i_1 \cdots i_q$). Obtain $w$ from $\Pi$ as before. Replace (2) by

$$\mathcal{J}(C)\equiv\sum_{n=0}^{\infty} n!^{-1}\int_\xi S_n(\xi)[C(\xi)]^n. \tag{8}$$

Retain (1), (3), and (4). Then (5), (6), (7) are still true.

This follows immediately from Theorem 1. For the values of $w(T)$ and $\mathcal{J}(C)$ so obtained are the same as would have followed from the old definitions if the $S_n$ had the form

$$S_n(\xi_1 \cdots \xi_n)=\delta(\xi_1 \cdots \xi_n)S_n(\xi_1) \tag{9}$$

where $\delta(\xi_1 \cdots \xi_n)$ is a suitable product of Kronecker or Dirac $\delta$'s, depending on whether $X$ is discrete or continuous.

## III. THE NUMBER OF CAYLEY TREES

As a simple application of Theorem 1, we derive a theorem due to Lee and Yang[2] on the number of doubly labeled Cayley trees of order $l$. Such a tree is just a simply connected graph composed of $l$ labeled points joined by $l-1$ labeled lines. Letting $A_l$ denote the number of ways to do this,[3] Lee and Yang showed that

$$-\sum_{l=1}^{\infty} A_l(-x)^l/(l-1)!^2\,l = D+\tfrac{1}{2}D^2 \tag{10}$$

where

$$De^D=x. \tag{11}$$

To reproduce this result from Theorem 1, we transform Cayley trees into dual trees by inserting a square into each line, and replacing each point by a circle, as shown in Fig. 2. In this way we obtain only those dual trees in which all squares have order 2. We insure that all *other* dual trees have no value by setting $S_n=0$ for $n\neq 2$. We let the domain $X$ consist of one point, so that all the $S_n$, $C_n$, $\bar{S}$, $\bar{C}$, are numbers instead of functions. We let $S_2=-1$, and $C_n=x$, independently of $n$. Then we have $n_C=l$, $n_S=l-1$, and hence

$$w(T)=\frac{1}{l!}\frac{1}{(l-1)!}(-)^{l-1}x^l \tag{12}$$

for any dual tree $T$ that corresponds to a Cayley tree of $l$ points. It follows that $\tau$ is just the left side of Eq. (10).

Putting our definitions of $C_n$, $S_n$ into (1) and (2), we have

$$\mathcal{D}(S)=\sum_0^{\infty} n!^{-1}xS^n=xe^S, \tag{13}$$

$$\mathcal{J}(C)=2!^{-1}(-1)C^2=-\tfrac{1}{2}C^2 \tag{14}$$

and so (6) becomes

$$\bar{S}=-\bar{C}, \quad \bar{C}=xe^{\bar{S}} \tag{15}$$

from which we immediately obtain (11) on setting $D\equiv\bar{C}$.

Now if we use (15) to write (13) as

$$\mathcal{D}(\bar{S})=\bar{C} \tag{16}$$

and substitute, with (14) and (15), into (7), we obtain

$$\begin{aligned}\tau&=\bar{C}+(-\tfrac{1}{2}\bar{C}^2)-(-\bar{C})\bar{C}\\&=\bar{C}+\tfrac{1}{2}\bar{C}^2\end{aligned} \tag{17}$$

which is equivalent to (10).

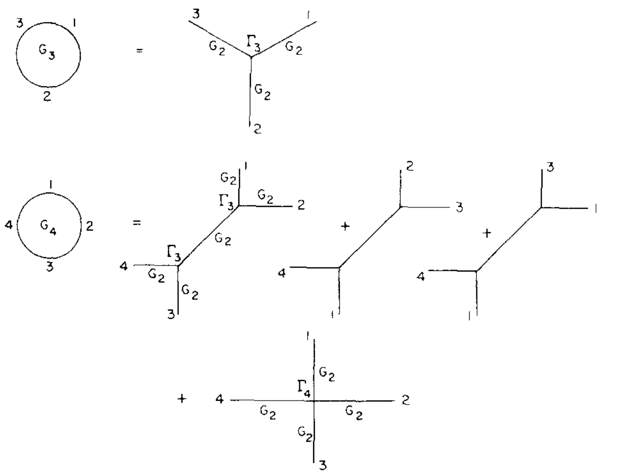



FIG. 3. Graphical expansion of $G_3$ and $G_4$ in terms of irreducible vertices $\Gamma_m$ and propagators $G_2$.

## IV. THE GENERATING FUNCTIONAL OF IRREDUCIBLE PARTS

For our second application, we have in mind any system in which a sequence of Green's functions is generated by some functional $\mathcal{G}$ on an external field $J$.

Suppose that we have symmetric functions

$$G_2(x_1, x_2), \quad G_3(x_1, x_2, x_3), \cdots$$

where the $x$'s may be position vectors in 3-space or in space—time, or may also contain intrinsic information (spin, etc.). We introduce a complementary sequence $\Gamma_2(x_1, x_2)$, $\Gamma_3(x_1, x_2, x_3)$, etc. The first has a special definition

$$\Gamma_2 = -G_2^{-1} \tag{18}$$

where the meaning of (18) is that

$$\int_{x_2} \Gamma_2(x_1, x_2) G_2(x_2, x_3) = -\delta(x_1, x_3). \tag{19}$$

The higher $\Gamma$'s are uniquely defined as the irreducible vertex parts in a graphical expansion for the higher $G$'s. That is, each $G_n$ $(n \geq 3)$ can be obtained from $G_2$ and $\Gamma_3$, $\Gamma_4$, $\cdots$ by the following prescription: Take all possible unlabeled Cayley trees with $n$ labeled endpoints, and with no vertices of order 2. Associate the variables $x_1 \cdots x_n$ with the endpoints in the specified order. In each tree, associate extra variables $x_i (i > n)$ with each insertion of a line into an internal vertex. To each vertex of order $m$, assign a factor $\Gamma_m(x_{i_1}, x_{i_2}, \ldots, x_{i_m})$. To each line, assign a factor $G_2(x_i, x_j)$. Integrate the product of all these factors over the $x_i$ $(i > n)$ and sum the result over all the permitted Cayley trees with $n$ endpoints. The result is $G_n(x_1, x_2, \ldots, x_n)$.

Thus

$$G_3(x_1, x_2, x_3) = \int_{x_4 x_5 x_6} \Gamma_3(x_4, x_5, x_6) G_2(x_1, x_4) G_2(x_2, x_5) \times G_2(x_3, x_6), \tag{20}$$

$$\begin{aligned} G_4(x_1, x_2, x_3, x_4) = \int_{x_5 \cdots x_{10}} & \Gamma_3(x_5, x_6, x_7) \Gamma_3(x_8, x_9, x_{10}) \\ & \times G_2(x_7, x_{10}) G_2(x_1, x_5) G_2(x_2, x_6) G_2(x_3, x_8) \\ & \times G_2(x_4, x_9) \\ & + \text{cyclic permutations on } x_1, x_2, x_3 \\ + \int_{x_5 \cdots x_8} & \Gamma_4(x_5, x_6, x_7, x_8) G_2(x_1, x_5) G_2(x_2, x_6) \\ \times & G_2(x_3, x_7) G_2(x_4, x_8) \end{aligned} \tag{21}$$

and so on. (See Fig. 3.)

It is easily seen that if $m$ is the order of a vertex,

$$\sum (m-2) = n - 2 \tag{22}$$

and therefore only a finite number of trees contribute to $G_n$, involving only $\Gamma_3 \cdots \Gamma_n$, and only one tree involves $\Gamma_n$. Therefore $\Gamma_3$, $\Gamma_4$, $\cdots$ are uniquely determined by (20), (21), etc.

Now let $\mathcal{G}$ and $\mathcal{K}$ be the generating functionals

$$\mathcal{G}(J) = \sum_2^\infty n!^{-1} \int_{x_1 \cdots x_n} G_n(x_1 \cdots x_n) \prod_1^n J(x_i), \tag{23}$$

$$\mathcal{K}(A) = \sum_2^\infty n!^{-1} \int_{x_1 \cdots x_n} \Gamma_n(x_1 \cdots x_n) \prod_1^n A(x_i). \tag{24}$$

It is a well-known theorem,[4] but one whose proof is usually indicated rather than given in detail, that $\mathcal{G}$ and $\mathcal{K}$ are related by a Legendre transformation. That is, if we let

$$A(x) = \frac{\delta \mathcal{G}(J)}{\delta J(x)}, \tag{25}$$

then

$$\mathcal{G}(J) = \mathcal{K}(A) + \int_x J(x) A(x). \tag{26}$$

We shall now obtain this result from Theorem 1.

By combining (23) with (20), (21), etc., we obtain $\mathcal{G}(J)$ as a sum over *all* Cayley trees. In this sum, each $m$-vertex $(m > 2)$ contributes a factor $\Gamma_m$, each line a $G_2$, and each endpoint a $J$. Any 2-vertex contributes a factor 0. Only the endpoints are labeled, and we are to divide by $n!$ where $n$ is the number of endpoints. The term of (23) with $n = 2$ is included naturally if we count the Cayley tree with one line and no vertex.

We convert each Cayley tree into a dual tree as follows. First we insert a square into each line as in the previous section. The variables $x_i$ are now associated with the newly formed lines.

The labeling of the endpoints renders all $n_S$ squares inequivalent. We may therefore label all the squares and compensate by dividing the value of each tree by $n_S!$, since there are now $n_S!$ as many labeled trees as before.

The labeling of the squares now renders the endpoints inequivalent. Therefore we may drop the labeling of endpoints and compensate by omitting the factor $n!^{-1}$.

We now replace each vertex *and* each endpoint by a circle. All the circles are inequivalent because the

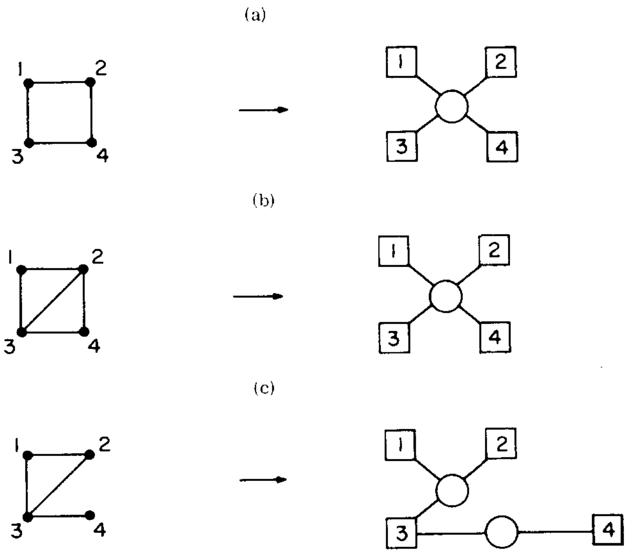



FIG. 4. Transformation of Mayer cluster graphs to dual trees. Note that (a) and (b) give the same dual tree.

squares are labeled. Hence we may label the circles and divide by $n_C!$ where $n_C$ is the total number of circles.

$\mathcal{G}$ is now given as a sum over dual trees, and the contribution of each dual tree $T$ can be identified with $w(T)$ as defined in Sec. II, provided that $X$ is the domain of the $x$'s and

$$S_2 = G_2, \quad S_{n\neq 2} = 0,$$
$$C_0 = C_2 = 0, \quad C_1 = J, \quad C_{n\geq 3} = \Gamma_n. \tag{27}$$

Hence we may identify $\tau$ with $\mathcal{G}$, and $\bar{S}$ with $A$. [Compare (25) with (4).]

Putting (27) into (1) and (2), we have

$$\begin{aligned}\mathcal{D}(\bar{S}) &= \int_x J(x)A(x) + \sum_3^\infty n!^{-1}\int_{x_1\cdots x_n}\Gamma_n(x_1\cdots x_n)\prod_1^n A(x_i)\\ &= \int_x J(x)A(x) + \mathcal{K}(A) - \tfrac{1}{2}\int_{x_1x_2}\Gamma_2(x_1,x_2)A(x_1)A(x_2)\end{aligned} \tag{28}$$

and

$$\mathcal{J}(C) = \tfrac{1}{2}\int_{x_1x_2} G_2(x_1,x_2)C(x_1)C(x_2), \tag{29}$$

whence

$$\bar{S}(x) = \int_{x'} G_2(x,x')\bar{C}(x') \tag{30}$$

by (6). Therefore,

$$\begin{aligned}\mathcal{J}(\bar{C}) - \int_x \bar{S}(x)\bar{C}(x) &= -\tfrac{1}{2}\int_x \bar{S}(x)\bar{C}(x)\\ &= -\tfrac{1}{2}\int_{x_1x_3}\bar{S}(x_1)\delta(x_1,x_3)\bar{C}(x_3)\\ &= \tfrac{1}{2}\int_{x_1x_2}\Gamma_2(x_1,x_2)\bar{S}(x_1)\bar{S}(x_2)\end{aligned} \tag{31}$$

on account of (19) with (30).

If we add (28) to (31) and use (7), identifying $\mathcal{G}$ with $\tau$ on the left side and $\bar{S}$ with $A$ on the right, we obtain the desired Eq. (26).

## V. MAYER CLUSTER EXPANSION

It is well known[5] that the pressure $p$ and density $\rho$ of a gas of short-range interacting particles obeying classical statistics are given in the thermodynamic limit by

$$p/kT = \sum_{l=1}^\infty b_l z^l, \quad \rho = \sum_{l=1}^\infty l b_l z^l \tag{32}$$

where $z$ is the fugacity and $b_l$ is defined as follows. Take $l$ distinct (labeled) points and connect them by lines. Each pair of points may be connected by at most one line. Enough lines must be drawn to bind all $l$ points into a connected graph or *cluster*; it need not be simply connected. Now assign a position $\mathbf{r}_i$ to the $i$th point. To each line (joining the $i$th and $j$th points) associate the quantity $f(\mathbf{r}_i - \mathbf{r}_j)$, where $f$ is a certain temperature-dependent even function of short range. Multiply the $f$'s together, integrate with respect to all the $\mathbf{r}$'s over the volume $V$ of the gas, sum over all clusters of $l$ points, and divide by $Vl!$. The result approaches $b_l$ as $V\to\infty$.

Mayer and coworkers[6] observed that any cluster can be uniquely resolved into (one or more) *irreducible clusters* joined together through common points in a simply connected manner. An irreducible cluster is one which cannot be severed by the removal of any single point and the lines attached to it. They showed that if $(k+1)^{-1}\beta_k$ is the contribution of single irreducible clusters to $b_{k+1}$ then the elimination of $z$ from (32) yields

$$p/kT = \rho\left(1 - \sum_{k=1}^\infty (k+1)^{-1}k\beta_k\rho^k\right). \tag{33}$$

The same result can be obtained easily from the corollary to Theorem 1.

We first show how each cluster of $l$ points yields a dual tree with $n_S = l$, unique except for the labeling of circles. Simply resolve the cluster into irreducible parts; for each irreducible part introduce a circle connected by lines to each of the constituent points; erase all the original lines between points, and replace the points by squares with the same labels. (See Fig. 4.) There are now $n_C!$ distinct ways to label the circles—distinct because the squares are labeled already.

To a given tree may correspond more than one distinct cluster, because a circle with $>3$ lines attached may represent any of a plurality of irreducible clusters. (See Fig. 4. a, b.) All the clusters that reduce to the same tree may be grouped into a *species*.

To apply the Corollary, we take $\xi_i$ as $\mathbf{r}_i$, the position of the $i$th particle; $X$ as the space occupied by the gas; $S_n(\mathbf{r}) = z$ for all $n$, $\mathbf{r}$; and $C_n(\mathbf{r}_1,\ldots,\mathbf{r}_n) = \sum\prod f(\mathbf{r}_i - \mathbf{r}_j)$, where the sum goes over all irreducible clusters of $n$ points, and the product is over those pairs $i$, $j$ joined by a line of the selected irreducible cluster. It is easily seen that for any tree $T$, the contribution of the corresponding species to $z^{n_S}b_{n_S}$ is just $(n_C!/V)w(T)$. Since the species yields $n_C!$ distinct trees, the comparison of (3) and (5) with (32) yields

$$\tau = pV/kT, \quad \mathcal{J}(\bar{C}) = \rho V. \tag{34}$$

On the other hand, $\mathcal{J}$ is given explicitly through (8) as

$$\mathcal{J}(C)=\int_V d^3rz\exp[C(\mathbf{r})] \tag{35}$$

and the use of (6) gives

$$\bar{S}(\mathbf{r})=z\exp[\bar{C}(\mathbf{r})]. \tag{36}$$

In the limit $V\to\infty$, the range of $f$ remaining finite, the functions $\bar{S}$ and $\bar{C}$ as determined by (4) or (6) will be r-independent except for a negligible region near the boundary of $V$. Therefore (35) and (36) yield

$$\mathcal{J}(\bar{C})=V\bar{S} \tag{37}$$

and, from (34),

$$\bar{S}=\rho. \tag{38}$$

Finally, by comparing the definition of $\beta_k$ with that of $C_{k+1}$ and using (1), we obtain

$$\mathcal{D}(S)=V\sum_{k=1}^{\infty}(k+1)^{-1}\beta_k S^{k+1} \tag{39}$$

for any constant function $S$. Using (38) we have

$$\mathcal{D}(\bar{S})=V\sum_{k=1}^{\infty}(k+1)^{-1}\beta_k\rho^{k+1} \tag{40}$$

and using (6) we get

$$\bar{C}=\sum_k\beta_k\rho^k. \tag{41}$$

The substitution of (34), (38), (40), and (41) into (7) yields (33) directly. Also from (36), (38), and (41) we may solve for $z$, obtaining

$$\ln z=\ln\rho-\sum_k\beta_k\rho^k. \tag{42}$$

The proof of Theorem 1 as applied to this example may be regarded as a relatively inefficient version of a derivation of (33) and (42) given by Uhlenbeck and Ford.[7] The advantage of the present treatment is that it places the result in a more general perspective.

## VI. OCCUPATION-NUMBER EXPANSION IN QUANTUM STATISTICS

Lee and Yang[8] found that the logarithm of the grand partition function for a system of interacting Bose particles could be expressed as the sum of connected diagrams ("primary 0-graphs") formed as follows:

Every line is directed; it begins at a vertex and ends at a vertex, possibly the same one. Every vertex has $n$ lines entering and $n$ lines leaving, where $n$ is a positive integer pertaining to the vertex; a 1-vertex has $n=1$, a 2-vertex has $n=2$, etc. Each line is associated with a momentum variable $k$.

To evaluate a graph, we take a factor $z$ for each line ($z=$ fugacity) and a factor $Y_n(k_{i_1}\cdots k_{i_n};k_{j_1}\cdots k_{j_n})$ for each $n$-vertex where $i_1\cdots i_n$ are the indices of the incoming lines and $j_1\cdots j_n$ are those of the outgoing lines, and $Y_n$ is a function whose form need not concern us except that it is symmetric in the $k_i$'s, symmetric in the $k_j$'s, and vanishes unless momentum is conserved, that is unless

$$\sum_{\nu=1}^{n}k_{i_\nu}=\sum_{\nu=1}^{n}k_{j_\nu}. \tag{43}$$

We multiply all these factors together and divide by the symmetry number of the graph. Graphs differing only in the labeling of lines are not regarded as distinct.

Let us regard $z$ as a bare propagator for these lines. In an obvious way, Lee and Yang introduced a clothed propagator $M(k)$, obtained by modifying a line of momentum $k$ with all possible insertions [see their Eq. (IV.21)]. They also defined a half-clothed propagator $m(k)$ by allowing only 1-vertex insertions; thus

$$m(k)^{-1}=z^{-1}-Y_1(k,k). \tag{44}$$

Physically, the quantity $z^{-1}M(k)-1$ is the mean occupation number of the particle state $k$, and $z^{-1}m(k)-1$ is that occupation number in the absence of interaction, but still under the influence of Bose statistics.

After various manipulations, they found that the logarithm of the partition function could be written as

$$\mathcal{P}=\sum_k\ln[z^{-1}M(k)]-\sum_k m(k)^{-1}[M(k)-m(k)]+\mathcal{P}' \tag{45}$$

where $\mathcal{P}'$ is the sum over all irreducible graphs, excluding 1-vertices and replacing $z$ with $M(k)$ for each line. [See their Eq. (IV.33).] By an irreducible graph is meant one that cannot be severed by removing *two* lines.

Our purpose in this section is to derive (45) from the Corollary to Theorem 1. The treatment can be adapted to Fermi statistics by trivial modifications.

Other authors[9,10] have developed a field-theoretic graphical expansion for the logarithm of the partition function, more closely related to perturbation theory. From that expansion one obtains a formula [Eq. (47) of Ref. (9)] just like Eq. (45). However, each vertex then corresponds to a single application of the interaction Hamiltonian, whereas the vertices of Lee and Yang correspond to modified scattering matrices; and the variable $k$ has not only the three momentum components but also a fourth energylike component that varies discretely in steps proportional to the temperature.

The differences, however, between the formalism just described and that of Lee and Yang are *completely inessential* to what follows. Therefore the work of this section is equally applicable to both, although we shall use the notation of Lee and Yang.

It is, on the other hand, essential here that $k$ be conserved at each vertex as expressed by Eq. (43). This restriction distinguishes the subject matter of this section from that of Sec. IV, in which $k$-conservation was not assumed for the Fourier-transformed vertices.

Before describing the reduction of graphs to dual trees, we need some preliminary results concerning the graphs themselves.

Let us say that two vertices $V_1$ and $V_2$ are *well-connected* if the removal of any two lines from the graph leaves $V_1$ and $V_2$ still joined by a continuous path. Well-connectedness is clearly an equivalence relation.

Let us say that two lines $i$ and $j$ are *related* if the graph falls into two parts when $i$ and $j$ are both removed. By $k$-conservation, any two related lines must have the same $k$. But the converse also holds. For if $i$ and $j$ are *not* related, there must be a continuous path connecting the head of $i$ to its tail, not traversing either $i$ or $j$. Therefore it is possible to increment $k_i$ without incrementing $k_j$, and still conserve $k$ at each vertex by applying the same increment everywhere along this path.

Since $i$ and $j$ are related if and only if they are required to have the same $k$, relatedness is also an equivalence relation. We shall call the equivalence classes *families* of lines.

We now state a crucial topological lemma.

*Lemma*: The lines in a family have a unique cyclical ordering such that the head of each is well-connected to the tail of the one following.

The truth of this is intuitively evident, and it is used without proof in most treatments. For completeness we give a proof in Appendix C.

With the help of this lemma, we can express the logarithm of the partition function as a sum over dual trees. The transformation from graph to dual tree will have two parts: first we shall transform each graph $G$ into a *circuit graph* $J$, and then we shall transform each circuit graph $J$ into a dual tree $T$.

We start by labeling all the lines in $G$. This yields $n_l!/\sigma$ distinct labeled graphs, where $n_l$ is the number of lines and $\sigma$ is the symmetry number. To compensate we divide by $n_l!$ instead of $\sigma$ in evaluating each graph. Thus the value of a labeled graph is $n_l!z^{n_l}\sum_{\{k\}}V_G(\{k\})$ where $V_G(\{k\})$ is the product of the functions $Y_n$ for each $n$-vertex and $\{k\}$ represents the $n_l$ variables $k$.

Next, we form $J$ by identifying all vertices in $G$ that are well connected, without destroying any lines. It follows from our lemma that the lines of each family now form a connected circuit, and for this reason we call $J$ a circuit graph.

We now define the value of $J$ to be the sum of the values of all labeled $G$'s that yield $J$ under the foregoing operation. This sum may be evaluated from the structure of $J$ in the following way.

If we let $\bar{P}$ be the set of vertices in $G$ that condense into one vertex $P$ in $J$, then the topology of $\bar{P}$ and of all the lines attached to it is completely contained in the graph $I$ that is formed by deleting all vertices of $G$ not in $\bar{P}$ and all lines of $G$ not attached to $\bar{P}$, and connecting each line that leaves $\bar{P}$ with the related line that enters $P$. $I$ is an irreducible graph, since each family has only one line. (The reverse transformation, from $I$ to $G$, may be regarded as the placing of insertions on the lines of $I$.)

It follows that $G$ may be reconstructed from $J$ if for each $n$-vertex $P$ in $J$ we are given the corresponding $n$-line irreducible graph $I$ in $G$. Since $v(G)$ is just the product of $v(I)$ over all $P$ in $J$, we can write the value of $J$ as $n_l!^{-1}z^{n_l}\sum_{\{k\}}u_J(\{k\})$, where $u_J(\{k\})$ is the product of factors $C_n(k_1\cdots k_n)$ for each $n$-vertex $P$ of $J$. It is now understood that each *circuit* (i.e., family) is assigned a variables $k$, and $k_1\cdots k_n$ are the variables assigned to the $n$ circuits passing through $P$. The function $C_n(k_1\cdots k_n)$ is just the sum of $v_I(k_1\cdots k_n)$ over all *irreducible* graphs $I$.

The reduction of $G$ to $J$ is equivalent to that described by Bloch[11] in treating this problem. We now proceed to transform $J$ into a dual tree $T$.

First, we label all the vertices (let their number be $n_C$) of $J$. This yields $n_C!$ different labeled circuit graphs, since each vertex is already specified uniquely as the head of some labeled line. We compensate by introducing an additional factor $n_C!^{-1}$ into the value. The vertices of $J$ will be the circles of $T$.

Next, we place a new element, a square, at the "center" of each circuit of $J$, and connect it by "new" lines to each vertex of the circuit. The variable $k$ associated with the circuit can now be regarded as belonging to the square. Let $n_S$ be the number of squares—that is, the number of families in the original graph $G$. We label the squares, increasing the multiplicity of structures by $n_S!$ since each square is already distinguished by the labeled lines of its circuit. We compensate by introducing a factor $n_S!^{-1}$ into the value, which is now given by $n_S!^{-1}n_C!^{-1}n_l!^{-1}z^{n_l}\sum_{\{k\}}u_J(\{k\})$.

The final step is to erase the $n_l$ "old" lines, those forming the circuits of $J$. The remaining structure $T$ is a dual tree since the "new" lines only join squares to circles and the removal of any line of $T$ (corresponding to the removal of two successive members of a family of $G$) splits $T$ into two parts. However, the erasure has reduced the multiplicity of structures by a factor calculated as follows.

If we examine a square of $n$th order in $T$, we find there are $(n-1)!$ ways to order the $n$ neighboring circles cyclically, and hence $(n-1)!$ ways to draw in the $n$ "old" lines of the circuit corresponding to this square. After all the "old" lines have been restored, there are $n_l!$ ways to label them, since each is already specified by the labeled square associated with its circuit and the labeled circle at its head. So the multiplicity has been reduced by $(n-1)!$ for each square of $n$th order, as well as by an overall factor $n_l!$. To compensate, we omit the factor $n_l!^{-1}$ from the value and introduce a factor $(n-1)!$ for each square of $n$th order in $T$, in addition to the $n$ factors $z$ inherited from the "old" lines of the corresponding circuit.

The value assigned to $T$ is now identical with $w_T$, provided that we define

$$S_n(k_1\cdots k_n)=(n-1)!\,z^n \tag{46}$$

for $n=1,2,\cdots$, and $S_0=0$. The cyclical ordering factor $(n-1)!$ is the distinguishing feature of this application and is the origin of the logarithm in (45). The whole transition from $G$ to $J$ to $T$ is depicted in Fig. 5.

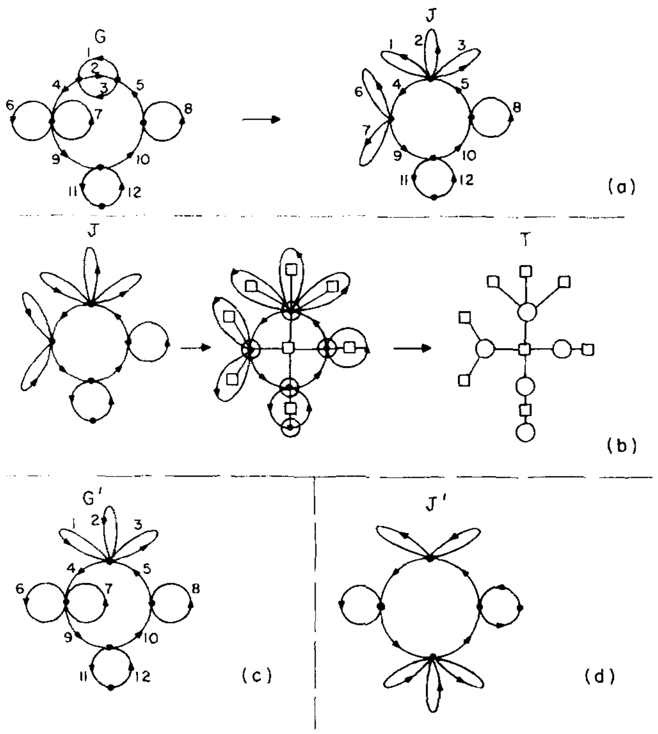



FIG. 5. Transformation of Lee and Yang's primary 0-graphs to dual trees. (a) From labeled graph $G$ to circuit graph $J$. The upper two vertices in $G$ are well connected and coalesce in $J$. (b) From circuit graph $J$ to dual tree $T$ (labeling not shown). (c) Another graph $G'$ that yields the same $J$. (d) Another circuit graph $J'$ that yields the same $T$.

The derivation of Eq. (45) is now straightforward. From (2) with (46) we have

$$\mathcal{J}(C)=\sum_{n=1}^{\infty}\sum_{k} z^n C(k)^n/n=\sum_{k}\ln[1-zC(k)]^{-1} \tag{47}$$

and hence, from (6),

$$\bar{S}(k)=z/[1-z\bar{C}(k)]. \tag{48}$$

Now, $\bar{S}$ is just the $M$ of Lee and Yang. This can be seen either from Eq. (A3) of Appendix A or from Eq. (4), noting that $M(k)$ is the functional derivative of the logarithm of the partition function with respect to a $k$-dependent point insertion in the propagator.

It follows, from (1) with our definition of $C_n$, that

$$\mathcal{D}(\bar{S})=\sum_{n}\frac{1}{n!}\sum_{I}^{n}\sum_{k_1\cdots k_n} v_I(k_1\cdots k_n)\prod_{i=1}^{n}M(k_i) \tag{49}$$

where $\sum_I^n$ denotes the sum over all irreducible graphs $I$ having $n$ lines. But (49) is just the quantity $\rho'$ appearing in (45), except that $\rho'$ is restricted to graphs with no 1-vertex. However, the only *irreducible* graph that has 1-vertex is the one consisting of one line, and its contribution to $\mathcal{D}(\bar{S})$ is

$$\sum_{k} M(k)Y_1(k)=\sum_{k} M(k)[z^{-1}-m(k)^{-1}] \tag{50}$$

on account of (44). Therefore

$$\mathcal{D}(\bar{S})=\rho'+\sum_{k} M(k)[z^{-1}-m(k)^{-1}]. \tag{51}$$

If we write (47) and (48) in terms of $\bar{S}$—that is, $M$—we obtain

$$\mathcal{J}(\bar{C})=\sum_{k}\ln[z^{-1}M(k)] \tag{52}$$

and

$$\bar{C}(k)=z^{-1}-\bar{S}(k)^{-1}=z^{-1}-M(k)^{-1}. \tag{53}$$

Substituting the last three equations into (7), we have

$$\begin{aligned}\tau&=\sum_{k}\ln[z^{-1}M(k)]+\rho'+\sum_{k}M(k)[z^{-1}-m(k)^{-1}]\\&\quad-\sum_{k}M(k)[z^{-1}-M(k)^{-1}]\\&=\sum_{k}\ln[z^{-1}M(k)]+\rho'+\sum_{k}[1-m(k)^{-1}M(k)]\end{aligned} \tag{54}$$

which is equivalent to (45) since $\tau$ is the logarithm of the partition function.

In the discussion of Lee and Yang, a preliminary reduction was carried out by which the 1-vertices were eliminated. We have deliberately bypassed this reduction in order to show the power of the dual tree method. However, we can obtain Eq. (45) just as well by applying the Corollary to Theorem 1 after the reduction has been carried out.

The effect of the reduction is to replace $Y_1$ with 0 and $z$ with $m(k)$ in evaluating the graphs. The sum of graphs found in this way must be augmented by a term we shall call $\rho_1$ to obtain the logarithm of the partition function. $\rho_1$ is the contribution of those original graphs containing only 1-vertices. Thus

$$\begin{aligned}\rho_1&=\sum_{1}^{\infty}\sum_{k}z^nY_1(k)^n=\sum_{k}\ln[1-zY_1(k)]^{-1}\\&=\sum_{k}\ln[z/m(k)]^{-1}\end{aligned} \tag{55}$$

on account of (44).

In applying the Corollary to Theorem 1, we note that $\mathcal{D}(S)$ no longer receives the contribution of Eq. (50) since $Y_1$ has been replaced with 0. Therefore (51) becomes

$$\mathcal{D}(\bar{S})=\rho'. \tag{56}$$

Likewise, we must replace $z$ with $m(k)$ in (52) and (53). However, $\bar{S}$ is still the same as $M$. The expression for the logarithm of the partition function is now

$$\begin{aligned}\rho_1+\tau&=\sum_{k}\ln[z/m(k)]^{-1}+\sum_{k}\ln[M(k)/m(k)]\\&\quad+\rho'-\sum_{k}M(k)[m(k)^{-1}-M(k)^{-1}]\\&=\sum_{k}\ln[M(k)/z]+\rho'+\sum_{k}[1-m(k)^{-1}M(k)]\end{aligned} \tag{57}$$

which is the same as obtained before.

## VII. SUMMARY

It must be understood that we do not claim to have obtained results unknown before, or even to have found shorter proofs *ab initio*. The frequent references to earlier treatments should make it clear that the proofs given here are not original.

What is original is the explicit unification of diverse theorems into a single statement which is proved once and for all. The statement of Theorem 1 is aesthetically simple and easily memorized. Once it has been mastered and the flexibility of the dual tree concept is understood, one may derive a variety of results with great facility, each of which would otherwise require a separate argument of some complexity. The applications include not only those exhibited in the preceding four sections, but others known to the author, and doubtless unknown theorems as well, pertaining to graphical expansions yet to be studied.


## ACKNOWLEDGMENTS

My thanks are due to J.M. Luttinger for introducing me to this subject and to G.C. Wick for suggesting the application in Sec. IV. I am grateful also to M. Weinstein for an interesting discussion and to G.C. Wick and G. Parisi for critical comments.


## APPENDIX A

We shall first prove an auxiliary statement analogous to Eq. (5), namely

$$\int_\xi \bar{C}(\xi)\bar{S}(\xi)=\sum_T n_L w_T \tag{A1}$$

where $n_L$ is the number of lines in $T$. By a similar argument, slightly more complicated, we shall prove Eq. (5). Once (A1) and (5) are established, Eq. (7) follows immediately because of the relation

$$n_S+n_C-n_L=1 \tag{A2}$$

true of any dual tree.

[For completeness, we note that (A2) can be proved by induction on $n_L$. If $n_L=0$ it is obviously true. If $n_L>0$, remove an arbitrary line, and the tree will fall into two parts, having $n_L^{(1)}$ and $n_L^{(2)}$ lines, respectively, where $n_L^{(1)}+n_L^{(2)}+1=n_L$. Since (A2) can be assumed true for each part, the number of squares and circles in the first part is $n_L^{(1)}+1$, and in the second is $n_L^{(2)}+1$, so that for the whole tree $n_S+n_C=(n_L^{(1)}+1)+(n_L^{(2)}+1)=n_L^{(1)}+n_L^{(2)}+2=n_L+1$. QED

This way of establishing Eq. (7) is the one used by Bloch[11] for the system treated in Sec. VI of this paper.

In order to prove Theorem 1, then, we need only establish (A1), (5), (6). For this purpose we introduce the *rooted tree*, which is a dual tree in which one square or circle (the root) has an extra line projecting from it, the other end of which is unattached. The value of a rooted tree is computed in the same way as that of an ordinary dual tree, except that the $\xi$-variable associated with the extra line is not summed over. Hence the value is a function of one variable $\xi$.

Thus, if $T$ is an unrooted dual tree containing a square with three lines attached (call them $i_1$, $i_2$, $i_3$) and $T_S$ is the rooted tree formed from $T$ by attaching an extra line to this square, then $w_{T_S}(\xi)$ is formed from $w_T$ by replacing the factor $S_3(\xi_{i_1},\xi_{i_2},\xi_{i_3})$ with $S_4(\xi,\xi_{i_1},\xi_{i_2},\xi_{i_3})$.

Let $T_S$ represent any rooted tree having a square for root, and $T_C$ any one having a circle for root. Then we assert that

$$\bar{S}(\xi)=\sum_{T_S} w_{T_S}(\xi),\quad \bar{C}(\xi)=\sum_{T_C} w_{T_C}(\xi) \tag{A3}$$

where $\bar{S}$ and $\bar{C}$ are defined by (4).

To prove this, let $T$ be any dual tree, containing $n_C$ circles in all. Then $\delta w_T/\delta C_1(\xi)$ is the sum of terms, one for each circle of order 1, obtained by replacing the corresponding factor $C_1(\xi_i)$ in $w_T$ with $\delta(\xi_i,\xi)$. (We intend a Kronecker or Dirac $\delta$ function according to the meaning of $\int_\xi$.) But the result of this replacement is precisely the function $(n_C)^{-1}w_{T_S}(\xi)$ where $T_S$ is the rooted tree obtained by *deleting* the circle in question, but leaving the line to which it was attached. The factor $(n_C)^{-1}=(n_C-1)!/n_C!$ arises from the fact that $T_S$ has only $n_C-1$ circles.

In forming $T_S$, it was necessary to relabel the remaining circles; let us specify that this was done so as to preserve their order. Thus, if the deleted circle bore the label $j$, then all circles with labels $<j$ in $T$ have the same labels in $T_S$, while all circles with labels $>j$ in $T$ have those labels decreased by 1 in $T_S$, so that the labeling in $T_S$ is consecutive. Then, if we add up all the terms of $\delta w_T/\delta C_1(\xi)$, and sum over all possible $T$, we shall form every possible $T_S$ just $n_C$ times, one for each value of $j$ from 1 to $n_C$. (By $n_C$, now, we mean one *more* than the number of circles in $T_S$.) Thus

$$\sum_T \frac{\delta w_T}{\delta C_1(\xi)}=\sum_{T_S} n_C\frac{1}{n_C}w_{T_S}(\xi) \tag{A4}$$

which is equivalent to the first part of (A3). The second part is proved similarly.

We now prove (A1). Take any dual tree $T$, and any line in it. If this line is cut in two, we have two rooted trees $T_S$ and $T_C$. Say that $T$ contains $n_S$ squares and $n_C$ circles; $T_S$, $n_S^S$ squares and $n_C^S$ circles; and $T_C$, $n_S^C$ squares and $n_C^C$ circles. Then it is manifest that

$$w_T=\frac{n_S^S!\,n_S^C!}{n_S!}\cdot\frac{n_C^S!\,n_C^C!}{n_C!}\int_\xi w_{T_S}(\xi)w_{T_C}(\xi). \tag{A5}$$

However, $T_S$ and $T_C$ had to be relabeled so as to make each one consecutive. Say that this was done so as to preserve ordering within each set $n_S^S$, $n_S^C$, $n_C^S$, $n_C^C$. Then, if we let $T$ range over all dual trees and cut each line of $T$ in turn, we shall generate each pair $(T_S, T_C)$ just $(n_S!/n_S^S!\,n_S^C!)\cdot(n_C!/n_C^S!\,n_C^C!)$ times, since that is the number of ways to divide the $n_S$ square labels and the $n_C$ circle labels each into subsets of the right size (once this is done, $T$ is uniquely determined by $T_S$ and $T_C$). Hence the prefactors in (A5) cancel out and we have

$$\sum_T n_L w_T=\sum_{T_S}\sum_{T_C}\int_\xi w_{T_S}(\xi)w_{T_C}(\xi), \tag{A6}$$

which in view of (A3) is equivalent to (A1).

The proof of (5) is very similar. Let us define an $S_m$-structure as a dual tree in which a square of order $m$ has been specially marked, and the lines attached to it have been numbered from 1 to $m$. Let the value of an $S_m$-structure be just that of the original tree. Since a dual tree $T$, containing $n_{S_m}$ squares of order $m$, can be converted into an $S_m$-structure in $n_{S_m} m!$ ways, we have

$$\sum_T n_S w_T = \sum_T \sum_m n_{S_m} w_T = \sum_m \sum_T n_{S_m} w_T = \sum_m F_m / m! \tag{A7}$$

where $F_m$ is the sum of the values of all $S_m$-structures.

To find $F_m$, we note that each $S_m$-structure consists of $m$ rooted trees $T_C^i$ $(i = 1 \cdots m)$ whose extra lines are inserted into the marked square. The value of the $S_m$-structure is manifestly

$$\frac{\Pi_1^m n_S^{(i)}!}{n_S!} \frac{\Pi_1^m n_C^{(i)}!}{n_C!} \int_{\xi_1 \cdots \xi_n} S_m(\xi_1 \cdots \xi_m) \prod_1^m w_{T_C^i}(\xi_i) \tag{A8}$$

in an obvious notation. As in the previous arguments, the prefactors are canceled by the number of $S_m$-structures that reduce to the same $m$-tuple $(T_C^1 \cdots T_C^m)$ upon consecutive relabeling. No complication arises from any topological identity among the various $T_C^i$, since they are distinguished from one another once for all by the numbering of the $m$ root-lines.

Aside from this relabeling factor, the sum over $S_m$-structures is equivalent to an independent summation over all the $T_C^i$, and we have

$$F_m = \int_{\xi_1 \cdots \xi_m} S_m(\xi_1 \cdots \xi_m) \prod_1^m \bar{C}(\xi_i). \tag{A9}$$

Substituting into (A7), and comparing with Eq. (1), we have the first half of (5). The second half is proved in the same way.

It remains only to prove (6). We define a rooted $S_m$-structure as a rooted tree whose root is a square of order $m+1$ ($m$, without the extra line) and in which the $m$ interior lines attached to the root have been numbered from 1 to $m$. Its value is that of the rooted tree. It is then manifest, from (A3), that

$$\bar{S}(\xi) = \sum_m F'_m(\xi)/m! \tag{A10}$$

where $F'_m$ is the sum of the values of all rooted $S_m$-structures. But an argument identical to the one that led to (A9) gives

$$F'_m(\xi) = \int_{\xi_1 \cdots \xi_m} S_{m+1}(\xi, \xi_1 \cdots \xi_m) \prod_1^m \bar{C}(\xi_i). \tag{A11}$$

Substituting (A11) into (A10), and replacing $m$ by $n-1$, we have

$$\bar{S}(\xi) = \sum_{n=1}^{\infty} (n-1)! \int_{\xi_1 \cdots \xi_{n-1}} S_n(\xi, \xi_1 \cdots \xi_{n-1}) \prod_1^{n-1} \bar{C}(\xi_i). \tag{A12}$$

On the other hand, if we evaluate $\delta \mathcal{J}(C)/\delta C(\xi)$ directly from Eq. (2) and set $C = \bar{C}$, we obtain the right side of (A12). This proves half of (6); the other half is proved in the same way. The proof of Theorem 1 is now complete.

## APPENDIX B

Here we shall give an alternate proof of Theorem 1, not using rooted trees but emphasizing the concept of infinitesimal variation. With respect to Eqs. (5) and (6), this proof may be regarded as a rewording of the one given in Appendix A, but with respect to Eq. (7) it is essentially different. In Appendix A the proof of (7) depended on (5) and (A1) but not on (6). Here it will depend on (6) but not directly on (5) and not at all on (A1).

Our proofs rest on two preliminary assertions:

$$\frac{\delta \tau}{\delta C_1(\xi)} = \sum_m \int_{\xi_1 \cdots \xi_m} S_{m+1}(\xi, \xi_1 \cdots \xi_m) \frac{\delta \tau}{\delta S_m(\xi_1 \cdots \xi_m)} \tag{B1}$$

and similarly with $S$, $C$ interchanged; and

$$\frac{\delta \tau}{\delta S_m(\xi \cdots \xi_m)} = m!^{-1} \prod_1^m \bar{C}(\xi_i) \tag{B2}$$

and similarly with $S$, $C$ interchanged.

It must be understood, since $S_m$ is restricted to be symmetric in its arguments, that $\delta\tau/\delta S_m(\xi_1 \cdots \xi_m)$ is also to be symmetric in the $\xi$'s and satisfy

$$\delta \tau = \int_{\xi_1 \cdots \xi_m} \delta S_m(\xi_1 \cdots \xi_m) \frac{\delta \tau}{\delta S_m(\xi_1 \cdots \xi_m)} \tag{B3}$$

for arbitrary *symmetric* infinitesimal functions $\delta S_m$.

To prove (B1), suppose that $C_1(\xi)$ is incremented infinitesimally by $\delta C_1(\xi)$. Then the first-order increment in $w_T$ is the sum of terms, each obtained by choosing a circle of order 1 and replacing $C_1$ with $\delta C_1$ for that circle. Each such term could also be obtained from a smaller tree $T'$, from which the circle in question and its attached line have been deleted, by replacing $S_m(\xi_1 \cdots \xi_m)$ for the square at the other end of that line with $(n_C)^{-1} \int_\xi S_{m+1}(\xi, \xi_1 \cdots \xi_m) \delta C_1(\xi)$ where $n_C$ is the number of circles in $T$.

Now if one sums over all terms coming from all choices of $T$, one obtains each $T'$, with each square singled out, just $n_C$ times [see the proof of (A3) in the previous appendix] so that the factor $n_C$ cancels out and one has

$$\delta \tau = \sum_m \int_{\xi, \xi_1 \cdots \xi_m} \frac{\delta \tau}{\delta S(\xi_1 \cdots \xi_m)} S_{m+1}(\xi, \xi_1 \cdots \xi_m) \delta C_1(\xi) \tag{B4}$$

which yields (B1) upon elimination of the arbitrary increment $\delta C_1$.

To prove (B2), we note that $S_m$ can have only symmetric increments and that any symmetric function can be expressed as a linear combination of terms of the form $\Pi_1^m f(\xi_i)$. Therefore we need only consider an increment

$$\delta S_m(\xi_1 \cdots \xi_m) = \prod_1^m f(\xi_i) \tag{B5}$$

where $f$ is infinitesimal.

The resulting increment in $w_T$ has a term for each square of order $m$, consisting of a product of $m$ factors. Each factor is (apart from factorials) the value of a tree obtained by severing one line from the square, retaining everything on the other side of the line, and attaching the factor $f(\xi_i)$ to the severed end.

In summing over $T$ and over the choice of square, the factorials cancel out and leave a factor $1/m!$, as in the proof of (5) in Appendix A. However, the sum over each one-line factor is of the form $\int_\xi f(\xi) P(\xi)$ where $P$ is independent of $m$. It follows that

$$\frac{\delta\tau}{\delta S_m(\xi_1\cdots\xi_m)}=m!^{-1}\prod_1^m P(\xi_i) \tag{B6}$$

and by setting $m=1$ and using (4), we have $P=\bar{C}$, so that (B2) follows.

With (B1) and (B2) established, we can easily prove Theorem 1. First, substitute (B2) into (B1) and replace $m$ with $n-1$. This yields (A12) from which half of (6) follows as in Appendix A. The other half is proved similarly.

Next, we introduce the short notation $S_m(\delta/\delta S_m)$ for $\int_{\xi_1\cdots\xi_m} S_m(\xi_1\cdots\xi_m)[\delta/\delta S_m(\xi_1\cdots\xi_m)]$. Multiplying (B2) by $S_m$, summing over the $\xi$'s and summing over $m$, we have, with the use of (2),

$$\sum_m S_m\frac{\delta\tau}{\delta S_m}=\mathcal{J}(\bar{C}). \tag{B7}$$

On the other hand, if a tree $T$ has $n_S$ squares then $w_T$ is homogeneous of degree $n_S$ in $S_0, S_1, \cdots$, so that by Euler's theorem

$$\sum_m S_m\frac{\delta w_T}{\delta S_m}=n_S w_T. \tag{B8}$$

Summing over $T$, and comparing with (B7), we have half of Eq. (5); the other half is proved similarly.

Finally, let us define the quantity

$$\sigma\equiv\mathcal{J}(\bar{C})+\mathcal{D}(\bar{S})-\int_\xi \bar{S}(\xi)\bar{C}(\xi) \tag{B9}$$

and consider its variation when the $S_m$ are changed infinitesimally. This variation has two parts: one part due to the variation of $\bar{S}$ and $\bar{C}$, which depend on the $S_m$ through Eqs. (3) and (4), and the other part due to the direct role of the $S_m$ in the definition of $\mathcal{J}$, Eq. (2). But the first part is just

$$\int_\xi\frac{\delta\mathcal{J}}{\delta\bar{C}(\xi)}\delta\bar{C}(\xi)+\int_\xi\frac{\delta\mathcal{D}}{\delta\bar{S}(\xi)}\delta\bar{S}(\xi)-\int_\xi[\bar{S}(\xi)\delta\bar{C}(\xi)+\bar{C}(\xi)\delta\bar{S}(\xi)]$$
$$=0 \tag{B10}$$

on account of (6). In other words, (B9) is completely stationary when regarded as a functional on $\bar{C}$ and $\bar{S}$, the $S_m$ and $C_m$ being fixed and Eq. (4) being disregarded in the variation but satisfied at the stationary point.

It follows that for all $m$ we have

$$\begin{aligned} S_m\frac{\delta\sigma}{\delta S_m}&=S_m\left(\frac{\delta\sigma}{\delta S_m}\right)_{\bar{S},\bar{C}}\\ &=S_m\left(\frac{\delta\mathcal{J}(\bar{C})}{\delta S_m}\right)_{\bar{S},\bar{C}}. \end{aligned} \tag{B11}$$

We sum over $m$, noting that $\mathcal{J}(\bar{C})$ for fixed $\bar{C}$ is a linear homogeneous functional on the $S_m$, and obtain

$$\begin{aligned} \sum_m S_m\frac{\delta\sigma}{\delta S_m}&=\mathcal{J}(\bar{C})\\ &=\sum_m S_m\frac{\delta\tau}{\delta S_m} \end{aligned} \tag{B12}$$

on account of (B7).

Now let all the $S_m$ be multiplied by the parameter $u$, varying from 0 to 1. Let the resulting values of Eqs. (3) and (B9) be called $\tau(u)$ and $\sigma(u)$. Then clearly

$$u\frac{d}{du}=\sum_m S_m\frac{\delta}{\delta S_m}$$

and (B12) becomes

$$\frac{d\tau(u)}{du}=\frac{d\sigma(u)}{du} \tag{B13}$$

for all $u$.

If $u=0$ then all the $S_m$ vanish, and the only tree that contributes to $\tau$ is the one with a circle and no square:

$$\tau(0)=C_0. \tag{B14}$$

Putting (B14) into (4), we find that $\bar{S}=0$ and hence

$$\sigma(0)=\mathcal{D}(0)=C_0 \tag{B15}$$

so that

$$\tau(0)=\sigma(0). \tag{B16}$$

Integrating (B13) from 0 to 1, and adding (B16), we have

$$\tau(1)=\sigma(1) \tag{B17}$$

which is equivalent to Eq. (7) since putting $u=1$ restores the original values of the $S_m$.

The proof of Theorem 1 is now complete. This method of deriving Eq. (7) is essentially the same as that given by Lee and Yang[8] for the quantum statistical system treated in Sec. VI of this paper, and adapted by Luttinger and Ward[9] for the field theoretic formulation. A proof along these lines was given by Uhlenbeck and Ford[7] for the Mayer cluster expansion dealt with in Sec. V.

## APPENDIX C

We shall prove the topological lemma stated in Sec. VI.

Let $i$ be a line in the family $F$. If $F$ contains no other line, the head of $i$ is well connected to its tail, and the lemma is satisfied. Otherwise, let $j$ be another member of $F$. When $i$ and $j$ are removed, the graph falls into two parts $A$ and $B$, one of which (say $A$) contains the head of $i$ (call it $h_i$) and the tail of $j$ (call it $t_j$).

Now $h_i$ and $t_j$ are connected through $A$ and also outside of $A$. Hence if they are not well connected, they can be separated by the removal of two lines just one of which (call it $j'$) lies in $A$. Now if we remove only $j'$, $h_i$ and $t_j$ are connected neither through $A$ (since $j'$ is removed) nor outside of $A$ (since $i$ is removed). Therefore $j'$ is also related to $i$, and when $i$ and $j'$ are removed the graph falls into two parts one of which $(A')$ is a subset of $A$ and contains $h_i$ and $t_{j'}$. Replacing $j$, $A$ with $j'$, $A'$ in the foregoing argument, we continue until we arrive at a line $i'$ that belongs to $F$ and whose tail is well connected to the head of $i$.

We may now replace $i$ with $i'$ in the foregoing and develop a sequence $i, i', i'', \cdots$ of members of $F$, each having its head well connected with the tail of the next. Eventually the sequence must repeat, and therefore it contains a circuit $H$. It remains only to prove that $H$ contains all the members of $F$, and that its ordering is unique.

Let $p$ be a member of $H$, and $q$ be a line not in $H$. The head of $p$ is well connected to the tail of $p'$, which follows $p$ in $H$. Therefore this connection can be made so that it does not pass through $p$ or $q$. Likewise the head of $p'$ can be connected to the tail of $p''$ without passing through $p$ or $q$. In this way we can find a continuous path from the head of $p$ to its tail, passing neither through $p$ nor through $q$. But if $p$ and $q$ were related, this would be impossible, since it permits us to increment $k_p$ and not $k_q$. Hence $H$ contains all the members of $F$.

The ordering in $H$ is unique apart from cyclic permutations. For if the head of $p$ were well connected to the tail of $p^{(n)}$, a circuit $H'$ could be formed by deleting the lines $p', p'', \ldots, p^{(n-1)}$ from $H$. Then, if $p^{(n)} \neq p'$, $H'$ would not contain all the members of $F$, which has been proved impossible.

*This research was supported in part by the U.S. Atomic Energy Commission.

[1]Equivalently, we could drop the labels on the squares and circles—thereby reducing the number of distinct trees—and calculate $w(T)$ by dividing not by $n_S!n_C!$ but by $\sigma(T)$, the symmetry number of the tree. For example, the twelve trees in Fig. 1 would reduce to two, each having $\sigma=2$, and their total contribution to $\tau$ would be the same as in our definition. That method is more convenient for evaluating $\tau$, but the one we choose is equivalent and seems better adapted to proving theorems.

Our definitions allow trees consisting of a single square or circle and having the value $S_0$ or $C_0$. However, the theorem loses no generality in practice if it is restricted to $S_0 = C_0 = 0$, which means that these two degenerate trees are omitted.

[2]T.D. Lee and C.N. Yang, Phys. Rev. 117, 12 (1960), Appendix C.

[3]It is known that $A_2 = (l-1)!\,l^{l-2}$, but this fact is not used either here or in Ref. 2.

[4]G. Jana-Lasinio, Nuovo Cimento **34**, 1790 (1964).

[5]H. Ursell, Proc. Camb. Phil. Soc. **23**, 685 (1927); J.E. Mayer, J. Chem. Phys. **5**, 67 (1937).

[6]J.E. Mayer and P.F. Ackermann, J. Chem. Phys. **5**, 74 (1937), especially Eqs. (9)ff, (17), and Note in Proof; J.E. Mayer and S.F. Harrison, J. Chem. Phys. **6**, 87 (1938), especially Appendix.

[7]G.E. Uhlenbeck and G.W. Ford, *Studies in Statistical Mechanics* edited by DeBoer and Uhlenbeck (Interscience, New York, 1962), Vol. 1, pp. 123ff, Secs. III. 2, 3.

[8]T.D. Lee and C.N. Yang, Phys. Rev. **117**, 22 (1960).

[9]J.M. Luttinger and J. Ward, Phys. Rev. **118**, 1417 (1960).

[10]C. Bloch, *Studies in Statistical Mechanics* edited by DeBoer and Uhlenbeck (Interscience, New York, 1964), Vol. 3, pp. 7ff, and references cited therein.

[11]See Ref. 10, Sec. 4.21. The connection with the Mayer cluster expansion is indicated on p. 133.

# A new derivation of some fluctuation theorems in statistical mechanics

David J. Vezzetti

*Physics Department, University of Illinois at Chicago Circle, Chicago, Illinois 60680*

(Received 17 June 1974)

A simple derivation is given of some of the fluctuation theorems of statistical mechanics which relate integrals of molecular distribution functions to thermodynamic properties. The derivation employs the generating function for the probability $P_\omega(n)$ that a domain $\omega$ contains $n$ particles. Various forms of the generating function are derived, and each leads to a different form of the fluctuation theorems.

## I. INTRODUCTION

In classical statistical mechanics there exist a number of well-known identities, commonly called fluctuation theorems, which relate integrals of various molecular distribution functions to the thermodynamic properties of the system. Derivations of these theorems usually proceed by functional expansions (or the use of testing functions), by graphical techniques or by combinatorial analysis. [1]

In this brief note, we give a new derivation of these theorems which utilizes only the most elementary ideas of probility theory. The derivation is based on the fact that the probability $P_\omega(n)$ that a domain $\omega$ of space contains exactly $n$ particles is simply expressed in terms of the various distribution functions. The generating function for the $P_\omega(n)$ is then related to the grand partition function of the system and the theorems follow by differentiation.

Recently, Kac and Luttinger[2] gave a derivation of an expression which relates the pressure of a system to the probability that a domain is empty, i.e., $P_\omega(0)$. The present note essentially extends that work in that the fluctuation theorems require the use of the $P_\omega(n)$ for all $n$.

## II. BASIC DERIVATION

Consider a system of particles described by a grand canonical ensemble at temperature $(k\beta)^{-1}$, absolute activity $z$ and confined to a domain $\Omega$ of space. The modified molecular distribution functions[3] of the system are defined by

$$\hat{\rho}_l(\mathbf{x}_1,\ldots,\mathbf{x}_l\mid[\Omega],\beta,z)\equiv\left\langle\sum_{i_1,i_2,\ldots}\delta(\mathbf{x}_1-\mathbf{r}_{i_1})\ldots\delta(\mathbf{x}_l-\mathbf{r}_{i_l})\right\rangle, \tag{1}$$

where the average is taken over the grand canonical ensemble. Now let $\omega$ be a subdomain of $\Omega$. Then, from Eq. (1), we have immediately[4]

$$\int_\omega \hat{\rho}_l(\mathbf{x}_1,\ldots,\mathbf{x}_l\mid z)\,d^3x_1\ldots d^3x_l=\langle n_\omega^l\rangle, \tag{2}$$

where $n_\omega$ is the number of particles in $\omega$. Next, we let $P_\omega(n)$ be the probability that exactly $n$ particles are in $\omega$ and form the generating function of the $P_\omega(n)$. Thus,

$$\begin{aligned}f_\omega(\xi)&\equiv\sum_{n=0}^\infty \xi^n P_\omega(n)=\sum_{n=0}^\infty\sum_{l=0}^\infty\frac{(n\ln\xi)^l}{l!}P_\omega(n)\\&=\sum_{l=0}^\infty\frac{(\ln\xi)^l}{l!}\sum_{n=0}^\infty n^l P_\omega(n)\\&=\sum_{l=0}^\infty\frac{(\ln\xi)^l}{l!}\langle n_\omega^l\rangle=1+\sum_{l=1}^\infty\frac{(\ln\xi)^l}{l!}\\&\quad\times\int_\omega\hat{\rho}_l(\mathbf{x}_1,\ldots,\mathbf{x}_l\mid z)\,d^3x_1\ldots d^3x_l.\end{aligned} \tag{3}$$

This gives the generating function in terms of the modified molecular distribution functions.

We now introduce the modified Ursell functions $\hat{\mathcal{J}}_l$ by their relation to the modified molecular distribution functions:

$$\begin{aligned}&\hat{\rho}_1(\mathbf{x}_1\mid z)=\hat{\mathcal{J}}_1(\mathbf{x}_1\mid z);\;\hat{\rho}_2(\mathbf{x}_1,\mathbf{x}_2\mid z)=\hat{\mathcal{J}}_2(\mathbf{x}_1,\mathbf{x}_2\mid z)+\hat{\mathcal{J}}_1(\mathbf{x}_1\mid z)\\&\quad+\hat{\mathcal{J}}_1(\mathbf{x}_2\mid z);\;\hat{\rho}_3(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3\mid z)=\hat{\mathcal{J}}_3(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3\mid z)\\&\quad+\hat{\mathcal{J}}_2(\mathbf{x}_1,\mathbf{x}_2\mid z)\hat{\mathcal{J}}_1(\mathbf{x}_3\mid z)+\hat{\mathcal{J}}_2(\mathbf{x}_1,\mathbf{x}_3\mid z)\hat{\mathcal{J}}_1(\mathbf{x}_2\mid z)\\&\quad+\hat{\mathcal{J}}_2(\mathbf{x}_2,\mathbf{x}_3\mid z)\hat{\mathcal{J}}_1(\mathbf{x}_1\mid z)\\&\quad+\hat{\mathcal{J}}_1(\mathbf{x}_1\mid z)\hat{\mathcal{J}}_1(\mathbf{x}_2\mid z)\hat{\mathcal{J}}_1(\mathbf{x}_3\mid z),\text{ etc.}\end{aligned} \tag{4}$$

In general, $\hat{\rho}_l$ is a sum of products of the form $\hat{\mathcal{J}}_{i_1}\hat{\mathcal{J}}_{i_2}\ldots$, such that the indices $i_1, i_2,\ldots$ sum to $l$ and the $l$ coordinates $\mathbf{x}_1,\ldots,\mathbf{x}_l$ are distributed among the $\hat{\mathcal{J}}_i$ in all possible ways. Since, however, we require only the integral of $\hat{\rho}_l$ for our derivation, we have

$$\begin{aligned}&\int_\omega\hat{\rho}_l(\mathbf{x}_1,\ldots,\mathbf{x}_l\mid z)\,d^3x_1,\ldots d^3x_l\\&\quad=\sum_{\substack{i_1,i_2,\ldots=0\\(\sum_k k i_k=l)}}^{l}\frac{l!}{i_1!\,i_2!\ldots}\left(\frac{1}{1!}\int_\omega\hat{\mathcal{J}}_1(\mathbf{x}_1\mid z)\,d^3x_1\right)^{i_1}\\&\quad\times\left(\frac{1}{2!}\int_\omega\hat{\mathcal{J}}_2(\mathbf{x}_2,\mathbf{x}_3\mid z)\,d^3x_2\,d^3x_3\right)^{i_2}\ldots\end{aligned} \tag{5}$$

Using this in Eq. (3), the sum over $l$ yields the exponential function and we thus obtain for the generating function

$$f_\omega(\xi)=\exp\left(\sum_{l=1}^\infty\frac{(\ln\xi)^l}{l!}\int_\omega\hat{\mathcal{J}}_l(\mathbf{x}_1,\ldots,\mathbf{x}_l\mid z)\,d^3x_1\ldots d^3x_l\right). \tag{6}$$

It is now only a brief step to the first fluctuation theorem, since if we set $\omega=\Omega$ then the $P_\Omega(n)$ are given by

$$P_\Omega(n)=\frac{z^n Z_n([\Omega],\beta)}{Q(z,[\Omega],\beta)} \tag{7}$$

where $Z_n$ and $Q$ are the canonical and grand canonical partition functions. Now forming the generating function for the $P_\Omega(n)$ as given by Eq. (7) and using Eq. (6), we have

$$f_\Omega(\xi)=\frac{Q(\xi z)}{Q(z)}$$
$$=\exp\left(\sum_{l=1}^{\infty}\frac{(\ln\xi)^l}{l!}\int_\Omega \hat{\mathcal{J}}_l(\mathbf{x}_1,\ldots,\mathbf{x}_l\mid z)\,d^3x_1\ldots d^3x_l\right). \tag{8}$$

Differentiating the logarithm of Eq. (8) with respect to $\ln\xi$ and setting $\xi=1$, we obtain

$$\frac{\partial^s \ln Q(z)}{\partial(\ln z)^s}=\int_\Omega \hat{\mathcal{J}}_s(\mathbf{x}_1,\ldots,\mathbf{x}_s\mid z)\,d^3x_1\ldots d^3x_s, \tag{9}$$

which is the desired theorem. For $s=1$ and 2, Eq. (9) yields the mean number of particles and the compressibility equation of state in the well-known way. For $s=3,\ldots,$ higher derivatives of the compressibility are expressed in terms of the integrals of higher molecular distribution functions.

A related theorem can be derived by using the form of the generating function given by Eq. (3). Thus, differentiating with respect to $\ln\xi$ and setting $\xi=1$, we obtain

$$\frac{1}{Q(z)}\frac{\partial^s Q(z)}{\partial(\ln z)^s}=\int_\Omega \hat{\rho}_s(\mathbf{x}_1,\ldots,\mathbf{x}_s\mid z)\,d^3x_1\ldots d^3x_s. \tag{10}$$

## III. OTHER FORMS OF THE GENERATING FUNCTION

Two additional forms of the generating function and the associated fluctuation theorems can be derived by using the ordinary molecular distribution functions and Ursell functions. For these, it is easiest to begin in a canonical ensemble of $N$ particles. The ordinary molecular distribution functions are defined by[5]

$$\rho_l^{(N)}(\mathbf{x}_1,\ldots,\mathbf{x}_l)=\left\langle \sum_{i_1,\ldots,i_l=1}^{N}{}' \delta(\mathbf{x}_1-\mathbf{r}_{i_1})\ldots\delta(\mathbf{x}_l-\mathbf{r}_{i_l})\right\rangle. \tag{11}$$

Here, the prime on the summation indicates that no two of the indices $i_1, i_2,\ldots$ are equal. We now introduce the characteristic function of the domain $\omega$ by

$$G_\omega(\mathbf{x})=\begin{cases}1, & \mathbf{x}\in\omega\\ 0, & \mathbf{x}\in\omega.\end{cases} \tag{12}$$

We then find for the probability of $n(\leq N)$ particles in $\omega$:

$$P_\omega^{(N)}(0)=\left\langle\prod_{i=1}^{N}(1-G_\omega(\mathbf{x}_i))\right\rangle=1+\sum_{l=1}^{N}\frac{(-1)^l N!}{l!(N-l)!}\times\left\langle G_\omega(\mathbf{x}_1)\ldots G_\omega(\mathbf{x}_l)\right\rangle$$
$$=1+\sum_{l=1}^{N}\frac{(-1)^l}{l!}\int_\omega \rho_l^{(N)}(\mathbf{x}_1,\ldots,\mathbf{x}_l)\,d^3x_1\ldots d^3x_l \tag{13a}$$

and

$$P_\omega^{(N)}(n)=\frac{N!}{n!(N-n)!}\left\langle\prod_{i=1}^{n}G_\omega(\mathbf{x}_i)\prod_{j=n+1}^{N}(1-G_\omega(\mathbf{x}_j))\right\rangle$$
$$=\sum_{l=n}^{N}\frac{(-1)^{l+n}}{n!(l-n)!}\int_\omega \rho_l^{(N)}(\mathbf{x}_1,\ldots,\mathbf{x}_l)\,d^3x_1\ldots d^3x_l \quad (n=1,\ldots,N). \tag{13b}$$

The generating function (still in the canonical ensemble) is now

$$f_\omega^{(N)}(\xi)=\sum_{n=0}^{N}\xi^n P_\omega^{(N)}(n)=1+\sum_{l=1}^{N}\frac{(\xi-1)^l}{l!}\times\int_\omega \rho_l^{(N)}(\mathbf{x}_1,\ldots,\mathbf{x}_l)\,d^3x_1\ldots d^3x_l. \tag{14}$$

We can now go over to a grand canonical ensemble by multiplying Eq. (14) by $P_\Omega(N)$ [as given by Eq. (7)] and summing over $N$. Thus,

$$f_\omega(\xi)=1+\sum_{l=1}^{\infty}\frac{(\xi-1)^l}{l!}\int_\omega \rho_l(\mathbf{x}_1,\ldots,\mathbf{x}_l\mid z)\,d^3x_1\ldots d^3x_l. \tag{15}$$

Now the Ursell functions $\mathcal{J}_l$ are defined from the $\rho_l$ in the same manner as the $\hat{\mathcal{J}}_l$ are defined from the $\hat{\rho}_l$. Thus, using the same procedure as before, we have

$$f_\omega(\xi)=\exp\left(\sum_{l=1}^{\infty}\frac{(\xi-1)^l}{l!}\int_\omega \mathcal{J}_l(\mathbf{x}_1,\ldots,\mathbf{x}_l\mid z)\,d^3x_1\ldots d^3x_l\right). \tag{16}$$

From these two forms of the generating function, two fluctuation theorems are easily obtained by setting $\omega=\Omega$, differentiating with respect to $\xi$ and setting $\xi=0$. Thus, from Eq. (15)

$$\frac{z^s}{Q(z)}\frac{\partial^s Q(z)}{\partial z^s}=\int_\Omega \rho_s(\mathbf{x}_1,\ldots,\mathbf{x}_s\mid z)\,d^3x_1\ldots d^3x_s, \tag{17}$$

and from Eq. (16)

$$z^s\frac{\partial^s \ln Q(z)}{\partial z^s}=\int_\Omega \mathcal{J}_s(\mathbf{x}_1,\ldots,\mathbf{x}_s\mid z)\,d^3x_1\ldots d^3x_s. \tag{18}$$

Thus, each of the four forms of the generating function [Eqs. (3), (6), (15), and (16)] leads to a fluctuation theorem.

As a final brief comment, we might add that two expansions for the pressure of the system can be obtained from Eqs. (15) and (16). For, by the usual prescription, we have

$$\beta p(z)=\lim_{V\to\infty}\frac{1}{V}\ln Q(z), \tag{19}$$

where $V$ is the volume of $\Omega$. However, from Eq. (7) we see

$$\ln Q(z)=-\ln P_\Omega(0)=-\ln f_\Omega(\xi=0). \tag{20}$$

Thus, we obtain

$$\beta p(z)=-\lim_{V\to\infty}\frac{1}{V}\sum_{l=1}^{\infty}\frac{(-1)^l}{l!}\int_\Omega \mathcal{J}_l(\mathbf{x}_1,\ldots,\mathbf{x}_l\mid z)\,d^3x_1\ldots d^3x_l \tag{21}$$

or

$$\beta p(z) = -\lim_{V\to\infty} \frac{1}{V}$$

$$\ln \left\{1 + \sum_{l=1}^{\infty} \frac{(-1)^l}{l!} \int_{\Omega} \rho_l(\mathbf{x}_1, \ldots, \mathbf{x}_l \mid z)\, d^3x_1 \ldots d^3x_l. \right. \qquad (22)$$

Equation (20) relating the pressure to the probability that the domain $\Omega$ is empty has been derived recently by Kac and Luttinger[2] in the canonical ensemble. The result in the grand canonical ensemble is, as we see, completely trivial. The resulting derivation of Eqs. (21) and (22), however, appears to be considerably simpler than many of the alternative treatments found in the literature.

[1]See, e.g., J.L. Lebowitz and J.K. Percus, J. Math. Phys. **4**, 1495 (1963); S.A. Rice and P. Gray, *The Statistical Mechanics of Simple Liquids* (Interscience, New York, 1965); J.K. Percus in, *The Equilibrium Theory of Classical Fluids*, edited by H.L. Frisch and J.L. Lebowitz (Benjamin, New York, 1964); M. Green in, *Lectures in Theoretical Physics* (Interscience, New York, 1960), Vol. 3.

[2]M. Kac and J.M. Luttinger, J. Math. Phys. **14**, 583 (1973).

[3]Definitions of the various distribution functions are given by J.K. Percus, Ref. 1 or J.L. Lebowitz and J.K. Percus, Ref. 1.

[4]The dependence of $\hat{\rho}_l$ on $\Omega$ and $\beta$ is surpressed for notational convenience.

[5]The superscript $(N)$ is used to denote functions in a canonical ensemble of $N$ particles. When no superscript is used, as in Sec. II, the functions refer to a grand canonical ensemble.

# Gravitational and electromagnetic radiation in Kerr–Maxwell spaces

Robert W. Lind*†

*Department of Physics, Florida State University, Tallahassee, Florida 32306*
(Received 24 June 1974)

The class of Kerr–Maxwell spaces is defined. This class consists of regular electrovac spacetimes in which a geodesic, diverging, shear-free principal null vector field of the Weyl tensor coincides with a principal null vector field of the Maxwell tensor. It is shown that this class admits no Petrov–Penrose type III or type N solutions. It is also shown that the most general nonradiative solution is the Kerr–Newman metric.

## 1. INTRODUCTION

We have recently shown[1] that a class of solutions to the *real* Maxwell equations exists which can be viewed as arising from a monopole moving along a *complex* world line in *complex* Minkowski space. These solutions are called complexified Lienard—Wiechert solutions (CLW) and are characterized geometrically by the fact that the Maxwell tensor possesses a principal null vector field (p. n. v. f.) which satisfies the following conditions:

(i) The p. n. v. f. is the tangent field to a congruence of null geodesics;

(ii) the p. n. v. f. has non-vanishing divergence;

(iii) the shear of the p. n. v. f. vanishes.

We have also shown[1] that the class of regular, algebraically special type II, twisting (Kerr-type[2]) metrics in Einstein's general theory of relativity is the natural analog of the class of CLW solutions in that a p. n. v. f. of the Weyl tensor satisfies conditions (i)—(iii).

Because of the strong analogy between the CLW Maxwell fields and the Kerr-type gravitational fields it is only natural to unite the two and form the following class of electrovac spacetimes.

*Definition*: A Kerr—Maxwell space is a *regular* electrovac spacetime in which the Maxwell tensor possesses a p. n. v. f. satisfying conditions (i)—(iii).

By a corollary of the Goldberg—Sachs theorem[3] the Weyl tensor is then algebraically special with a degenerate p. n. v. f. coincident with the p. n. v. f. of the Maxwell tensor. If the p. n. v. f. of the Kerr—Maxwell space satisfies a fourth condition in the real space that

(iv) the twist (or curl) of the p. n. v. f. vanishes, then then one obtains the class of regular Robinson—Trautman—Maxwell electrovac solutions.[4]

In the next section we present a summary of the Kerr—Maxwell metrics in the Newman—Penrose[5] formalism and define the condition of regularity. In Sec. 3 we show that the Kerr—Maxwell class possesses no type III (or type N) solutions. In Sec. 4 we show that with the exception of the Kerr-Newman[6] metric (the Kerr, Reissner—Nordström, and Schwarzschild metrics are also included, being special cases of the Kerr—Newman metric) systems described by Kerr—Maxwell solutions must be radiative.

The Kerr—Maxwell spaces can be viewed as representing the gravitational and electromagnetic fields produced by a charged source moving along an arbitrary complex timelike world line in a complex space.[1] However, all of the work in this paper will be done entirely in the real space. Although the complexified Lienard—Wiechert solutions to the Maxwell equations in complex Minkowski space are referred to briefly in Sec. 2, this is only for the purpose of providing additional physical insight into the regularity condition, which is imposed in the real space.

## 2. THE KERR-MAXWELL METRICS

In this section we present a brief review of the spin coefficient formulation of the Kerr—Maxwell metrics.

In a four-dimensional Riemannian manifold with signature $(+,-,-,-)$ a null tetrad $Z_{a\mu}=(l_\mu, n_\mu, m_\mu, \overline{m}_\mu)$ is introduced composed of two real $(l, n)$ and two complex $(m, \overline{m})$ null vectors satisfying

$$l\cdot n = -m\cdot\overline{m}=1, \tag{2.1}$$

all other scalar products vanishing. Equation (2.1) implies the completeness relation

$$g_{\mu\nu}=2(l_{(\mu}n_{\nu)}-m_{(\mu}\overline{m}_{\nu)}). \tag{2.2}$$

From the tetrad one can define the Ricci rotation coefficients

$$\gamma^{abc}=Z^a{}_{\mu;\nu}Z^{b\mu}Z^{c\nu} \tag{2.3}$$

and the spin coefficients

$$\begin{aligned}
&\kappa=l_{\mu;\nu}m^\mu l^\nu, && \nu=-n_{\mu;\nu}\overline{m}^\mu n^\nu,\\
&\rho=l_{\mu;\nu}m^\mu\overline{m}^\nu, && \mu=-n_{\mu;\nu}\overline{m}^\mu m^\nu,\\
&\sigma=l_{\mu;\nu}m^\mu m^\nu, && \lambda=-n_{\mu;\nu}\overline{m}^\mu\overline{m}^\nu,\\
&\tau=l_{\mu;\nu}m^\mu n^\nu, && \pi=-n_{\mu;\nu}\overline{m}^\mu l^\nu,\\
&\alpha=\tfrac{1}{2}(l_{\mu;\nu}n^\mu\overline{m}^\nu-m_{\mu;\nu}\overline{m}^\mu\overline{m}^\nu),\\
&\beta=\tfrac{1}{2}(l_{\mu;\nu}n^\mu m^\nu-m_{\mu;\nu}\overline{m}^\mu m^\nu),\\
&\gamma=\tfrac{1}{2}(l_{\mu;\nu}n^\mu n^\nu-m_{\mu;\nu}\overline{m}^\mu n^\nu),\\
&\epsilon=\tfrac{1}{2}(l_{\mu;\nu}n^\mu l^\nu-m_{\mu;\nu}\overline{m}^\mu l^\nu).
\end{aligned} \tag{2.4}$$

The tetrad components of the Weyl tensor are defined by

$$\begin{aligned}
\Psi_0&=-C_{\mu\nu\rho\sigma}l^\mu m^\nu l^\rho m^\sigma,\\
\Psi_1&=-C_{\mu\nu\rho\sigma}l^\mu n^\nu l^\rho m^\sigma,\\
\Psi_2&=-C_{\mu\nu\rho\sigma}\overline{m}^\mu n^\nu l^\rho m^\sigma,
\end{aligned} \tag{2.5}$$

$$\Psi_3 = -C_{\mu\nu\rho\sigma}\overline{m}^\mu n^\nu l^\rho n^\sigma,$$
$$\Psi_4 = -C_{\mu\nu\rho\sigma}\overline{m}^\mu n^\nu \overline{m}^\rho n^\sigma.$$

Similarly for the Maxwell tensor we have

$$\Phi_0 = F_{\mu\nu} l^\mu m^\nu,$$
$$\Phi_1 = \tfrac{1}{2}F_{\mu\nu}(l^\mu n^\nu + \overline{m}^\mu m^\nu), \tag{2.6}$$
$$\Phi_2 = F_{\mu\nu}\overline{m}^\mu n^\nu.$$

Directional derivatives have the form

$$D\phi \equiv \phi_{;\mu} l^\mu, \quad \Delta\phi \equiv \phi_{;\mu} n^\mu, \quad \delta\phi \equiv \phi_{;\mu} m^\mu, \quad \overline{\delta}\phi \equiv \phi_{;\mu}\overline{m}^\mu. \tag{2.7}$$

The spin coefficient formalism then consists of four sets of first order differential equations (equivalent to the coupled vacuum Einstein–Maxwell equations) for the four sets of variables: the spin coefficients, the Weyl tensor components, the Maxwell tensor components and the tetrad or metric components.

The formalism can be readily adapted to yield the Kerr–Maxwell solutions by simply choosing the null vector field $l$ to be a p.n.v.f. of the Maxwell tensor satisfying conditions (i)–(iii). Thus,

$$(F^{\mu\nu} + i^* F^{\mu\nu})\, l_\nu \propto l^\mu$$

or, equivalently,

$$\Phi_0 = 0, \tag{2.8}$$

and, in terms of spin coefficients, conditions (i)–(iii) become

$$\text{(i) } \kappa = 0, \quad \text{(ii) } \rho + \overline{\rho} \neq 0, \quad \text{(iii) } \sigma = 0. \tag{2.9}$$

These assumptions [(2.8) and (2.9)] in turn lead to the result that $l$ must be a degenerate p.n.v.f. of the Weyl tensor satisfying

$$C_{\mu\nu\rho[\sigma} l_{\tau]} l^\nu l^\rho = 0,$$

or, equivalently,

$$\Psi_0 = \Psi_1 = 0. \tag{2.10}$$

We also introduce a null coordinate system $x^\mu = (u, r, \zeta, \overline{\zeta})$, associated with the null tetrad such that $r$ is an affine parameter along the null geodesics (labelled by $u, \zeta, \overline{\zeta}$) to which $l$ is tangent.

The four sets of equations may now be integrated under the above assumptions. Because the details already appear elsewhere,[7] we simply quote the results here.

The tetrad components of the Weyl tensor have the form

$$\Psi_0 = \Psi_1 = 0, \tag{2.11a}$$
$$\Psi_2 = \Psi_2^0\rho^3 + 2k\Phi_1^0\overline{\Phi}_1^0\overline{\rho}\rho^3, \tag{2.11b}$$
$$\Psi_3 = \Psi_3^0\rho^2 + \Psi_3^1\rho^3 + \Psi_3^2\rho^4 + k\overline{\Phi}_1^0\overline{\rho}(D\Phi_2), \tag{2.11c}$$
$$\Psi_4 = \Psi_4^0\rho + \Psi_4^1\rho^2 + \tfrac{1}{2}\Psi_4^2\rho^3 + \tfrac{1}{3}\Psi_4^3\rho^4 + \tfrac{1}{4}\Psi_4^4\rho^5 + k\overline{\Phi}_1^0\overline{\rho}[\Psi_4^5\rho^2 + \Psi_4^6\rho^3 + \Psi_4^7\rho^4 + \Psi_4^8\rho^5], \tag{2.11d}$$

with

$$\Psi_3^0 = \delta_0 R + L\dot{R}, \tag{2.12a}$$
$$\Psi_3^1 = \overline{\delta}_0\Psi_2^0 + \overline{L}\dot{\Psi}_2^0 + 3\dot{\overline{L}}\Psi_2^0, \tag{2.12b}$$
$$\Psi_3^2 = 3i\Psi_2^0[\overline{\delta}_0\Sigma + (\overline{L}\Sigma)^\cdot], \tag{2.12c}$$
$$\Psi_4^0 = \dot{R}, \tag{2.12d}$$
$$\Psi_4^1 = \overline{\delta}_0\Psi_3^0 + \overline{L}\dot{\Psi}_3^0 + 4\dot{\overline{L}}\Psi_3^0, \tag{2.12e}$$
$$\Psi_4^2 = \overline{\delta}\Psi_3^1 + \overline{L}\dot{\Psi}_3^1 + 5\dot{\overline{L}}\Psi_3^1 + 4i\Psi_3^0[\overline{\delta}_0\Sigma + (\overline{L}\Sigma)^\cdot], \tag{2.12f}$$
$$\Psi_4^3 = \overline{\delta}_0\Psi_3^2 + \overline{L}\dot{\Psi}_3^2 + 6\dot{\overline{L}}\Psi_3^2 + 6i\Psi_3^1[\overline{\delta}_0\Sigma + (\overline{L}\Sigma)^\cdot], \tag{2.12g}$$
$$\Psi_4^4 = 8i\Psi_3^2[\overline{\delta}_0\Sigma + (\overline{L}\Sigma)^\cdot], \tag{2.12h}$$
$$\Psi_4^5 = \overline{\delta}_0\Phi_2^0 + \overline{L}\dot{\Phi}_2^0 + 3\dot{\overline{L}}\Phi_2^0, \tag{2.12i}$$
$$\Psi_4^6 = \overline{\delta}_0\Phi_2^1 + \overline{L}\dot{\Phi}_2^1 + 4\dot{\overline{L}}\Phi_2^1 + 2i\Phi_2^0[\overline{\delta}_0\Sigma + (\overline{L}\Sigma)^\cdot], \tag{2.12j}$$
$$\Psi_4^7 = \overline{\delta}_0\Phi_2^2 + \overline{L}\dot{\Phi}_2^2 + 5\dot{\overline{L}}\Phi_2^2 + 4i\Phi_2^1[\overline{\delta}_0\Sigma + (\overline{L}\Sigma)^\cdot], \tag{2.12k}$$
$$\Psi_4^8 = 6i\Phi_2^2[\overline{\delta}_0\Sigma + (\overline{L}\Sigma)^\cdot]. \tag{2.12l}$$

The tetrad components of the Maxwell tensor are

$$\Phi_0 = 0, \tag{2.13a}$$
$$\Phi_1 = \Phi_1^0\rho^2, \tag{2.13b}$$
$$\Phi_2 = \Phi_2^0\rho + \Phi_2^1\rho^2 + \Phi_2^2\rho^3, \tag{2.13c}$$

with

$$\Phi_2^1 = \overline{\delta}_0\Phi_1^0 + \overline{L}\dot{\Phi}_1^0 + 2\dot{\overline{L}}\Phi_1^0, \tag{2.14a}$$
$$\Phi_2^2 = 2i\Phi_1^0[\overline{\delta}_0\Sigma + (\overline{L}\Sigma)^\cdot]. \tag{2.14b}$$

The spin coefficients become

$$\kappa = \epsilon = \pi = \sigma = \tau = \lambda = 0, \tag{2.15a}$$
$$\rho = -(r + i\Sigma)^{-1}, \tag{2.15b}$$
$$\alpha = \tfrac{1}{2}(\overline{\delta}_0 \log P_0 + 2\dot{\overline{L}})\,\rho, \tag{2.15c}$$
$$\beta = -\tfrac{1}{2}(\delta_0 \log P_0)\,\overline{\rho}, \tag{2.15d}$$
$$\gamma = \tfrac{1}{2}\Psi_2^0\rho^2 + k\Phi_1^0\overline{\Phi}_1^0\overline{\rho}\rho^2, \tag{2.15e}$$
$$\mu = (\delta_0 N + L\dot{N})\,\overline{\rho} + \tfrac{1}{2}\Psi(\rho^2 + \rho\overline{\rho}) + k\Phi_1^0\overline{\Phi}_1^0\overline{\rho}\rho^2, \tag{2.15f}$$
$$\nu = \dot{N} + \Psi_3^0\rho + \tfrac{1}{2}\Psi_3^1\rho^2 + \tfrac{1}{3}\Psi_3^2\rho^3 + k\overline{\Phi}_1^0\overline{\rho}(\Phi_2), \tag{2.15g}$$

with

$$L = -\delta_0\phi/\dot{\phi}, \tag{2.16a}$$
$$2i\Sigma = \delta_0\overline{L} + L\dot{\overline{L}} - \overline{\delta}_0 L - \overline{L}\dot{L}, \tag{2.16b}$$
$$N = \overline{\delta}_0 \log P_0 + \dot{\overline{L}}, \tag{2.16c}$$
$$R = \overline{\delta}_0 N + \overline{L}\dot{N} + N^2 - 2N\overline{\delta}_0 \log P_0, \tag{2.16d}$$

where

$$P_0 = \tfrac{1}{2}(1 + \zeta\overline{\zeta}), \tag{2.16e}$$
$$\phi = \phi(u, \zeta, \overline{\zeta}). \tag{2.16f}$$

The metric may be written in the form

$$ds^2 = 2(ln - m\overline{m}), \tag{2.17}$$

with

$$l = l_\mu dx^\mu = du - (L/2P_0)\, d\zeta - (\overline{L}/2P_0)\, d\overline{\zeta}, \tag{2.18a}$$
$$n = n_\mu dx^\mu$$

$$= dr + (\tfrac{1}{2}P_0)[\dot{\mathring{L}}(r - i\Sigma) + i(\eth_0\Sigma + L\mathring{\Sigma} + 2\mathring{L}\Sigma)]\,d\zeta$$
$$+ (\tfrac{1}{2}P_0)[\mathring{\bar{L}}(r + i\Sigma) - i(\bar{\eth}_0\Sigma + \bar{L}\dot{\mathring{\Sigma}} + 2\dot{\bar{L}}\Sigma)\,d\bar{\zeta}$$
$$+ [1 + \tfrac{1}{2}(\eth_0\dot{\bar{L}} + L\ddot{\bar{L}} + \bar{\eth}_0\mathring{L} + \bar{L}\ddot{L})$$
$$+ \tfrac{1}{2}(\Psi_2^0\rho + \bar{\Psi}_2^0\bar{\rho}) + k\Phi_1^0\bar{\Phi}_1^0\rho\bar{\rho}\,]\,l_\mu dx^\mu, \qquad (2.18\text{b})$$

$$m = m_\mu dx^\mu = -(1/2P_0\rho)\,d\bar{\zeta}, \qquad (2.18\text{c})$$

$$\bar{m} = \bar{m}_\mu dx^\mu = -(1/2P_0\bar{\rho})\,d\zeta. \qquad (2.18\text{d})$$

Finally $\Phi_1^0$, $\Phi_2^0$, and $\Psi_2^0$ must satisfy the differential equations

$$\eth_0\Phi_1^0 + L\mathring{\Phi}_1^0 + 2\mathring{L}\Phi_1^0 = 0, \qquad (2.19\text{a})$$

$$\dot{\Phi}_1^0 = \eth_0\Phi_2^0 + L\mathring{\Phi}_2^0 + \mathring{L}\Phi_2^0, \qquad (2.19\text{b})$$

$$\eth_0\Psi_2^0 + L\dot{\Psi}_2^0 + 3\dot{L}\Psi_2^0 = 2k\Phi_1^0\bar{\Phi}_2^0, \qquad (2.20\text{a})$$

$$\dot{\Psi}_2^0 = \eth_0\Psi_3^0 + L\dot{\Psi}_3^0 + 2\dot{L}\Psi_3^0 - k\Phi_2^0\bar{\Phi}_2^0, \qquad (2.20\text{b})$$

with

$$\text{Im}\left[\Psi_2^0 - \left(\eth_0 + L\frac{\partial}{\partial u}\right)\left(\eth_0 + L\frac{\partial}{\partial u}\right)(\bar{\eth}_0\bar{L} + \bar{L}\dot{\bar{L}})\right] = 0. \qquad (2.21)$$

In all of the preceding results a dot above a quantity denotes $(\partial/\partial u)$ [for example, $\dot{L} = (\partial L/\partial u)$], $\eth_0$ is the operator edth $(\eth)$[8,9] applied to the unit sphere and $k$ is related to the Newtonian gravitational constant $G$ by $k = 2G$.

Before we can obtain the class of Kerr–Maxwell metrics, the further condition of regularity must be imposed. First we introduce the quantity $X(\phi, \zeta, \bar{\zeta})$ defined implicitly from Eq. (2.16f) by

$$u = X(\phi, \zeta, \bar{\zeta}), \qquad (2.22)$$

Then we define the complex function $V(\phi, \zeta, \bar{\zeta})$ by

$$V(\phi, \zeta, \bar{\zeta}) \equiv X'(\phi, \zeta, \bar{\zeta}) \equiv \mathring{\phi}^{-1}(u, \zeta, \bar{\zeta}), \qquad (2.23)$$

where the prime denotes the partial derivative with respect to $\phi$. Now, the regularity condition may be expressed in terms of $V$ as follows:

(v) The function $V(\phi, \zeta, \bar{\zeta})$, defined by Eq. (2.23), must be a smooth, bounded function on the sphere, having no zeros in the real spacetime.

Now we are ready to give another definition of Kerr–Maxwell spaces equivalent to the one given in Sec. 1.

*Definition*: The Kerr–Maxwell space is obtained as follows: Choose one complex function $\phi(u, \zeta, \bar{\zeta})$, such that condition (v) is satisfied, and three complex functions $\Phi_1^0(u, \zeta, \bar{\zeta})$, $\Phi_2^0(u, \zeta, \bar{\zeta})$, and $\Psi_2^0(u, \zeta, \bar{\zeta})$, which are regular solutions to the differential equations (2.19)–(2.21). Then the Weyl tensor, Maxwell tensor, and metric are given by (2.11), (2.13), (2.17), and the auxilliary equations given above.

The regularity condition (v) can be interpreted as placing a physical restriction on the class of solutions, so as to only include those representing gravitational and electromagnetic radiation emanating from *bounded* sources moving along *timelike* world lines. The possibility that the world line in question may be complex is also allowed in order to include sources possessing intrinsic angular momentum.[1,10]

To get a better feeling for what the condition means, let us consider the class of complexified Lienard–Wiechert solutions[1] to the Maxwell equations. The class of CLW fields consists of regular solutions to Eqs. (2.19) in complex Minkowski space such that condition (v) is satisfied. Each CLW solution represents the electromagnetic radiation field produced by an electric monopole moving along an arbitrary *complex* "timelike" world line $\xi^\mu(\phi)$. In this case the function $X$ is directly related to the world line by $X(\phi, \zeta, \bar{\zeta}) = \xi^\mu(\phi) l_\mu(\zeta, \bar{\zeta})$. The condition that the complex world line be "timelike" has been provided by condition (v). For a real world line this definition coincides with the usual one, i.e., for a *real* world line, condition (v) $\longleftrightarrow \xi'^\mu(\phi)\xi'_\mu(\phi) > 0$.

In the Robinson–Trautman–Maxwell limit to the Kerr–Maxwell solutions (i.e., when the p.n.v.f. satisfies condition (iv) on the real space) condition (v) reduces to the regularity condition used by Derry, Isaacson, and Winicour[11] for the Robinson–Trautman solutions.

It is clear from Eq. (2.16a) that one still has the freedom in a Kerr–Maxwell space to choose a new function $\tilde{\phi}$ equal to any analytic function of the old $\phi$,

$$\tilde{\phi} = G(\phi). \qquad (2.24)$$

The Lorentz transformation freedom represented by the fractional linear transformations on $\zeta$

$$\tilde{\zeta} = (a\zeta + b)/(c\zeta + d), \quad ad - bc \neq 0, \qquad (2.25)$$

also remains at our disposal.

## 3. TYPE III (N) KERR-MAXWELL SPACES

In this section we prove the following theorem.

*Theorem I*: The most general Petrov–Penrose type III (type N) Kerr–Maxwell space is flat empty space.

*Proof*: Type III Kerr–Maxwell spaces are characterized in the spin coefficient formalism by

$$\Psi_2 = 0. \qquad (3.1)$$

From (2.11b) we see that this implies that

$$\Phi_1^0 = 0 = \Psi_2^0. \qquad (3.2)$$

It is convenient to rewrite the surviving differential equations (2.19b) and (2.20b) using the independent variables $\phi, \zeta, \bar{\zeta}$ rather than $u, \zeta, \bar{\zeta}$. The equations become

$$\eth_0'(\Phi_2^0 V) = 0, \qquad (3.3)$$

$$\eth_0'(V^2\partial_0' R) = k\Phi_2^0\bar{\Phi}_2^0 V^2, \qquad (3.4)$$

where $\eth_0'$ is $\eth_0$ holding $\phi$ constant. The regularity condition implies that the functions $V$, $\Phi_2^0$, and $R$ have the following expansion in spin-weighted spherical harmonics[8,9]:

$$V = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} a^{lm}(\phi)\,{}_0Y_{lm}(\zeta, \bar{\zeta}), \qquad (3.5\text{a})$$

$$\Phi_2^0 = \sum_{l=1}^{\infty}\sum_{m=-l}^{l} b^{lm}(\phi)\,{}_{-1}Y_{lm}(\zeta, \bar{\zeta}), \qquad (3.5\text{b})$$

$$R = \sum_{l=2}^{\infty}\sum_{m=-l}^{l} c^{lm}(\phi)\,{}_{-2}Y_{lm}(\zeta, \bar{\zeta}). \qquad (3.5\text{c})$$

By using the properties of spin weighted functions and the edth operator,[8] it is possible to show the following:

Given a function $\eta_s(\phi, \zeta, \bar{\zeta})$ expandable in spin $s$ sperhical harmonics,

$$\eta_s = \sum_{l=|s|}^{\infty} \sum_{m=-l}^{l} \eta_s^{lm}(\phi)\, {}_sY_{lm}(\zeta, \bar{\zeta}), \tag{3.6}$$

the equation

$$\eth_0' \eta_s = 0 \tag{3.7}$$

has the regular solution

$$\eta_s = 0 \quad (s < 0), \tag{3.8a}$$

$$\eta_s = \sum_{m=-s}^{s} \eta_s^{sm}(\phi)\, {}_sY_{sm}(\zeta, \bar{\zeta}) \quad (s \geq 0). \tag{3.8b}$$

Equation (3.7) can be rewritten as

$$\frac{\partial}{\partial \zeta}\left[\eta_s(1 + \zeta\bar{\zeta})^s\right]\Big|_{\phi = \mathrm{const}} = 0,$$

which has the solution

$$\eta_s(1 + \zeta\bar{\zeta})^s = F(\phi, \bar{\zeta}), \tag{3.9}$$

where $F$ is an analytic function of $\bar{\zeta}$. By (3.6) $\eta_s$ is a bounded function on the sphere. When $s < 0$, the left-hand side of (3.9) is also bounded on the sphere. The only choice for $F$ which satisfies this condition is zero and we have (3.8a). When $s \geq 0$, nonvanishing bounded solutions for $\eta_s$ of the form

$$\eta_s = \sum_{k=0}^{2s} A_k(\phi)\bar{\zeta}^k/(1 + \zeta\bar{\zeta})^s \tag{3.10}$$

are possible, and from the definition of the spin weighted sperhical harmonics we find that (3.10) is equivalent to (3.8b).

Applying these results to Eqs. (3.3) and (3.4) and using the fact from condition (v) that $V \neq 0$, we see immediately from (3.3) that

$$\Phi_2^0 = 0, \tag{3.11}$$

so that the electromagnetic field vanishes, and then from (3.4) that

$$\eth_0' R = 0. \tag{3.12}$$

Since $\Psi_3^0 = \eth_0 R + L\dot{R} = \eth_0' R = 0$, we see from (2.11) and (2.12) that the solutions must be type N. But $R$ has a spin weight $-2$ so that (3.12) has the solution

$$R = 0, \tag{3.13}$$

and we are left with flat empty space.

## 4. NONRADIATIVE SOLUTIONS

We have already shown[7] that the most general Kerr–Maxwell solution for which $R = 0$ (equivalent to the vanishing of the Bondi news function) and $\Phi_2^0 = 0$ is the Kerr–Newman metric.[6] In this section we obtain the same result using the weaker condition $\Psi_4^0 = \dot{R} = 0$ and at the same time generalize a result of Derry, Isaacson, and Winicour by proving the following theorem.

*Theorem II*: The most general Kerr–Maxwell solution having vanishing radiative $O(r^{-1})$ part of the Riemann tensor and Maxwell tensor is the Kerr–Newman metric. Thus, a Kerr–Maxwell solution is either radiative or one of the following exact solutions:

(a) Kerr–Newman metric [p.n.v.f. $l$ satisfies conditions (i)–(iii) and Maxwell field does not vanish],

(b) Kerr metric [$l$ satisfies conditions (i)–(iii), Maxwell field vanishes],

(c) Reissner–Nordström metric [$l$ satisfies conditions (i)–(iv), Maxwell field does not vanish],

(d) Schwarzschild metric [$l$ satisfies conditions (i)–(iv), Maxwell field vanishes],

(e) empty flat space.

*Proof*: The condition that the radiative parts of the Riemann and Maxwell tensors vanish is

$$\Psi_4^0 = \dot{R} = 0, \tag{4.1a}$$

$$\Phi_2^0 = 0. \tag{4.1b}$$

We wish to find the most general regular solution to Eqs. (2.19) and (2.20) such that (4.1) holds. Once again we change our independent variables from $u, \zeta, \bar{\zeta}$ to $\phi, \zeta, \bar{\zeta}$ so that the equations have the form

$$\eth_0'(\Phi_1^0 V^2) = 0, \tag{4.2a}$$

$$\Phi_1^{0\prime} = 0, \tag{4.2b}$$

$$\eth_0'(\Psi_2^0 V^3) = 0, \tag{4.2c}$$

$$\Psi_2^{0\prime} V = \eth_0'(V^2 \eth_0 R), \tag{4.2d}$$

where, as earlier the prime denotes differentiation with respect to $\phi$, $\eth_0'$ is $\eth_0$ holding $\phi$ constant and $\eth_0' R = \eth_0 R$ since $R = R(\zeta, \bar{\zeta})$ only.

The condition $\dot{R} = 0$ becomes

$$[(\eth_0'^2 V)/V]' = 0$$

or

$$\eth_0'[V^2 \eth_0'(V'/V)] = 0. \tag{4.3}$$

Since it has already been shown[7] that the case $\Phi_2^0 = R = 0$ leads to the Kerr–Newman metric, it will be sufficient for us here to simply show that (4.2) and (4.3) lead to $R = 0$.

Equation (4.2a) can be written as

$$\eth_0' \Phi_1^0 = -2\Phi_1^0 \eth_0' \log V.$$

Differentiating with respect to $\phi$ and using (4.2b), we obtain the result that

$$\Phi_1^0 \eth_0'(V'/V) = 0.$$

Either

$$\Phi_1^0 = 0 \tag{4.4}$$

or

$$\eth_0'(V'/V) = 0. \tag{4.5}$$

Suppose $\Phi_1^0 = 0$. Equation (4.2c) can be written as

$$\eth_0' \Psi_2^0 = -3\Psi_2^0 \eth_0' \log V \tag{4.6}$$

and (4.2d) as

$$\Psi_2^{0\prime} = V\eth_0^2 R + 2\eth_0' V \eth_0 R. \tag{4.7}$$

Differentiating (4.6) by $\phi$, operating on (4.7) with $\eth_0'$, and equating the two results yields

$$6(\eth_0' \log V)\eth_0^2 R + 6(\eth_0' \log V)^2 \eth_0 R + 3\Psi_2^0[\eth_0'(V'/V)]/V + \eth_0^3 R + 2\bar{R}\eth_0 R = 0. \tag{4.8}$$

Differentiating (4.8) with respect to $\phi$ and using (4.7) yields

$$H' = 0, \tag{4.9}$$

where

$$H \equiv (\Psi_2^0)^3[\eth_0'(V'/V)]/V = [\Psi_2^0 V^3]^3[V^2\eth_0'(V'/V)]\,V^{-12}. \tag{4.10}$$

Operating on $H$ with $\eth_0'$ and using (4.2c), (4.3), (4.9), and (4.10) then gives us

$$\eth_0' H = \eth_0 H = -12H\eth_0' \log V, \tag{4.11}$$

and finally after differentiating (4.11) by $\phi$ we obtain

$$-12H\eth_0'(V/V) = 0.$$

Either $\Psi_2^0 = 0$, which we reject since it leads to flat space by Theorem I, or (4.5) must hold whether $\Phi_1^0$ vanishes or not.

Equation (4.5) has the regular solution

$$V'/V = f(\phi),$$

or

$$V = \exp[g(\phi) + h(\zeta, \bar{\zeta})]. \tag{4.12}$$

Under the freedom (2.24)

$$V \to \tilde{V} = V/G'(\phi),$$

so that this can be used to put

$$V = V(\zeta, \bar{\zeta}). \tag{4.13}$$

Equation (4.2c) has the regular solution

$$\Psi_2^0 = M(\phi)/V^3, \tag{4.14}$$

and substitution of this into (4.2d) using (4.13) yields

$$M'/V^2 = \eth_0(V^2\eth_0 R). \tag{4.15}$$

Differentiating with respect to $\phi$ gives us

$$M''(\phi) = 0,$$

or

$$M'(\phi) = \text{const.} \tag{4.16}$$

The quantity $V^2\eth_0 R$ has spin weight minus one so that integration of (4.15) over the sphere[8,9] yields

$$M' \int \frac{d\zeta\, d\bar{\zeta}}{V^2(1+\zeta,\bar{\zeta})^2} = 0.$$

The integral cannot vanish by condition (v) so that

$$M' = 0, \tag{4.16'}$$

and (4.15) becomes

$$\eth_0(V^2\eth_0 R) = 0,$$

for which the only regular solution is

$$R = 0, \tag{4.17}$$

and the theorem is proved.

## ACKNOWLEDGMENTS

I wish to acknowledge useful discussions concerning the topics of this paper with Professor Ted Newman at the University of Pittsburgh and with Professor Peter Havas at Temple University, where this work was begun.

*Research supported in part by the U.S. Atomic Energy Commission.

†Present address: Department of Physics, West Virginia Institute of Technology, Montogmery, West Virginia 25136.

[1]R. Lind and E.T. Newman, J. Math. Phys. 15, 1103 (1974).
[2]G.C. Debney, R.P. Kerr, and A. Schild, J. Math. Phys. 10, 1842 (1969).
[3]J.N. Goldberg and R.K. Sachs, Acta Phys. Polon. 22, Suppl., 13 (1962).
[4]E.T. Newman and R. Posadas, Phys. Rev. 187, 1784 (1969).
[5]E.T. Newman and R. Penrose, J. Math. Phys. 3, 566 (1962).
[6]E.T. Newman, E. Couch, K. Chinnapered, A. Exton, A. Prakash, and K. Torrence, J. Math. Phys. 6, 918 (1965).
[7]R. Lind, General Relativity Gravitation J. 5, 25 (1974).
[8]E.T. Newman and R. Penrose, J. Math. Phys. 7, 863 (1966).
[9]J.N. Goldberg, A. MacFarlane, E.T. Newman, R. Rohrlich, and C.G. Sudarshan, J. Math. Phys. 8, 2155 (1967).
[10]R. Penrose and M.A.H. MacCallum, Phys. Rep. 66, (1973).
[11]L. Derry, R. Isaacson and J. Winicour, Phys. Rev. 185, 1647 (1969).

# Stationary Kerr–Maxwell spaces*

Robert W. Lind†

*Department of Physics, The Florida State University, Tallahassee, Florida 32306*
(Received 24 July 1974)

We consider algebraically special electrovac spacetimes in which a diverging, geodesic, shear-free (but twisting) degenerate principal null direction of the Weyl tensor coincides with a principal null direction of the Maxwell tensor. All stationary solutions of this class are solved exactly and found to be of Petrov–Penrose type D, the most general regular stationary solution being the Kerr–Newmann metric.

## 1. INTRODUCTION

In this paper we begin by considering the class of electrovac spacetimes in which a principal null vector field $l$ of the Maxwell tensor satisfies the following three conditions:

(i) $l$ is the tangent field to a congruence of null geodesics,
(ii) $l$ has nonvanishing divergence,
(iii) the shear of $l$ vanishes.

By a corollary of the Goldberg–Sachs theorem,[1] $l$ is coincident with a degenerate principal null vector field of the Weyl tensor, which is then algebraically special in the Petrov–Penrose[2] sense. As we have already called regular solutions to this class Kerr–Maxwell (KM) spaces,[3] we will refer to the general solutions as general Kerr–Maxwell (GKM) spaces. In Sec. 2 the metric for the class of GKM spaces is presented using the Newman–Penrose[4] spin coefficient formalism.

A stationary spacetime is by definition one that admits a global timelike Killing vector field, i.e., a vector field $k^\mu$ that satisfies

$$k^{(\mu;\nu)}=0, \tag{1.1}$$

and

$$k^\mu k_\mu > 0, \tag{1.2}$$

everywhere. In Sec. 3 we use the spin coefficient formalism to find the stationary GKM spaces. We show that they are all type D and that the most general stationary KM solution is the Kerr–Newman metric.[5]

## 2. THE GKM METRIC

In this section we present a brief review of the spin coefficient formulation[4,6] of the GKM metrics.

In a four-dimensional Riemmannian manifold with signature $(+,-,-,-)$, we choose the principal null vector $l$ to be one member of a null tetrad consisting of two real $(l, n)$ and two complex $(m, \bar{m})$ null vectors satisfying

$$l\cdot n = -m\cdot\bar{m} = 1, \tag{2.1}$$

all other scalar products vanishing. Equation (2.1) implies the completeness relation

$$g_{\mu\nu} = 2[l_{(\mu}n_{\nu)} - m_{(\mu}\bar{m}_{\nu)}].$$

The tetrad components of the Weyl tensor and Maxwell tensor are defined by

$$\begin{aligned}
\Psi_0 &= -C_{\mu\nu\rho\sigma}l^\mu m^\nu l^\rho m^\sigma,\\
\Psi_1 &= -C_{\mu\nu\rho\sigma}l^\mu n^\nu l^\rho m^\sigma,\\
\Psi_2 &= -C_{\mu\nu\rho\sigma}\bar{m}^\mu n^\nu l^\rho m^\sigma,\\
\Psi_3 &= -C_{\mu\nu\rho\sigma}\bar{m}^\mu n^\nu l^\rho n^\sigma,\\
\Psi_4 &= -C_{\mu\nu\rho\sigma}\bar{m}^\mu n^\nu \bar{m}^\rho n^\sigma,
\end{aligned} \tag{2.2}$$

and

$$\begin{aligned}
\Phi_0 &= F_{\mu\nu}l^\mu m^\nu,\\
\Phi_1 &= (1/2)F_{\mu\nu}(l^\mu n^\nu + \bar{m}^\mu m^\nu),\\
\Phi_2 &= F_{\mu\nu}\bar{m}^\mu n^\nu,
\end{aligned} \tag{2.3}$$

respectively.

The GKM spaces are characterized by the fact that $l$ is tangent to a degenerate principal null direction of the Weyl tensor coincident with a principal null direction of the Maxwell tensor. In terms of tetrad components this is equivalent to

$$\Psi_0 = \Psi_1 = \Phi_0 = 0. \tag{2.4}$$

The requirement that $l$ satisfy conditions (i)–(iii) can be expressed in terms of spin coefficients as

$$\begin{aligned}
&\text{(i)}\quad \kappa = 0,\\
&\text{(ii)}\quad \rho+\bar{\rho}\neq 0,\\
&\text{(iii)}\quad \sigma = 0.
\end{aligned} \tag{2.5}$$

Further simplifications, namely

$$\epsilon = \pi = \tau = \lambda = 0, \tag{2.6}$$

can be achieved by a proper choice of the remaining tetrad vectors $(n, m, \bar{m})$ and by introducing an associated null coordinate system $(u, r, \zeta, \bar{\zeta})$ such that $r$ is an affine parameter along the null geodesics to which $l$ is tangent.[6] The remaining spin coefficients are given by

$$\begin{aligned}
\rho &= l_{\mu;\nu}m^\mu\bar{m}^\nu,\\
\alpha &= (1/2)(l_{\mu;\nu}n^\mu\bar{m}^\nu - m_{\mu;\nu}\bar{m}^\mu\bar{m}^\nu),\\
\beta &= (1/2)(l_{\mu;\nu}n^\mu m^\nu - m_{\mu;\nu}\bar{m}^\mu m^\nu),\\
\gamma &= (1/2)(l_{\mu;\nu}n^\mu n^\nu - m_{\mu;\nu}\bar{m}^\mu n^\nu),\\
\mu &= -n_{\mu;\nu}\bar{m}^\mu m^\nu,\\
\nu &= -n_{\mu;\nu}\bar{m}^\mu n^\nu.
\end{aligned} \tag{2.7}$$

Directional derivatives acting on a scalar $S$ are defined by

$$DS \equiv S_{;\mu}l^\mu,\quad \Delta S \equiv S_{;\mu}n^\mu,$$
$$\delta S \equiv S_{;\mu}m^\mu,\quad \bar{\delta} S \equiv S_{;\mu}\bar{m}^\mu,$$

and obey the commutation relations

$$(\Delta D - D\Delta)S = (\gamma+\bar{\gamma})DS,$$

$$(\delta D - D\delta)S = (\bar{\alpha} + \beta)DS - \bar{\rho}\delta S,$$
$$(\delta\Delta - \Delta\delta)S = -\bar{\nu}DS - (\bar{\alpha} + \beta)\Delta S + (\mu - \gamma + \bar{\gamma})\delta S, \qquad (2.8)$$
$$(\bar{\delta}\delta - \delta\bar{\delta})S = (\bar{\mu} - \mu)DS + (\bar{\rho} - \rho)\Delta S - (\bar{\alpha} - \beta)\bar{\delta}S + (\alpha - \bar{\beta})\delta S.$$

Finally, the metric for a GKM space can be written in the form

$$ds^2 = 2(ln - m\bar{m}) \qquad (2.9)$$

with

$$l \equiv l_\mu dx^\mu = du - (L/2P_0V)\, d\zeta - (\bar{L}/2P_0V)\, d\bar{\zeta}, \qquad (2.10a)$$
$$n \equiv n_\mu dx^\mu = dr - (1/2P_0V)(\omega/\bar{\rho})\, d\zeta - (1/2P_0V)(\bar{\omega}/\rho)\, d\bar{\zeta} - Ul, \qquad (2.10b)$$
$$m \equiv m_\mu dx^\mu = -(1/2P_0V\rho)\, d\bar{\zeta}, \qquad (2.10c)$$
$$\bar{m} \equiv \bar{m}_\mu dx^\mu = -(1/2P_0V\bar{\rho})\, d\zeta, \qquad (2.10d)$$

where

$$\rho = -(r + i\Sigma)^{-1}, \qquad (2.11a)$$
$$2i\Sigma = V\bar{\eth}_0(\bar{L}/V) + L\dot{\bar{L}} - V\bar{\eth}_0(L/V) - \bar{L}\dot{L}, \qquad (2.11b)$$
$$P_0 = (1/2)(1 + \zeta\bar{\zeta}), \qquad (2.11c)$$

and $\eth_0$ is the edth ($\eth$) operator[8] acting on the unit sphere. The functions $U$ and $\omega$ satisfy the differential equations

$$DU = -(\gamma + \bar{\gamma}), \qquad (2.12a)$$
$$D\omega = \bar{\rho}\omega - (\bar{\alpha} + \beta), \qquad (2.12b)$$
$$\delta U - \Delta\omega = -\bar{\nu} - (\bar{\alpha} + \beta)U + (\mu - \gamma + \bar{\gamma})\omega, \qquad (2.12c)$$
$$\bar{\delta}\omega - \delta\bar{\omega} = (\bar{\mu} - \mu) + (\bar{\rho} - \rho)U - (\bar{\alpha} - \beta)\bar{\omega} + (\alpha - \bar{\beta})\omega. \qquad (2.12d)$$

The spin coefficient formulation of our problem then consists of the Eqs. (2.12) together with additional equations involving the tetrad components of the Weyl (2.2) and Maxwell (2.3) tensors and the spin coefficients (2.7). The entire set of equations (which are completely equivalent to the coupled vacuum Einstein—Maxwell equations corresponding to the same problem) can be easily obtained by simply applying the specilizations, (2.4)—(2.6) and (2.10), to the general equations given in Ref. 4. Therefore, we present here only those Newman—Penrose equations which we shall use later on:

$$D\alpha = \rho\alpha, \qquad (2.13a)$$
$$D\beta = \bar{\rho}\beta, \qquad (2.13b)$$
$$D\gamma = \Psi_2 + k\Phi_1\bar{\Phi}_1, \qquad (2.13c)$$
$$D\mu = \bar{\rho}\mu + \Psi_2, \qquad (2.13d)$$
$$D\nu = \Psi_3 + k\bar{\Phi}_1\Phi_2, \qquad (2.13e)$$
$$\Delta\rho = -\rho(\bar{\mu} - \gamma - \bar{\gamma}) - \Psi_2, \qquad (2.13f)$$
$$\Delta\bar{\alpha} - \bar{\delta}\gamma = \bar{\rho}\nu + \bar{\alpha}(\gamma - \mu) + \beta\bar{\gamma} - \bar{\Psi}_3, \qquad (2.13g)$$
$$\delta\gamma - \Delta\beta = -(\bar{\alpha} + \beta)\gamma - \beta(\gamma - \bar{\gamma} - \mu) + k\Phi_1\bar{\Phi}_2, \qquad (2.13h)$$
$$D\Psi_2 = 3\rho\Psi_2 + 2k\Phi_1\bar{\Phi}_1\rho, \qquad (2.13i)$$
$$D\Psi_3 = 2\rho\Psi_3 + \bar{\delta}\Psi_2 + k\bar{\Phi}_1 D\Phi_2, \qquad (2.13j)$$

where $k$ is the gravitational constant.

The form of the GKM metric remains unchanged under the following coordinate-tetrate trnasformations:

$$\tilde{u} = u, \quad \tilde{r} = r, \quad \tilde{\zeta} = \tilde{\zeta}(\zeta),$$
$$\tilde{l} = l, \quad \tilde{n} = n, \qquad (2.14)$$
$$\tilde{m} = \exp(i\lambda)m \equiv [(\partial\bar{\tilde{\zeta}}/\partial\bar{\zeta})/(\partial\tilde{\zeta}/\partial\zeta)]^{1/2}m,$$

and

$$\tilde{u} = G(u, \zeta, \bar{\zeta}), \quad \tilde{r} = \dot{G}^{-1}r, \quad \tilde{\zeta} = \zeta,$$
$$\tilde{l} = \dot{G}l, \quad \tilde{n} = \dot{G}^{-1}n, \quad \tilde{m} = m, \qquad (2.15)$$

where a dot denotes the partial derivative with respect to $u$.

Under (2.15), $\tilde{V} = \dot{G}^{-1}V$, so that (2.15) is often used to put

$$V = 1, \qquad (2.16)$$

which is Bondi's coordinate condition.

## 3. STATIONARY SOLUTIONS

In this section we solve for the class of stationary GKM metrics exactly. These solutions are all Petrov—Penrose[2] type D and the most general regular one is the Kerr—Newman metric.[5]

We now make the assumption that a well-defined time-like Killing vector field $k^\mu$ exists everywhere, in particular, in the asymptotic region in which the affine parameter $r$ becomes large.

By resolving $k^\mu$ on the tetrad

$$k^\mu = Al^\mu + Bn^\mu + \bar{C}m^\mu + C\bar{m}^\mu, \qquad (3.1)$$

a set of ten real equations can be written in terms of spin coefficients involving the three functions $A$, $B$, $C$, which are equivalent to the Killing equations (1.1). (For completeness, although they are never used here, the Killing equations for a general spacetime are given in spin coefficient form in the Appendix.)

These equations are obtained by substituting (3.1) into (1.1), contracting with various combinations of tetrad vectors and then substituting for the spin coefficients (2.7) where ever they occur. For example, contraction of (1.1) with $l_\mu l_\nu$ yields

$$DB = 0. \qquad (3.2)$$

Thus,

$$B = B_0(u, \zeta, \bar{\zeta}), \qquad (3.3)$$

where, in order that $k^\mu$ be everywhere well defined and timelike, $B_0$ must be a regular function on the sphere with no zeros, i.e., $B_0$ is expandable in spherical harmonics

$$B_0 = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} b^{lm}(u)_0 Y_{lm}(\zeta, \bar{\zeta}).$$

Under the freedom (2.15), $\tilde{B}_0 = \dot{G}B_0$. Since $B_0$ is regular with no zeros we can use (2.15) to put

$$B_0 = 1, \qquad (3.4)$$

without affecting the possible regularity of any other functions such as $V$, for instance.

The condition (1.2) that the Killing vector field be everywhere timelike (in particular, in the asymptotic region) also leads to the results that

$$DC = 0, \tag{3.5}$$

and

$$K \equiv V^2 \eth_0 \bar{\eth}_0 \log P_0 V > 0, \tag{3.6}$$

the details of which are given in the Appendix.

The result (3.6) implies that the real function $V(u, \zeta, \bar{\zeta})$ is, in fact, also a positive regular function on the sphere with no zeros. Thus, we are free to impose Bondi's coordinate condition (2.16) at anytime without affecting the character of $B_0$. To preserve the regularity of $V$ the transformation (2.14) must now be restricted to the fractional linear (Lorentz) transformation given by

$$\bar{\zeta} = \frac{a\zeta + b}{c\zeta + d}, \tag{2.14'}$$

where $a$, $b$, $c$, $d$ are complex constants such that $ad - bc = 1$.

Making use of (3.4) and (3.5), the remaining Killing equations become

$$DA - (\gamma + \bar{\gamma}) = 0, \tag{3.7a}$$

$$\bar{\rho}C + (\bar{\alpha} + \beta) = 0, \tag{3.7b}$$

$$\Delta A + (\gamma + \bar{\gamma})A + \nu C + \bar{\nu}\bar{C} = 0, \tag{3.7c}$$

$$\delta A - \Delta C + (\bar{\alpha} + \beta)A + (\mu + \gamma - \bar{\gamma})C - \bar{\nu} = 0, \tag{3.7d}$$

$$\delta C + (\bar{\alpha} - \beta)C = 0, \tag{3.7e}$$

$$\bar{\delta}C + \delta\bar{C} - (\alpha - \bar{\beta})C - (\bar{\alpha} - \beta)\bar{C} - (\rho + \bar{\rho})A + (\mu + \bar{\mu}) = 0. \tag{3.7f}$$

After applying the operator $D$ to (3.7$d$), subtracting from this $\delta$ applied to (3.7a) and $\Delta$ applied to (3.7b), a straightforward calculation involving the commutators (2.8) and the Eqs. (2.13) then yields the result

$$2\Psi_2 C = 0. \tag{3.8}$$

Thus,

$$C = 0. \tag{3.9}$$

For suppose $C \neq 0$. Then $\Psi_2 = 0$ and we have a type III solution. From (2.13i) we see that $\Phi_1 = 0$ and from (2.13c) that $D\gamma = 0$. Applying the operator $D[\bar{\rho}^{-3}D^2]$ to Eq. (3.7c) then yields

$$(6C\Psi_3\rho/\bar{\rho}^3)(\rho - \bar{\rho}) = 0, \tag{3.10}$$

where Eqs. (2.13e) and (2.13j) have been used. Now $\rho \neq 0$ by condition (ii), so that either $(\rho - \bar{\rho}) = 0$ or $\Psi_3 = 0$. If $(\rho - \bar{\rho}) = 0$, we are dealing with a regular type III Robinson–Trautman–Maxwell solution, the most general one of which is flat empty space.[3,9] In the Appendix we show that a stationary solution with $\Psi_2 = \Psi_3 = 0$ must also be flat and empty.

Hence, (3.9) must hold and the Eqs. (3.7) are further simplified to

$$(\bar{\alpha} + \beta) = 0, \tag{3.11a}$$

$$A = (\mu + \bar{\mu})/(\rho + \bar{\rho}), \tag{3.11b}$$

$$DA = (\gamma + \bar{\gamma}), \tag{3.11c}$$

$$\Delta A = -(\gamma + \bar{\gamma})A, \tag{3.11d}$$

$$\delta A = \bar{\nu}. \tag{3.11e}$$

For a GKM space[6]

$$(\alpha + \beta) = V(L/V)^{\cdot}\bar{\rho}.$$

Thus,

$$(L/V)^{\cdot} = 0. \tag{3.12}$$

Incorporating (3.12) into other results from Ref. 6, we find

$$\gamma = -(1/2)(\dot{V}/V) + (1/2)\Psi_2^0\rho^2 + k\Phi_1^0\bar{\Phi}_1^0\bar{\rho}\rho^2, \tag{3.13}$$

and

$$\mu = [K + V^2\eth_0(\bar{L}\dot{V}/V^2) + LV\bar{\eth}_0(\dot{V}/V) + L\bar{L}(\dot{V}/V)^{\cdot}]\bar{\rho} + (1/2)\Psi_2^0(\rho^2 + \rho\bar{\rho}) + k\Phi_1^0\bar{\Phi}_1^0\bar{\rho}\rho^2, \tag{3.14}$$

with $K$ given by (3.6).

Equations (3.11b) and (3.11c) together then yield the results that

$$\dot{V} = 0 \tag{3.15}$$

and

$$A \equiv -U = K + \mathrm{Re}\psi_2^0\rho + k\bar{\Phi}_1^0\Phi_1^0\rho\bar{\rho}. \tag{3.16}$$

From (3.11d) we obtain

$$\dot{A} = -\dot{U} = 0, \tag{3.17}$$

and, finally, by substituting (3.11e) into (2.12c) and using (2.12b), we find that

$$\dot{\omega} = 0, \tag{3.18a}$$

where

$$\omega = -i(V\eth_0\Sigma)\bar{\rho}. \tag{3.18b}$$

Thus, the metric is independent of the coordinate $u$ and the Killing vector field is given simply by

$$k^\mu = \partial x^\mu/\partial u. \tag{3.19}$$

The equations remaining to be solved at this stage[6] are

$$\eth_0\Phi_1^0 = 0, \tag{3.20a}$$

$$V^2\eth_0(\Phi_2^0/V) + L\mathring{\Phi}_2^0 = 0, \tag{3.20b}$$

$$\eth_0\Psi_2^0 = 2k\Phi_1^0\bar{\Phi}_2^0, \tag{3.20c}$$

$$\eth_0\bar{\eth}_0 K = k|\Phi_2^0/V|^2, \tag{3.20d}$$

$$\mathrm{Im}\Psi_2^0 = V^2\eth_0\bar{\eth}_0\Sigma + 2K\Sigma, \tag{3.20e}$$

where [from (3.16), (3.17), and (3.20e)] the fact that

$$\dot{\Psi}_2^0 = 0 = \dot{\Phi}_1^0, \tag{3.21}$$

has been used.

The regularity of $V$ implies that $\bar{\eth}_0 K$ is a spin weight[8] minus one guantity and integration of (3.20d) over the sphere yields

$$0 = k\int |\Phi_2^0/V|^2 \frac{d\zeta d\bar{\zeta}}{P_0^2},$$

or

$$\Phi_2^0 = 0, \tag{3.22}$$

which satisfies (3.20b) automatically.

Equations (3.20a) and (3.20c) then have the general solutions

$$\Phi_1^0 = E(\bar{\zeta}) \tag{3.23}$$

and

$$\Psi_2^0 = M(\bar{\zeta}), \tag{3.24}$$

respectively.

Again using the regularity of $V$, the general solution of (3.20d) becomes

$$K = V^2\eth_0\bar{\eth}_0 \log P_0 V = \text{positive const.}$$

Thus, $V$ is expandable in $l=0$, 1 spherical harmonics only and the transformation (2.14)′ can be used to put

$$V=1. \tag{3.25}$$

After substituting for $\Sigma$ from (2.11b), the last equation has the general solution[10]

$$L = -(1/2)[\bar{M}(\zeta)/\zeta] + \mathcal{L}(\bar{\zeta})/(1+\zeta\bar{\zeta}). \tag{3.26}$$

The class of stationary GKM metrics has been solved exactly, the entire class being completely determined by the three analytic functions $E$, $M$, and $\mathcal{L}$ of $\bar{\zeta}$. The Weyl tensor components $\Psi_3$ and $\Psi_4$ both vanish[6] so that we have proved the following.

*Theorem* 1: Stationary GKM spaces are all of Petrov–Penrose type D. The metric for this class of solutions can be given locally in the form (2.9) and (2.10), together with (2.11), (3.16), (3.18b), (3.23), (3.24), (3.25), and (3.26). The nonvanishing components of the Weyl and Maxwell tensors are given by

$$\Psi_2 = M(\bar{\zeta})\rho^3 + 2kE(\bar{\zeta})\bar{E}(\zeta)\bar{\rho}\rho^3, \tag{3.27}$$

and

$$\Phi_1 = E(\bar{\zeta})\rho^2, \tag{3.28}$$

respectively.

The most general regular solution of this type is the Kerr–Newman metric. Hence, we have also proved

*Theorem* 2: The most general stationary KM solution is the Kerr–Newman metric.

Finally, we have the following corollary to Theorem 1.

*Corollary* 1: A stationary vacuum spacetime is either algebraically general (Petrov–Penrose type I) or algebraically special type D.

## APPENDIX

In a general space, the tetrad components of the Killing vector (3.1) satisfy the equations

$$DB=0, \tag{A1}$$

$$DA + \Delta B - (\gamma+\bar{\gamma})B + (\pi - \bar{\tau})C + (\bar{\pi} - \tau)\bar{C} = 0, \tag{A2}$$

$$DC - \delta B + (\bar{\alpha}+\beta)B + \bar{\rho}C + \sigma\bar{C} + \bar{\pi}B = 0, \tag{A3}$$

$$\Delta A + (\gamma+\bar{\gamma})A + \nu C + \bar{\nu}\bar{C} = 0, \tag{A4}$$

$$\Delta C - \delta A - (\tau + \bar{\alpha} + \beta)A - (\mu + \gamma - \bar{\gamma})C + \bar{\nu}B - \bar{\lambda}\bar{C} = 0, \tag{A5}$$

$$\delta C + (\bar{\alpha} - \beta)C + \bar{\lambda}B - \sigma A = 0, \tag{A6}$$

$$\bar{\delta}C + \delta\bar{C} - (\alpha - \bar{\beta})C - (\bar{\alpha} - \beta)\bar{C} - (\rho+\bar{\rho})A + (\mu+\bar{\mu})B = 0. \tag{A7}$$

Equations (A3), (A5), and (A6) are complex so that the set (A1)–(A7) consists of ten real equations completely equivalent to Eqs. (1.1).

Using the results of Ref. 6, Eqs. (A1)–(A3) can be integrated immediately to yield

$$A = A_0 - (\dot{V}/V)r + \mathrm{Re}\Psi_2^0\rho + k\Phi_1^0\bar{\Phi}_1^0\rho\bar{\rho}, \tag{A8}$$

and

$$C = C_0 r - V(L/V)^{\cdot} - i\Sigma C_0, \tag{A9}$$

where $A_0$ and $C_0$ are independent of $r$ and (2.15) has been used to put

$$B_0(u,\zeta,\bar{\zeta}) = 1. \tag{A10}$$

The timelike condition (1.2) becomes

$$\begin{aligned} k^\mu k_\mu &= 2(AB - C\bar{C}) \\ &= 2[A_0 - (\dot{V}/V)r + \mathrm{Re}\Psi_2^0 + k\Phi_1^0\bar{\Phi}_1^0\rho\bar{\rho} - C_0\bar{C}_0 r^2 \\ &\quad + \mathrm{Re}[\bar{C}_0(V(L/V)^{\cdot} + i\Sigma C_0)r] - |V(L/V)^{\cdot} + i\Sigma C_0|^2] \\ &> 0, \end{aligned} \tag{A11}$$

from which we see immediately that $C_0 = 0$, or, equivalently, that

$$DC = 0. \tag{A12}$$

Substitution of (A8) and (A9) into (A7), again using the results of Ref. 6, yields

$$\dot{V} = 0 \tag{A13}$$

and

$$A_0 = K + \dot{L}\dot{\bar{L}}, \tag{A14}$$

so that (A11) also gives us

$$K \equiv V^2\eth_0\bar{\eth}_0 \log P_0 V > 0, \tag{A15}$$

which is (3.6).

Finally, if $\Psi_2 = \Psi_3 = 0$, then $\Phi_1 = \Phi_2 = 0$ and

$$\Psi_4 = \Psi_4^0\rho,$$

where

$$\Psi_4^0 = \dot{R}V^2.$$

But if $C \neq 0$, then (A6) yields

$$R = (\bar{\eth}^2 V)/V,$$

so that $\Psi_4^0 = 0$ and we have flat, empty space.

*Research supported by the U.S. Atomic Energy Commission.
†Present address: Department of Physics, West Virginia, Institute of Technology, Montgomery, West Virginia, 25136.
[1]J.N. Goldberg and R.K. Sachs, Acta Phys. Polon. 22, Suppl 13, (1962).
[2]Z. Petrov, Sci. Not. Kazan State Univ. 114, 55 (1954). R. Penrose, Ann. Phys. 10, 171 (1960).
[3]R. Lind, J. Math. Phys. 16, 34 (1975).
[4]E.T. Newman and R. Penrose, J. Math. Phys. 3, 566 (1962).
[5]E.T. Newman, E. Couch, K. Chinnapered, A. Exton, A. Prakash, and R. Torrence, J. Math. Phys. 6, 918 (1965).
[6]R. Lind, General Relativity Gravitation 5, 25 (1974).
[7]B. Aronson, R. Lind, J. Messmer, and E.T. Newman, J. Math. Phys. 12, 2462 (1971).
[8]J.N. Goldberg, A. MacFarlane, E.T. Newman, F. Rohrlich, and C.G. Sudarshan, J. Math. Phys. 8, 2155 (1967).
[9]L. Derry, R. Isaacson, and J. Winicour, Phys. Rev. 185, 1647 (1969).
[10]I. Robinson, J.R. Robinson, and J.D. Zund, J. Math. Mech. 18, 881 (1969).

# Propagation of transients in a random medium

Alan R. Wenzel

*NASA-Ames Research Center, Moffett Field, California 94035*
(Received 25 June 1974)

The propagation of transient scalar waves in a three-dimensional random medium is considered. The analysis is based on the smoothing method. An integro-differential equation for the coherent (or average) wave is derived and solved for the case of a statistically homogeneous and isotropic medium and a delta-function source. This yields the coherent Green's function of the medium. It is found that the waveform of the coherent wave depends generally on the distance from the source measured in terms of a certain dimensionless parameter. Based on the magnitude of this parameter, three propagation zones, called the near zone, the far zone, and the intermediate zone, are defined. In the near zone the evolution of the waveform is determined primarily by attenuation of the high-frequency components of the wave, whereas in the far zone it is determined mainly by dispersion of the low-frequency components. The intermediate zone is a region of transition between the near and far zones. The results show that, in general, the randomness of the medium causes a gradual smoothing and broadening of the waveform, as well as a decrease in amplitude of the wave, with propagation distance. In addition, the propagation speed of the wave is reduced. It is also found that an oscillating tail appears on the waveform as the propagation distance increases.

## INTRODUCTION

The subject of wave propagation in random media has been studied extensively over the past two decades (see, e.g., the review article by Frisch[1]). Most of the work in this area has been concerned with time-harmonic waves rather than with transient phenomena. Recently, however, interest in transient waves has been stimulated by a desire to understand how sonic booms propagate through atmospheric turbulence, and a number of theoretical papers dealing with this phenomenon have appeared.[2–7] A review of current research on this topic has been given by Pierce and Maglieri.[8]

Our main objective here is to study the propagation of transient waves in random media from a general viewpoint. Consequently, we have adopted a more general (but also more idealized) analytical model than those which have been used previously to study sonic-boom propagation. Our approach is based on the smoothing method, which has been discussed by Frisch.[1] This leads to what is essentially a linear treatment; i.e., effects such as nonlinear steepening of the wave, considered in Refs. 4, 5, and 7, are neglected. Our analysis does include multiple-scattering effects, however.

In Sec. I we formulate the problem in terms of an integro-differential equation for the coherent wave. In Sec. II we solve this equation by transform methods for the case in which the medium is statistically homogeneous and isotropic and the source term is a space-time delta function. This yields the coherent Green's function of the medium. The main result of Sec. II is given by Eq. 35, which is an integral expression for the Green's function. In Secs. IIA and IIB we evaluate this expression approximately for two propagation regions which we call the near and far zones. In addition, in Sec. IIC we evaluate it numerically for the region which we call the intermediate zone. The results of the latter calculation are presented in Figs. 1–5.

As noted in Sec. IIA, our near-zone results are similar to those of Cole and Friedman.[6] However, our results for the intermediate and far zones do not appear to have been obtained previously.

## I. FORMULATION OF THE PROBLEM

We wish to consider the propagation of transient scalar waves in an unbounded, three-dimensional, random medium. As our mathematical model of this phenomenon we choose the scalar wave equation

$$(c^{-2}\partial_t^2 - \nabla^2)u(\mathbf{x}, t) = f(\mathbf{x}, t), \tag{1}$$

where $\mathbf{x} = (x_1, x_2, x_3)$, and the "acoustic speed" $c(\mathbf{x}, t)$ is assumed to be a random function of space and time. The problem we are concerned with can be formulated, generally, as follows: Given the (nonrandom) source function $f$ and some appropriate statistical properties of $c$, find some specified statistical properties of $u$. Here we shall be concerned with the ensemble average of $u$, denoted by $\langle u\rangle$, which is called the coherent wave.

We now proceed to get an equation for the coherent wave. We begin by assuming that the random inhomogeneities of the medium are small; i.e., we write

$$c(\mathbf{x}, t) = c_0[1 + \epsilon\mu(\mathbf{x}, t)], \tag{2}$$

where $c_0$ is the average acoustic speed (assumed constant), and $\mu(\mathbf{x}, t)$ is a random function defined so that $\langle\mu\rangle = 0$ and $\langle\mu^2\rangle = 1$. The parameter $\epsilon$ is a measure of the deviation of $c$ from its average, and is assumed to be small.

By substituting for $c$ from Eq. 2 we can write Eq. 1 in the form

$$[L_0 + \epsilon L_1 + \epsilon^2 L_2 + O(\epsilon^3)]u = f, \tag{3}$$

where the operators $L_0$, $L_1$, and $L_2$ are given by

$$L_0 = c_0^{-2}\partial_t^2 - \nabla^2, \tag{4}$$

$$L_1 = -2c_0^{-2}\mu\,\partial_t^2, \tag{5}$$

$$L_2 = 3c_0^{-2}\mu^2\partial_t^2. \tag{6}$$

Keller[9] has shown that the ensemble-averaged solution of Eq. 3, i.e., the coherent wave, satisfies the equation

$$\{L_0 + \epsilon^2[\langle L_2\rangle - \langle L_1 L_0^{-1} L_1\rangle] + O(\epsilon^3)\}\langle u\rangle = f. \tag{7}$$

(In deriving Eq. 7 it has been assumed that $\langle L_1\rangle = 0$, which is the case here.) Thus, for our case the equation for the coherent wave is obtained by substituting Eqs. 4–6 into Eq. 7 and noting that the inverse operator $L_0^{-1}$ can be written in the form

$$L_0^{-1}\phi_1(\mathbf{x},t)=(4\pi)^{-1}\int r^{-1}\phi_1(\mathbf{x}+\mathbf{r},t-c_0^{-1}r)\,d\mathbf{r}. \tag{8}$$

(Here, and henceforth, an integral sign without limits denotes an integral taken over all of three-dimensional space.) The result, after terms of order $\epsilon^3$ are dropped, is

$$\begin{aligned}(c_0^{-2}\partial_t^2-\nabla^2)w(\mathbf{x},t)+\epsilon^2\{3c_0^{-2}w_{tt}(\mathbf{x},t)\\ -\pi^{-1}c_0^{-4}\int r^{-1}[R_{\tau\tau}(\mathbf{r},c_0^{-1}r)w_{tt}(\mathbf{x}+\mathbf{r},t-c_0^{-1}r)\\ -2R_\tau(\mathbf{r},c_0^{-1}r)w_{ttt}(\mathbf{x}+\mathbf{r},t-c_0^{-1}r)+R(\mathbf{r},c_0^{-1}r)\\ \times w_{tttt}(\mathbf{x}+\mathbf{r},t-c_0^{-1}r)]\,d\mathbf{r}\}=f(\mathbf{x},t).\end{aligned} \tag{9}$$

The letter subscripts denote differentiation.

In deriving Eq. 9 we have set $w\equiv\langle u\rangle$. Also, we have assumed that the medium is statistically homogeneous, and we have introduced the correlation function

$$R(\mathbf{r},\tau)=\langle\mu(\mathbf{x},t)\mu(\mathbf{x}+\mathbf{r},t-\tau)\rangle. \tag{10}$$

We now assume that the random fluctuations of the medium are independent of time. (This is equivalent to assuming that the characteristic time associated with the wave is much less than that associated with the fluctuations of the medium. In the case of waves propagating in real media, this condition is generally satisfied. In particular, it is satisfied in the case of sonic-boom propagation through atmospheric turbulence.) Then $\mu=\mu(\mathbf{x})$, and Eq. 9 simplifies to

$$\begin{aligned}(c_0^{-2}\partial_t^2-\nabla^2)w(\mathbf{x},t)+\epsilon^2\{3c_0^{-2}w_{tt}(\mathbf{x},t)\\ -\pi^{-1}c_0^{-4}\int r^{-1}R(\mathbf{r})w_{tttt}(\mathbf{x}+\mathbf{r},t-c_0^{-1}r)\,d\mathbf{r}\}=f(\mathbf{x},t),\end{aligned} \tag{11}$$

where now

$$R(\mathbf{r})=\langle\mu(\mathbf{x})\mu(\mathbf{x}+\mathbf{r})\rangle. \tag{12}$$

The procedure by which Eq. 9 is obtained from Eq. 1 is due essentially to Keller,[9] and is referred to as the smoothing method by Frisch.[1] It has been used previously to study the propagation of time-harmonic waves in various types of random media.[9–13]

## II. THE COHERENT GREEN'S FUNCTION

We now proceed to solve Eq. 11 when $f(\mathbf{x},t)=\delta(\mathbf{x})\delta(t)$. This will yield the free-space coherent Green's function of the medium [i.e., if $w(\mathbf{x},t)$ is the solution of Eq. 11 with $f(\mathbf{x},t)=\delta(\mathbf{x})\delta(t)$, then the Green's function $G$ is given by $G(\mathbf{x},t;\mathbf{x}',t')=w(\mathbf{x}-\mathbf{x}',t-t')$]. This function describes the propagation of a spherical delta-function pulse.

We begin by introducing the Fourier time and space transforms, defined by

$$Tf_1(\omega)\equiv\bar{f}_1(\omega)=\int_{-\infty}^{\infty}\exp(i\omega t)f_1(t)\,dt, \tag{13}$$

$$Sg_1(\mathbf{k})\equiv\hat{g}_1(\mathbf{k})=\int\exp(-i\mathbf{k}\cdot\mathbf{x})g_1(\mathbf{x})\,d\mathbf{x}, \tag{14}$$

respectively, where $f_1(t)$ and $g_1(\mathbf{x})$ are any square-integrable functions. The inverse transforms are given by

$$T^{-1}\bar{f}_1(t)=f_1(t)=(2\pi)^{-1}\int_{-\infty}^{\infty}\exp(-i\omega t)\bar{f}_1(\omega)\,d\omega, \tag{15}$$

$$S^{-1}\hat{g}_1(\mathbf{x})=g_1(\mathbf{x})=(2\pi)^{-3}\int\exp(i\mathbf{k}\cdot\mathbf{x})\hat{g}_1(\mathbf{k})\,d\mathbf{k}. \tag{16}$$

Applying the operator $T$ to Eq. 11 yields

$$\begin{aligned}-(\nabla^2+k_0^2)\bar{w}(\mathbf{x},\omega)-\epsilon^2[3k_0^2\bar{w}(\mathbf{x},\omega)\\ +\pi^{-1}k_0^4\int r^{-1}\exp(ik_0r)R(\mathbf{r})\bar{w}(\mathbf{x}+\mathbf{r},\omega)\,d\mathbf{r}]=\delta(\mathbf{x}),\end{aligned} \tag{17}$$

where $k_0=\omega/c_0$. Operating on Eq. 17 with $S$ gives

$$D(\mathbf{k},\omega)\hat{\bar{w}}(\mathbf{k},\omega)=1, \tag{18}$$

where

$$\begin{aligned}D(\mathbf{k},\omega)=k^2-k_0^2-\epsilon^2[3k_0^2+\pi^{-1}k_0^4\int r^{-1}\exp(ik_0r)R(\mathbf{r})\\ \times\exp(i\mathbf{k}\cdot\mathbf{r})\,d\mathbf{r}].\end{aligned} \tag{19}$$

We now assume that the medium is isotropic, so that $R(\mathbf{r})=R(r)$. Then we can carry out the angular integration in Eq. 19, after which we find that

$$\begin{aligned}D(\mathbf{k},\omega)&=D(k,\omega)\\ &=k^2-k_0^2-\epsilon^2k_0^2[3+4k_0^2k^{-1}\int_0^\infty\exp(ik_0r)R(r)\sin kr\,dr].\end{aligned} \tag{20}$$

Hence, from Eq. 18,

$$\hat{\bar{w}}(\mathbf{k},\omega)=\hat{\bar{w}}(k,\omega)=[D(k,\omega)]^{-1}. \tag{21}$$

The function $\bar{w}(\mathbf{x},\omega)$ is obtained by operating on Eq. 21 with $S^{-1}$ and carrying out the resulting angular integration in k space. The result, after some algebra, is

$$\bar{w}(\mathbf{x},\omega)=\bar{w}(x,\omega)=(2\pi^2x)^{-1}\int_0^\infty[D(k,\omega)]^{-1}k\sin kx\,dk. \tag{22}$$

We can evaluate the integral in Eq. 22 by means of contour integration, after which the expression for $\bar{w}$ becomes

$$\bar{w}(x,\omega)=(2\pi x)^{-1}[D_k(k_1,\omega)]^{-1}k_1\exp(ik_1x). \tag{23}$$

Here $D_k$ is the derivative of $D$ with respect to $k$, and $k_1$ is the root of the dispersion equation $D(k,\omega)=0$ which lies in the upper half of the complex $k$ plane, and which has the property that $k_1\to k_0$ as $\epsilon\to 0$. To lowest order in $\epsilon$, it is given by

$$k_1=k_0+(1/2)\epsilon^2[3k_0+4k_0^2\int_0^\infty\exp(ik_0r)R(r)\sin k_0r\,dr]. \tag{24}$$

That $k_1$, as given by Eq. 24, indeed has a positive imaginary part is easily proved using an analysis similar to that of Keller (Ref. 9, p. 152). The same analysis shows that $|\mathrm{Re}k_1|>|k_0|$, and that the root near $-k_0$ has a negative imaginary part.

In deriving Eq. 23 we have neglected all other zeros of $D(k,\omega)$ lying in the upper half-plane. The justification for neglecting these zeros (if they exist) is that, as shown in Appendix A, they have large imaginary parts and hence correspond to rapidly-attenuating waves. Such waves will be important only in a region near the source. Since we are not interested in the behavior of the solution in this region, we disregard these waves.

We now rewrite Eq. 23 by inserting the formula for $k_1$ given by Eq. 24 into the expression for $D_k(k,\omega)$ obtained from Eq. 20. Upon neglecting higher-order terms in $\epsilon$, we obtain

$$\bar{w}(x,\omega)=[1+2\epsilon^2\Psi(k_0)](4\pi x)^{-1}\exp(ik_1x), \tag{25}$$

where we have defined

$$\Psi(k_0)=k_0^2\int_0^\infty \exp(ik_0r)rR(r)\cos k_0r\,dr$$
$$-k_0\int_0^\infty \exp(ik_0r)R(r)\sin k_0r\,dr. \qquad (26)$$

Note that the quantity $\bar{w}(\mathbf{x},\omega)\exp(-i\omega t)$, with $\bar{w}$ given by Eq. 25, is just the coherent wavefield radiated by a time-harmonic point source; i. e., it is the solution of Eq. 11 when $f(\mathbf{x},t)=\delta(\mathbf{x})\exp(-i\omega t)$. Since, as noted above, $\mathrm{Im}k_1>0$ and $|\mathrm{Re}k_1|>|k_0|$, we see that this wave decays exponentially with distance from the source, and also that its phase speed is less than $c_0$. These properties of the coherent wave are similar to those found in previous studies of plane, time-harmonic waves.[9,10]

The solution of Eq. 11 can now be obtained by operating on Eq. 25 with $T^{-1}$. This yields

$$w(\mathbf{x},t)=w(x,t)=(8\pi^2x)^{-1}\int_{-\infty}^{\infty}[1+2\epsilon^2\Psi(k_0)]$$
$$\times\exp[ik_0x(1+\tfrac{3}{2}\epsilon^2)+2i\epsilon^2k_0x\Phi(k_0)-i\omega t]\,d\omega. \qquad (27)$$

Here we have defined

$$\Phi(k_0)=k_0\int_0^\infty \exp(ik_0r)R(r)\sin k_0r\,dr, \qquad (28)$$

and we have substituted for $k_1$ from Eq. 24.

We can write Eq. 27 in a more convenient form as follows. We first note that

$$1+2\epsilon^2\Psi(k_0)=\exp[2\epsilon^2\Psi(k_0)]+O(\epsilon^4). \qquad (29)$$

Next, after substituting Eq. 29 into Eq. 27, we introduce $\kappa$, a new integration variable, which is defined so that $\kappa=k_0(1+\frac{3}{2}\epsilon^2)$. Upon dropping terms of order $\epsilon^4$ in the resulting expression for $w$, we obtain

$$w(x,t)=c_*(8\pi^2x)^{-1}\int_{-\infty}^{\infty}\exp\{i\kappa(x-c_*t)$$
$$+2\epsilon^2[\Psi(\kappa)+i\kappa x\Phi(\kappa)]\}\,d\kappa, \qquad (30)$$

where

$$c_*=(1+\tfrac{3}{2}\epsilon^2)^{-1}c_0. \qquad (31)$$

As a check on our results, we note that by setting $\epsilon=0$ in Eq. 30 we obtain the well-known Green's function solution for the case of a uniform medium.

In order to simplify Eq. 30 further, we now assume that $x$ is so large that we can neglect the term $\Psi(\kappa)$ compared to the term $i\kappa x\Phi(\kappa)$ in the square brackets in that equation. In dimensionless terms this means that we must have $x/l\gg1$, where $l$ is the correlation length of $R$. Then Eq. 30 becomes

$$w(x,t)=c_*(8\pi^2x)^{-1}\int_{-\infty}^{\infty}\exp[i\kappa y+2i\epsilon^2\kappa x\Phi(\kappa)]\,d\kappa, \qquad (32)$$

where we have defined $y=x-c_*t$.

We now write the integral of Eq. 32 in dimensionless form. We begin by introducing the normalized correlation function $S(s)$, defined by

$$R(r)=S(r/l). \qquad (33)$$

Next we define the length scale $\delta$ by

$$\delta=\epsilon(lx)^{1/2}. \qquad (34)$$

Then in terms of the integration variable $p=\kappa\delta$, Eq. 32 becomes

$$w(x,t)=c_*(8\pi^2\delta x)^{-1}\int_{-\infty}^{\infty}\exp[i\eta p-p^2\phi(\alpha^{-1}p)]\,dp, \qquad (35)$$

where we have defined $\eta=y/\delta$,

$$\alpha=\epsilon(x/l)^{1/2}, \qquad (36)$$

and

$$\phi(q)=\int_0^\infty(1-\exp(2iqs))S(s)\,ds. \qquad (37)$$

Equation 35 shows that the waveform associated with the Green's function, expressed in terms of dimensionless coordinates, is determined by the parameter $\alpha$. Depending on the magnitude of this parameter, we can identify three propagation zones which we call the near zone, the far zone, and the intermediate zone. They are discussed in detail below.

### A. The near zone

The near zone is defined by the condition that $\alpha\ll1$. Referring to Eq. 35 we see that, in this case, $|\alpha^{-1}p|\gg1$ over the entire range of integration, except for a small interval near the origin which we shall neglect. Hence, we can evaluate the integral of Eq. 35 approximately by substituting for the function $\phi$ its asymptotic expansion for large values of the argument. This is obtained by integrating by parts in Eq. 37, and can be written

$$\phi(q)=m_0+(2iq)^{-1}+O(q^{-3}), \qquad (38)$$

where

$$m_n\equiv\int_0^\infty s^nS(s)\,ds, \quad n=0,1,2,\cdots. \qquad (39)$$

Upon inserting Eq. 38 into Eq. 35 (after dropping the term of order $q^{-3}$ in Eq. 38) and noting that the resulting integral is tabulated, we obtain, after some manipulation,

$$w(x,t)=c_1(8\pi x\delta_1)^{-1}(m_0\pi)^{-1/2}\exp[-(x-c_1t)^2/4m_0\delta_1^2], \qquad (40)$$

where

$$\delta_1=(1+\tfrac{1}{2}\epsilon^2)^{-1}\delta, \qquad (41)$$

and

$$c_1=(1+2\epsilon^2)^{-1}c_0. \qquad (42)$$

Equation 40 describes the waveform associated with the Green's function in the near zone. It shows that, near the wavefront, i. e., near $x=c_1t$, the waveform is given approximately by a Gaussian curve. We see from Eqs. 34 and 41 that the waveform broadens in proportion to the square root of the propagation distance, while the amplitude of the wave decreases (in addition to the decrease due to spherical spreading) as the inverse square root of the propagation distance. Equations 40 and 42 show that the wave propagates with a speed equal to $c_1$, and that $c_1$ is less than $c_0$. Thus, the propagation speed of the wave is reduced by the randomness of the medium.

Replacing the function $\phi$ in Eq. 35 by the approximate form given by Eq. 38 is equivalent to considering the effect of the randomness of the medium on only the high-frequency components of the wave. This effect consists of an attenuation which is proportional to the square of the frequency, as well as a frequency-independent reduc-

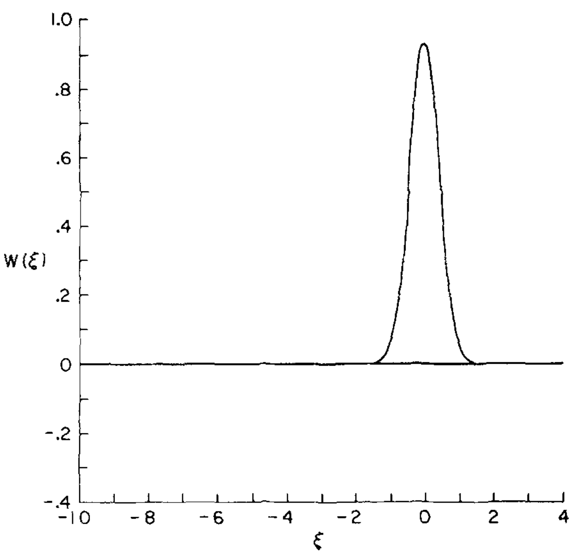



FIG. 1. The waveform associated with the Green's function, computed using Eq. 49, for the case $\alpha = 0.1$. The function $W(\xi)$ is related to the Green's function by Eq. 48. The stretched coordinate $\xi$ is defined so that $\xi = \delta_3^{-1}(x - c_* t)$, where $\delta_3$ and $c_*$ are given by Eqs. 50 and 31. The wave is propagating from left to right.

tion in propagation speed. Hence, in the near zone, the effect of the randomness of the medium on transient waves, insofar as the waveform is concerned, is essentially that of a pseudoviscosity, with the waveform being determined mainly by attenuation of the high-frequency components of the wave.

Our near-zone solution (Eq. 40) is similar to one which was obtained by Cole and Friedman (Ref. 6, p. 71, Eq. 11) for the case of a plane wave propagating in a turbulent medium. (Their solution is expressed in terms of an error function since they considered a step-function, rather than a delta-function, pulse). These authors obtained their result by, in effect, solving the linearized Burgers' equation. Since it is well known that this equation governs (approximately) the propagation of small-amplitude sound waves in a viscous fluid,[14] their formulation better illustrates the pseudoviscous character of the medium noted above.

### B. The far zone

The far zone is defined by the condition that $\alpha \gg 1$. In this case $|\alpha^{-1}p| \ll 1$ over that range of $p$ which yields the major contribution to the integral of Eq. 35. Hence we can evaluate that integral approximately by substituting for $\phi$ the first few terms of its power series expansion. This is obtained by expanding the function $\exp(2iqs)$ in Eq. 37 in a power series and integrating term by term. Upon dropping all but the first term of the expansion we find that the integral of Eq. 35 can be evaluated in terms of the Airy function. After some manipulation, the resulting expression for $w$ can be written

$$w(x, t) = c_*(4\pi x \delta_3)^{-1}\mathrm{Ai}(\xi) \tag{43}$$

where Ai denotes the Airy function,

$$\delta_3 = (6\epsilon^2 m_1 l^2 x)^{1/3}, \tag{44}$$

and $\xi = \delta_3^{-1} y$.

Equation 43 describes the waveform associated with the Green's function in the far zone. We see that, near the wavefront, i.e., near $x = c_* t$, the waveform is given approximately by an Airy function. Equations 43 and 44 show that the waveform broadens in proportion to the cube root of the propagation distance, while the amplitude of the wave decreases (in addition to the decrease due to spherical spreading) as the inverse cube root of the propagation distance. Note that the propagation speed of the wave is equal to $c_*$. This is slightly greater than the propagation speed in the near zone. However, from Eq. 31, $c_*$ is less than $c_0$; hence, as in the near zone, the propagation speed is reduced by the randomness of the medium.

More detailed information regarding the waveform defined by Eq. 43 can be obtained from any standard reference work on the Airy function (see, e.g., Ref. 15).

Replacing the function $\phi$ in Eq. 35 by the first few terms of its power-series expansion is equivalent to neglecting all but the low-frequency components of the wave. Moreover, in keeping only the first term of this expansion we are, in effect, ignoring attenuation of the low-frequency components, and considering only dispersion. Hence, in the far zone the high-frequency components of the wave have effectively attenuated to zero, with the waveform being determined primarily by dispersion of the low-frequency components.

### C. The intermediate zone

The intermediate zone is defined by the condition that $\alpha \approx 1$. In this case it is necessary to evaluate the integral of Eq. 35 numerically. In order to do this, we must first assume a particular form for the correlation function of the medium. Here we shall assume a Gaussian correlation function; i.e., we write

$$S(s) = \exp(-s^2). \tag{45}$$

By substituting Eq. 45 into Eq. 37 we find that the function $\phi(q)$ for this case can be written

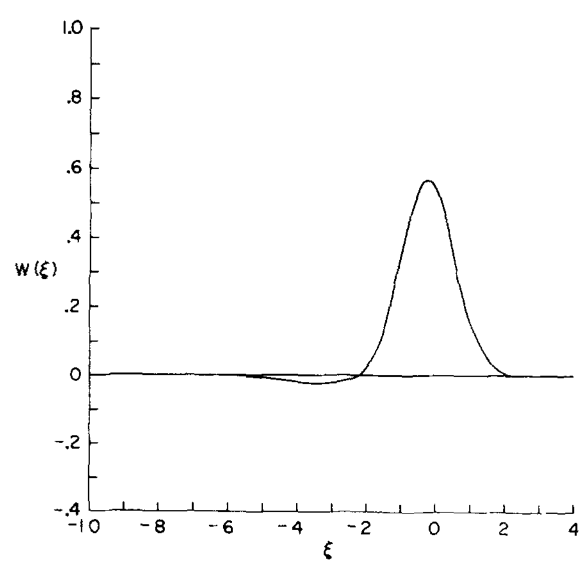



FIG. 2. Same as Fig. 1, except that $\alpha = 0.5$.

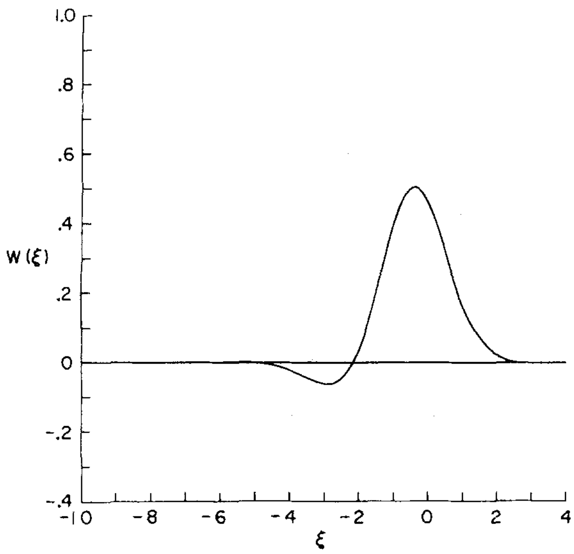



FIG. 3. Same as Fig. 1, except that $\alpha = 1.0$.

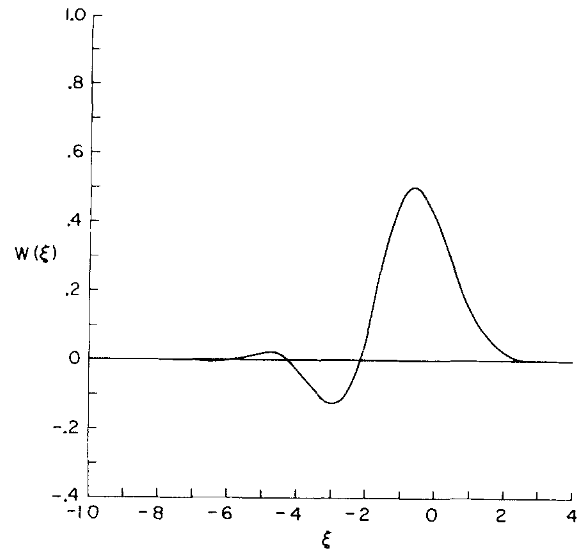



FIG. 4. Same as Fig. 1, except that $\alpha = 2.0$.

$$\phi(q) = (\pi^{1/2}/2)[1 - \exp(-q^2)] - iD(q), \tag{46}$$

where $D(q)$ is Dawson's integral:

$$D(q) = \exp(-q^2) \int_0^q \exp(z^2)\, dz. \tag{47}$$

With the aid of Eq. 46, Eq. 35 can be put into the form

$$4\pi c_*^{-1} \delta_3 x w(x, t) = W(\xi), \tag{48}$$

where

$$W(\xi) = \pi^{-1} \int_0^\infty \cos[\xi p + \tfrac{1}{3}\gamma p^2 D(p/\gamma)] \exp[-\tfrac{1}{6}\pi^{1/2}\gamma p^2(1 - \exp(-p^2/\gamma^2))]\, dp, \tag{49}$$

$\gamma = (3\alpha^2)^{1/3}$, and, in this case,

$$\delta_3 = (3\epsilon^2 l^2 x)^{1/3}. \tag{50}$$

The function $W(\xi)$ describes the waveform associated with the Green's function in the intermediate zone. Using Eqs. 47 and 49, we have made numerical calculations of $W(\xi)$ for a range of $\xi$ near the wavefront (i.e., near $\xi = 0$), and for several values of $\alpha$. These results are plotted in Figs. 1–5. The figures describe the transition of the wave profile from the near-zone to the far-zone form as the wave propagates through the intermediate zone (i.e., as $\alpha$ increases). The most prominent feature of this transition is the development of an oscillating "tail" on the profile which becomes more pronounced with propagation distance. This is the result of dispersion of the low-frequency components of the wave which, as we have seen, becomes increasingly important in determining the wave profile as the wave propagates into the far zone.

## ACKNOWLEDGMENTS

Research supported by the Atmospheric Sciences Section, National Science Foundation, NSF Grant GA-31561, and by the National Research Council. Part of this work was done while the author held an NRC Resident Research Associateship at NASA-Ames Research Center. The author is grateful to Henry Poor of the Division of Atmospheric Science, University of Miami, and to Margaret Johnson of the Computer Sciences Corporation, Mountain View, California, for programming the numerical calculations.

## APPENDIX A

We wish to show here that all the roots of the dispersion equation

$$D(k, \omega) = 0 \tag{A1}$$

which lie in the upper half of the complex $k$ plane, and which are bounded away from $\pm k_0$ as $\epsilon \to 0$ (i.e., all the roots except $k_1$) have the property that $\operatorname{Im} k \to +\infty$ as $\epsilon \to 0$.

To see this, we assume that $k = k(\epsilon)$ is such a root. By using the definition of $D(k, \omega)$ given by Eq. 20, we can write Eq. A1 in the form

$$k^2 - k_0^2 = \epsilon^2 k_0^2 [3 + 4k_0^2 k^{-1} \int_0^\infty \exp(ik_0 r) R(r) \sin kr\, dr]. \tag{A2}$$

Since the left-hand side of Eq. A2 is bounded away from

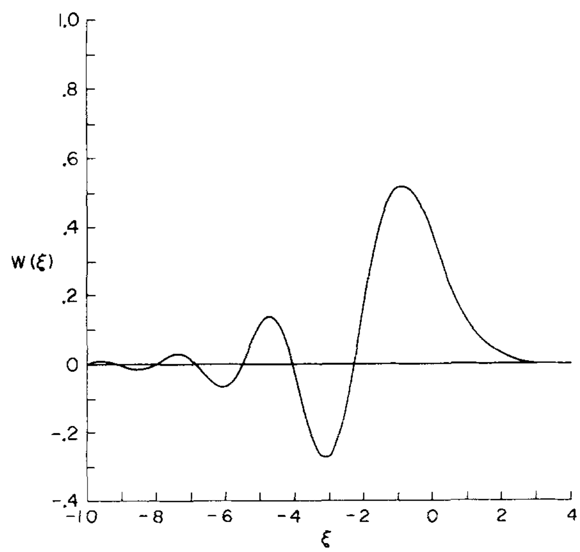



FIG. 5. Same as Fig. 1, except that $\alpha = 10.0$.

zero as $\epsilon \to 0$, the right-hand side must be similarly bounded. This requires that the second term in the brackets in Eq. A2 be unbounded as $\epsilon \to 0$. This term, however, is an entire function of $k$; hence we must have $|k| \to \infty$ as $\epsilon \to 0$. As a consequence, the integral of Eq. A2 must be unbounded as $\epsilon \to 0$. But this is possible only if $\mathrm{Im}\, k \to \infty$ as $\epsilon \to 0$, as can be seen by writing the term $\sin kr$ in exponential form. Thus, the result is established.

[1] U. Frisch, "Wave Propagation in Random Media," in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968).
[2] S.C. Crow, J. Fluid Mech. 37, 529 (1969).
[3] A. D. Pierce, J. Acoust. Soc. Am. 49, 906 (1971).
[4] A. R. George and K. Plotkin, Phys. Fluids 14, 548 (1971).
[5] K. J. Plotkin and A. George, J. Fluid Mech. 54, 449 (1972).
[6] W. J. Cole and M. B. Friedman, "Analysis of Multiple Scattering of Shock Waves by a Turbulent Atmosphere," in *Third Conference on Sonic Boom Research*, edited by I. R. Schwartz, NASA SP-255 (1971).
[7] M. S. Howe, J. Sound Vib. 24, 269 (1972).
[8] A. D. Pierce and D. Maglieri, J. Acoust. Soc. Am. 51, 702 (1972).
[9] J. B. Keller, Proc. Symp. Appl. Math. 16, 145 (1964).
[10] A. R. Wenzel and J. B. Keller, J. Acoust. Soc. Am. 50, 911 (1971).
[11] R. E. Collin, Radio Sci. 4, 279 (1969).
[12] F. C. Karal and J. B. Keller, J. Math. Phys. 5, 537 (1964).
[13] J. B. Keller and F. Karal, J. Math. Phys. 7, 661 (1966).
[14] D. T. Blackstock, J. Acoust. Soc. Am. 41, 1312 (1967).
[15] H. A. Antosiewicz, "Bessel Functions of Fractional Order," in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegun (National Bureau of Standards, Washington, D.C., 1964), p. 446.

# Discrete state perturbation theory via Green's functions*

William Rubinson

*Department of Chemistry, Brookhaven National Laboratory, Upton, New York 11973*
(Received 14 June 1974)

The exposition of stationary state perturbation theory via the Green's function method in Goldberger and Watson's *Collision Theory* is reworked in a way that makes explicit its mathematical basis. It is stressed that the theory consists of the construction of, and manipulations on, a mathematical identity. The perturbation series fall out of the identity almost immediately. The logical status of the method is commented on.

This note is a reworking of the version of stationary state perturbation theory given in Sec. 8.1 of Goldberger and Watson's book *Collision Theory*.[1] Readers of that section may be pleased to learn how simple it can be made. The account presented here involves only a few simple manipulations and, I believe, makes the nature of the mathematical argument more evident. Even people who are constitutionally repelled by apparently unmotivated mathematical procedures may be impressed by the way this one deftly produces the sought-for perturbation series, as if by sleight-of-hand, out of only the definition of a Green's function and a related mathematical identity.

## I. THE GREEN'S FUNCTION AND MATHEMATICAL IDENTITIES RELATED TO IT

The notation used in the following is essentially Goldberger and Watson's.

The operator Green's function $G(E)$ for the Schrödinger equation

$$H|\lambda\rangle = E_\lambda|\lambda\rangle \tag{I.1}$$

is an operator function of a numerical variable $E$ defined in terms of the Hamiltonian $H$ by

$$G(E) = 1/(E-H). \tag{I.2}$$

For any eigenket $|\lambda\rangle$ of $H$ belonging to eigenvalue $E_\lambda$

$$G(E)|\lambda\rangle = [1/(E-E_\lambda)]|\lambda\rangle = G_\lambda(E)|\lambda\rangle, \tag{I.3}$$

which shows that for all $E \neq E_\lambda$ the eigenket $|\lambda\rangle$ of $H$ is also an eigenket of $G(E)$, belonging to eigenvalue $G_\lambda(E) = (E-E_\lambda)^{-1}$. If we take $E$ to be a complex variable, $G_\lambda(E)$ is an analytic function of $E$ defined everywhere in the $E$ plane except at $E=E_\lambda$, where it has a singularity. If $E_\lambda$ is a discrete eigenvalue of $H$, the singularity is a simple pole, so that

$$G_\lambda(E) \to \infty \quad \text{as } E \to E_\lambda. \tag{I.4}$$

If $E_\lambda$ is a point of continuum of eigenvalues of $H$, the singularity is essential, i.e., $G_\lambda(E)$ tends to no unique value as $E \to E_\lambda$.

We treat here only the case of discrete $E_\lambda$, so that we can appeal to (I.4). In addition we will assume that $E_\lambda$ is nondegenerate. It should be remarked that the equations deduced below without appeal to (I.4) are valid for continuum $E_\lambda$ and so can be used to treat perturbations of a continuum.

Let $|a\rangle$ be any ket whatever that is expandable in eigenkets of $H$. Then [cf. (I.3)]

$$\begin{aligned} G(E)|a\rangle &= \sum_{\lambda'} \frac{|\lambda'\rangle\langle\lambda'|a\rangle}{E-E_{\lambda'}} \\ &= \frac{1}{E-E_\lambda}\left(|\lambda\rangle\langle\lambda|a\rangle + (E-E_\lambda)\sum_{\lambda'\neq\lambda}\frac{|\lambda'\rangle\langle\lambda'|a\rangle}{E-E_{\lambda'}}\right). \end{aligned} \tag{I.5}$$

Define

$$F(E)|a\rangle = G(E)|a\rangle/\langle a|G(E)|a\rangle\,. \tag{I.6}$$

Then, in view of (I.5),

$$F(E)|a\rangle = (|\lambda\rangle\langle\lambda|/|\langle\lambda|a\rangle|^2)|a\rangle + O(E-E_\lambda), \tag{I.7}$$

which, in the limit as $E \to E_\lambda$, gives

$$F(E_\lambda)|a\rangle = (|\lambda\rangle\langle\lambda|/|\langle\lambda|a|^2)|a\rangle|^2)|a\rangle, \tag{I.8}$$

showing, since $|a\rangle$ is arbitrary, that $F(E_\lambda)$ is proportional to $|\lambda\rangle\langle\lambda|$, the operator that projects from the state space of $H$ onto the eigenstate $|\lambda\rangle$. Note that this is basically a mathematical triviality; it is the result, essentially, of taking $|\lambda\rangle\langle\lambda|/(E-E_\lambda)$ plus a function of $E$ that is nonsingular at $E \approx E_\lambda$, multiplying this sum by const $x(E-E_\lambda)$, and then setting $E=E_\lambda$.

Next, from (I.2) we have the operator identity

$$(E-H)G(E) = 1 \tag{I.9}$$

which, multiplied by $|a\rangle/\langle a|G(E)|a\rangle$ gives the identity [cf. (I.6)]

$$(E-H)F(E)|a\rangle = [1/G_a(E)]|a\rangle, \tag{I.10}$$

where we have written

$$G_a(E) = \langle a|G(E)|a\rangle. \tag{I.11}$$

Inspection of the expression for $\langle a|G(E)|a\rangle$ from (I.5) shows that

$$\lim_{E\to E_\lambda} G_a(E) = \infty. \tag{I.12}$$

Therefore, as $E \to E_\lambda$ the identity (I.10) reduces to

$$(E_\lambda - H)\; F(E_\lambda)|a\rangle = 0, \tag{I.13}$$

showing that $F(E_\lambda)|a\rangle$ is an (unnormalized) eigenket of $H$ belonging to eigenvalue $E_\lambda$. This is another trivial mathematical truth, in view of (I.1) and (I.8).

Note well that (I.10) is an identity. It holds for all values of $E$, including $E_\lambda$, for all kets $|a\rangle$ that are expandable in eigenkets of $H$, and for all analytic functions $G_a(E)$ that have a simple pole at $E=E_\lambda$. One would expect, therefore, that nothing but further mathematical trivialities could be deduced from it. Remarkably enough, it can serve as the basis for a rapid and easy deduction of formal perturbation series for the eigenvalue $E_\lambda$ and the eigenket $|\lambda\rangle$ of $H$.

## II. DEDUCTION OF PERTURBATION SERIES FOR $E_\lambda$ AND $|\lambda\rangle$

This can be accomplished by specializing the general ket $|a\rangle$ of Sec I to be an eigenket, belonging to eigenvalue $\epsilon_a$ of the Hamiltonian $K$ of an auxiliary Schrödinger equation

$$K|a\rangle=\epsilon_a|a\rangle. \tag{II.1}$$

The Hamiltonians $H$ and $K$ differ by an operator $V$:

$$H=K+V. \tag{II.2}$$

$K$ is completely arbitrary, except that at least one of its eigenkets $|a\rangle$ must be such that $\langle\lambda|a\rangle\neq 0$. It is idle to insert at this point the customary remarks about $V$ being sufficiently restricted to insure convergence of the perturbation series that ensue, because no appeal is made to such restrictions in the course of deducing the series. See the remarks in Sec. IV below.

Insertion of (II.2) into the identity (I.10) gives the key identity

$$(E-K-V)F(E)|a\rangle=[1/G_a(E)]|a\rangle. \tag{II.3}$$

Here the representative of the ket $F(E)|a\rangle$ in the representation with $K$ diagonal is [cf. (I.6) and (I.11)]

$$\langle a'|F(E)|a\rangle=\langle a'|G(E)|a\rangle/G_a(E). \tag{II.4}$$

In particular, the diagonal element is, in view of (I.11),

$$\langle a|F(E)|a\rangle=1, \tag{II.5}$$

identically, a fact that will be of great service.

Our goal, formal perturbation series for the eigenvalue $E_\lambda$ and the eigenket $|\lambda\rangle$ of $H$, is now rapidly attained from the key identity (II.3) with the help of the identity (II.5) and the property (I.12).

First note that by (I.12), at $E=E_\lambda$ Eq. (II.3) reduces to

$$(E_\lambda-K-V)F(E_\lambda)|a\rangle=0, \tag{II.6}$$

showing [cf. (I.13) and the remarks that follow it] that $F(E_\lambda)|a\rangle$ is an (unnormalized) eigenket of $H$ belonging to eigenvalue $E_\lambda$. Expressed as a perturbation series and normalized in the way described later, it is the sought-for eigenket solution of (I.1).

Next, multiply (II.3) by $\langle a|$ to get, noting (II.1) and (II.5),

$$E-\epsilon_a-\langle a|VF(E)|a\rangle=1/G_a(E), \tag{II.7}$$

and substitute this expression for $1/G_a(E)$ back into (II.3) to get

$$(E-K-V)F(E)|a\rangle=(E-\epsilon_a)|a\rangle-|a\rangle\langle a|VF(E)|a\rangle, \tag{II.8}$$

which can be written as

$$(E-K)F(E)|a\rangle=(E-K)|a\rangle+(1-|a\rangle\langle a|)VF(E)|a\rangle. \tag{II.9}$$

Note that this form is valid only when used with ket $|a\rangle$.

From (II.9) we have $F(E)$ expressed in terms of known operators as

$$F(E)=1+[1/(E-K)](1-|a\rangle\langle a|)VF(E). \tag{II.10}$$

This can be expanded in an infinite series by iteration, starting with the zeroth-order approximation $F(E)=1$. The result, which is immediate, is

$$F(E)=1+\sum_{n=1}^{\infty}\left[\frac{1}{E-K}(1-|a\rangle\langle a|)V\right]^n \tag{II.11}$$

where the bracketed expression to the $n$th power means

$$\left[\frac{1}{E-K}(1-|a\rangle\langle a|)V\right]\left[\frac{1}{E-K}(1-|a\rangle\langle a|)V\right]\cdots \quad n\text{ times.}$$

With Eqs. (II.7) and (II.11) we have essentially reached our goal of a perturbation series for the computation of the eigenvalues $E_\lambda$ of $H$. The explicit series, given as Eq. (II.16) below, follows immediately from the fact that, by (I.12), at $E=E_\lambda$ Eq. (II.7) reduces to

$$E_\lambda=\epsilon_a+\langle a|VF(E_\lambda|a\rangle, \tag{II.12}$$

the right side of which is now completely known, $\epsilon_a$ and $|a\rangle$ being known as solutions of (II.1), $V$ being known from (II.2), and $\langle a|VF(E_\lambda)|a\rangle$ being known from (II.11) as the diagonal matrix element of

$$VF(E_\lambda)=V+V\frac{1}{E_\lambda-K}(1-|a\rangle\langle a|)V+V\frac{1}{E_\lambda-K}(1-|a\rangle\langle a|)V\frac{1}{E_\lambda-K}(1-|a\rangle\langle a|)V+\cdots, \tag{II.13}$$

namely, expanded in the eigenkets of $K$ [and recognizing that $\langle a'|(1-|a\rangle\langle a|)=\langle a'|(1-\delta_{a'a})$],

$$\langle a|VF(E_\lambda)|a\rangle=\langle a|V|a\rangle+\sum_{a'\neq a}\langle a|V|a'\rangle\times\frac{1}{E_\lambda-\epsilon_{a'}}\langle a'|V|a\rangle+\cdots. \tag{II.14}$$

Equation (II.12) with Eq. (II.14) is the sought-for series for $E_\lambda$. It can be evaluated by successive approximations, as follows.

Introduce the abbreviation

$$R_a(E)=\langle a|VF(E)|a\rangle. \tag{II.15}$$

Then, by (II.12) and (II.14),

$$E_\lambda=\epsilon_a+R_a(E_\lambda)=\epsilon_a+\langle a|V|a\rangle+\sum_{a'\neq a}\frac{|\langle a|V|a'\rangle|^2}{E_\lambda-\epsilon_{a'}}+\sum_{a'\neq a}\sum_{a''\neq a}\frac{\langle a|V|a'\rangle\langle a'|V|a''\rangle\langle a''|V|a\rangle}{(E_\lambda-\epsilon_{a'})(E_\lambda-\epsilon_{a''})}+\cdots. \tag{II.16}$$

Denote the $\nu$th approximation value of $E_\lambda$ by $E_\lambda^{(\nu)}$. As the zeroth approximation take $E_\lambda^{(0)}=\epsilon_a$ which, inserted into (II.16) gives the first-approximation value

$$E_\lambda^{(1)}=\epsilon_a+R_a(\epsilon_a)=\epsilon_a+\langle a|V|a\rangle+\sum_{a'\neq a}\frac{|\langle a|V|a'|^2}{\epsilon_a-\epsilon_{a'}}+\cdots, \tag{II.17}$$

which is just the expression given by Rayleigh–Schrödinger perturbation theory. Substitution of $E^{(1)}$ for $E_\lambda$ in the rightmost side of (II.16) gives

$$E_\lambda^{(2)} = \epsilon_a + R_a(E_\lambda^{(1)}),$$

etc., the $\nu$th approximation value being

$$E_\lambda^{(\nu)} = \epsilon_a + R_a(E_\lambda^{(\nu-1)}). \tag{II.18}$$

This completes our deduction of $E_\lambda$. The problem of deducing the corresponding eigenket is already essentially solved. For, as remarked in connection with Eq. (II.6), all that is required is to normalize $F(E_\lambda)|a\rangle$. Let $1/C$ be the normalization constant. Then, by Eq. (II.11)

$$|\lambda\rangle = \frac{1}{C} F(E_\lambda)|a\rangle$$

$$= \frac{1}{C}\left(|a\rangle + \sum_{a'\neq a} \frac{\langle a'|V|a\rangle}{E_\lambda - \epsilon_{a'}}|a'\rangle + \cdots\right) \tag{II.19}$$

with $C$ computed from ($\bar{F}$ denotes the Hermitian conjugate of $F$)

$$|C|^2 = \langle a|\bar{F}(E_\lambda)F(E_\lambda)|a\rangle = 1 + \sum_{a'\neq a} \frac{|\langle a'|V|a\rangle|^2}{(E_\lambda - \epsilon_{a'})^2} + \cdots, \tag{II.20}$$

using the value of $E_\lambda$ from the preceding perturbation calculation.

## III. DEDUCTION OF AN ALTERNATIVE FORM OF PERTURBATION SERIES

An alternative form of perturbation series can be obtained from our general equations, as follows. Define $\delta E$ by

$$E = \epsilon_a + \delta E \tag{III.1}$$

and insert this expression for $E$ into (II.9) to get

$$(\epsilon_a - K + \delta E)[F(E)|a\rangle - |a\rangle] = (1 - |a\rangle\langle a|)VF(E)|a\rangle,$$

which, in view of [cf. (II.5)],

$$F(E)|a\rangle - |a\rangle = (1 - |a\rangle\langle a|)F(E)|a\rangle,$$

can be written

$$(\epsilon_a - K)[F(E) - 1]|a\rangle = (1 - |a\rangle\langle a|)(V - \delta E)F(E)|a\rangle.$$

Consequently,

$$F(E) = 1 + [1/(\epsilon_a - K)](1 - |a\rangle\langle a|)(V - \delta E)F(E). \tag{III.2}$$

Expansion of expression (III.2) in an infinite series by iteration gives [compare (II.11)]

$$F(E) = 1 + \sum_{n=1}^{\infty}\left[\frac{1}{\epsilon_a - K}(1 - |a\rangle\langle a|)(V - \delta E)\right]^n \tag{III.3}$$

and, therefore, as in the transition from (II.13) to (II.14) (and taking account of the fact that $\langle a'|\delta E|a\rangle = \delta_{a'a}\cdot\delta E$),

$$R_a(E) = \langle a|VF(E)|a\rangle = \langle a|V|a\rangle + \sum_{a'\neq a}\frac{|\langle a|V|a'\rangle|^2}{\epsilon_a - \epsilon_{a'}}$$
$$- \delta E \sum_{a'\neq a}\frac{|\langle a|V|a'\rangle|^2}{(\epsilon_a - \epsilon_{a'})^2}$$
$$+ \sum_{a'\neq a}\sum_{a''\neq a}\frac{\langle a|V|a'\rangle\langle a'|V|a''\rangle\langle a''|V|a\rangle}{(\epsilon_a - \epsilon_{a'})(\epsilon_a - \epsilon_{a''})} + \cdots. \tag{III.4}$$

Now note that for the case $E = E_\lambda$ we have from (III.1), (II.15), and (II.12) that

$$R_a(E_\lambda) = E_\lambda - \epsilon_a = \delta E, \tag{III.5}$$

which, inserted into (III.4), gives the sought-for second form of perturbation series as

$$E_\lambda = \epsilon_a + \langle a|V|a\rangle + \sum_{a'\neq a}\frac{|\langle a|V|a'\rangle|^2}{\epsilon_a - \epsilon_{a'}}$$
$$- (E_\lambda - \epsilon_a)\sum_{a'\neq a}\frac{|\langle a|V|a'\rangle|^2}{(\epsilon_a - \epsilon_{a'})^2} + \cdots. \tag{III.6}$$

We remark that the expression (III.6) is simply the Taylor expansion of the series (II.16) about the value $\epsilon_a$, for

$$E_\lambda - \epsilon_a = R_a(E_\lambda) = R_a(\epsilon_a + E_\lambda - \epsilon_a)$$
$$= R_a(\epsilon_a) + R_a'(\epsilon_a)(E_\lambda - \epsilon_a) + \tfrac{1}{2}R_a''(\epsilon_a)(E_\lambda - \epsilon_a)^2 + \cdots,$$

and $R_a'(\epsilon_a) = [dR_a(E_\lambda)/dE_\lambda]_{E_\lambda = \epsilon_a}$, as evaluated from the series (II.16), is identical with the coefficient of $E_\lambda - \epsilon_a$ in (III.6).

## IV. COMMENTS ON THE GREEN'S FUNCTION METHOD

From the above account the essence of the Green's function method can be seen to consist of the following:

1. Construction of the operator identity $(E - H)G(E) = 1$, observation that for any normalized state whatever, $|\chi\rangle$, its mean value is necessarily $\langle\chi|(E - H)G(E)|\chi\rangle = 1$, and further observation that division of this numerical identity by any one function $f(E)$ of a certain broad class of analytic functions gives

$$\langle\chi|(E - H)G(E)|\chi\rangle/f(E) = 1/f(E), \tag{IV.1}$$

a trivial mathematical truth, holding for all values of $E$, all normalized kets $|\chi\rangle$, and all functions $f(E)$ in the given class.

2. Recognition of the fact that if $|\chi\rangle$ is restricted to be an eigenket $|a\rangle$ of any arbitrary Hermitian operator $K$ that has a complete set of eigenkets, and if, simultaneously, $f(E)$ is restricted to be $G_a(E) = \langle a|G(E)|a\rangle$, and if, further, $V$ is defined by $H = K + V$, then the trivial identity (IV.1) reduces to $\langle a|$ times the key identity (II.3) from which the perturbation series follow almost immediately.

This analysis makes it clear that the method by which the key identity was obtained can provide no help in ascertaining the applicability of the result in any practical problem. The method provides formal perturbation series; the question of their applicability is a completely independent one (essentially untouched in our exposition). The question can be stated as: Given a Schrödinger equation with Hamiltonian $H$, what $K$'s can be used to effect a solution?

At first sight it is puzzling that something as nontrivial as perturbation series can be generated from an identity, a feat transcending any powers that inhere in an identity. The power for the feat comes, not from the identity, but from the Hilbert space of a Hamiltonian $K$ into which the identity is inserted. The role played here by the identity is more or less like the role played with such great effect, in certain arguments in mathematics, by the trick of multiplying by a complicated form of the number 1; e.g., in the theory of the $\Gamma$-function a multiplication by $\exp[\sum_n(1/n - 1/n)]$ is used in arriving at Weierstrass' expression for $\Gamma(z)$.

In the theory of the $\Gamma$-function it is advantageous to take Weierstrass' expression as the *definition* of $\Gamma(z)$. One must then ascertain the set of $z$'s for which the definition is meaningful [simple inspection suffices for this in the case of $\Gamma(z)$]. Similarly, in the present theory it would be logically advantageous to use the expression for $F(E)|a\rangle$ given in Eq. (II.9) to *define* the sought-for eigenket $|\lambda\rangle$ of $H$ by

$$|\lambda\rangle = \lim_{E \to E_\lambda} F(E)|a\rangle. \tag{IV.2}$$

Analogously to the case of the $\Gamma$ function, one would then have to ascertain the set of $K$'s, or, alternatively, the set of $V = H - K$, for which the definition is meaningful. In addition to its logical advantages such a procedure has a psychological one, namely, it makes it perfectly clear that the Green's function method brings no new content into perturbation theory.

We remark that all the equations in this note, except those obtained by setting $E = E_\lambda$, are valid for continuum states. They are the basis for the treatment of continuum state perturbation theory in Sec. 8.2 of Goldberger and Watson's book. Consequently, the important expressions obtained there, too, are the results of manipulation of identities. Some of the manipulations are very unobvious, but they are nonetheless manipulations of identities.

The problem of the conditions on $V = H - K$ needed to insure meaningful results in continuum state perturbation theory has been studied by the mathematicians. A set of such conditions is given in a book by Friedrichs.[2]

*Note added in proof*: For the deduction of discrete state perturbation series the limit process in (IV.2) and the Green's function machinery that suggested it are superfluous. The series can be derived with a minimum of work in the following way. Let $F(E_\lambda)$ be an unknown operator defined by

$$|\lambda\rangle = \langle a|\lambda\rangle F(E_\lambda)|a\rangle$$

which, note, entails $\langle a|F(E_\lambda)|a\rangle = 1$. Insertion of this expression for $|\lambda\rangle$ into $(E_\lambda - K - V)|\lambda\rangle = 0$ gives (II.6) which, left-multiplied by $\langle a|$ gives (II.12), and left-multiplied by $1 - |a\rangle\langle a|$ gives (II.9) with $E_\lambda$ in place of $E$. The remainder of the derivation follows unchanged.

*Work performed under the auspices of the U.S. Atomic Energy Commission.

[1]Marvin L. Goldberger and Kenneth M. Watson, *Collision Theory* (Wiley, New York, 1964), p. 425.
[2]Kurt O. Friedrichs, *Perturbation of Spectra in Hilbert Space* (Am. Math. Soc., Providence, R.I., 1965).

# An exact determination of the gravitational potentials $g_{ij}$ in terms of the gravitational fields $R^l_{ijk}$

Edwin Ihrig

*Department of Mathematics, University of Toronto, Toronto, Ontario, Canada*
(Received 15 April 1974)

Under a certain asymmetry assumption, the gravitational potentials $g_{ij}$ are determined up to a conformal factor from the field $R^l_{ijk}$. This constitutes a partial solution to Einstein's equations in the general case. Moreover, the solutions are simple in that the $g_{ij}$ are expressed as polynomial functions of the $R^l_{ijk}$.

## 1. INTRODUCTION

Since Einstein first proposed the general theory, a major task has been to find the solutions to his field equations. These equations are second order nonlinear partial differential equations in several unknowns and as such have been next to impossible to solve, except in special cases.

In this work we will give a totally different emphasis to the field equations. We have determined an exact solution for the potentials $g_{ij}$ (within a conformal factor) in terms of the gravitational field $R^l_{ijk}$ in very general circumstances. The condition we need on the metric is that there is not too much symmetry. This changes the emphasis in Einstein's equations to that of finding the gravitational field given the energy momentum tensor.

## 2. THE SOLUTION

We first define the conditions we need on the $R^l_{ijk}$ to get our solution. We assume that $R^l_{ijk}$ is the Riemann tensor of some unknown pseudometric $g_{ij}$. We will then find the $g_{ij}$. In other words, we do not show existence of a solution, we just show how to obtain the solution if it exists.

*Definition* 1: Let $V_m$ be the vector space generated by the following set:

$$\{R_m(v,w)\,|\,v,w\in T_m(M)\},$$

where $R$ is the Riemann tensor of some pseudometric. $R$ is called *total at m* if

$$\dim V_m = n(n-1)/2, \quad n=\dim M.$$

$R$ being total for every $m$ says two things. It says that $R$ does not collapse at any point, that is, that $V_m$ is the lie algebra of the holonomy group. Secondly, it says that there are no strong symmetries—no totally geodesic submanifolds; that is, it says that the holonomy group must be the whole Lorentz group. Most space—times satisfy both these criteria.

Now we proceed to find our solutions. Using the fact that $R_{ij,kl} = -R_{ij,lk}$ we have

$$g_{ll'}R^{l'}_{ij,k} = -g_{kl'}R^{l'}_{ij,l}$$

which are constraints on the $g_{ij}$ in terms of the $R^l_{ij,k}$. The following theorem says that these equations completely determine the $g_{ij}$ up to a conformal factor.

*Theorem* 1: Suppose $R$ is total at $m$. Then $g$ is determined to within a conformal factor by

$$g_{ll'}R^{l'}_{ij,k} + g_{kl'}R^{l'}_{ij,l} = 0 \tag{1}$$

at the point $m$.

*Proof*: By a theorem of Ambrose and Singer[1] we have that the holonomy group of a connection has its Lie algebra spanned by elements of the form $\tau^{-1}R_m(v,w)\tau$, where $\tau$ is parallel transport along an arbitrary path in $M$. So

$$\begin{aligned} V_m &\subset \text{holonomy Lie algebra}\\ &\subset \text{Lorentz Lie algebra, and}\\ \dim(V_m) &= \dim\,(\text{Lorentz Lie algebra}). \end{aligned}$$

Thus we know $V_m$ is the Lie algebra of the Lorentz group. Also $g$ is specified to within a conformal factor by

$$LgL^+ = g, \tag{2}$$

where $L$ is allowed to be any Lorentz transformation. So

$$Lg = g(L^{-1})^+. \tag{3}$$

Now let $L$ be replaced by a one parameter subgroup $L(t)$, and, by taking derivatives at $t=0$, we find the following relationship in the Lie algebra of the Lorentz group:

$$Tg = -gT^+, \text{ where } T = \left.\frac{dL(t)}{dt}\right|_{t=0}. \tag{4}$$

Since $L$ in (2) only has to be taken from the identity component of the Lorentz group, (4) will imply (3) for all the needed transformations. We find that (1) is nothing more than

$$R^{\cdot}_{ij\cdot}\, g_{\cdot\cdot} = -g_{\cdot\cdot}\,(R^{\cdot}_{ij\cdot})^+,$$

where the dots stand for the matrix indices, $i$ and $j$ being fixed. Thus Eq. (1) implies Eq. (4) for all $T$ in the lie algebra of the Lorentz group, and we are finished.[2]

We proceed to give the exact solution. Now that we know that (1) determines the solution we need only solve (1). This problem is more notational than anything else. Let $\chi_{ij}$, $1 \leq \leq j \leq n$, be $n(n+1)/2$ independent vectors in a $n(n+1)/2$ dim vector space. Let $\chi_{ij} = \chi_{ji}$ if $i>j$. Define

$$v_{ij,k,l} = \chi_{ll'}R^{l'}_{ij,k} + \chi_{kl'}R^{l'}_{ij,l}$$

Now relable $v_{ij,k,l}$ so that they are indexed by one index $\alpha$ which goes from 1 to $(n-1)n^3+1$. Define $v_\alpha$ to be $\chi_{ij}$ for $\alpha$ ranging between $(n-1)n^3+1$ and $(n-1)n^3+(n+1)n/2$. We now have all the vectors defined that we need. We need only define the inner product in the space to be able to begin. The inner product ( , ) is defined by the rule

$(x_{i_1 j_1}, x_{i_2 j_2}) = \delta_{i_1 i_2}\delta_{j_1 j_2}$.

Now we will give the solution using the Gramm Schmidt method, since our solution will be a vector which is orthogonal to all the $v_{ij,k,l}$. We added the extra $v_\alpha$ to get the one more vector we needed to complete the process. We define

$$\bar{\omega}_\alpha = v_\alpha - \sum_{\beta<\alpha} (v_\alpha, \omega_\beta)\omega_\beta,$$

where $\omega_\alpha = \bar{\omega}_\alpha/(\bar{\omega}_\alpha, \bar{\omega}_\alpha)^{1/2}$, and also define

$$\omega = \sum_{\alpha=a+1}^{b} \omega_\alpha, \quad \begin{cases} a = (n-1)n^3/2 \\ b = a + n(n+1)/2. \end{cases}$$

We observe only one $\omega_\alpha$ will not be zero for $\alpha > a$.

Thus the above sum will be equal to the vector which will be orthogonal to all the $v_\beta$, $\beta < a$, and we have

$$(\omega, x_{ij}) = \lambda g_{ij}, \quad \lambda \text{ a conformal factor.}$$

Although $\omega$ will contain a lot of redundancy, a lot of terms which turn out to be zero, our construction has the following advantage. Given a particular dimension for our space–time (say 4), one can, using the above formula, write out the $g_{ij}$ explicitly as functions of the $R^l_{ijk}$. We summarize this solution in the following theorem.

*Theorem* 2: Let $R^l_{ijk}$ be total at $m$. Then $g_{ij}$ is determined to within a conformal factor $\lambda$ by the following formulas:

(i) $\lambda g_{ij} = (\omega, x_{ij})$;

(ii) $\omega = \sum_{\alpha=a+1}^{b} \omega_\alpha, \quad \begin{cases} a = (n-1)n^3/2 \\ b = a + n(n+1)/2; \end{cases}$

(iii) $\omega = \bar{\omega}/(\bar{\omega}_\alpha, \bar{\omega}_\alpha)^{1/2}$,

$\bar{\omega} = v_\alpha - \sum_{\beta<\alpha} (v_\alpha, \omega_\beta)\omega_\beta$;

(iv) $v_\alpha = \begin{cases} x_{ll'}R^{l'}_{ij,k} + x_{kl'}R^{l'}_{ij,l}, & \alpha < a, \\ x_{ij}, & \alpha \geq a; \end{cases}$

(v) $(x_{i_1 j_1}, x_{i_2 j_2}) = \delta_{i_1 i_2}\delta_{j_1 j_2}$.

## 3. DISCUSSION

Now having an explicit algebraic formula for $g_{ij}$ in terms of $R^l_{ij,k}$ one has reduced the problem of solving Einstein's equations for $g_{ij}$ to solving for $R^l_{ij,k}$. Since Einstein's equations specify $R^i_j$ in terms of energy quantities alone but not $R_{ij}$, one might wish to determine $R^{j,l}_{i,k}$ instead of $R^l_{ij,k}$. Because the $i$ and $j$ indices play no role in our equations, one can show exactly the same theorems with the $j$ index raised.

In specific cases, the formulas given in Theorem 2 may be considerably simplified. Suppose, for example, one has some reason to specify that the metric is diagonal (as in the Schwarzschild case). Then we may get the $g_{ij}$ in a very simple way from the $R^l_{ij,k}$. Suppose we want to find $g_{ii}$. We have

$$g_{ii}R^i_{jk,l} = -g_{ll}R^l_{jk,i}$$

or

$$g_{ii} = -g_{ii}(R^i_{jk,i}/R^i_{jk,l}) \text{ if } R^i_{jk,l} \neq 0.$$

There must be at least one $l \neq i$ and $j$, $k$ for which $R^i_{jk,l} = 0$, for otherwise the holonomy group would be $(n-1)$-dimensional. So we may start with a given index, 0 say, and express $g_{00}$ in terms of $g_{l_0 l_0}$. Then we find $g_{l_0 l_0}$ in terms $g_{l_1 l_1}$, etc. If we can not cover all the $g_{ll}$'s in this fashion, then the set $\{\partial/\partial x_{l_i}\}$ will form an invariant subspace under the action of the holonomy group. This would mean that the holonomy group is a proper subgroup of the Lorentz group contradicting our assumption that $R$ is total. Thus our final solution will look like

$$g_{ii} = \prod_{a\in I_i} \pm [R^{l_a}_{j_a k_a, i_a}/R^{i_a}_{j_a k_a, l_a}]g_{00}$$

So we see that we not only have a solution, but a manageable one at that. One might point out, however, that it may be nearly as difficult to get the $R^l_{ijk}$ from $T_{ij}$ as to completely solve Einstein's equations. Although this could be so, it is not of too much importance, since in practice the $T_{ij}$ and the $R^l_{ijk}$ are equally hard to determine in the nonempty case. They also enjoy approximately equal importance as physical objects—the one being the gravitational field, the other the energy tensor. Thus a solution of the $R^l_{ijk}$ in terms of the $T_{ij}$ would constitute more a translation of the problem than a solution. A real solution of the problem would be partially obtained by an existence theorem to go with our uniqueness theorem. An existence theorem would say when a tensor $R^l_{ijk}$ was actually the gravitational field for some configuration. If one had such a theorem, then one could pick an admisible $R^l_{ijk}$ that satisfied appropriate physical conditions and boundary conditions involving the value of $T_{ij}$ on a past-light cone. Then, using Theorem 2, one could find $g_{ij}$ to within a conformal factor. One would then solve the differential equation determining the conformal factor to get $g_{ij}$, thus determining the motion of particles in the field. So it appears the next step is to find existence theorems for Lorentz metrics having a given tensor $R^l_{ijk}$ as curvature.

[1]W. Ambrose and I. M. Singer, Trans. Amer. Math. Soc. 75, 428 (1953).

[2]A more elegant proof can be given from the bundle of frames point of view. An easy application of the concepts in B. G. Schmidt, Comm. Math. Phys. 19, 55 (1973), can be used to construct such a proof.

# On resonance and stability of conservative systems*

K. B. Dysthe and O. T. Gudmestad

*The University of Tromsø, The Auroral Observatory, P.O. Box 953, N-9001, Tromsø, Norway*
(Received 9 April 1974)

Resonance and stability of conservative systems are considered by means of a perturbation method similar to the averaging method of Bogoliubov. The accuracy of the method is tested by numerical simulations and by comparing the conditions for stability derived here with the well-known conditions given by Moser and Arnold.

## I. INTRODUCTION

In the following we consider conservative Hamiltonian systems of $n$ degrees of freedom. Assuming that the system is near an equilibrium, a number of perturbation techniques are available in dealing with the nonlinear differential equations describing the system. They are all based on expansions in terms of a small parameter $\epsilon$ characterizing the smallness of the initial deviation from equilibrium.

To the lowest order (regardless of the particular perturbation scheme) one obtains the so-called linearized equations. Assuming that the system is stable according to these equations, standard theory shows that there exist $n$ modes of harmonic vibrations with frequencies $\omega = \{\omega_1, \omega_2, \omega_3 \ldots, \omega_n\}$.

Carrying the approximation further, some of the perturbation schemes are confronted with the following difficulty. When the eigenfrequencies are commensurable, i.e., when a vector $\mathbf{n} = \{n_1, n_2, \ldots, n_n\}$ of integer components exists such that

$$\omega \cdot \mathbf{n} = \omega_1 n_1 + \omega_2 n_2 + \ldots + \omega_n n_n = 0,$$

the perturbation expansion may break down due to the occurrence of terms with $\omega \cdot \mathbf{n}$ in the denominator.

Astronomers long ago (Poincaré, 1881 [1]) noted that these "small denominators" lead to considerable mathematical difficulties. The name given to the problem seemed to indicate that it was mainly due to the lack of an adequate mathematical method, and not related to a physical phenomenon. Already Poincaré[1], however, noted that the problem was due to the very nature of the physical system. A more adequate name for the phenomenon is resonance, indicating that it relates to an observable effect.

In this report we consider a particular perturbation scheme (a variant of the "averaging method," Bogoliubov[2]). Our aim has been to test the accuracy of its predictions with regard to resonance and stability.

To this end we have followed two lines of approach. First we have compared by numerical simulations the predictions given by the original equations and the "averaged" equations emerging from the perturbation scheme. In particular we have investigated the accuracy of the method in describing a "sharp" phenomenon like the periodic solutions of a resonant system. Secondly we have investigated whether, by means of this perturbation method, one could indicate stability criteria for the system. The simplicity of the well known criteria given by Moser, 1962 [3] and Arnol'd 1963, [4] has been a great challenge.

As it turns out, the stability criteria found through our perturbation method are the same as those of Moser and Arnol'd for a system of two degrees of freedom ($n=2$). For a system with $n>2$ Arnol'd only gives conditions sufficient for stability for a "majority of initial conditions." We show that his conditions cannot be sufficient for all initial conditions. Further we derive conditions that should be sufficient for stability for all initial conditions.

## II. MATHEMATICAL MODEL AND GENERAL EQUATIONS

We consider a conservative mechanical system with $n$ degrees of freedom. The motion of this system is described by a Hamiltonian $H = H(\mathbf{q}, \mathbf{p})$, where $\mathbf{q}, \mathbf{p}$ are generalized coordinates and momentum. The system is assumed to be near an equilibrium position $\mathbf{q} = \mathbf{p} = 0$ which is stable according to a linearized theory. We introduce new coordinates by the canonical transformation

$$\mathbf{q} \to \epsilon \mathbf{q}, \quad \mathbf{p} \to \epsilon \mathbf{p}, \tag{2.1}$$

where $0 < \epsilon \ll 1$ is a measure of smallness of the deviation from equilibrium. The new Hamiltonian can be expanded about equilibrium as

$$H = \sum_{i=1}^{n} \frac{\omega_i}{2}(p_i^2 + q_i^2) + \sum_{m=3}^{\infty} \epsilon^{m-2} H_m. \tag{2.2}$$

Further $\omega_i$, $i = 1, 2, \ldots, n$, are real numbers and $H_m$ are homogeneous polynoms of order $m$ in $\mathbf{q}$ and $\mathbf{p}$.

The linear approximation is obtained from (2.2) when $\epsilon = 0$. The system then executes harmonic oscillations with frequencies $\omega_1, \ldots, \omega_n$, and the coordinates are normal coordinates.

It is useful to introduce angle-action variables. We then introduce the canonical transformation

$$(\mathbf{q}, \mathbf{p}) \xrightarrow{\mathrm{Tr}} (\theta, \mathbf{J}), \tag{2.3}$$

given by the generating function

$$F_1(\mathbf{q}, \theta) = \sum_{i=1}^{n} \frac{1}{2} q_i^2 \operatorname{cotg} \theta_i; \tag{2.4}$$

$$q_i = (2J_i)^{1/2} \sin\theta_i,$$
$$p_i = (2J_i)^{1/2} \cos\theta_i. \tag{2.5}$$

The Hamiltonian can now be written

$$H(\theta, \mathbf{J}) = \omega \cdot \mathbf{J} + \epsilon H_3(\theta, \mathbf{J}) + \epsilon^2 H_4(\theta, \mathbf{J}) = \ldots, \tag{2.6}$$

and the corresponding canonical equations of motion become

$$\dot{\theta} = \omega + \epsilon \frac{\partial H_3}{\partial \mathbf{J}} + \epsilon^2 \frac{\partial H_4}{\partial \mathbf{J}} + \dots,$$

$$\dot{\mathbf{J}} = -\epsilon \frac{\partial H_3}{\partial \theta} - \epsilon^2 \frac{\partial H_4}{\partial \theta} + \dots . \tag{2.7}$$

Introducing (1.5) into the Taylor expansion (2.2), one obtains

$$H = \overline{H}(\mathbf{J}) + \sum_{k=3}^{\infty} \epsilon^{k-2} A_i^k(\mathbf{J}) \sin(\mathbf{n}_i^k \cdot \theta + \kappa_i^k), \tag{2.8}$$

where

$$\overline{H}(\mathbf{J}) = \omega \cdot \mathbf{J} + \epsilon^2 \overline{H}_4 + \epsilon^4 \overline{H}_6 + \dots, \tag{2.9}$$

are the contributions to $H$ independent of $\theta$ (the symbol $^{-}$ represents an average over all components of $\theta$).

Further $\mathbf{n}_i^k$ are $n$-dimensional vectors with integer components such that $|\mathbf{n}_i^k| \equiv |n_{i1}^k| + \dots + |n_{in}^k| = k$, $\kappa_i^k$ are constants,

$$\mathbf{J} = (J_1, J_2, \dots, J_n), \quad J_i > 0,$$

$$\omega = (\omega_1, \omega_2 \dots, \omega_n),$$

$$\theta = (\theta_1, \theta_2, \dots, \theta_n),$$

$$A_i^k(\mathbf{J}) = \prod_{s=1}^{n} J_s^{|n_{is}^k|/2} \times \text{constant}.$$

Finally if we denote the sum in (2.8) $\widetilde{H}$, (2.7) can be recast to the form

$$\dot{\theta} = \frac{\partial \overline{H}}{\partial \mathbf{J}} + \frac{\partial \widetilde{H}}{\partial \mathbf{J}},$$

$$\dot{\mathbf{J}} = -\frac{\partial \widetilde{H}}{\partial \theta}. \tag{2.10}$$

## III. INTERNAL RESONANCE

In the linearized case ($\epsilon = 0$) the solution of (2.10) obviously is

$$\mathbf{J} = \text{const}, \text{ and } \theta = \omega t + \text{const}. \tag{3.1}$$

Thus the energy of the normal mode $k$, $J_k \omega_k$ ($k = 1, 2, \dots, n$), is a constant of motion.

If a small nonlinearity ($0 < \epsilon \ll 1$) is introduced, one does not expect anything drastic to happen on the short time scale $\tau_0 = 1/\omega$ (where $\omega$ is a characteristic frequency).

On some longer time scale $\tau_k = \tau_0/\epsilon^k (k \geq 1)$, however, the total energy of the system may be redistributed between the different modes.

To obtain more insight into this process of slow energy transfer, let us consider the equations (2.10). The first thing to notice, is that the rate of change of the angle variable $\theta$ is dominated by the term $\omega$. Since $\dot{\mathbf{J}}$ is of the order $\epsilon$ or smaller, it takes at least a time interval of the order of $\tau_1 = \tau_0/\epsilon$ to bring about a variation in J to zero order (in $\epsilon$). The angle-dependent part $\widetilde{H}$ of the Hamiltonian consists of terms whose angle dependence are like $\sin(\mathbf{n}_i^k \cdot \theta + \kappa_i^k)$. As $\theta$ varies on the time scale $\tau_0$, so will in general the terms of $\widetilde{H}$. The only exception occurs when $\mathbf{n}_i^k \cdot \omega$ becomes small. The rapidly varying terms of $\widetilde{H}$ cannot bring about any cumulative (zero order) change in J. This can only be achieved by slowly varying terms. Let the lowest order term of the latter category (of the order $\epsilon^k$, say) have an angular dependence $\sin(\mathbf{n}^{k+2} \cdot \theta + \kappa_i^{k+2})$. If the term had been a constant of the order $\epsilon^k$, it would take a time of the order of $\tau_k = \tau_0/\epsilon^k$ to change J to zero order in $\epsilon$. To obtain the same result with an angular dependent term, we must require that the variation of the angle $\mathbf{n}^{k+2} \cdot \theta$ does not take place on a shorter time scale than $\tau_k$, i.e.,

$$\frac{d}{dt}(\mathbf{n}^{k+2} \cdot \theta) \lesssim O(\epsilon^k). \tag{3.2}$$

The relation (3.2) can be shown to imply that

$$\omega \cdot \mathbf{n}^{k+2} \lesssim \begin{cases} O(\epsilon) & \text{when } k = 1 \\ O(\epsilon^2) & \text{when } k > 1. \end{cases} \tag{3.3}$$

When there exists an n in (2.8) such that (3.2) is satisfied, it is denoted by internal resonance.

## IV. EQUATIONS OF MOTION

In studying internal resonances, it seems quite natural to apply some variant of the method of averaging of Bogoliubov[2] (extended by Besjes, 1969 [5]). As shown by Burshtein and Solovev, 1962,[6] it is possible to find an "averaged" Hamiltonian, such that the corresponding canonical equations are equivalent with the averaged original equations. In both cases one introduces the expansion

$$\theta = \theta_0 + \epsilon\theta_1 + \epsilon^2\theta_2 + \dots,$$

$$\mathbf{J} = \mathbf{J}_0 + \epsilon\mathbf{J}_1 + \epsilon^2\mathbf{J}_2 + \dots . \tag{4.1}$$

Here $\mathbf{J}_0$ and $\theta_0 - \omega t$, do not vary on the time scale $\tau_0$, while $\theta_i$, $\mathbf{J}_i$ ($i > 0$) do. The equations of motion of the slowly varying quantities are

$$\dot{\theta}_0 = \frac{\partial \mathcal{H}}{\partial \mathbf{J}_0},$$

$$\dot{\mathbf{J}}_0 = -\frac{\partial \mathcal{H}}{\partial \theta_0}. \tag{4.2}$$

The "averaged" Hamiltonian to order $\epsilon^2$ is given by the expression (c.f. Ref. 6)

$$\mathcal{H} = \omega \cdot \mathbf{J}_0 + \epsilon H_3(\theta_0, \mathbf{J}_0)_{\text{res}} + \epsilon^2[K_4(\theta_0, \mathbf{J}_0)_{\text{res}} + K(\mathbf{J}_0)] + O(\epsilon^3). \tag{4.3}$$

Here $H_{3\text{res}}$ denotes the resonant terms of $H_3$, i.e., the terms where

$$\mathbf{n}_i^3 \cdot \omega \lesssim O(\epsilon). \tag{4.4}$$

The quantity $K_4(\theta_0, \mathbf{J}_0)$ is defined

$$K_4(\theta_0, \mathbf{J}_0) = H_4(\theta_0, \mathbf{J}_0) + \frac{1}{2}\left[\mathbf{J}_1(\theta_0, \mathbf{J}_0) \cdot \frac{\partial H_3}{\partial \mathbf{J}_0} + \theta_1(\theta_0, \mathbf{J}_0) \cdot \frac{\partial H_3}{\partial \theta_0}\right], \tag{4.5}$$

and $K_{4\text{res}}$ represents the resonant terms from $K_4$. These consist of resonant terms from $H_4$, i.e., terms where

$$\mathbf{n}_i^4 \cdot \omega \lesssim O(\epsilon^2), \tag{4.6}$$

and resonant terms originating from the last two terms of (4.5), where the resonances

$$(\mathbf{n}_i^3 \pm \mathbf{n}_j^3)\cdot\omega \lesssim (\epsilon^2), \quad (i \neq j), \tag{4.7}$$

may occur.

The quantity $K(\mathbf{J}_0)$ in (3.3) is given by

$$K(\mathbf{J}_0) = \overline{K_4(\theta_0, \mathbf{J}_0)}, \tag{4.8}$$

where again $\bar{}$ denotes averaging over all components of $\theta_0$. According to (4.5) and (4.8) both $K_4$ and $K$ are quadratic functions of $\mathbf{J}_0$. To lowest significant order the Eqs. (4.2) become

$$\mathring{\theta}_0 = \omega + \epsilon \sum_{i\,\mathrm{res}} \frac{\partial A_i^3}{\partial \mathbf{J}_0} \sin(\mathbf{n}_i^3 \cdot \theta_0 + \kappa_i^3),$$

$$\mathring{\mathbf{J}}_0 = -\epsilon \sum_{i\,\mathrm{res}} \mathbf{n}_i^3 A_i^3 \cos(\mathbf{n}_i^3 \cdot \theta_0 + \kappa_i^3). \tag{4.9}$$

The corresponding equations for $\theta_1$ and $\mathbf{J}_1$ become

$$\epsilon\mathring{\theta}_1 = \frac{\partial}{\partial \mathbf{J}_0}(H - \mathcal{H}) = \epsilon \sum_{i\,\not{\mathrm{res}}} \frac{\partial A_i^3}{\partial \mathbf{J}_0} \sin(\mathbf{n}_i^3 \cdot \theta_0 + \kappa_i^3),$$

$$\epsilon\mathring{\mathbf{J}}_1 = -\frac{\partial}{\partial \theta_0}(H - \mathcal{H}) = -\epsilon \sum_{i\,\not{\mathrm{res}}} \mathbf{n}_i^3 A_i^3 \cos(\mathbf{n}_i^3 \cdot \theta_0 + \kappa_i^3). \tag{4.10}$$

In (4.10) the summation is carried out over nonresonant terms. The equations can be integrated directly on the short time scale to give

$$\theta_1 = \sum_{i\,\not{\mathrm{res}}} \frac{-1}{\omega \cdot \mathbf{n}_i^3} \frac{\partial A_i^3}{\partial \mathbf{J}_0} \cos(\mathbf{n}_i^3 \cdot \theta_0 + \kappa_i^3) + O(\epsilon),$$

$$\mathbf{J}_1 = -\sum_{i\,\not{\mathrm{res}}} \frac{\mathbf{n}_i^3}{\omega \cdot \mathbf{n}_i^3} A_i^3 \sin(\mathbf{n}_i^3 \cdot \theta_0 + \kappa_i^3) + O(\epsilon). \tag{4.11}$$

From (4.11) it is apparent that $\theta_1$ and $\mathbf{J}_1$ are rapidly oscillating quantities. When (4.11) is inserted into the last two terms of (4.5), one obtains $K(\mathbf{J}_0)$ as

$$K(\mathbf{J}_0) = \overline{H}_4 - \frac{1}{4} \sum_{i\,\not{\mathrm{res}}} \frac{\mathbf{n}_i^3}{\mathbf{n}_i^3 \cdot \omega} \cdot \frac{\partial}{\partial \mathbf{J}_0} (A_i^3)^2. \tag{4.12}$$

The averaging process leading from (2.10) to 4.2) is a time average over some interval of time $\Delta t$ such that $\tau_0 \ll \Delta t \ll \tau_1$. This should not be confused with the process denoted by the symbol $\bar{}$, of averaging over all components of $\theta$.

## V. RESONANCES AND INVARIANTS

### A. First order resonance

A resonance of the order $\epsilon$ occurs when there exists one or more $\mathbf{n}^3$ such that (4.4) is satisfied, i.e., $H_{3\mathrm{res}}$ in (4.3) is nonzero. As a result of such a resonance, one generally has a transfer of energy to zero order between the modes.

It is in general not possible to solve the equations of motion (4.2) exactly, except for the special case of only one resonance, where the equations can be solved to the order $\epsilon$ (see Appendix).

As the "averaged" Hamiltonian does not depend on time explicitly, $\mathcal{H}$ is an invariant. In the case of only "exact" resonances, i.e., $\omega \cdot \mathbf{n}^3 = 0$, there are two additional invariants, namely $\omega \cdot \mathbf{J}_0$ and $H_{3\mathrm{res}}$. This is readily shown by taking the scalar product of the last of the equations (4.9) with $\omega$, and recalling that $\mathcal{H}$ is an invariant.

Even in the case where the resonance is not exact, the quantity $\omega \cdot \mathbf{J}_0$, which is the only zero order contribution to $\mathcal{H}$, can only have a variation of the order $\epsilon$ while $\mathbf{J}_0$ changes to the zero order.

### B. Second order resonance. Frequency shift

If there does not exist any first order resonances (i.e., $\omega \cdot \mathbf{n}_i^3 \neq 0$ for all $\mathbf{n}_i^3$), $\theta_0 - \omega t$ and $\mathbf{J}_0$, according to (4.2) and (4.3), do not have a variation on the time scale $\tau_1$. $\theta_1$ and $\mathbf{J}_1$ as given by (4.11) are still the relevant solution to the order $\epsilon$. As the sum in (4.11) now extends over all $\mathbf{n}_i^3$, the last equation gives

$$H_3(\theta, \mathbf{J}) = -\omega \cdot \mathbf{J}_1(\theta, \mathbf{J}). \tag{5.1}$$

To obtain the equations for the $\theta_0$ and $\mathbf{J}_0$, one can use (4.2) and (4.3). For later use, however, we want to point out another possible approach. This method consists in finding a transformation near to the identity, that will enable us to get rid of $H_3$ altogether. If $\hat{\theta}$ and $\hat{\mathbf{J}}$ are the new canonical variables, one can choose a transformation characterized by the generating function

$$S(\theta, \hat{\mathbf{J}}) = \theta \cdot \hat{\mathbf{J}} + \epsilon \sum_i \frac{1}{\omega \cdot n_i^3} A_i^3(\hat{\mathbf{J}}) \cos(\mathbf{n}_i^3 \cdot \theta + \kappa_i^3). \tag{5.2}$$

$S$ generates the relation

$$\hat{\theta} = \frac{\partial S}{\partial \hat{\mathbf{J}}} = \theta - \epsilon\theta_1(\theta, \hat{\mathbf{J}}),$$

$$\mathbf{J} = \frac{\partial S}{\partial \theta} = \hat{\mathbf{J}} + \epsilon \mathbf{J}_1(\theta, \hat{\mathbf{J}}). \tag{5.3}$$

As $S$ does not depend explicitly on time, the new Hamiltonian $H$ is given by

$$H = H[\theta(\hat{\theta}, \hat{\mathbf{J}}), \mathbf{J}(\hat{\theta}, \hat{\mathbf{J}})]$$

$$= \omega \cdot \mathbf{J} + \epsilon H_3(\theta, \mathbf{J}) + \epsilon^2 H_4(\theta, \mathbf{J}) + \dots . \tag{5.4}$$

To the second order in $\epsilon$ one obtains from (5.1), (5.3) and (5.4)

$$H = \omega \cdot \hat{\mathbf{J}} + \epsilon^2 \left[ H_4(\hat{\theta}, \hat{\mathbf{J}}) + \mathbf{J}_1(\hat{\theta}, \hat{\mathbf{J}}) \cdot \frac{\partial H_3}{\partial \hat{\mathbf{J}}} \right]. \tag{5.5}$$

Again one introduces an expansion of the type (4.1)

$$\hat{\theta} = \hat{\theta}_0 + \epsilon^2 \hat{\theta}_2 + \dots,$$

$$\hat{\mathbf{J}} = \hat{\mathbf{J}}_0 + \epsilon^2 \hat{\mathbf{J}}_2 + \dots, \tag{5.6}$$

where the first order terms are absent for obvious reasons, and $\hat{\theta}_0 - \omega t$ and $\hat{\mathbf{J}}_0$ do not vary on the time scale $\tau_0$ and $\tau_1$.

The equations governing $\hat{\theta}_0$ and $\hat{\mathbf{J}}_0$ to the second order in $\epsilon$ are found to be

$$\dot{\hat{\theta}}_0 = \frac{\partial \mathcal{H}}{\partial \hat{\mathbf{J}}_0} = \omega + \Delta\omega + \epsilon^2 \left( \frac{\partial K_4}{\partial \hat{\mathbf{J}}_0} \right)_{\mathrm{res}},$$

$$\dot{\hat{\mathbf{J}}}_0 = -\frac{\partial \mathcal{H}}{\partial \hat{\theta}_0} = \epsilon^2 \left( \frac{\partial K_4}{\partial \hat{\theta}_0} \right)_{\mathrm{res}}, \tag{5.7}$$

where

$$\Delta\omega = \epsilon^2 \frac{\partial K}{\partial \hat{\mathbf{J}}_0}.$$

Here $K_4(\hat{\theta}_0, \hat{\mathbf{J}}_0)$ is given by (4.5) and $K(\hat{\mathbf{J}}_0)$ by (4.12).

Like the case of first order resonance, the equations (5.7) cannot in general be solved exactly. An exception is again the case where there is only one resonance present (see Appendix).

Again $\mathcal{H}$ is an invariant, and the zero order contribution to $\mathcal{H}$, $\omega \cdot \mathbf{J}_0$, can only have a variation of the order $\epsilon^2$.

When no second order resonances are present, (5.7) reduces to the form

$$\mathring{\hat{\theta}}_0 = \boldsymbol{\omega} + \Delta\boldsymbol{\omega}, \quad \mathring{\hat{\mathbf{J}}}_0 = 0, \tag{5.8}$$

which shows $\hat{\mathbf{J}}_0$ to be constant on the time scale $\tau_2$. Another important fact to notice is the *frequency shift* $\Delta\omega$, which is a linear function of $\hat{\mathbf{J}}_0$, of second order in $\epsilon$. The finite amplitude vibrations change slightly (to the order $\epsilon^2$) the average properties (equilibrium) of the system, which again influences the eigenfrequencies.

### C. Periodic solutions

When first or second order resonances occur, there is generally a continuous flow of energy between the interacting modes. When only one such resonance occurs, the Hamiltonian generating the equations of motion for $\theta_0$ and $\mathbf{J}_0$ can be written

$$\mathcal{H} = \omega \cdot \mathbf{J}_0 + \epsilon^2 K + \epsilon A(\mathbf{J}_0)\sin(\mathbf{n}\cdot\theta_0 + \kappa), \tag{5.9}$$

where we have assumed that there are no resonances of the second order. For completeness we have kept the $\epsilon^2$ term giving rise to a frequency shift.

The equations of motion are

$$\mathring{\theta}_0 = \omega + \epsilon^2 \frac{\partial K}{\partial \mathbf{J}_0} + \epsilon \frac{\partial A}{\partial \mathbf{J}_0}\sin(\mathbf{n}\cdot\theta_0 + \kappa),$$

$$\mathring{\mathbf{J}}_0 = -\epsilon A \mathbf{n}\cos(\mathbf{n}\cdot\theta_0 + \kappa). \tag{5.10}$$

If a solution can be found where neither $\mathbf{n}\cdot\theta_0$ nor $\mathbf{J}_0$ varies, then no energy is transferred, and the resonant modes have periods that are exactly commensurable. Such a solution therefore represents a periodic motion and occurs when $\mathbf{n}\cdot\mathring{\theta}_0 = \mathring{\mathbf{J}}_0 = 0$, i.e., when

$$\mathbf{n}\cdot\theta_0 + \kappa = (\pi/2) + N\pi, \quad (N = 0, 1),$$

and

$$\mathbf{n}\cdot\omega + \mathbf{n}\cdot\left[\epsilon^2 \frac{\partial K}{\partial \mathbf{J}_0} + \epsilon \frac{\partial A}{\partial \mathbf{J}_0}(-1)^N\right] = 0. \tag{5.11}$$

Taking into account the structure of

$$A \ (A \propto \prod_{i=1}^{n} J_i^{|n_i|/2}),$$

one obtains instead of the last equation

$$\mathbf{n}\cdot(\omega + \Delta\omega) + \epsilon \frac{A}{2}\sum_{i=1}^{n} \frac{|n_i| n_i}{J_{0i}} = 0, \tag{5.12}$$

which gives for exact resonance ($\mathbf{n}\cdot\omega = 0$)

$$\sum_{i=1}^{n} \frac{n_i |n_i|}{J_{0i}} = O(\epsilon). \tag{5.13}$$

As to the stability of the periodic solutions, an analysis shows them to be stable provided

$$A^2 \sum_{i=1}^{n} \frac{|n_i|^3}{J_{0i}^2} - (-1)^N \frac{(\mathbf{n}\cdot\omega)^2}{\epsilon^2} > 0. \tag{5.14}$$

Thus for exact resonance, (5.14) is always satisfied.

In Sec. IX we have tested a particular system by numerical solution of the basic equations. In particular we wanted to find out by what degree of accuracy the averaged equations would describe a "sharp" phenomenon like the periodic solutions. The results indicated a surprising accuracy even for such high values of $\epsilon$ as 0.1.

## VI. HIGHER ORDER RESONANCES

In the following one considers the case where first and second order resonances are absent. The transformation procedure outlined in the previous section can be repeated to get rid of the $\epsilon^2$ terms of the angle-dependent part of the Hamiltonian. The relevant canonical transformation is now

$$\hat{\hat{\theta}} = \hat{\theta} - \epsilon^2 \hat{\theta}_2(\hat{\theta}, \hat{\hat{\mathbf{J}}}),$$

$$\hat{\mathbf{J}} = \hat{\hat{\mathbf{J}}} + \epsilon^2 \hat{J}_2(\hat{\theta}, \hat{\hat{\mathbf{J}}}). \tag{6.1}$$

Here $\hat{\theta}_2$ and $\hat{\mathbf{J}}_2$ are found by direct integration (on the timescale $\tau_0$) of the canonical equations for these variables.

The transformation process can be repeated to get rid of $\epsilon^3$, $\epsilon^4$, ... etc. contributions to $H - \bar{H}$ (i.e., the angle dependent part of the Hamiltonian), as long as no resonances occur to these orders (c.f. Birkhoff, 1966 [7]).

Let the first resonance occur to the order $\epsilon^k$, and assume for simplicity that there is only one resonance to this order, $\mathbf{N}\cdot\omega$ say.

The Hamiltonian becomes

$$H = \bar{H} + \epsilon^k F(\theta', \mathbf{J}'), \tag{6.2}$$

where

$$\bar{H} = \omega\cdot\mathbf{J}' + \epsilon^2 K(\mathbf{J}') + O(\epsilon^4),$$

and the variables $\theta'$, $\mathbf{J}'$ are the end products of the chain of transformations near to the identity, mentioned above.

One introduces the expansions

$$\theta_0' = \theta_0 + \epsilon^k \theta_K + \ldots,$$

$$\mathbf{J}' = \mathbf{J}_0 + \epsilon^k J_K + \ldots, \tag{6.3}$$

where again $\theta_0$, $\mathbf{J}_0$ have no variation on the time scales $\tau_0$, $\tau_1$, ..., and $\tau_{k-1}$.

The equations governing the variation of $\theta_0$ and $\mathbf{J}_0$ become

$$\mathring{\theta}_0 = \frac{\partial \bar{H}}{\partial \mathbf{J}_0} + \epsilon^k \left(\frac{\partial F}{\partial J_0}\right)_{\text{res}},$$

$$\mathring{\mathbf{J}}_0 = -\epsilon^k \left(\frac{\partial F}{\partial \theta_0}\right)_{\text{res}}, \tag{6.4}$$

which are on a canonical form like (4.2) and (5.7) with $\mathcal{H}$ given by

$$\mathcal{H} = \bar{H} + \epsilon^k F_{\text{res}}. \tag{6.5}$$

Let $F_{\text{res}}$ be given by

$$F_{res} = A \sin(\mathrm{N} \cdot \theta_0 + \kappa). \tag{6.6}$$

From (6.4) and (6.6) it is seen that $\dot{\mathrm{J}}_0 \| N$. Introducing the new variables $J$ and $\psi$ by

$$\psi = \mathrm{N} \cdot \theta_0 \text{ and } \mathrm{J}_0 = \mathrm{h} + \mathrm{N}J,$$

where h is a constant, one obtains the equations

$$\mathring{\psi} = \frac{\partial \bar{H}}{\partial J} + \epsilon^k \frac{\partial A}{\partial J} \sin(\psi + \kappa),$$

$$\mathring{J} = -\epsilon^k A \cos(\psi + \kappa). \tag{6.7}$$

A necessary condition for a resonance to bring about a zero order change in $\mathbf{J}_0$, is that the variation of $\psi$ does not take place on a shorter time scale than $\tau_k$. Referring to Eq. (6.7), this condition can be stated explicitly as

$$\frac{d\bar{H}}{dJ} \lesssim O(\epsilon^k). \tag{6.8}$$

Taking into account that an expansion of $\bar{H}$ in terms of $J$ correct to the order $\epsilon^k$, only contains terms of power $m \leq k$ in $J$, the above condition can be replaced by the following set of conditions

$$\left.\frac{d^m \bar{H}}{dJ^m}\right|_{J=0} \lesssim O(\epsilon^k) \text{ for } \begin{cases} \text{all } m \leq k \ (k \text{ even}) \\ \text{or } m \leq k-1 \ (k \text{ odd}). \end{cases} \tag{6.9}$$

As $k > 2$ for a higher order resonance, a necessary condition for these resonances to transfer energy to zero order is

$$\left.\frac{d\bar{H}}{dJ}\right|_{J=0} = [\omega + \Delta\omega(0)] \cdot \mathrm{N} \lesssim O(\epsilon^{\min(k,4)}), \tag{6.10}$$

and

$$\left.\frac{d^2\bar{H}}{dJ^2}\right|_{J=0} = \epsilon^2 \mathrm{N} \cdot \frac{\partial^2 K}{\partial \mathrm{J}_0 \partial \mathrm{J}_0} \cdot \mathrm{N} \lesssim O(\epsilon^{\min(k,4)}). \tag{6.11}$$

For $k=3$ the conditions (6.10)—(6.11) are equivalent to (6.8). For $k > 3$ (6.8) is more restrictive than (6.10) and (6.11).

It can also be shown directly from the invariance of $H$, that (6.10) and (6.11) are necessary conditions for a zero order variation of $\mathrm{J}_0$. If $\mathrm{J}_0$ is given by h initially $[\mathrm{J}_0(0) = \mathrm{h}]$, $J$ is a direct measure of the variation in $\mathrm{J}_0$. From the constancy of $H$ one obtains

$$\left[\omega + \epsilon^2 \left.\frac{\partial K}{\partial \mathrm{J}_0}\right|_{J=0}\right] \cdot \mathrm{N}J + \frac{\epsilon^2}{2} \mathrm{N} \cdot \frac{\partial^2 K}{\partial \mathrm{J}_0 \partial \mathrm{J}_0} \cdot \mathrm{N}J^2 + O(\epsilon^4)$$
$$+ \epsilon^k F_{res} = \text{const}, \tag{6.12}$$

where

$$\epsilon^2 \left.\frac{\partial K}{\partial \mathrm{J}_0}\right|_{J=0} = \Delta\omega(0).$$

If the relations (6.10)—(6.11) are satisfied, the first two terms in (6.12) become of the order $\epsilon^{\min(k,4)}$. For $k=3$ this may be sufficient for $J$ (and thereby the variation of $\mathrm{J}_0$) to be of the order zero, as shown by simulation of a particular example in Sec. VIII. For $k > 3$ the term of order $\epsilon^4$ in (6.12) must be taken into account, and the condition (6.9) becomes much more restrictive.

In any case, if only the relation (6.10) is satisfied (i.e., the resonance condition is satisfied initially), (6.12) tells us that the variation of $\mathrm{J}_0$ will be bounded like

$$|J| \lesssim O(\epsilon^{\min(k/2,2)-1}). \tag{6.13}$$

The relation above tells us for example that for $k=3$, $J_0$ can only have a variation less than or of the order $\sqrt{\epsilon}$.

## VII. RELATION TO STABILITY

If the equilibrium $\mathbf{q} = \mathbf{p} = 0$ (Sec. II) is a true equilibrium, the eigenfrequencies $\omega_1, \omega_2 \ldots, \omega_n$ are positive quantities. This is so, because the total energy associated with an arbitrary small perturbation must be positive.

If the "equilibrium" is not a true one (e.g., the restricted three body problems and the gravity gradient satellite problem—see Alfriend *et al.*, 1972 [8]) one may still end up with a meaningful expansion like (2.6). In this case, however, there is no reason to expect that the Hamiltonian of the perturbed motion (2.6) is positive definite. That is, some of the quantities $\omega_i$ may be negative.

In problems of nonlinear interaction of waves one may have a similar situation with "negative energy waves" (see, e.g., Dysthe, 1970 [9]).

In the case where all $\omega_i$ are positive, equilibrium is always stable in the sense that small perturbations from the equilibrium remain small.

If, however, some of the $\omega_i$ are negative (and some positive), a "kinematical" possibility for an instability exists. This is so because it is now possible to increase the energy (or rather the absolute value of the energy $|\omega_i J_i|$ of the different modes without violating the conservation of total energy.

As can be seen from the previous sections, the only dynamical effect that can bring about an appreciable change in J are the internal resonances. For a given resonance $\mathbf{n} \cdot \omega \approx 0$ one has $\dot{\mathbf{J}} \| \mathbf{n}$. Since the components $J_i > 0$, an "unlimited" growth can only occur if all components of n are positive. The resonance condition can still be satisfied, as some of the components of $\omega$ are negative.

As shown in the previous sections a first or second order resonance $\mathrm{n} \cdot \omega \approx 0$, may transfer energy between the different modes, if the components of n are all positive, this means that positive energy will be transferred from the negative energy modes to the positive energy modes. Thus the absolute value of the energy increases for the modes involved in the resonance, and we have an unstable situation.

For the higher order resonances, however, the condition (6.8) must be satisfied in order to have transfer of energy. A sufficient condition for stability can be indicated as follows.

(a) No first and second order resonances should exist.

(b) While higher order resonances may exist, the relation (6.11) should not be satisfied.

Arnold, 1963 [4] gives somewhat different conditions.

Instead of (a) he introduces the condition

(A) $\omega \cdot \mathbf{n} \neq 0$ for all n such that $|\mathbf{n}| \leq 4$,

which is more restrictive since a violation of (A) does not imply that a resonance actually occurs in the system.

Instead of (b) he gives the conditions

(B)

(i) $\mathrm{Det}\left|\frac{\partial^2 K}{\partial \mathbf{J}_0 \partial \mathbf{J}_0}\right| \neq 0,$

or

(ii) $\mathrm{Det}\begin{vmatrix} \frac{\partial^2 K}{\partial \mathbf{J}_0 \partial \mathbf{J}_0} & \omega \\ \omega & 0 \end{vmatrix} \neq 0.$

For a system of two degrees of freedom ($n=2$), it is easily shown that (b) and [B (ii)] are equivalent. For $n > 2$, however, (b) is more restrictive than both [B (ii)] and [B (i)].

To illustrate this point we have constructed an example with $n=3$, where [B (i) and (ii)] are satisfied and (b) is not.

Such a system should according to Arnold be stable for a majority of initial conditions. We find (c.f. Sec. VIII) that when the initial data is chosen such that (6.10) is satisfied, the system is unstable. To satisfy (6.10) J(0) must be chosen from a section of J space of measure (volume) $\epsilon$, so Arnolds conclusions are not contradicted.

Our results show, however, that in order to obtain general criteria of stability (not excluding a minority of initial conditions) one should apply a condition of the type (b) rather than [B (i) and (ii)]. It seems that a sufficient condition would be (A) together with the requirement that $K(\mathbf{J})$ be definite.

## VIII. AN EXAMPLE

In studying the effect of higher order internal resonances, we have considered a system with the Hamiltonian

$$H = \sum_{i=1}^{3} \frac{\omega_i}{2}(p_i^2 + q_i^2) + \epsilon H_3 + \epsilon^2 H_4 + \epsilon^3 H_5 + 0(\epsilon^4), \qquad (8.1)$$

where

$$H_3 = 0,$$
$$H_4 = Aq_1^4 + Bq_2^4 + Cq_3^4,$$
$$H_5 = q_1^2 q_2^2 q_3. \qquad (8.2)$$

For one internal resonance to order $\epsilon^3$ we obtain the averaged Hamiltonian

$$\mathcal{H} = \omega \cdot \mathbf{J} + \tfrac{3}{2}\epsilon^2(AJ_1^2 + BJ_2^2 + CJ_3^2) + \epsilon^3 \frac{\sqrt{2}}{4} J_1 J_2 (J_3)^{1/2} \sin \mathbf{n} \cdot \theta + O(\epsilon^4), \qquad (8.3)$$

that is

$$\mathcal{H} = \omega \cdot \mathbf{J} + \epsilon^2 \mathbf{J} \cdot \mathbf{K} \cdot \mathbf{J} + \epsilon^3 D(\mathbf{J}) \sin \mathbf{n} \cdot \theta + O(\epsilon^4), \qquad (8.4)$$

where

$$\mathbf{K} = \frac{3}{2}\begin{Bmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{Bmatrix}. \qquad (8.5)$$

From (6.10)–(6.11) a necessary condition for a transfer of energy to zero order is

$$(\omega + 2\epsilon^2 \mathbf{h} \cdot \mathbf{K}) \cdot \mathbf{n} \lesssim O(\epsilon^3),$$

and

$$\mathbf{n} \cdot \mathbf{K} \cdot \mathbf{n} \lesssim O(\epsilon). \qquad (8.6)$$

The exact equations of motion are given by

$$\ddot{q}_1 + \omega_1^2 q_1 + \epsilon^2 \omega_1 (4Aq_1^3 + \epsilon 2 q_1 q_2^2 q_3) = 0,$$
$$\ddot{q}_2 + \omega_2^2 q_2 + \epsilon^2 \omega_2 (4Bq_2^3 + \epsilon 2 q_1^2 q_2 q_3) = 0, \qquad (8.7)$$
$$\ddot{q}_3 + \omega_3^2 q_3 + \epsilon^2 \omega_3 (4Cq_3^3 + \epsilon q_1^2 q_2^2) = 0.$$

### A. Simulation

We have simulated (8.7) for two internal resonances to order $\epsilon^3$, using an integration routine described by Bulirsh *et al.*, 1967.[10]

(A) When $\mathbf{n} = (2, 2, 1)$,

with constants

$\omega_1 = -1,91, \;\; \omega_2 = 0,82, \;\; A = \frac{2}{3}, \;\; B = -\frac{8}{15}, \;\; \epsilon = 0,1,$

$\omega_3$ and $C$ satisfying (8.6),

and initial values

$q_1 = \frac{1}{2}, \;\; \mathring{q}_1 = \frac{\sqrt{3}}{2}\omega_1, \;\; q_2 = \frac{1}{2}, \;\; \mathring{q}_2 = \frac{\sqrt{3}}{2}\omega_2, \;\; q_3 = \frac{\sqrt{3}}{2}, \;\; \mathring{q}_3 = \frac{1}{2}\omega_3.$

This gives

$$\theta(0) = 2\theta_1 + 2\theta_2 + \theta_3 = 2 \cdot \frac{\pi}{6} + 2 \cdot \frac{\pi}{6} + \frac{\pi}{3} = \pi, \;\; J_i(0) = \tfrac{1}{2}\omega_i.$$

Arnold's conditions for stability, VII (B), are satisfied

$$\left|\frac{\partial^2 K}{\partial \mathbf{J}_0 \partial \mathbf{J}_0}\right| = |\mathbf{K}| \approx 0.28 \neq 0,$$

$$\begin{vmatrix} \frac{\partial^2 K}{\partial \mathbf{J}_0 \partial \mathbf{J}_0} & \omega \\ \omega & 0 \end{vmatrix} = \begin{vmatrix} \mathbf{K} & \omega \\ \omega & 0 \end{vmatrix} \approx 2.02 \neq 0.$$

(B) When $\mathbf{n} = (-2, 2, 1)$,

with constants

$\omega_1 = 1,91, \;\; \omega_2 = 0,82, \;\; A = \frac{2}{3}, \;\; B = -\frac{3}{5}, \;\; \epsilon = \frac{1}{9},$

$\omega_3$ and $C$ satisfying (8.6),

and initial values

$q_1 = 1, \;\; \mathring{q}_1 = 0, \;\; q_2 = 1, \;\; \mathring{q}_2 = 0, \;\; q_3 = 0, \;\; \mathring{q}_3 = \omega_3.$

This gives

$$\theta(0) = 2\theta_1 + 2\theta_2 + \theta_3 = 2 \cdot (\pi/2) + 2 \cdot (\pi/2) + 0 = 2\pi,$$
$$J_i(0) = \tfrac{1}{2}\omega_i.$$

Arnold's conditions for stability are satisfied

$$|\mathbf{K}| \approx 0.16 \neq 0,$$

$$\begin{vmatrix} \mathbf{K} & \omega \\ \omega & 0 \end{vmatrix} \approx 3.4 \neq 0.$$

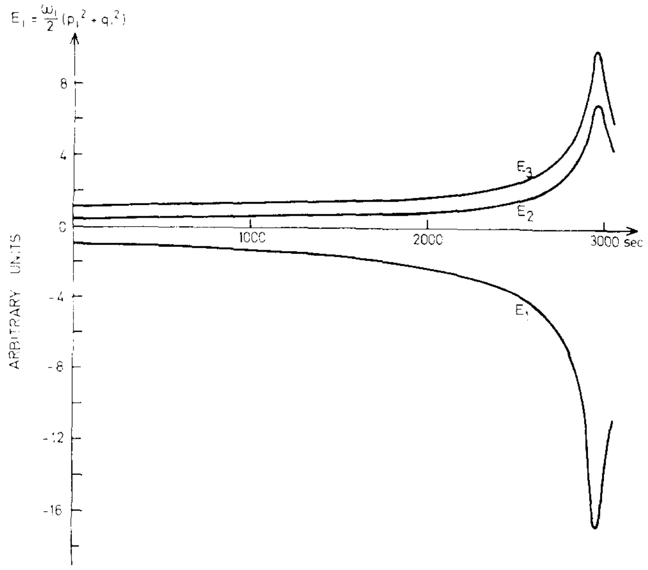



FIG. 1. Evolution of average energy of oscillators, positive and negative energy oscillators.

### B. Results

For the system (A) an "explosive" growth is obtained on the time scale $\tau_3$, the results are given in Fig. 1. After a certain time, however, the growth is saturated, and energy is transferred the oppoiste way due to the presence of higher order terms.

For the system (B) a transfer of energy is obtained on the time scale $\tau_3$. Energy from mode 2 and 3 is transferred to mode 1. The results are given in Fig. 2.

## IX. A TEST OF THE METHOD OF AVERAGING—ELASTIC PENDULUM

We consider the elastic pendulum (a two degree of freedom system) c.f. Van der Burgh, 1968.[11]

The system consists of a linear spring with zero mass, length $l_0$, and string constant $k$. With a load of mass $m$, the length in the equilibrium position is $l$, c.f. Fig. 3.

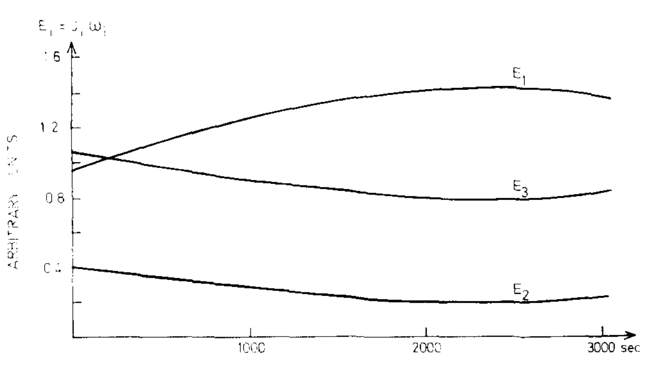



FIG. 2. Evolution of average energy of positive energy oscillators.

The Hamiltonian is given by

$$H=\tfrac12 m\left(\mathring{q}_1^2+\frac{k}{m}q_1^2\right)+\frac{m}{2}(l+q_1)^2\mathring{q}_2^2+mg(l+q_1)(1-\cos q_2), \tag{9.1}$$

and the exact equations of motion are

$$\ddot q_1+\omega_1^2 q_1-\cos q_2-(l+q_1)\mathring{q}_2^2+g=0,$$
$$\ddot q_2+\omega_2^2\frac{\sin q_2}{1+q_1/l}+\frac{2}{l}\frac{\mathring{q}_1\mathring{q}_2}{1+q_1/l}=0, \tag{9.2}$$

where

$$q_1=r,\quad q=\theta,\quad \omega_1^2=k/m,\quad \omega_2^2=g/l. \tag{9.3}$$

Choosing $k/m=4(g/l)$, there is a resonance to order $\epsilon$, i.e.,

$$\omega_1=2\omega_2. \tag{9.4}$$

Defining $\theta=2\theta_2-\theta_1$, the averaged Hamiltonian is found to be

$$\mathcal{H}=\omega\cdot\mathbf{J}-\left[\frac34\frac1l\left(\frac{\omega_2}{m}\right)^{1/2}\right](J_1)^{1/2}J_2\sin\theta+\frac{1}{ml^2}\left(\tfrac{39}{64}J_1J_2-\tfrac{33}{256}J_2^2\right), \tag{9.5}$$

the equations of motion to the order $\epsilon^2$ are given by

$$\mathring{\theta}_1=\omega_1+\frac{A}{2}\frac{J_2}{(J_1)^{1/2}}\sin\theta+\frac{39}{64}\frac{1}{ml^2}J_2,$$
$$\mathring{\theta}_2=\omega_2+A(J_1)^{1/2}\sin\theta+\frac{1}{ml^2}\left(\tfrac{39}{64}J_1-\tfrac{33}{128}J_2\right), \tag{9.6}$$
$$\mathring{J}_1=A(J_1)^{1/2}J_2\cos\theta,$$
$$\mathring{J}_2=-2A(J_1)^{1/2}J_2\cos\theta,$$

where

$$A=\frac{3}{4l}\left(\frac{\omega_2}{m}\right)^{1/2}.$$

Periodic solutions occur when $\mathring{\theta}=J_1=J_2=0$. From (9.6) one obtains the following conditions (expressed in the variables $q$) correct to order $\epsilon^2$

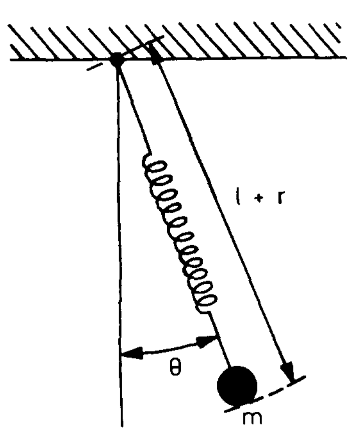



FIG. 3. The elastic pendulum.

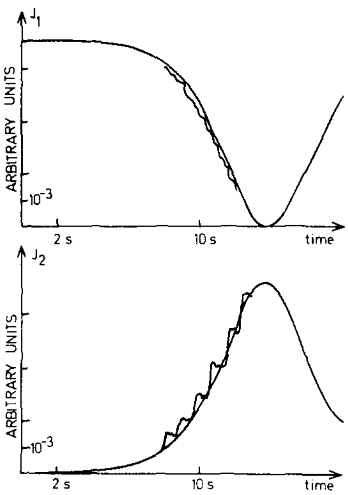



FIG. 4. Action variables for the elastic pendulum, considerable energy transfer. Exact and averaged equations.

$$q_{20} = \pm \frac{2\sqrt{2}}{1} q_{10} \left(1 - \frac{35}{32} \frac{q_{10}}{l}\right). \tag{9.7}$$

As shown in Sec. V these oscillations should be stable.

## Simulation

We have simulated the equations of motion (9.2), (9.6), and the resulting action variables are given by Fig. 4 and Fig. 5 for the cases of considerable energy-transfer and for periodic oscillations of the order $\epsilon$ (in this example $\epsilon$ is of the order 0.1). The rapidly oscillating parts result from simulation of (9.2). This shows that simulation of the averaged equations (9.6) are highly effective and laboursaving.

We further searched for the periodic solutions and chose $q_{10} = 0.06m$, $\omega_2 = 4.0$, $g = 9.819^{m}/\sec^2$. From (9.7) periodic solutions are obtained for $q_{20} = \pm\ 0.2470$. The simulation indicates a periodic orbit for $q_{20} \simeq \pm\ 0.2436$ (Fig. 6). Thus the difference between our numerical estimates and those calculated from the averaged equations are of the order $\epsilon^3$.

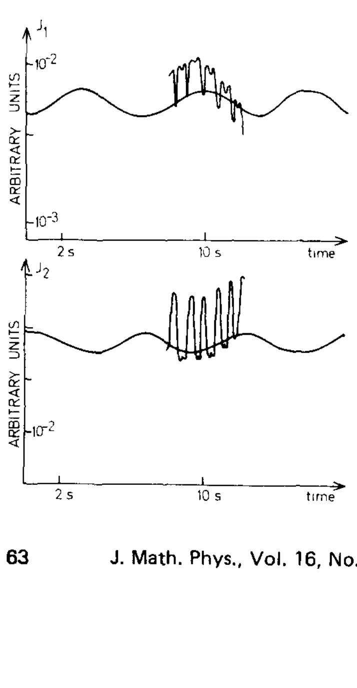



FIG. 5. Action variables for the elastic pendulum, periodic oscillations of the order $\epsilon$. Exact and averaged equations.

## APPENDIX

### 1. Exact solution to order $\epsilon$

Suppose there exists only *one* exact resonance to order $\epsilon$. To this order the equations of motion are given by

$$\overset{\circ}{\theta}_0 = \omega + \epsilon \frac{\partial A}{\partial J_0} \sin(\mathbf{n} \cdot \theta_0 = \kappa),$$

$$\overset{\circ}{J}_0 = -\epsilon \mathbf{n} A \cos(\mathbf{n} \cdot \theta_0 + \kappa). \tag{A1}$$

From Sec. V we have the following invariant to order $\epsilon$

$$H_{3\mathrm{res}} = A(\mathbf{J}_0) \sin(\mathbf{n} \cdot \theta_0 + \kappa) = E_1 = \text{const.} \tag{A2}$$

(A1) gives $\mathbf{J}_0 \| \mathbf{n}$, therefore we introduce $\mathbf{h}$ and $\alpha$

$$\mathbf{J}_0 = \mathbf{h} + \mathbf{n}\alpha \tag{A3}$$

where $\mathbf{h}$ is a constant vector giving the intial values of $\mathbf{J}_0$. Then we obtain

$$\frac{d\alpha}{dt} = \mp \epsilon (A^2 - E^2)^{1/2} = \mp \epsilon (C_1 \alpha^3 + C_2 \alpha^2 + C_3 \alpha + C_4)^{1/2}, \tag{A4}$$

where

$$C_1 = \mathbf{K} \vdots \mathbf{nnn},$$

$$C_2 = 3\mathbf{K} \vdots \mathbf{nnh},$$

$$C_3 = 3\mathbf{K} \vdots \mathbf{nhh},$$

$$C_4 = \mathbf{K} \vdots \mathbf{hhh} - E_1^2, \tag{A5}$$

and $\mathbf{K}$ is a symmetric tensor of order 3 given by

$$A^2(\mathbf{J}_0) = \mathbf{K} \vdots \mathbf{J}_0 \mathbf{J}_0 \mathbf{J}_0. \tag{A6}$$

The solution $\alpha$ of (A4) can be found by using elliptic functions.

### 2. Exact solution to order $\epsilon^2$

Suppose there is only *one* exact resonance to order $\epsilon^2$. To this order the equations of motion are given by

$$\overset{\circ}{\theta}_0 = \omega + \epsilon^2 \frac{\partial}{\partial \mathbf{J}_0} [K(\mathbf{J}_0) + S(\mathbf{J}_0) \sin(\mathbf{n} \cdot \theta_0 + \kappa)],$$

$$\overset{\circ}{\mathbf{J}}_0 = -\epsilon^2 \mathbf{n} S(\mathbf{J}_0) \cos(\mathbf{n} \cdot \theta_0 + \kappa), \tag{A7}$$

where $S(\mathbf{J}_0)$ is a homogeneous polynomial of order four

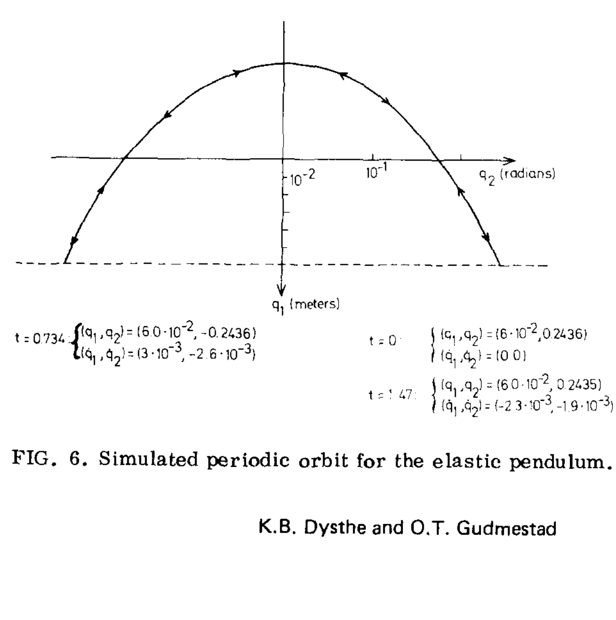



FIG. 6. Simulated periodic orbit for the elastic pendulum.

in the components $(J_1)^{1/2}$ and $K(\mathbf{J}_0)$ is given by (4.12). The following invariant to order $\epsilon^2$ is obtained:

$$K(\mathbf{J}_0)+S(\mathbf{J}_0)\sin(\mathbf{n}\cdot\boldsymbol{\theta}_0+\kappa)=E_2=\text{const.} \tag{A8}$$

By introducing h and $\alpha$ as in (A3) we obtain

$$\frac{d\alpha}{dt}=\pm\,\epsilon^2\{[S(\alpha)-K(\alpha)+E_2][S(\alpha)+K(\alpha)-E_2]\}^{1/2}, \tag{A9}$$

where $S^2(\alpha)-[E_2-K(\alpha)]^2$ is a polynomial of order four in $\alpha$. The solution $\alpha$ of (A9) can be found by using elliptic functions.

*This work has been supported by the Norwegian Research Council for Science and the Humanities.

[1]H. Poincaré, J. Math. 7 (1881).
[2]N.N. Bogoliubov and Y.A. Mitropolsky, *Asymptotic Methods in the Theory of Non-Linear Oscillations* (Hindustan Publ. Corp., Delhi, India 1961), p. 387.
[3]J. Moser, Nachr. Akad. Wiss. Gött. Math.-Phys. 2 Vol. 1 (1962).
[4]V.I. Arnold, Russ. Math. Surv. 18, 85 (1963).
[5]J.G. Besjes, J. Mec. 8, 357 (1969).
[6]E.L. Burshtein and L.S. Solovev, Sov. Phys.—Dokl. 6, 731 (1962).
[7]C.D. Birkhoff, *Dynamical Systems* (Am. Math. Soc., Providence, R.I., 1966).
[8]K.T. Alfriend and D.L. Richardson, Celestial Mech. 7, 408 (1973).
[9]K.B. Dysthe, Int. J. Electron. 29, 401 (1970).
[10]R. Bulirsh and J. Stoer, Numer. Math. 9, 271 (1967).
[11]A. Van der Burg, J. Mec. 7, 505 (1968).

# Motion of a body in general relativity*

Robert Geroch and Pong Soo Jang

*Department of Physics, University of Chicago, Chicago, Illinois*
(Received 26 August 1974)

A simple theorem, whose physical interpretation is that an isolated, gravitating body in general relativity moves approximately along a geodesic, is obtained.

## 1. INTRODUCTION

It is a consequence of Einstein's equation in general relativity that the divergence of the stress—energy tensor of matter vanishes, i.e., that, in physical terms, "locally, energy and momentum are conserved." One might expect, therefore, that it should be true in some sense that the motion of a body in the theory must be along a geodesic. One would like to prove some theorem in general relativity to this effect. The difficult part of obtaining such a theorem is apparently the formulation of its statement. A physical body is described in the theory by a four-dimensional region of space—time: What, then, is to be meant by "move along a geodesic?" Even the passage to an infinitesimal body does not immediately resolve the difficulty, for in this case, although one indeed obtains a unique world line for the body, the metric would be expected to become singular there: What, then, is to be meant by "this world line is a geodesic?"

A number of results suggestive of geodesic motion are known.

It is easily shown that, if the matter consists only of dust, then the world line of each dust particle must be a geodesic. This result suggests the following conjecture: The world tube of any body contains a timelike geodesic. Indeed, this conjecture is known[1] to be true for the case of a perfect fluid with isotropic pressure. Unfortunately, the conjecture is apparently false for more general sources.[2]

In an alternative approach to the problem of motion, due to Newman and his co-workers,[3] the motion of the body is described in terms of the asymptotic behavior of its gravitational field. The final equations governing this asymptotic field are indeed suggestive of geodesic motion. It appears, however, to be difficult to interpret these equations directly in terms of the appearance of the body to observers in its local neighborhood. Furthermore, the method is not immediately applicable to the case of one body moving under the influence of another, since the asymptotic analysis would require that both bodies in this case be regarded as a single system. It has been suggested[4] that both of these difficulties can be avoided, at least for the case of a black hole, by reinterpreting the equations as representing the behavior of the gravitational field near the hole.

A third approach[5] involves the passage to the limit of an infinitesimal body, i.e., the replacement of the physical body by a "line singularity" in an otherwise smooth space—time. One wishes to show, by analyzing the structure of such a singular world line, that it represents, in some sense, a geodesic. Since the metric itself is singular on this world line, one is forced to introduce some sort of averaging procedure. Apparently, the procedures available at present may not be independent of the choice of coordinates. Furthermore, recent work[6] suggests that there may even be ambiguities already in the attachment, to a smooth space—time, of the "world line of singular points."

Finally, we mention an approach due to Dixon,[7] in which one introduces a certain world-line within a gravitating body, a line which suitably generalizes the Newtonian center of mass. The acceleration of this world line is expressed as a sum of integrals over the body, where these integrals represent the interaction of the mass multipoles of the body with the curvature of space—time. Geodesic motion arises as follows: One would expect that, for the case of a "small body, with little multipole structure," these integrals will also be small, whence the center-of-mass line will be nearly a geodesic. Of course, this formulation gives, not only this geodesic limit, but also the motion of a body in detail. Consider, for example, an isolated body which is spherical and homogeneous, except for a small region of higher density, slightly displaced from the center. One expects (e.g., from the Newtonian limit) that the center-of-mass world line of such a body will not be a geodesic; the present formulation would express this acceleration in terms of integrals over the body. Yet external observers would see the body as a whole moving approximately on a geodesic. What one might like to do for this example, and what is apparently difficult to do in detail, is introduce an "average acceleration" of the entire body, rather than an "acceleration of its average position."

The purpose of this paper is to introduce still another approach to the problem of the motion of a body in general relativity. Our approach differs from those discussed above in one, apparently minor, respect: We first introduce a world line, and only then the gravitating body, rather than the other way around. One is thus able to obtain a theorem which suggests geodesic motion, which is general, and yet which is extremely simple, both to state and to prove. The disadvantages of our approach are, first, that the physical interpretation of the theorem is somewhat less direct, and, second, that the method itself is not well-suited to obtaining any further details about the motion of the body.

## 2. MOTION OF BODIES

We first recall some facts about the motion of a body in special relativity.

We represent our body by a nonzero, symmetric tensor field $T^{ab}$, its stress—energy, on Minkowski space

$M$, where this $T$ is conserved: $\nabla_b T^{ab}=0$. Denote by $P_a$ and $J_{ab}$ $(=J_{[ab]})$ those tensor fields[8] on $M$ with the following property: for any Killing field $\xi^a$ on $M$,

$$-P_a\xi^a + J_{ab}\nabla^a\xi^b = \int_S T^{ab}\xi_b\, dS_a, \tag{1}$$

where the integral on the right extends over any spacelike 3-surface $S$ cutting the world tube of the body, i.e., cutting the support of $T$. By conservation of $T$ and Killing's equation, this integral is independent of the choice of $S$. Physically, $P_a$ and $J_{ab}$, evaluated at a point of $M$, represent the momentum and angular momentum, respectively, of the body about this point as origin. From the fact that the left side of (1) must be independent of position, it follows that

$$\nabla_a P_b = 0,$$

$$\nabla_a J_{bc} = g_{a[b}P_{c]}. \tag{2}$$

This, of course, is the dependence one would expect of the momentum and angular momentum on the choice of origin.

Now suppose that our $T^{ab}$ satisfies the following (strong) energy condition: For $t_a$ and $t'_a$ any future-directed timelike vectors at a point at which $T^{ab}$ is nonzero, $T^{ab}t_a t'_b$ is positive. It follows in this case from (1) (choosing for $\xi^a$ a time-translation) that $P_a$ is also timelike and future-directed. Define the center-of-mass world line $\gamma$ of the body as the set of points of $M$ at which $P^a J_{ab}=0$. It is easily checked from (2) (which can be integrated explicitly) that this $\gamma$ is a timelike geodesic, with tangent vector $P^a$.

There remains only to show that, in some sense, this center-of-mass world line $\gamma$ remains "near the world tube of the body." Define the (spatial) *convex hull* of $T$ to be the union of all segments of spacelike geodesics having both endpoints in the world tube. Consider now (1), evaluating the left side at a point $p$ of $\gamma$, using for the $S$ on the right the spacelike 3-plane through $p$ orthogonal to $P^a$, and using for $\xi^a$ a boost about $P^a$ at $p$. For these choices, the left side of (1) vanishes. But the integral on the right is a positively weighted average, over the support of $T$, of position relative to $p$. Hence, $p$ must lie within the convex hull of $T$. We conclude that the geodesic $\gamma$ lies entirely within the convex hull of $T$. In this sense, then, a body in special relatively "moves on a geodesic."

Of course, the above result is not available in the presence of curvature, for one does not normally have enough Killing fields in that case. Our result is the following:

*Theorem*: Let $M$, $g_{ab}$ be a space–time. Let $\Gamma$ be a curve on $M$ satisfying the following condition: For any neighborhood $U$ of $\Gamma$, there exists a nonzero, symmetric, conserved tensor field $T^{ab}$ on $M$ which satisfies the energy condition, and whose support is in $U$. Then $\Gamma$ is a timelike geodesic.

The proof consists of noting that "the nearer one is to $\Gamma$, the more nearly is the result of special relativity applicable." Fix,[9] once and for all, a flat metric $\tilde{g}_{ab}$ in some neighborhood of $\Gamma$, such that the metrics $g_{ab}$ and $\tilde{g}_{ab}$, as well as their derivative operators $\nabla_a$ and $\tilde{\nabla}_a$, coincide on $\Gamma$. Consider a symmetric $T^{ab}$ having support in this neighborhood. For each spacelike 3-plane (with respect to $\tilde{g}$) $S$, define $P_a(S)$ and $J_{ab}(S)$ by (1), where the Killing fields therein refer to $\tilde{g}$, and where the integral on the right is to be carried out over $S$. For each $S$, this $P_a(S)$ and $J_{ab}(S)$ satisfy (2), and so we obtain as before a geodesic (with respect to $\tilde{g}$), $\gamma(S)$ at a point of the convex hull (with respect to $\tilde{g}$) of $T$.

Now suppose that $T^{ab}$ is conserved with respect to $g$. Then $T^{ab}$ will not in general be conserved with respect to $\tilde{g}$. However, since the derivative operators coincide on $\Gamma$, $\tilde{\nabla}_b T^{ab} = (\tilde{\nabla}_b - \nabla_b)T^{ab}$ can be made as small as we wish (relative to the size of $T^{ab}$) by choosing the support of $T^{ab}$ to be sufficiently small. Since the difference between the right sides of (1) for two surfaces, $S$ and $S'$, is given by $\int_V (\tilde{\nabla}_b T^{ab})\xi_a\, dV$ where the integral extends over the region $V$ between $S$ and $S'$, this right side can also be made as small as we please. That is, the geodesics $\gamma(S)$, as $S$ ranges over 3-planes, can all be made to be as close to each other as we wish. From this and the fact that the intersection of each $S$ with the convex hull of the world tube contains a point of some $\gamma(S)$, we conclude that the curve $\Gamma$ is as close as we wish to some geodesic (with respect to $\tilde{g}$). But this is possible only if $\Gamma$ is itself a geodesic with respect to $\tilde{g}$. Since $\nabla_a = \tilde{\nabla}_a$ on $\Gamma$, $\Gamma$ must therefore be a geodesic also with respect to $g$.

Of course, the physical interpretation of the theorem is that, for any body, "insofar as that body is sufficiently small compared with the curvature that it may be regarded as a realization of the limit implicit in the theorem, then to that extent so may it be regarded as following some geodesic $\Gamma$."

Finally, we remark that the theorem does not conflict with the standard (nongeodesic) equations for the motion of a spinning body, or of a body with a quadrupole moment. For a body satisfying the energy condition, and with spatial extension of the order of $\delta$, its angular momentum per unit mass and quadrupole moment per unit mass cannot exceed the order of $\delta$ and $\delta^2$, respectively. Thus, for such a body, the effects of angular momentum and quadrupole moment on its motion can be made to be as small as one wishes by choosing the body itself to be sufficiently small. The theorem, however, asserts only that $\Gamma$ is a geodesic if "arbitrarily small bodies follow $\Gamma$."

*Supported in part by the National Science Foundation, Contract No. GP-34721X1, and by the Sloan Foundation.

[1]A. Taub, Proc. Natl. Acad. Sci. (USA) **48**, 1570 (1962).

[2]In Fact, it is easy to construct counterexamples even to the Newtonian version of this conjecture.

[3]See, for example, E.T. Newman, R. Posadas, Phys. Rev. **187**, 1784 (1969); J. Math. Phys. **12**, 2319 (1971); R.W. Lind, J. Messmer, E.T. Newman, J. Math. Phys. **13**, 1884 (1972).

[4]E.T. Newman, private communication.

[5]See, for example, J. Goldberg, article in *Gravitation, an Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), and references therein.

[6]See, for example, R. Geroch, J. Math. Phys. **9**, 450 (1968); B. Schmidt, J. Rel. Grav. **1**, 269 (1971).

[7]G. Dixon, Nuovo Cimento **34**, 317 (1964).

[8]Such fields exist. First note that the right-hand side of (1) defines a linear mapping from the ten-dimensional vector space of Killing fields on $M$ to the reals. Since a Killing field is completely determined by its value, together with the value of its derivative, at any one point, there exists, for each point $p$, a $P_a$ and $J_{ab}$ at $p$ such that (1) holds with the left-hand side evaluated at $p$.

[9]E. Fermi, Atti. Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat. **31**, 21, 51 (1922).

# A dispersion series for nonlocal potentials

T. H. Yao

*Mathematics Department, Bedford College, Regent's Park, London NW1, England*
(Received 12 July 1974)

For a class of short-range nonlocal potentials, and for the energy variable $E$ in a certain part of the complex plane, we obtain a generalized subtracted dispersion relation which relates the forward scattering amplitude to contributions from negative energy pole terms, a usual dispersion integral along the positive real axis of the complex energy plane, and a uniformly convergent infinite series, apart from subtraction terms, subject to the condition that no bound state exists with energy less than $-\gamma^2$, where $\gamma$ is some parameter of the potential.

## I. THE RESULT

We consider short-ranged rotationally invariant nonlocal potentials $V(\mathbf{x},\mathbf{x}')$ satisfying the following conditions:

(1) $V(\mathbf{x},\mathbf{x}')$ is real, $V(\mathbf{x},\mathbf{x}')=V(\mathbf{x}',\mathbf{x})$

(2) $V(\mathbf{x},\mathbf{x}')$ is rotationally invariant:

$$V(\mathbf{x},\mathbf{x}')=V(x,x',\cos\nu),$$

$$x=|\mathbf{x}|>0,\quad x'=|\mathbf{x}'|>0,\quad 1\geqslant\cos\nu\geqslant-1,$$

where $\nu$ is the angle between $\mathbf{x}$ and $\mathbf{x}'$,

(3)

$$V(x,x',\cos\nu)=\frac{(x+a)^m\exp(-\gamma x)}{x^\alpha}\,\tilde{V}(x,x',\cos\nu)\,\frac{(x'+a)^m\exp(-\gamma x')}{x'^\alpha}$$

$$\gamma>0,\quad a>0,\quad m=0,1,2,\cdots,\ \text{etc.},\quad \tfrac{3}{2}>\alpha\geqslant0,$$

where $\tilde{V}(x,x',\cos\nu)$ is continuous in $x$, $x'$, and $\cos\nu$ in $x>0$, $x'>0$, and $1\geqslant\cos\nu\geqslant-1$, and bounded in this region.

(4) The system has no negative energy bound state with energy less than $-\gamma^2$.

For any potential belonging to this class, the forward scattering amplitude $F(k)=f(E)$, where $E=k^2$, is holomorphic in $\gamma>\operatorname{Im}k>-(\gamma-\epsilon)$, for any $\epsilon$ in $\gamma\geqslant\epsilon>0$, perhaps with the exception of a finite number of poles at the nonreal zeroes of $\Delta(k)$,[1] where $\Delta(k)$ is the Fredholm determinant of the kernel

$$K(k;\mathbf{x},\mathbf{x}')=\frac{-1}{4\pi}\int d\mathbf{x}''\times\frac{\exp(ik|\mathbf{x}-\mathbf{x}''|)}{|\mathbf{x}-\mathbf{x}''|}\,V(\mathbf{x}'',\mathbf{x}').$$

We introduce a region $I(s,\tau)$ in the $E$ plane as follows. We know that[3] $\Delta(k)$ is holomorphic in $\operatorname{Im}k>-\gamma$ whose zeroes in $\operatorname{Im}k>0$ are finite in number and are all situated on the upper imaginary axis $\operatorname{Re}k=0$, $\operatorname{Im}k>0$ and correspond to negative bound state energies of the system. We define $U_+(s)$ to be the set $\{k\,|\,s>\operatorname{Im}k>0\}$ where $\gamma>s>0$ and s is sufficiently small so that $\bar{U}_+(s)$, the closure of $U_+(s)$, does not include any nonreal zero of $\Delta(k)$. We next define $U_+(s,\tau)$ to be the set of all points in $U_+(s)$ which are at distances more than some positive number $\tau$ from the real zeroes $\pm k_1,\ \pm k_2,\ \ldots,\ \pm k_n$, with $k_n>k_{n-1}>\cdots>k_1>0$, and the point $k=0$, if $\Delta(0)=0$, of $\Delta(k)$.[4] We then define a region $W_+(s,\tau)$ in the $E$ plane to be $W_+(s,\tau)=\{E=k^2\,|\,k\in U_+(s,\tau)\}$. Then we have the following result, valid for $E\in W_+(s,\tau)$:

$$f(E)=D_0+D_1(E-E_r)+(E-E_r)^2\sum_{i=1}^{N^-}\frac{B_i}{E-E_i^-}+(E-E_r)^2\sum_{p=1}^{\infty}\sum_{q=4}^{Q_p}\frac{G_{pq}}{[E+(\gamma+p\lambda)^2]^q}+\frac{(E-E_r)^2}{\pi}\int_0^\infty dE'\,\frac{\operatorname{Im}f(E')}{(E'-E_r)^2}\,\frac{1}{E'-E},\quad Q_p<\infty,\qquad(1)$$

where $\lambda>0$ is an arbitrary positive number, $E_i^-$, $i=1,2,\ldots,N^-$, are the negative bound state energies of the system (which are assumed to be all greater than $-\gamma^2$), $E_r$ is any constant not on the interval $[0,\infty)$ and not equal to $E_i^-$ and $-(\gamma+p\lambda)^2$ for $i=1,2,\ldots,N^-$ and $p=1,2,\ldots$, and $D_0$, $D_1$, $B_i$, for $i=1,2,\ldots,N^-$, and $G_{pq}$ are constants which are likely to depend on $\lambda$ and $E_r$. Here the infinite series is uniformly convergent with respect to $E$ for $E\in W_+(s,\tau)$. We call this series a dispersion series. This relation is also valid for $E\in I(s,\rho)=\{E=k^2\,|\,k\in U_+(s),\ |E-k_i^2|>\rho,\ i=1,2,\ldots,n,$ for some arbitrary $\rho$, and $E$ satisfying further $|E|>\rho$ if $\Delta(0)=0\}$. We represent the regions $U_+(s,\tau)$ and $I(s,\rho)$ in Figs. 1 and 2, respectively, for the case $n=1$, $\Delta(0)\neq0$.

We now write down a similar relation valid for $E$

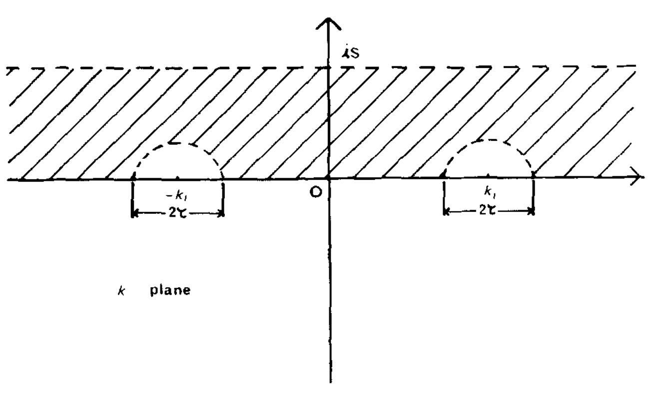



FIG. 1. $U_+(s,\tau)$ is the shaded region, for $n=1$, $\Delta(0)\neq0$.

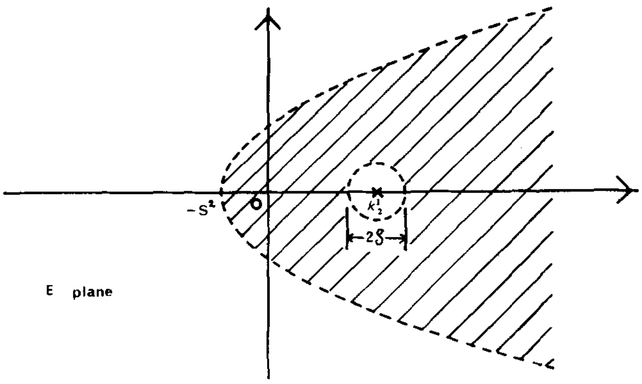



FIG. 2. $I(s,\rho)$ is the shaded region for $n=1$, $\Delta(0)\neq 0$.

belonging to part of the positive real axis. We define, for any $\rho>0$, the set $I(\rho)=\{E\,|\,E>\rho,\ |E-E_i^+|>\rho,\ i=1,2,\ldots,N^+\}$, where the set $\{E_i\,|\,i=1,2,\ldots,N^+\}$ are the positive energy eigenvalues of the system. Then for $E\in I(\rho)$, we have the relation

$$\begin{aligned}
\operatorname{Re}f(E) &= D_0+D_1(E-E_r)+(E-E_r)^2\sum_{i=1}^{N^-}\frac{B_i}{E-E_i^-} \\
&+(E-E_r)^2\sum_{p=1}^{\infty}\sum_{q=4}^{Q_p}\frac{G_{pq}}{[E+(\gamma+p\lambda)^2]^q} \\
&+\frac{(E-E_r)^2}{\pi}\,P\int_0^\infty dE'\,\frac{\operatorname{Im}f(E')}{(E'-E_r)^2} \\
&\times\frac{1}{(E'-E)},\qquad Q_p<\infty,
\end{aligned}\tag{2}$$

with the same constants as before.

## II. METHOD OF PROOF

We now outline the proof of the result in Sec. I.

To start with, we introduce the Hilbert space $\mathcal{L}^2(0,\infty)$ of measurable functions $f(x)$ on $(0,\infty)$ satisfying

$$\int_0^\infty dx\,x^2\,|f(x)|^2<\infty$$

with scalar product

$$(f_1,f_2)=\int_0^\infty dx\,x^2 f_1(x)^* f_2(x)$$

for $f_1$, $f_2$ belonging to the space. We also introduce the following set of functions:

$$\psi_{lp}(x)=x^l\exp[-(\mu_1+p\lambda)x],\quad \infty>x>0,$$

$$l=0,1,2,\ldots,\quad p=1,2,\ldots,\quad \lambda>0,\quad \mu_1=\gamma-\mu,$$

where $\gamma>\mu>0$ if the system has no negative energy bound states and $\gamma>\mu>(E_{\min}^-)^{1/2}>0$ if there is at least one negative energy bound state, with $E_{\min}^-$ being the lowest of the negative energy eigenvalues.

By using the transformation

$$y=\exp(-\lambda x)$$

and using the Weierstrass approximation theorem,[5] we may show that for any given $l$ any continuous function on $(0,\infty)$ vanishing outside some interval $(\eta,\xi)$, $\xi>\eta>0$, can be approximated arbitrarily closely in $\mathcal{L}^2(0,\infty)$ norm by linear combinations of $\psi_{lp}(x)$, $p=1,2,\cdots$, for fixed $l$. Consequently, for any given $l$, the linear span of the set $\{\psi_{lp}(x)\,|\,p=1,2,\cdots\}$ is dense in $\mathcal{L}^2(0,\infty)$. If we let $\{\phi_{lm}(x)\,|\,m=1,2,\cdots\}$ denote the set of orthonormal functions in $\mathcal{L}^2(0,\infty)$ obtained from the set $\{\psi_{lp}(x)\,|\,p=1,2,\cdots\}$ by the Schmidt orthonormalization procedure, then it forms a complete orthonormal basis in $\mathcal{L}^2(0,\infty)$, for any given $l$.

We introduce the Hilbert space $L^2(R^3)$ of measurable functions $f(\mathbf{x})$ on $R^3$ satisfying

$$\int d\mathbf{x}\,|f(\mathbf{x})|^2<\infty$$

equipped with the scalar product

$$\langle f_1,f_2\rangle=\int d\mathbf{x}\,f_1(\mathbf{x})^*f_2(\mathbf{x})$$

for $f_1$, $f_2\in L^2(R^3)$.

Putting

$$V(\mathbf{x},\mathbf{x}')=\exp(-\mu x)\,\hat V(\mathbf{x},\mathbf{x}')\exp(-\mu x'),\quad x=|\mathbf{x}|,\ x'=|\mathbf{x}'|,$$

and introducing

$$C_{lmn}=\int_0^\infty\int_0^\infty\int_{-1}^{+1}dx\,dx'\,d\cos\nu\,\hat V(x,x',\cos\nu)\,\phi_{lm}(x)\times\phi_{ln}(x')P_l(\cos\nu),\quad l=0,1,2,\cdots,\ m,n=1,2,\cdots$$

with $\mathbf{x}\cdot\mathbf{x}'=xx'\cos\nu$, we define

$$V_A(\mathbf{x},\mathbf{x}')=\exp(-\mu x)\,\hat V_A(\mathbf{x},\mathbf{x}')\exp(-\mu x'),$$

$$\hat V_A(\mathbf{x},\mathbf{x}')=\sum_{l=0}^{L}\sum_{m=1}^{M}\sum_{n=1}^{N}C_{lmn}\,\phi_{lm}(x)\,\phi_{ln}(x')\,P_l(\cos\nu),$$

where $A$ denotes the triplet $\{L,M,N\}$. From our result on the completeness of the set $\{\phi_{lm}(x)\,|\,m=1,2,\cdots\}$ for arbitrary $l$, we obtain

$$\lim_{A\to\infty}\|\hat V_A-\hat V\|=0,\qquad \lim_{A\to\infty}\|V_A-V\|=0,$$

where $\|\ \|$ is the Hilbert–Schmidt norm for Hilbert–Schmidt operators in $L^2(R^3)$, $A\to\infty$ means $L,M,N\to\infty$ simultaneously, and $\hat V$, $\hat V_A$, $V$, $V_A$ denote operators in $L^2(R^3)$ with the respective kernels $\hat V(\mathbf{x},\mathbf{x}')$, etc.

We reintroduce the kernel

$$K(k;\mathbf{x},\mathbf{x}')=\frac{-1}{4\pi}\int d\mathbf{x}''\,\frac{\exp(ik|\mathbf{x}-\mathbf{x}''|)}{|\mathbf{x}-\mathbf{x}''|}\,V(\mathbf{x}'',\mathbf{x}')$$

for $\operatorname{Im}k>-\mu$, and introduce the following other kernels, also for $\operatorname{Im}k>-\mu$:

$$K_A(k;\mathbf{x},\mathbf{x}')=\frac{-1}{4\pi}\int d\mathbf{x}''\,\frac{\exp(ik|\mathbf{x}-\mathbf{x}''|)}{|\mathbf{x}-\mathbf{x}''|}\,V_A(\mathbf{x}'',\mathbf{x}'),$$

$$\hat K(k;\mathbf{x},\mathbf{x}')=\frac{-1}{4\pi}\int d\mathbf{x}''\,\frac{\exp(ik|\mathbf{x}-\mathbf{x}''|)}{|\mathbf{x}-\mathbf{x}''|}\,\hat V(\mathbf{x}'',\mathbf{x}'),$$

$$\hat K_A(k;\mathbf{x},\mathbf{x}')=\frac{-1}{4\pi}\int d\mathbf{x}''\,\frac{\exp(ik|\mathbf{x}-\mathbf{x}''|)}{|\mathbf{x}-\mathbf{x}''|}\,\hat V_A(\mathbf{x}'',\mathbf{x}').$$

If we let $\Delta(k)$, $\Delta_A(k)$, $\hat\Delta(k)$, and $\hat\Delta_A(k)$ be the Fredholm determinants, for $\operatorname{Im}k>-\mu$, for the kernels $K(k;\mathbf{x},\mathbf{x}')$, $K_A(k;\mathbf{x},\mathbf{x}')$, $\hat K(k;\mathbf{x},\mathbf{x}')$, and $\hat K_A(k;\mathbf{x},\mathbf{x}')$ respectively, defined as in Ref. 2, we have the relations

$$\Delta(k)=\hat\Delta(k),\quad \Delta_A(k)=\hat\Delta_A(k),\quad \operatorname{Im}k>-\mu.$$

If we further let $\hat{K}(k)$ and $\hat{K}_A(k)$ be the integral operators in $L^2(R^3)$ with kernels $\hat{K}(k;\mathbf{x},\mathbf{x}')$ and $\hat{K}_A(k;\mathbf{x},\mathbf{x}')$, then they belong to the Hilbert–Schmidt class. We can prove, for any $\mu_2$ satisfying $\mu > \mu_2 > 0$,

$$\|\hat{K}(k)\| < \text{const}, \quad \operatorname{Im} k > -\mu_2,$$

and

$$\lim_{A\to\infty} \|\hat{K}_A^n(k) - \hat{K}^n(k)\| = 0, \quad n = 1, 2, \cdots,$$

uniformly with respect to $k$ in $\operatorname{Im} k > -\mu_2$, by induction.

Hence, using a result of Ref. 6, we have

$$\operatorname{Tr} \hat{K}_A{}^n(k) \underset{A\to\infty}{\to} \operatorname{Tr} \hat{K}^n(k), \quad n = 2, 3, \cdots,$$

uniformly with respect to $k$ in $\operatorname{Im} k > -\mu_2$. We also have

$$\operatorname{Tr} \hat{K}_A(k) \underset{A\to\infty}{\to} \operatorname{Tr} \hat{K}(k)$$

uniformly with respect to $k$ in $\operatorname{Im} k > -\mu_2$.

Hence we obtain, using a formula of Ref. 7 for the Fredholm determinant of a Hilbert–Schmidt operator, the result

$$\hat{\Delta}_A(k) \underset{A\to\infty}{\to} \hat{\Delta}(k)$$

uniformly with respect to $k$ in $\operatorname{Im} k > -\mu_2$, and consequently the result

$$\Delta_A(k) \underset{A\to\infty}{\to} \Delta(k)$$

uniformly with respect to $k$ in $\operatorname{Im} k > -\mu_2$. Similarly, using a formula of Ref. 7 for the Fredholm minor of a Hilbert–Schmidt operator, we obtain

$$\lim_{A\to\infty} \|\hat{\Delta}_A(k;\cdot,\cdot) - \hat{\Delta}(k;\cdot,\cdot)\| = 0$$

uniformly with respect to $k$ in $\operatorname{Im} k > -\mu_2$, where $\hat{\Delta}(k;\cdot,\cdot)$ and $\hat{\Delta}_A(k;\cdot,\cdot)$, for $\operatorname{Im} k > -\mu$, are the Hilbert–Schmidt operators in $L^2(R^3)$ whose kernels are the Fredholm minors of the operators $\hat{K}(k)$ and $\hat{K}_A(k)$ respectively. We have

$$\Delta(k;\mathbf{x},\mathbf{x}') = \exp(\mu x)\,\hat{\Delta}(k;\mathbf{x},\mathbf{x}')\exp(-\mu x'),$$

$$\Delta_A(k;\mathbf{x},\mathbf{x}') = \exp(\mu x)\,\hat{\Delta}_A(k;\mathbf{x},\mathbf{x}')\exp(-\mu x'), \quad \operatorname{Im} k > -\mu,$$

where $\Delta(k;\mathbf{x},\mathbf{x}')$ and $\Delta_A(k;\mathbf{x},\mathbf{x}')$ are the Fredholm minors for $\operatorname{Im} k > -\mu$ for the kernels $K(k;\mathbf{x},\mathbf{x}')$ and $K_A(k;\mathbf{x},\mathbf{x}')$ respectively, defined as in Ref. 2.

We now let $F(k)$ and $F_A(k)$ be the forward scattering amplitude corresponding to the potentials $V(\mathbf{x},\mathbf{x}')$ and $V_A(\mathbf{x},\mathbf{x}')$ respectively, which can be shown[8] to be holomorphic in $|\operatorname{Im} k| < \gamma$, perhaps with the exception of a finite number of nonreal poles in each case. Using the results obtained above, we can demonstrate

$$F_A(k) \underset{A\to\infty}{\to} F(k) \tag{3}$$

uniformly in $W(s,\tau,t)$, for any sufficiently small positive $t$, where $W(s,\tau,t)$ is the set of all points in $s > \operatorname{Im} k > -t$ which are at distances more than $\tau > 0$ from the real zeroes of $\Delta(k)$, where $\tau$ is arbitrary.

For the scattering amplitude $F_A(k)$, which is actually holomorphic in $\operatorname{Im} k > -\epsilon$ for some sufficiently small $\epsilon > 0$, perhaps with the exception of poles along the upper imaginary axis $k = i\kappa$, $\kappa > 0$, we have the following subtracted dispersion relation, valid for $E$ in the complex $E$-plane cut from 0 to $\infty$:

$$\begin{aligned} f_A(E) &= D_0 + D_1(E - E_r) + (E - E_r)^2 \\ &\quad \times \sum_{i=1}^{N_A^-} \frac{B_{iA}}{E - E_{iA}^-} + (E - E_r)^2 \\ &\quad \times \sum_{p=1}^{M} \sum_{q=4}^{Q_p} \frac{G_{pq}}{[(E + (\gamma + p\lambda)^2]^q} \\ &\quad + \frac{(E - E_r)^2}{\pi} \int_0^\infty dE' \, \frac{\operatorname{Im} f_A(E')}{(E' - E_r)^2} \, \frac{1}{E' - E}, \quad Q_p < \infty, \end{aligned} \tag{4}$$

for all $A$ sufficiently large (i.e., $L, M, N$ all sufficiently large), where $\lambda$, $E_r$, $D_0$, $D_1$, and $G_{pq}$ are constants described in Sec. I, $E_{iA}^-$ are the negative bound state energies of the system described by the potential $V_A(\mathbf{x},\mathbf{x}')$, with $i = 1, 2, \cdots, N_A^-$, and $B_{iA}$ are constants depending on $i$, $\mu_1$, $\lambda$, $A$, and $E_r$, with $i = 1, 2, \cdots, N_A^-$. Here we have used the condition $E_i^- > -\gamma^2$, $i = 1, 2, \cdots, N^-$, and the result that given arbitrary disjoint open intervals $\delta_i$ around $E_i^-$, $i = 1, 2, \cdots, N^-$, the eigenvalues $E_{iA}^-$, $i = 1, 2, \cdots, N_A^-$, are all situated inside the intervals $\delta_i$, $i = 1, 2, \cdots, N_A^-$, for all sufficiently large $A$.[9]

Using the expression of the transition operator in terms of the potential and the resolvent of the Hamiltonian, and the eigenfunction expansion of the resolvent kernel,[10] we obtain

$$B_{iA} = \frac{1}{(E_r - E_{iA}^-)^2} \times \sum_{j=1}^{J_{iA}} \left| \int d\mathbf{x} \, \mathcal{V}_A^* (\sqrt{-1}\;\kappa_{iA}\,\hat{e};\mathbf{x})\,\chi_{iA}^{(j)}(\mathbf{x}) \right|^2,$$

where

$$\begin{aligned} \mathcal{V}_A(\sqrt{-1}\;\kappa_{iA}\hat{e};\mathbf{x}) &= \int d\mathbf{x}' \; V_A(\mathbf{x},\mathbf{x}') \\ &\quad \times \exp(\sqrt{-1} k_{iA}\hat{e}\cdot\mathbf{x}'), \quad k_{iA} = \sqrt{-1}\;\kappa_{iA} \\ &\quad = \sqrt{-1}\;\sqrt{|E_{iA}^-|}, \quad \hat{e} \text{ is any unit vector} \end{aligned}$$

and $\chi_{iA}^{(j)}(\mathbf{x})$, $j = 1, 2, \ldots, J_{iA}$, form the set of orthonormalised energy eigenfunctions of the system described by the potential $V_A(\mathbf{x},\mathbf{x}')$, corresponding to the energy $E_{iA}^-$.

Using the result mentioned above on the distribution of the eigenvalues $E_{iA}^-$, $i = 1, 2, \ldots, N_A^-$, as $A \to \infty$, and also the following relation,[9]

$$P_A(\delta_i) \underset{A\to\infty}{\to} P_i, \quad i = 1, 2, \ldots, N^-,$$

in operator norm, where $P_A(\delta_i)$ is the projection onto the direct sum of the characteristic subspaces of the Hamiltonian operator associated with the potential $V_A(\mathbf{x},\mathbf{x}')$ corresponding to its eigenvalues in the interval $\delta_i$ and where $P_i$ is the projection onto the characteristic subspace of the Hamiltonian operator associated with the potential $V(\mathbf{x},\mathbf{x}')$ corresponding to the eigenvalue $E_i^-$, we obtain

$$\sum_{i=1}^{N_A^-} \frac{B_{iA}}{k^2 - E_{iA}^-} \underset{A\to\infty}{\to} \sum_{i=1}^{N^-} \frac{B_i}{k^2 - E_i^-} \tag{5}$$

uniformly with respect to $k$ in $W(s,t)$, for some sufficiently small $t > 0$, the region in the $k$ plane consisting of all points belonging to $U_+(s)$ and all points in $0 \geq \mathrm{Im}k > -t$, where $B_i$ are some constants which may depend on $\lambda$ and $E_r$.

The relations (1) and (2) follow from (3), (4), and (5) by contour deformation in $k'$ plane, with $E' = k'^2$.

## ACKNOWLEDGMENT

The author would like to thank Professor L. Castillejo for interest in the work.

[1]We may use the method of Ref. 2 to show this. We use also

$$\lim_{|k|\to\infty} \Delta(k) = 1, \quad \mathrm{Im}k > -(\gamma - \epsilon).$$

[2]T.H. Yao, "On analytic nonlocal potentials. I. A forward dispersion relation," J. Math. Phys. 14, 1141 (1973).

[3]We may use the method of Ref. 2.

[4]The members of the set $\{k_i \mid i = 1, 2, \cdots n\}$ correspond to the positive bound state energies of the system.

[5]J. Dieudonne, *Foundations of Modern Analysis* (Academic, New York, 1960).

[6]F. Smithies, *Integral Equations* (Cambridge, U.P., Cambridge, 1958), p. 80, Theorem 6.1.2.

[7]Dunford and Schwartz, *Linear Operators*, *Part II* (Interscience, New York, 1963), p. 1084–86, Ex 47–48; see also Ref. 6.

[8]We may use the method of Ref. 2 to show this.

[9]F. Riesz and B. Sz-Nagy, *Functional Analysis* (Ungar, New York, 1955), pp. 372–73.

[10]T.H. Yao, "The $T$-operator and an inverse problem for nonlocal potentials," J. Math. Phys. 15, 1227 (1974).

# Casimir operators of complementary unitary groups*

M. C. K. Aguilera-Navarro

*Instituto de Matemática e Estatística, USP, São Paulo, Brazil*

V. C. Aguilera-Navarro

*Instituto de Física Teórica, São Paulo, Brazil*

(Received 2 July 1974; revised manuscript received 3 September 1974)

A relationship between all the generalized Casimir operators of complementary unitary groups is derived, both in the fermion and boson realizations of the corresponding Lie algebras. It is shown that the number of independent Casimir operators of unitary groups reduces essentially to half the number for self-conjugate irreducible representations.

## 1. INTRODUCTION

The notion of complementary subgroups within a given irreducible representation (IR) of the larger group is defined by Moshinsky and Quesne in Ref. 1. They used this concept in applications to the treatment of many-body forces, the state-labeling problem and the quasiparticle picture. [1,2]

In this article, we shall consider the one-row IR's of unitary groups and derive a connection between all the generalized Casimir operators of unitary complementary subgroups within those irreducible representations.

In the next section, we give well-known realizations of the relevant Lie algebras in terms of boson and fermion operators. The notion of generalized Casimir operators is given within our adopted convention for the index contractions.

In Sec. 3, we derive a relationship between all the generalized Casimir operators of complementary unitary groups both in the fermion and boson cases.

Some consequences of our results are discussed in Sec. 4, where we consider the particular case of complementary unitary groups of the same dimension.

The idea presented in this article are being extended by the authors to other physically interesting groups such as the symplectic and orthogonal ones.

## 2. THE GENERALIZED CASIMIR OPERATORS

As is well known, the generators of unitary groups can be realized in terms of boson or fermion operators. Here we discuss some aspects of this subject with the purpose of introducing notation and conventions.

In terms of the boson operators $a_\rho^\dagger$ and $a^\rho$ we can construct the following $N^2$ operators

$$\mathbf{A}_\rho^{\rho'} = a_\rho^\dagger a^{\rho'}, \quad \rho, \rho' = 1, 2, \ldots, N. \tag{2.1}$$

These operators are generators of the unitary group in $N$ dimensions, $\mathbf{U}(N)$, inducing one-row irreducible representations. [3]

The index $\rho$ labeling the boson (or fermion) operators will stand for a couple of indices $(\mu s)$, each of them associated with different subspaces of the original $N$-dimensional space. In applications, $\mu$ may refer to the orbital characterization of $s$-particle states or these indices may refer to spin—isospin states, just to mention two instances.

With such a splitting of $\rho$, we can define two new sets of operators $\mathcal{A}_\mu^{\mu'}$ and $A_s^{s'}$ by contracting on $s$ or $\mu$, in the following way:

$$\mathcal{A}_\mu^{\mu'} = \sum_{s=1}^{r} a_{\mu s}^\dagger a^{\mu' s}, \quad \mu, \mu' = 1, 2, \ldots, p, \tag{2.2}$$

$$A_s^{s'} = \sum_{\mu=1}^{p} a_{\mu s}^\dagger a^{\mu s'}, \quad s, s' = 1, 2, \ldots, r. \tag{2.3}$$

The $p^2$ operators $\mathcal{A}_\mu^{\mu'}$ are generators of the unitary group $\mathcal{U}(p)$ while the $r^2$ operators $A_s^{s'}$ generate $U(r)$.

From the commutation relations for the boson operators, it is easy to see that the generators (2.1), (2.2), and (2.3) satisfy the following commutation relation:

$$[X_i^j, X_k^l] = \delta_k^j X_i^l - \delta_i^l X_k^j, \tag{2.4}$$

where $X$ stands for any of the generators.

Since the operators $\mathcal{A}_\mu^{\mu'}$ and $A_s^{s'}$ commute for all values of the indices, we see that $\mathbf{U}(N)$ contains the direct product of $\mathcal{U}(p)$ and $U(r)$, i. e.,

$$\mathbf{U}(N) \supset \mathcal{U}(p) \otimes U(r). \tag{2.5}$$

Now, the one-row IR of $\mathbf{U}(N)$ contains only the IR's of $\mathcal{U}(p)$ and $U(r)$ characterized by the same pattern. [3] Thus, there is a one-to-one correspondence between the IR's of those subgroups of $\mathbf{U}(N)$, and they are therefore complementary within the one-row IR of $\mathbf{U}(N)$.

The generalized Casimir operators of $\mathcal{U}(p)$ and $U(r)$, $\mathcal{C}_n$ and $C_m$, respectively, are defined by (from now on we shall use the usual convention that repeated indices are summed over the whole range of their values)

$$\mathcal{C}_n = \mathcal{A}_{\mu_1}^{\mu_n}\mathcal{A}_{\mu_2}^{\mu_1}\cdots\mathcal{A}_{\mu_n}^{\mu_{n-1}}, \quad n = 1, 2, \ldots, p, \tag{2.6}$$

and

$$C_m = A_{s_1}^{s_m}A_{s_2}^{s_1}\cdots A_{s_m}^{s_{m-1}}, \quad m = 1, 2, \ldots, r. \tag{2.7}$$

In the next section, it will be shown that each $\mathcal{C}_n$ can be expressed in terms of the $C_m$'s and *vice-versa*.

Similarly, we can obtain realizations of the infinitesimal generators of unitary groups in terms of the fermion operators $b_\rho^\dagger$ and $b^\rho$. In this case, the subgroup $\mathcal{U}(p) \otimes U(r)$ is embedded in a totally antisymmetric IR of $\mathbf{U}(N)$ and the IR's of $\mathcal{U}(p)$ and $U(r)$ are characterized by conjugate patterns.[4] Again, a one-to-one correspondence is then established so that $\mathcal{U}(p)$ and $U(r)$ are complementary within one-column IR's of $\mathbf{U}(N)$.

In the case of the fermion realization of the relevant algebras we denote by $\mathcal{D}_n$ and $D_m$ the generalized Casimir operators of $\mathcal{U}(p)$ and $U(r)$, respectively. Formally, these invariants are identical to (2.6) and (2.7) but, as will be seen later on, the relations among them are somewhat different from those for $\mathcal{C}_n$ and $C_m$. This difference will provide us with new information.

A few words about the way the generalized Casimir operators were defined follow. Obviously, we will get an invariant regardless of the way of contracting all the indices. For instance, Perelomov and Popov[5] and Louck and Biedenharn[6] define the invariants by contracting "up—down" instead of "down—up" as we did here, i.e., they put

$$C'_m = A^{s_1}_{s_m} A^{s_2}_{s_1} \cdots A^{s_m}_{s_{m-1}}. \tag{2.8}$$

Clearly, $C'_1 \equiv C_1$ and $C'_2 \equiv C_2$, but for $m > 2$ they are related through expressions which get more and more complicated. We chose the "down—up" criterion *a posteriori* since the relationships we were looking for among the invariants led to simpler results.

## 3. RELATIONSHIP BETWEEN THE GENERALIZED CASIMIR OPERATORS

In this section, we will derive a relationship between the generalized Casimir operators of $\mathcal{U}(p)$ and $U(r)$. As the introduced algorithm is the same for both the boson and fermion realizations of the corresponding algebras, we shall discuss, in some detail, only the boson case, limiting ourselves to presenting the results for the fermion case.

Coming back to the definitions of $\mathcal{A}^{\mu'}_{\mu}$, $A^{s'}_{s}$, $\mathcal{C}_n$, and $C_m$ given by Eqs. (2.2), (2.3), (2.6), and (2.8), we see that we can express the generalized Casimir operators entirely in terms of boson creation $a^\dagger_{\mu s}$ and annihilation $a^{\mu s}$ operators in the following way:

$$C_m = a^\dagger_{\mu_1 s_1} a^{\mu_1 s_m} a^\dagger_{\mu_2 s_2} a^{\mu_2 s_1} \cdots a^\dagger_{\mu_m s_m} a^{\mu_m s_{m-1}} \tag{3.1}$$

and

$$\mathcal{C}_m = a^\dagger_{\mu_1 s_1} a^{\mu_m s_1} a^\dagger_{\mu_2 s_2} a^{\mu_1 s_2} \cdots a^\dagger_{\mu_m s_m} a^{\mu_{m-1} s_m}. \tag{3.2}$$

As all the indices are dummy ones, one sees that, by convenient permutations of the $a$'s, we can transform, say, (3.2) into (3.1) plus extra terms coming from the boson commutation relations. In this way, we will get a relation between $C_m$ and $\mathcal{C}_m$. Such a naive procedure, however, turns out to be cumbersome, as can be seen by just trying to apply it to the $m = 4$ case. To avoid tedious calculations, we show instead a connection between the two algebras involved. With the help of this connection, desired relationship will split out directly.

From the boson commutation relations and the definitions (2.2) and (2.3), it is easy to see that

$$a^\dagger_{\mu s} A^{s}_{s'} = a^\dagger_{\mu' s'} \mathcal{A}^{\mu'}_{\mu}. \tag{3.3}$$

In the fermion case, this bridge between the two algebras is given by

$$b^\dagger_{\mu s} B^{s}_{s'} = -\, b^\dagger_{\mu' s'} \mathcal{B}^{\mu'}_{\mu}. \tag{3.4}$$

where $\mathcal{B}^{\mu'}_{\mu}$ and $B^{s'}_{s}$ are defined as in (2.2) and (2.3) in terms of fermion operators.

Now substituting $A^{s_m}_{s_1}$ in (2.8) through its definition (2.3), we get

$$C_m = a^{\mu_1 s_m} a^\dagger_{\mu_1 s_1} A^{s_1}_{s_2} A^{s_2}_{s_3} \cdots A^{s_{m-1}}_{s_m} - p A^{s_m}_{s_2} A^{s_2}_{s_3} \cdots A^{s_{m-1}}_{s_m}, \tag{3.5}$$

where use was made of the boson commutation relations.

Using the bridge relation (3.3) and the definition (2.8), we can write (3.5) as

$$C_m + pC_{m-1} = a^{\mu_1 s_m} a^\dagger_{\mu_2 s_2} A^{s_2}_{s_3} \cdots A^{s_{m-1}}_{s_m} \mathcal{A}^{\mu_2}_{\mu_1}. \tag{3.6}$$

Next, we apply repeatedly (3.3) and (2.8) until all the $A$'s are converted to $\mathcal{A}$'s. In this process, no extra term is generated and (3.6) becomes

$$C_m + pC_{m-1} = a^{\mu_1 s_m} a^\dagger_{\mu_m s_m} \mathcal{A}^{\mu_m}_{\mu_{m-1}} \mathcal{A}^{\mu_{m-1}}_{\mu_{m-2}} \cdots \mathcal{A}^{\mu_2}_{\mu_1}. \tag{3.7}$$

The algorithm is completed by commuting the remaining $a$'s. The definitions (2.2) and (2.6) lead to the fundamental relation

$$C_m + pC_{m-1} = \mathcal{C}_m + r\mathcal{C}_{m-1}. \tag{3.8}$$

The corresponding relation, arising when one uses the fermion realization of the algebras, is

$$D_m - pD_{m-1} = (-1)^{m+1}[\mathcal{D}_m - r\mathcal{D}_{m-1}]. \tag{3.9}$$

Now consider the following relation:

$$C_{m+1} = \mathcal{C}_{m+1} + (r-p)\sum_{n=1}^{m} (-p)^{m-n}\mathcal{C}_n, \tag{3.10}$$

which can be proved to be true by successive iterations of the result (3.8). However, instead of doing so we sacrifice elegance in favour of comfort and suppose it to be true for a given $m$. Since it holds for $m = 1$ [use $C_1 = \mathcal{C}_1$ from the definitions and (3.8)], we show that it holds for $m+1$ also. From (3.8) we can write

$$\begin{aligned} 0 &= C_{m+2} + p\left[\mathcal{C}_{m+1} + (r-p)\sum_{n=1}^{m} (-p)^{m-n}\mathcal{C}_n\right] - \mathcal{C}_{m+2} - r\mathcal{C}_{m+1} \\ &= C_{m+2} - \mathcal{C}_{m+2} - (r-p)\sum_{n=1}^{m+1} (-p)^{m+1-n}\mathcal{C}_n, \end{aligned} \tag{3.11}$$

where we have introduced the inductive hypothesis into the brackets. This result shows that (3.10) is true for $m+1$ and, therefore, holds for all $m$.

For the fermion case, as in (3.9), the relationship (3.10) is affected by phases originating from the *anti* rather than commuting character of the fermion operators. From (3.9), we can similarly show that

$$D_{m+1} = (-1)^m\mathcal{D}_{m+1} - (r+p)\sum_{n=1}^{m} (-1)^n p^{m-n}\mathcal{D}_n. \tag{3.12}$$

Clearly, the mathematical symmetry of the problem allows us to write down, without further details, the following results:

$$\mathcal{C}_{m+1} = C_{m+1} + (p-r)\sum_{n=1}^{m} (-r)^{m-n}C_n, \tag{3.13}$$

and

$$\mathcal{D}_{m+1} = (-1)^m D_{m+1} - (r+p)\sum_{n=1}^{m} (-1)^n r^{m-n}D_n. \tag{3.14}$$

So, we have obtained relationships among the generalized Casimir operators of $\mathcal{U}(p)$ and $U(r)$. Particular cases of these relations have already appeared in the physics literature.[7,8]

## 4. DISCUSSION OF SOME PARTICULAR CASES

In this section, we shall consider the particular case $r=p$, i.e., when $\mathcal{U}$ and $U$ have the same dimension, and discuss some consequences of the previous results.

First of all, we note that, from the mathematical point of view, the groups $\mathcal{U}$ and $U$ are the same.

Since, in the boson case, the IR's of $\mathcal{U}$ and $U$ are characterized by the same Young tableau, the eigenvalues of the generalized Casimir operators $C_m$ and $\mathcal{C}_m$ must be the same,[9] i.e.,

$$C_m = \mathcal{C}_m. \tag{4.1}$$

This information is contained in (3.10) for the factor $r-p=0$ eliminates possible contributions from the summation.

Now let us see what we can learn if we consider the fermion case.

We know that within one-column IR's of U($N$), the IR's of $\mathcal{U}$ are the conjugates of $U$. If we further consider self-conjugate[10] IR's, there will be no significance in distinguishing between $D_m$ and $\mathcal{D}_m$, so that Eq. (3.14) can be put in the form

$$[1-(-1)^m]D_{m+1} = 2\sum_{n=1}^{m} (-1)^{m-n} r^{m-n+1} D_n. \tag{4.2}$$

So, all generalized Casimir operators of even order are given in terms of lower order ones through

$$D_m = \sum_{n=1}^{m+1} (-1)^{m-n+1} r^{m-n} D_n, \quad m \text{ even}. \tag{4.3}$$

This result reduces to $r/2$, or $(r-1)/2$, depending on whether $r$ is even or odd, the maximum number of independent invariants in self-conjugate irreducible representations of $U(r)$.

Indeed, we can derive a stronger constraint than the one given by (4.3). Since that relation holds for all even $m$, it is easy to see that

$$D_m = rD_{m-1}, \quad m \text{ even}. \tag{4.4}$$

Since $D_1 = h =$ the number of boxes in the Young tableau characterizing a given IR of $U(r)$, we see, for instance, from (4.4) that

$$D_2 = rh, \tag{4.5}$$

for self-conjugate IR's of $U(r)$.


## ACKNOWLEDGMENTS

The authors wish to express their gratitude to Professor M. Moshinsky and Professor L. C. Biedenharn for valuable discussions. One of us (MCKAN) is grateful to the FINEP for financial support in the beginning of this work and to the Instituto de Física Teórica, São Paulo, for their kind hospitality.

*Work partially supported by FINEP, Brazil.



[1] M. Moshinsky and C. Quesne, Phys. Lett. B **29**, 482 (1969).
[2] M. Moshinsky and C. Quesne, J. Math. Phys. **11**, 1631 (1970).
[3] M. Moshinsky, J. Math. Phys. **4**, 1128 (1963).
[4] B. F. Bayman, Lectures on "Groups and Their Applications to Spectroscopy," Nordita, Copenhagen, 1960.
[5] A. M. Perelomov and V. S. Popov, Sov. J. Nucl. Phys. **3**, 676 (1966).
[6] J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).
[7] E. Chacón and M. de Llano, Rev. Méx. Fis. **13**, 87 (1964).
[8] V. C. Aguilera-Navarro, J. A. Castilho Alcarás, and P. Leal Ferreira, An. Acad. Bras. Ciências **39**, 49 (1967).
[9] J. D. Louck, J. Math. Phys. **6**, 1786 (1965).
[10] M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1962).

# Ghost neutrinos in plane-symmetric spacetimes

Talmadge M. Davis and John R. Ray

*Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29631*
(Received 5 June 1974)

An exact solution to the Einstein–Dirac equations is presented for a plane-symmetric spacetime generated by neutrinos. The neutrino field is nonzero and corresponds to a neutrino current along the symmetry axis of the space. The neutrinos yield a nonzero energy–momentum tensor, which we specialize to $T_{ij} = 0$ for "ghost" neutrinos. We show that, since the energy–momentum tensor vanishes, the time-dependent "ghost" neutrino metric reduces to the static case. The time-dependent "ghost" current is then reduced to the static current through a Lorentz transformation and the "ghost" wavefunction reduced to the static wavefunction through a spinor transformation. The "ghost" neutrino current is geodesic and the spacetime is classified by the expansion, rotation, and shear of these geodesics. From previous results it follows that our plane-symmetric "ghost" solution is the most general solution to the Einstein–Dirac equations for a vanishing energy–momentum tensor and a neutrino current that is expanding. The solution is Petrov type D.

## I. INTRODUCTION

In a previous paper[1] we found the exact solution to the *static*, plane symmetric Einstein–Dirac equations. The interesting results were a vanishing energy–momentum tensor and a nonvanishing neutrino current. Our solution is not the most general solution for plane symmetry in that we have required the solution to be static. It is possibly this *ad hoc* static requirement that forces the energy–momentum tensor to vanish.

In this paper we return to the general case of plane symmetry and calculate the solution of the Einstein–Dirac equations for the general case. In Sec. II we present the solution to the Einstein–Dirac equations. The notation is that of our previous paper. The energy–momentum tensor no longer vanishes and results in two cases—$T_{ij}$ is a function of $x+t$ and $T_{ij}$ is a function of $x-t$.

In Sec. III we consider the special case of a zero energy–momentum tensor. We find the metric, neutrino wavefunction, and current for this special case. The vanishing energy–momentum tensor allows a coordinate transformation to the static metric.[2] Since the metrics are equivalent, one would also expect the neutrino wavefunction, current, etc., to reduce to the static forms. This we prove in Sec. IV.

In Sec. V we show that the current is a null geodesic and calculate the shear, rotation, and expansion of the null congruence.

We define a null-tetrad based on the neutrino current and determine the shear, rotation, and expansion in terms of these tetrad vectors.[3] We also use the tetrad to obtain the Petrov classification of the spacetime—Petrov type D.

Collinson and Morris[4] proceed in a different manner by solving the Neuman—Penrose field equations for ghost neutrinos. Finally, we show in Sec. V that our static metric is equivalent to the Collinson—Morris expanding metric through several coordinate transformations.

## II. GENERAL SOLUTION

We consider the time-dependent plane-symmetric spacetime defined by

$$ds^2 = e^{2u}(dx^2 - dt^2) + e^{2v}(dy^2 + dz^2), \qquad (2.1)$$

where

$$u = u(x, t),$$
$$v = v(x, t).$$

We use the orthonormal Cartan frame

$$\omega^1 = e^u\, dx,$$
$$\omega^2 = e^v\, dy,$$
$$\omega^3 = e^v\, dz, \qquad (2.2)$$
$$\omega^4 = e^u\, dt.$$

Since the trace of the energy–momentum tensor for neutrinos vanishes, the Einstein field equations become

$$R_{ij} = \frac{8\pi\kappa}{c^4} T_{ij}. \qquad (2.3)$$

The nonzero elements of the Ricci tensor are

$$R_{11} = \exp(-2u)(2u_{,4}v_{,4} + 2u_{,1}v_{,1} - 2v_{,1}{}^2 - 2v_{,11} - u_{,11} + u_{,44}), \qquad (2.4)$$

$$R_{22} = R_{33} = \exp(-2u)(v_{,44} - v_{,11} + 2v_{,4}{}^2 - 2v_{,1}{}^2), \qquad (2.5)$$

$$R_{44} = \exp(-2u)(2u_{,4}v_{,4} + 2u_{,1}v_{,1} - 2v_{,4}{}^2 - 2v_{,44} + u_{,11} - u_{,44}), \qquad (2.6)$$

$$R_{14} = 2\exp(-2u)(u_{,1}v_{,4} + u_{,4}v_{,1} - v_{,14} - v_{,1}v_{,4}), \qquad (2.7)$$

where the comma denotes partial differentiation.

We now solve the massless Dirac equation for the neutrino wavefunction in the geometry described by (2.1). The spin coefficients are given by

$$\Gamma_1 = -\tfrac{1}{2}u_{,4}e^{-u}\gamma^1\gamma^4, \qquad (2.8)$$

$$\Gamma_2 = -\tfrac{1}{2}e^{-u}(v_{,1}\gamma^2\gamma^1 + v_{,4}\gamma^2\gamma^4), \qquad (2.9)$$

$$\Gamma_3 = -\tfrac{1}{2}e^{-u}(v_{,1}\gamma^3\gamma^1 + v_{,4}\gamma^3\gamma^4), \qquad (2.10)$$

$$\Gamma_4 = -\tfrac{1}{2}u_{,1}e^{-u}\gamma^1\gamma^4. \qquad (2.11)$$

Using these spin coefficients, we find the Dirac equation becomes

$$\gamma^1\psi_{,1} + \gamma^4\psi_{,4} = -[(v_{,1} + u_{,1}/2)\gamma^1 + (v_{,4} + u_{,4}/2)\gamma^4]\psi. \qquad (2.12)$$

This can be simplified to

$$\chi_{,1} = -\gamma^1\gamma^4\chi_{,4}, \tag{2.13}$$

where

$$\chi = \exp(v + u/2)\psi.$$

For neutrinos $\chi$ has the form

$$\chi_\nu = \begin{pmatrix} \chi_1 \\ \chi_2 \\ i\chi_1 \\ i\chi_2 \end{pmatrix}. \tag{2.14}$$

In terms of the components of $\chi$ the Dirac equation yields

$$\chi_{1,1} = \chi_{2,4},$$
$$\chi_{2,1} = \chi_{1,4}. \tag{2.15}$$

The solutions are

$$\chi_1 = f(x+t) + g(x-t),$$
$$\chi_2 = f(x+t) - g(x-t) + c, \tag{2.16}$$

where $f(x+t)$ and $g(x-t)$ are arbitrary functions and $c$ is an arbitrary constant. We will denote differentiation with respect to $x+t$ by a prime ($'$) and with respect to $x-t$ by a dot ($\cdot$). Also, we define the functions $F$ and $G$ by

$$F = 2f + c = F_r + iF_i,$$
$$G = 2g - c = G_r + iG_i, \tag{2.17}$$

where $F_r$, $G_r$ are the real parts of $F$ and $G$, respectively, and $F_i$, $G_i$ are the imaginary parts.

The wavefunction now yields the energy–momentum tensor

$$T_{11} = T_{44} = \tfrac{1}{2}\hbar c \exp[-2(u+v)](F_rF_i' - F_iF_r' - G_r\dot{G}_i + G_i\dot{G}_r), \tag{2.18}$$

$$T_{14} = \tfrac{1}{2}\hbar c \exp[-2(u+v)](F_rF_i' - F_iF_r' + G_r\dot{G}_i - G_i\dot{G}_r), \tag{2.19}$$

$$T_{12} = -\tfrac{1}{4}\hbar c \exp[-2(u+v)][2(v-u)_{,4}(F_rG_r + F_iG_i) + F_r'G_r - F_r\dot{G}_r + F_i'G_i - F_i\dot{G}_i], \tag{2.20}$$

$$T_{13} = \tfrac{1}{4}\hbar c \exp[-2(u+v)][2(v-u)_{,4}(F_iG_r - F_rG_i) - F_r'G_i + F_r\dot{G}_i + F_i'G_r - F_i\dot{G}_r], \tag{2.21}$$

$$T_{24} = -\tfrac{1}{4}\hbar c \exp[-2(u+v)][2(v-u)_{,1}(F_rG_r + F_iG_i) + F_r'G_r + F_r\dot{G}_r + F_i'G_i + F_i\dot{G}_i], \tag{2.22}$$

$$T_{34} = \tfrac{1}{4}\hbar c \exp[-2(u+v)][2(v-u)_{,1}(F_iG_r - F_rG_i) - F_r'G_i - F_r\dot{G}_i + F_i'G_r + F_i\dot{G}_r]; \tag{2.23}$$

and the neutrino current

$$s^1 = -\exp[-(2v+u)](F_r^2 + F_i^2 - G_r^2 - G_i^2), \tag{2.24}$$

$$s^2 = -2\exp[-(2v+u)](F_iG_r - F_rG_i), \tag{2.25}$$

$$s^3 = -2\exp[-(2v+u)](F_iG_i + F_rG_r), \tag{2.26}$$

$$s^4 = \exp[-(2v+u)](F_r^2 + F_i^2 + G_r^2 + G_i^2). \tag{2.27}$$

Since $R_{12}$, $R_{13}$, $R_{24}$, and $R_{34}$, vanish, the field equations yield

$$T_{12} = T_{13} = T_{24} = T_{34} = 0. \tag{2.28}$$

The fact that the spacetime is plane-symmetric requires

$$s^2 = s^3 = 0. \tag{2.29}$$

From Eqs. (2.25), (2.26), and (2.29) we find

$$F_iG_r - F_rG_i = 0, \tag{2.30}$$

$$F_iG_i + F_rG_r = 0, \tag{2.31}$$

Applying these results to Eqs. (2.20)–(2.23) and (2.28), we find two cases:

*Case I*:

$$G_i = G_r = 0, \quad F_r, F_i \text{ arbitrary}. \tag{2.32}$$

*Case II*:

$$F_i = F_r = 0, \quad G_r, G_i \text{ arbitrary}. \tag{2.33}$$

We will discuss Case I and give the results of Case II in Sec. IV.

The field equations now become

$$R_{11} = \frac{2h\kappa}{c^3}\exp[-2(v+u)]F_r^2(F_i/F_r)', \tag{2.34}$$

$$R_{22} = 0, \tag{2.35}$$

$$R_{33} = 0, \tag{2.36}$$

$$R_{44} = \frac{2h\kappa}{c^3}\exp[-2(v+u)]F_r^2(F_i/F_r)', \tag{2.37}$$

$$R_{14} = \frac{2h\kappa}{c^3}\exp[-2(v+u)]F_r^2(F_i/F_r)'. \tag{2.38}$$

Equations (2.35) and (2.36) yield

$$e^{2v} = \alpha(x+t) + \beta(x-t), \tag{2.39}$$

where $\alpha$ and $\beta$ are arbitrary functions. Subtracting (2.37) from (2.34) and substituting (2.35) into this difference, we obtain

$$v + 2u = A(x+t) + B(x-t), \tag{2.40}$$

where $A$ and $B$ are arbitrary functions. Substitution of (2.39) and (2.40) into the field equations yields conditions on $A(x+t)$ and $B(x-t)$ which can be written as

$$A = \ln(2\alpha'/k) + C(x+t), \tag{2.41}$$

$$B = \ln(2\dot{\beta}/k), \tag{2.42}$$

where $k$ is a constant and $C(x+t)$ is defined by the integral

$$C(x+t) = \frac{2h\kappa}{c^3}\int \frac{F_r^2}{\alpha'}(F_i/F_r)'\,d(x+t). \tag{2.43}$$

The metric is now given by

$$ds^2 = \frac{4}{k^2}\alpha'\mathring{\beta}(\alpha+\beta)^{-1/2}\exp[C(x+t)](dx^2 - dt^2) + (\alpha+\beta)(dy^2 + dz^2). \tag{2.44}$$

This is the most general form of the metric for a plane-symmetric spacetime which allows neutrinos.

## III. "GHOST" SOLUTION ($T_{ij} = 0$)

Since there are no more equations to solve, we assume special forms for $F_r$ and $F_i$. In the remainder of the paper we consider the case of a vanishing energy—momentum tensor, or

$$F_r^2(F_i/F_r)'=0. \tag{3.1}$$

We have given the name "ghost" neutrinos to the neutrino solutions for which $T_{ij}=0$ and refer the reader to Ref. 1 for a discussion of "ghost" neutrinos. Equation (3.1) can be satisfied by several choices of $F_r$ and $F_i$. We select

$$F_r=a\gamma(x+t), \tag{3.2}$$

$$F_i=b\gamma(x+t), \tag{3.3}$$

where $a$ and $b$ are arbitrary real constants and $\gamma(x+t)$ is an arbitrary real function of $x+t$. Note that the other choices are either $a=0$ or $b=0$.

Taub[2] has shown that for a general plane-symmetric spacetime if the Ricci tensor vanishes (i.e., empty spacetime), then the metric can be reduced through a coordinate transformation to the static plane-symmetric metric given in Ref. 1. Since $C(x+t)=0$ by Eqs. (2.43) and (3.1), the general metric (2.44) takes the form

$$ds^2=\frac{4}{k^2}\alpha'\dot{\beta}(\alpha+\beta)^{-1/2}(dx^2-dt^2)+(\alpha+\beta)(dy^2+dz^2). \tag{3.4}$$

The coordinate transformation which reduces this general metric to the static form is

$$X^1+X^4=\frac{2}{k}[\alpha(x+t)-\tfrac{1}{2}], \tag{3.5}$$

$$X^1-X^4=\frac{2}{k}[\beta(x-t)-\tfrac{1}{2}], \tag{3.6}$$

$$X^2=y, \tag{3.7}$$

$$X^3=z. \tag{3.8}$$

Performing the transformation, we obtain

$$\begin{aligned}ds^2&=(kX^1+1)^{-1/2}[(dX^1)^2-(dX^4)^2]\\&\quad+(kX^1+1)[(dX^2)^2+(dX^3)^2],\end{aligned} \tag{3.9}$$

and the metric is the same as the static case presented in Ref. 1.

## IV. "GHOST" NEUTRINO WAVEFUNCTION AND CURRENT

Although we have an equivalence between the time-dependent and static "ghost" metrics, it is not obvious that the "ghost" wavefunction and current reduce to their static forms. The wavefunction is

$$\psi_\nu=\tfrac{1}{2}\exp[-(v+u/2)](a+ib)\gamma(x+t)\begin{pmatrix}1\\1\\i\\i\end{pmatrix}, \tag{4.1}$$

and the current is

$$s^1=-\exp[-(2v+u)](a^2+b^2)\gamma^2(x+t), \tag{4.2}$$

$$s^2=0, \tag{4.3}$$

$$s^3=0, \tag{4.4}$$

$$s^4=\exp[-(2v+u)](a^2+b^2)\gamma^2(x+t). \tag{4.5}$$

$\psi_\nu$, $s^1$, and $s^4$ are still arbitrary functions of $x+t$ and $x-t$, but they should be transformable to the static form of Ref. 1.

Up to this point, we have constructed everything with respect to the orthonormal frame

$$\begin{aligned}\omega^1&=\frac{2}{k}(\alpha'\dot{\beta})^{1/2}(\alpha+\beta)^{-1/4}\,dx,\\ \omega^2&=(\alpha+\beta)^{1/2}\,dy,\\ \omega^3&=(\alpha+\beta)^{1/2}\,dz,\\ \omega^4&=\frac{2}{k}(\alpha'\dot{\beta})^{1/2}(\alpha+\beta)^{-1/4}\,dt.\end{aligned} \tag{4.6}$$

After the coordinate transformation (3.5)—(3.8), we must construct the new orthonormal frame

$$\begin{aligned}W^1&=(kX^1+1)^{-1/4}\,dX^1,\\ W^2&=(kX^1+1)^{1/2}\,dX^2,\\ W^3&=(kX^1+1)^{1/2}\,dX^3,\\ W^4&=(kX^1+1)^{-1/4}\,dX^4.\end{aligned} \tag{4.7}$$

These two orthonormal frames are connected through a Lorentz transformation

$$W^k=L_l{}^k(x,t)\omega^l, \tag{4.8}$$

where

$$L_l{}^k(x,t)=\begin{pmatrix}(\alpha'+\dot{\beta})(\alpha'\dot{\beta})^{-1/2}/2 & 0 & 0 & (\alpha'-\dot{\beta})(\alpha'\dot{\beta})^{-1/2}/2\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ (\alpha'-\dot{\beta})(\alpha'\dot{\beta})^{-1/2}/2 & 0 & 0 & (\alpha'+\dot{\beta})(\alpha'\dot{\beta})^{-1/2}/2\end{pmatrix}. \tag{4.9}$$

In an appendix we show that for zero velocity this Lorentz transformation reduces to the identity transformation.

Applying the Lorentz transformation to the current

$$S^k=L_l{}^k s^l, \tag{4.10}$$

we find

$$S^j=\frac{k}{2}(\alpha+\beta)^{-3/4}(a^2+b^2)\frac{\gamma^2(x+t)}{\alpha'}(-\delta^j{}_1+\delta^j{}_4). \tag{4.11}$$

If this current allows the same timelike Killing vector as the static metric, we must have

$$\gamma(x+t)=2k_1(2\alpha'/k)^{1/2}, \tag{4.12}$$

where $k_1$ is an arbitrary real constant. The current becomes

$$S^j=4|c|^2(kX^1+1)^{-3/4}(-\delta^j{}_1+\delta^j{}_4),\quad |c|^2=k_1^2(a^2+b^2). \tag{4.13}$$

This is the "ghost" neutrino current in the $W_i$ frame and is in the static form of Ref. 1. Note that $S^j$ is a null vector as is required for neutrinos.

The wavefunction is transformed through a spinor transformation defined by

$$S=\exp(\gamma^1\gamma^4\theta/2), \tag{4.14}$$

where

$$\tanh\theta=v/c, \tag{4.15}$$

and

$$\Psi_\nu=S\psi_\nu. \tag{4.16}$$

Using the $\theta$ from the Lorentz transformation (4.9) and the requirement (4.12), we obtain

$$\Psi_\nu=c(kX^1+1)^{-3/8}\begin{pmatrix}1\\1\\i\\i\end{pmatrix}, \quad c=k_1(a+ib). \tag{4.17}$$

This is also in the static form as given in Ref. 1.

In Ref. 1 we found two solutions for the wavefunction and current. One of these solutions is given by (4.13) and (4.17). By choosing Case II in Sec. II (i.e, $F_i=F_r=0$), we obtain the other solution. Namely, the wavefunction is

$$\Psi_\nu=c(kX^1+1)^{-3/8}\begin{pmatrix}1\\-1\\i\\-i\end{pmatrix}, \tag{4.18}$$

and the current is

$$S^j=4|c|^2(kX^1+1)^{-3/4}(\delta^j{}_1+\delta^j{}_4), \tag{4.19}$$

with the requirement

$$\sigma(x-t)=2k_1(2\dot\beta/k)^{1/2}, \tag{4.20}$$

which arises in the same way as (4.12). $\sigma(x-t)$ is an arbitrary function of $x-t$ such that

$$G=(a+ib)\sigma(x-t). \tag{4.21}$$

## V. EXPANSION, ROTATION, SHEAR AND CLASSIFICATION

Since everything in our "ghost" solution can be reduced to the static solution through transformations, we will consider the static current. The static current $S^j$ obeys the equations

$$S^j_{;l}S^l=\pm 4|c|^2k(kX^1+1)^{-3/2}S^j, \tag{5.1}$$

in the coordinate basis defined by

$$E_i=\frac{\partial}{\partial X^i}.$$

(5.1) shows that $S^j$ is the tangent vector to null geodesics. Changing to an affine parameter results in a new form for the current

$$S^j=\frac{2}{k}(kX^1+1)^{1/2}(\mp\delta^j{}_1+\delta^j{}_4). \tag{5.2}$$

Using the notation of Sachs,[5] we find the following invariants for these null geodesics

$$\text{Expansion:}\quad \theta=(kX^1+1)^{-1/2}, \tag{5.3}$$

$$\text{Rotation:}\quad \omega=0, \tag{5.4}$$

and

$$\text{Shear:}\quad |\sigma|=0. \tag{5.5}$$

A null-tetrad based on the current (5.2) is given in the orthonormal frame (4.7) by

$$\begin{aligned}
l_j&=\frac{2}{k}(kX^1+1)^{1/4}(-\delta^1_j-\delta^4_j),\\
n_j&=\frac{k}{4}(kX^1+1)^{-1/4}(\delta^1_j-\delta^4_j),\\
m_j&=-\frac{1}{\sqrt2}(i\delta_j{}^2+\delta_j{}^3),\\
\bar m_j&=-\frac{1}{\sqrt2}(-i\delta_j{}^2+\delta_j{}^3).
\end{aligned} \tag{5.6}$$

In terms of Newman–Penrose spin coefficient formalism[3] we find

$$\text{Expansion:}\quad \theta=-(\rho+\bar\rho)/2,$$
$$\theta=(kX^1+1)^{-1/2}, \tag{5.7}$$

$$\text{Rotation (twist):}\quad 2\omega^2=-(\rho-\bar\rho)^2/2,$$
$$\omega=0, \tag{5.8}$$

and

$$\text{Shear:}\quad |\sigma|=(\sigma\bar\sigma)^{1/2},$$
$$|\sigma|=0, \tag{5.9}$$

where $\rho$ and $\sigma$ are defined in terms of the null-tetrad vectors and, explicitly,

$$\rho=\bar\rho=-(kX^1+1)^{-1/2}. \tag{5.10}$$

The space–times are classified by Newman and Penrose according to the nonvanishing of the quantities $\Psi_0$, $\Psi_1$, $\Psi_2$, $\Psi_3$, and $\Psi_4$ which are contractions of the Weyl tensor with the vectors of the null-tetrad. We find

$$\Psi_0=\Psi_1=\Psi_3=\Psi_4=0, \tag{5.11}$$

and

$$\Psi_2=-\frac{k^2}{8}(kX^1+1)^{-3/2}. \tag{5.12}$$

This implies that the spacetime is Petrov type D with propagation vectors $l_k$ and $n_k$.

Griffiths has shown that ghost neutrinos are either Petrov type D for nonzero expansion or Petrov type N for zero expansion.[6] Collinson and Morris have integrated the Newman–Penrose field equations to determine the metric for ghost neutrinos.[4] This result for the case of nonzero expansion is

$$g^{ij}=\begin{pmatrix}0&-1&0&0\\-1&-2/y&0&0\\0&0&2/y^2&0\\0&0&0&2/y^2\end{pmatrix}, \tag{5.13}$$

which is Petrov type D. Note that we have changed the signature in order to conform with our convention. After a reflection we perform three coordinate transformations:

$$\begin{aligned}
x&=-(x'+y'),\\
y&=(x'+3y')^{1/2},\\
z&=\sqrt2\,z',\\
t&=\sqrt2\,t',
\end{aligned} \tag{5.14}$$

then

$$x''=x',\quad y''=t',\quad z''=z',\quad t''=y',\tag{5.15}$$

and, finally,

$$\begin{aligned}X^1+X^4&=\sqrt{2}\,(x''+t''-\tfrac{1}{4}),\\ X^1-X^4&=-\frac{1}{\sqrt{2}}(x''-t''+\tfrac{1}{2}),\\ X^2&=y'',\\ X^3&=z''.\end{aligned}\tag{5.16}$$

The transformation (5.14) reduces the metric (5.13) to diagonal form. The transformation (5.15) then redefines the timelike coordinate. Finally, the metric is brought to the form of our static metric with $k=\sqrt{8}$ by the transformation (5.16). Hence, our "ghost" solution is the most general solution to the Einstein–Dirac equations for expanding ghost neutrinos.

## VI. CONCLUSIONS

We have found the general solution to the Einstein–Dirac equations for the case of a plane-symmetric spacetime and discussed the special solution having a zero energy–momentum tensor. The special solution is the same as in the static case and in particular has the same "ghost" property. The "ghost" solution is in fact the most general solution to the Einstein–Dirac equations for a vanishing energy–momentum tensor and a neutrino current that is expanding.

We have also shown that the static cylindrically-symmetric metric allows the expanding ghost neutrino solution.[7] It would be interesting to know all symmetry types that allow both the expanding and nonexpanding ghost neutrino solutions. Of course, if a spacetime allows expanding ghost neutrinos, it reduces to the plane-symmetric metric as in Ref. 7. Many different symmetry types could, however, still allow ghost neutrinos.

## APPENDIX

As a check on our Lorentz transformation (4.9), we can let $v$ go to zero. The $W^i$ frame should then reduce to the $\omega^i$ frame. Solving for $v/c$, we find

$$v/c=(\alpha'-\dot{\beta})/(\alpha'+\dot{\beta})=0.\tag{A1}$$

Integration yields

$$\alpha=c_1(x+t)+c_2,\tag{A2}$$

$$\beta=c_1(x-t)+c_3,\tag{A3}$$

where $c_1$, $c_2$, and $c_3$ are arbitrary constants. Substituting these into the coordinate transformation (3.5)–(3.8), we obtain

$$X^1+X^4=\frac{2}{k}[c_1(x+t)+c_2-1/2],\tag{A4}$$

$$X^1-X^4=\frac{2}{k}[c_1(x-t)+c_3-1/2],\tag{A5}$$

$$X^2=y,\tag{A6}$$

$$X^3=z.\tag{A7}$$

Now, if we let

$$c_1=k/2,\tag{A8}$$

$$c_2=c_3=1/2,\tag{A9}$$

then the two coordinate systems (and frames) are equal.


[1]T.M. Davis and J.R. Ray, Phys. Rev. D 9, 331 (1974).
[2]A. Taub, Ann. Math. 53, 472 (1951).
[3]E. Newman and R. Penrose, J. Math. Phys. 3, 566 (1962).
[4]C.D. Collinson and P.B. Morris, J. Phys. A 6, 915 (1973).
[5]R. Sachs, Proc. Roy. Soc. Lond. 264, 309 (1961).
[6]J.B. Griffiths, Commun. Math. Phys. 28, 295 (1972).
[7]T.M. Davis and J.R. Ray, J. Math. Phys. 16, 80 (1975).

# Neutrinos in cylindrically-symmetric spacetimes

Talmadge M. Davis and John R. Ray

*Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29631*

(Received 29 July 1974; final revised manuscript received 22 August 1974)

An exact solution to the Einstein–Dirac equations is presented for a static, cylindrically-symmetric spacetime. The neutrino field is nonzero and corresponds to a neutrino current in the radial direction. The neutrinos yield a zero energy–momentum tensor and therefore the gravitational field is the same as for the vacuum case. The neutrinos in these static cylindrically-symmetric spacetimes can exist only if the spacetimes are locally equivalent to static plane-symmetric spacetimes. This type of "ghost neutrino" solution is already known to exist in plane-symmetric spacetimes.

## I. INTRODUCTION

We will be considering solutions to the Einstein–Dirac equations in spacetimes which are static and cylindrically-symmetric where the metric is defined by

$$ds^2 = \exp[2(\nu-\lambda)](dr^2 - dt^2) + \exp(-2\lambda)\, r^2 d\phi^2 + \exp[2(\lambda+\mu)]\, dz^2, \tag{1.1}$$

where $\nu$, $\lambda$, and $\mu$ are functions of $r$ only as are all other functions. We shall carry out the calculations in the Cartan orthonormal frame defined by

$$\omega^1 = \exp(\nu-\lambda)\, dr, \tag{1.2}$$

$$\omega^2 = \exp(-\lambda)\, r\, d\phi, \tag{1.3}$$

$$\omega^3 = \exp(\lambda+\mu)\, dz, \tag{1.4}$$

$$\omega^4 = \exp(\nu-\lambda)\, dt. \tag{1.5}$$

In Sec. II we solve the Dirac equation for massless particles

$$\gamma^i \psi_{;i} = 0, \tag{1.6}$$

for the neutrino wave function which is required to have the form

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ i\psi_1 \\ i\psi_2 \end{pmatrix}. \tag{1.7}$$

Using this solution, we find the energy–momentum tensor and apply it to the Einstein equations

$$R_{ij} = \frac{8\pi\kappa}{c^4} T_{ij}, \tag{1.8}$$

in Sec. III. Solving these equations gives $\nu$, $\lambda$, and $\mu$ as functions of $r$. It also restricts the metric to only one specific form which, we show, is the same as in static plane-symmetric spacetimes. For our notation we refer the reader to our previous paper.[1]

## II. SOLUTION OF THE DIRAC EQUATION AND THE ENERGY-MOMENTUM TENSOR

The nonzero spin coefficients $\Gamma_i$ are given by

$$\Gamma_2 = \tfrac{1}{2}\exp(\lambda-\nu)(1/r - \lambda_{,1})\gamma^1\gamma^2, \tag{2.1}$$

$$\Gamma_3 = \tfrac{1}{2}\exp(\lambda-\nu)(\lambda_{,1} + \mu_{,1})\gamma^1\gamma^3, \tag{2.2}$$

$$\Gamma_4 = \tfrac{1}{2}\exp(\lambda-\nu)(\lambda_{,1} - \nu_{,1})\gamma^1\gamma^4, \tag{2.3}$$

where the comma denotes differentiation with respect to $r$. The Dirac equation becomes

$$\psi_{,1} + \tfrac{1}{2}(1/r + \nu_{,1} + \mu_{,1} - \lambda_{,1})\psi = 0, \tag{2.4}$$

with the solution

$$\psi = \psi_0 (r/r_0)^{-1/2} \exp[-\tfrac{1}{2}(\nu + \mu - \lambda)], \tag{2.5}$$

where $\psi_0$ is a constant spinor and $r_0$ is an arbitrary constant.

The nonzero components of the energy–momentum tensor are

$$T_{23} = \frac{i\hbar c}{4}\exp(\lambda-\nu)(2\lambda_{,1} + \mu_{,1} - 1/r)\psi^* \gamma^5 \psi, \tag{2.6}$$

$$T_{24} = \frac{i\hbar c}{4}\exp(\lambda-\nu)(\nu_{,1} - 1/r)\psi^* \gamma^1\gamma^2 \psi, \tag{2.7}$$

$$T_{34} = \frac{i\hbar c}{4}\exp(\lambda-\nu)(\nu_{,1} - 2\lambda_{,1})\psi^* \gamma^1\gamma^3 \psi. \tag{2.8}$$

All other components vanish identically or via the Dirac equation,

## III. THE EINSTEIN EQUATIONS

For the static metric with cylindrical symmetry the only nonzero components of the Ricci tensor $R_{ij}$ are

$$R_{11} = -\exp[2(\lambda-\nu)](\nu_{,11} - \lambda_{,11} + 2\lambda_{,1}{}^2 - \lambda_{,1}/r - \nu_{,1}/r + 3\mu_{,1}\lambda_{,1} + \mu_{,1}{}^2 + \mu_{,11} - \nu_{,1}\mu_{,1}), \tag{3.1}$$

$$R_{22} = -\exp[2(\lambda-\nu)](-\lambda_{,1}/r + \mu_{,1}/r - \lambda_{,11} - \lambda_{,1}\mu_{,1}), \tag{3.2}$$

$$R_{33} = -\exp[2(\lambda-\nu)](\lambda_{,11} + \lambda_{,1}/r + \mu_{,11} + \mu_{,1}{}^2 + \lambda_{,1}\mu_{,1} + \mu_{,1}/r), \tag{3.3}$$

$$R_{44} = \exp[2(\lambda-\nu)](\nu_{,11} - \lambda_{,11} - \lambda_{,1}/r + \nu_{,1}/r + \nu_{,1}\mu_{,1} - \lambda_{,1}\mu_{,1}). \tag{3.4}$$

Since the diagonal components of the energy–momentum tensor all vanish, the Einstein equations yield

$$R_{ii} = 0 \quad (\text{no sum on } i). \tag{3.5}$$

These are the vacuum field equations which were first solved by Weyl and Levi-Civita.[2] We will use the solutions as given by Witten[3];

$$\mu = 0, \tag{3.6}$$

$$\nu = d^2 \ln(r/r_0), \tag{3.7}$$

$$\lambda = d \ln(r/r_0), \tag{3.8}$$

where $d$ and $r_0$ are arbitrary constants. The nonzero components of the energy—momentum tensor now become

$$T_{23} = \frac{i\hbar c}{4r}(r/r_0)^{d-d^2}(2d-1)\psi^*\gamma^5\psi, \tag{3.9}$$

$$T_{24} = \frac{i\hbar c}{4r}(r/r_0)^{d-d^2}(d^2-1)\psi^*\gamma^1\gamma^2\psi, \tag{3.10}$$

$$T_{34} = \frac{i\hbar c}{4r}(r/r_0)^{d-d^2}d(d-2)\psi^*\gamma^1\gamma^3\psi. \tag{3.11}$$

These must vanish since $R_{23}=R_{24}=R_{34}=0$, resulting in a nonzero neutrino wavefunction only when

$$d = \tfrac{1}{2}. \tag{3.12}$$

Setting $d=\frac{1}{2}$ causes $T_{23}$ to vanish, whereas the vanishing of $T_{24}$ and $T_{34}$ force the neutrino wavefunction to have the form

$$\psi = \begin{pmatrix} 1 \\ \pm 1 \\ i \\ \pm i \end{pmatrix}\psi_1, \tag{3.13}$$

where $\psi_1$ is a scalar.

Substitution of Eqs. (3.6), (3.7), (3.8), and (3.12) into the solution (2.5) yields

$$\psi = a(r/r_0)^{-3/8}\begin{pmatrix} 1 \\ \pm 1 \\ i \\ \pm i \end{pmatrix}, \tag{3.14}$$

where $a$ is an arbitrary complex constant.

The neutrino current density

$$s^k = i\psi^\dagger\gamma^k\psi, \tag{3.15}$$

is

$$s^k = 4|a|^2(r/r_0)^{-3/4}(\mp\delta^k{}_1 + \delta^k{}_4). \tag{3.16}$$

This corresponds to a flow of neutrinos in the radial direction.

Substituting $d=\frac{1}{2}$ in the metric, we find it takes the form

$$ds^2 = (r/r_0)^{-1/2}(dr^2 - dt^2) + (r/r_0)(r_0{}^2 d\phi^2 + dz^2). \tag{3.17}$$

The coordinate transformation

$$kx^1 + 1 = r/r_0, \tag{3.18}$$

$$x^2 = r_0\phi, \tag{3.19}$$

$$x^3 = z, \tag{3.20}$$

$$x^4 = t, \tag{3.21}$$

transforms the metric to

$$ds^2 = (kx^1+1)^{-1/2}[(dx^1)^2 - (dx^4)^2] + (kx^1+1)[(dx^2)^2 + (dx^3)^2], \tag{3.22}$$

where we have set

$$k = 1/r_0.$$

The wavefunction is transformed to

$$\psi = a(kx^1+1)^{-3/8}\begin{pmatrix} 1 \\ \pm 1 \\ i \\ \pm i \end{pmatrix}. \tag{3.23}$$

Thus, we have exactly the same metric and wavefunction as in the static plane-symmetric case discussed in a previous paper.[1] We refer the reader to Ref. 1 for a discussion of this "ghost neutrino" solution together with a comparison with other solutions to the Einstein—Dirac equations.

## IV. CONCLUSIONS

We have presented an exact solution to the Einstein—Dirac equations for static cylindrically-symmetric spacetimes. The only neutrinos allowed have the "ghost" property of a vanishing energy—momentum tensor. Not only are these neutrinos "ghost" neutrinos, but they exist only in static cylindrically-symmetric spacetimes which are locally equivalent to static plane-symmetric spacetimes.

It has been proven by Madore[4] that in static, axially symmetric spacetimes the neutrino energy—momentum tensor vanishes. However, it was not pointed out that the wavefunction and current do not vanish. Griffiths[5] has shown that all Einstein—Dirac solutions which have a vanishing energy—momentum tensor are of Petrov type D or N. The spacetime presented here is type D. The general form of the metric for ghost neutrinos has been given by Collinson and Morris.[6] We have shown that our plane-symmetric ghost neutrino solution is equivalent to the type D solution obtained by Collinson and Morris.[7]


[1]T.M. Davis and J.R. Ray, Phys. Rev. D **9**, 331 (1974).
[2]H. Weyl, Ann. Physik **54**, 117 (1917) and T. Levi-Civita, Rend. Acc. Lincei **29**, 307 (1917).
[3]L. Witten, in *Colloquium on the Theory of Relativity*, Centre de Recherches Mathematiques, Free University (Universitaire, Louvain, Belgium, 1960), p. 59.
[4]J. Madore, Lett. Nuovo Cimento **5**, 48 (1972).
[5]J.B. Griffiths, Commun. Math. Phys. **28**, 295 (1972).
[6]C.D. Collinson and P.B. Morris, J. Phys. A **6**, 915 (1973).
[7]T.M. Davis and J.R. Ray, J. Math. Phys. 16, 75 (1975).

# Canonical transformation and accidental degeneracy. III. A unified approach to the problem

M. Moshinsky*, J. Patera†, and P. Winternitz†

*Centre de Recherches Mathématiques, Université de Montréal, Québec, Canada*
(Received 28 May 1974)

We continue the discussion of the groups of canonical transformations responsible for accidental degeneracy in quantum mechanical problems. A general unified treatment is provided for a wide class of two-dimensional physical systems, having an energy spectrum which is a linear combination of two quantum numbers. The general method involves the use of both nonorthonormal and orthonormal sets of states to construct groups of complex or real canonical transformations, mapping the problem under consideration onto the two-dimensional isotropic harmonic oscillator. The group responsible for the accidental degeneracy is then quite obviously $SU(2)$. The problem of an isotropic oscillator in a sector $\pi/q$ ($q$ integer) was discussed previously using a nonorthonormal basis. In the present paper we carry the analysis in an orthonormal basis to establish the general procedure mentioned above. We also analyze in detail the Calogero problem for three particles which has a spectrum of the type given above, and obtain explicitly the canonical transformation that maps it on the anisotropic oscillator whose ratio of frequencies is 2/3 and subsequently on the isotropic one.

## 1. INTRODUCTION AND SUMMARY

The purpose of this article is to continue the systematic study of physical systems (both quantum mechanical and classical) with "hidden" or "dynamical" symmetries that was initiated in the two previous articles of this series.[1,2] In particular we wish to provide a unified and general approach to the analysis of all such systems.

Specifically in this paper we consider nonrelativistic systems, described by a Schrödinger equation or a classical equation of motion, with a local potential $V(x_1, x_2)$ that is time and energy independent and which is a function of two variables only, with an energy spectrum of the type

$$E_{nN} = C(k_2 n + k_1 N) + D. \tag{1.1}$$

Here $C$ and $D$ are arbitrary real constants, $k_1$ and $k_2$ relatively prime integers, and $n$ and $N$ arbitrary integers. This type spectrum includes all two dimensional physical systems with accidental degeneracy that have so far occurred in the literature.[1-9]

As shown in the previous papers of this series[1,2] this set of energy levels can be split into subsets that show explicit accidental degeneracy. For this we only need to write

$$n = k_1 n_1 + \lambda_1, \quad N = k_2 n_2 + \lambda_2, \quad \lambda_i = 0, 1 \cdots k_i - 1, \quad i = 1, 2, \tag{1.2}$$

and, substituting in (1.1), we get

$$E_{n_1 n_2}^{(\lambda_1 \lambda_2)} = C k_1 k_2 (n_1 + n_2) + C(k_2 \lambda_1 + k_1 \lambda_2) + D. \tag{1.3}$$

Thus the $k_1 k_2$ subsets characterized by the pair of numbers $(\lambda_1 \lambda_2)$ have accidental degeneracy of the familiar type that we associate with the two-dimensional isotropic oscillator.[1]

We shall argue that for any such system it is possible to construct an algebra of invariants of motion transforming wave functions corresponding to a definite energy level irreducibly among themselves. What is more, it is possible to construct the group of canonical transformations, generated by the above "dynamical invariance algebra", through the mapping of the problem under consideration onto the two-dimensional isotropic harmonic oscillator. The group responsible for the accidental degeneracy and other interesting features of the system is then simply the $SU(2)$ group of the isotropic harmonic oscillator.

More generally, all known physical systems showing accidental degeneracy[1-9] (the hydrogen atom,[3] the isotropic harmonic oscillator,[4] the anisotropic harmonic oscillator,[1,4] the harmonic oscillator in a sector,[2] the linear three-body problem with two-body potentials proportional to the square and inverse square of the distance, i.e., the Calogero problem[5] and many others[7]) have several features in common. Among these features, which account for the physical interest of these problems, we have the following:

1. The energy spectrum for the quantum mechanical problem demonstrates "accidental" degeneracy, i.e., a degeneracy not associated with any obvious geometrical symmetry group.

2. The corresponding classical motion is nonergodic; in particular, all finite trajectories are closed.

3. The Schrödinger equation can be solved explicitly and analytically in terms of known functions and so can the corresponding classical equations of motion. All relevant partial differential equations allow the separation of variables in at least one coordinate system.

4. A dynamical invariance algebra can be constructed, i.e., a Lie algebra of operators, commuting with the Hamiltonian, such that the wavefunctions of the system corresponding to a given energy level transform among each other according to irreducible unitary representations of this algebra. The operators forming a geometrical invariance algebra are of first order in the derivatives[7] (first power in the momenta); those of a dynamical algebra include higher order derivatives (powers of the momenta larger than one).

5. The invariance algebra can be extended to a "dynamical noninvariance algebra," containing in addition to the invariance algebra further operators, acting as raising and lowering operators for the energy, i.e.,

transforming wavefunctions corresponding to one energy into those corresponding to a different energy.

6. It is possible to construct a group of canonical transformations, generated by the dynamical invariance algebra, which is directly responsible for the accidental degeneracy and other features of the problem. Point transformation groups, involving particle coordinates only, are generated by the geometric invariance algebra (first order derivatives in the operators). Canonical transformation groups, in which the new coordinates and momenta are functions of both the old coordinates, and momenta preserving the commutation relations (Poisson brackets) are generated by the dynamical invariance algebra (second and higher order derivatives in the operators).

A unified and systematic approach to accidental degeneracy problems should, according to our opinion, involve two aspects:

I. Provide a general method for finding physical systems with some or all of the properties listed above.

II. Provide a general method for constructing the dynamical invariance algebra of the problem and the group of canonical transformations responsible for the accidental degeneracy, once the system itself has been found and the degeneracy of the energy levels established.

The first aspect was considered in Refs. 7 where all two-dimensional potentials $V(x_1, x_2)$ were found for which the Schrödinger equation allows an invariance algebra of operators that are at most quadratic in the momenta. The same problem has also been studied[8] for three-dimensional potentials $V(x_1, x_2, x_3)$.

The present series of articles is mainly devoted to the second aspect of the problem, namely the construction of the group of canonical transformations explaining the accidental degeneracy. In this paper we illustrate the general method of tackling the problem through the analysis of the Hamiltonian introduced by Calogero.[6] As a first step in this analysis we again discuss the canonical transformation that maps the anisotropic oscillator whose ratio of frequencies is rational on an isotropic oscillator. This problem was discussed in Ref. 1, but in Sec. 2 of the present paper we perform the analysis using a nonnormal set of states. This simplifies the procedure for obtaining the canonical transformation mentioned above.

Once the previous point is achieved, we can tackle the problem of finding the canonical transformation that maps any physical system whose spectrum is given by (1.1) into an anisotropic oscillator whose ratio of frequencies is $k_2/k_1$. The procedure involves the following steps:

(i) Find the ground state solution $|0)$ of the Schrödinge equation.

(ii) Construct two independent raising operators $a^+, A^+$ such that a complete set of eigenstates of the Hamiltonian can be written as

$$|nN) = (n!N!)^{-1/2}(a^+)^n(A^+)^N|0). \tag{1.4}$$

In general this set of states will be nonorthonormal as seen for example in the sector problem discussed in Ref. 2.

(iii) Construct the lowering operators $\tilde{a}$, $\tilde{A}$ which will be canonically conjugate to $a^+$, $A^+$, i.e.,

$$[\tilde{a}, a^+] = [\tilde{A}, A^+] = 1, \quad [\tilde{a}, A^+] = [\tilde{A}, a^+] = 0, \tag{1.5}$$

but, in general, not Hermitian conjugate. From (1.4), (1.5) it follows that

$$a^+|nN) = (n+1)^{1/2}|n+1\ N), \tag{1.6a}$$

$$A^+|nN) = (N+1)^{1/2}|n\ N+1), \tag{1.6b}$$

$$\tilde{a}|nN) = n^{1/2}|n-1\ N), \tag{1.6c}$$

$$\tilde{A}|nN) = N^{1/2}|n\ N-1), \tag{1.6d}$$

and thus the Hamiltonians of the system whose spectrum is given by (1.1) is proportional to

$$k_1 a^+\tilde{a} + k_2 A^+\tilde{A} + (D/C), \tag{1.7}$$

which, in terms of the variables $a^+$, $A^+$, $\tilde{a}$, $\tilde{A}$ satisfying (1.5), implies that we have an anisotropic oscillator whose ratio of frequencies is $(k_2/k_1)$.

(iv) Introduce the usual definition of raising and lowering operators in terms of coordinates and momenta.[1] We have then the canonical transformation that maps the physical system whose spectrum is given by (1.1) into an anisotropic oscillator whose ratio of frequencies is $(k_2/k_1)$.

Steps (i)—(iv) where actually followed in Ref. 2 for the problem of anisotropic oscillator in the plane restricted to move in a sector of angle $\pi/q$ with $q$ integer: As indicated in that reference, since $\tilde{a}$, $\tilde{A}$ are not the Hermitian conjugates of $a^+$, $A^+$, the canonical transformations mentioned in (iv) are in general complex. This raises the question of how the classical orbits would transform under them. This problem, together with others,[2] suggests the importance of carrying out the analysis outlined in steps (i)—(iv) also for an orthonormal basis, and this requires further steps that we proceed to enumerate.

(v) Determine the orthonormal basis of the problem which depends on two quantum numbers which we still designate by $n, N$. The corresponding states will be denoted by the angular ket $|nN\rangle$ rather than by the round one $|nN)$ of the nonorthonormal basis (1.4). These orthonormal states will be eigenfunctions of the Hamiltonian $H$ and the other integral of motion of the problem which we designate by $M^2$ as its spectrum turns out to be positive in the examples to be discussed below.

(vi) Apply to the orthonormal states $|nN\rangle$ the operators $a^+, A^+$. We expect in general that we get linear combinations of these states corresponding to the raised energy rather than the simple expressions (1.6a, b). We proceed to show though that in many cases we can find new operators, which we designate by $\hat{a}^+$, $\hat{A}^+$, which are functions of $a^+$, $A^+$, $H$, $M^2$ and their commutators, that when applied to the orthonormal basis $|nN\rangle$ behave as raising operators in the sense (1.6a, b).

(vii) As we are now dealing with orthonormal basis the lowering operators $\hat{a}$, $\hat{A}$ are then just the Hermitian conjugate of $\hat{a}^+$, $\hat{A}^+$ and at the same time they are canonically conjugate. The Hamiltonian of the system whose energy spectrum is given by (1. 1) becomes then proportional to

$$k_1 \hat{a}^+\hat{a} + k_2 \hat{A}^+\hat{A} + (D/C), \tag{1.8}$$

which again indicates that we are dealing with an anisotropic oscillator whose ratio of frequencies is $(k_2/k_1)$.

(viii) Introducing for $\hat{a}^+$, $\hat{A}^+$, $\hat{a}$, $\hat{A}$ the usual definition of raising and lowering operators in terms of coordinate and momenta, [1] we would have then the *real* canonical transformation that maps the physical system whose spectrum is given by (1. 1) into an anisotropic oscillator whose ratio of frequencies is $(k_2/k_1)$

(ix) Once we achieved the step indicated in (viii) we can use the analysis developed in Ref. 1 to map the Hamiltonian (1. 8) onto an isotropic harmonic oscillator. Thus the dynamical symmetry group of a physical system whose spectrum is given by (1. 1) is then a certain realization of $SU(2)$.

In Sec. 3 of the present paper we implement steps (v)—(ix) for the case of a particle in an isotropic two-dimensional harmonic oscillator potential constrained to move in a sector of angle $\pi/q$ with $q$ integer. Combining this with the discussion of Ref. 2, we see that the full set of steps (i)—(ix) has been implemented for the sector problem. We are then in a position to extend the analysis to the Calogero[5] problem. Steps (i)—(iv) can be carried out using the nonorthonormal basis introduced by Perelomov. [6] Steps (v)—(ix) are performed by a procedure entirely parallel to the one used for the sector problem.

Once step (ix) is also implemented, we can explicitly construct the generators of the Lie algebra of the $SU(2)$ symmetry group of the problem. They will be of course complicated functions of operators such as $H$, $M^2$, but this is not important as the latter are diagonal in the basis $|nN\rangle$. Thus the generators of $SU(2)$ must be understood in the weak sense, i. e., as operators that are well defined only when acting on the basis $|nN\rangle$.

We proceed then to implement the above analysis starting with the discussion of the anisotropic oscillator in the nonnormal basis.

## 2. THE ANISOTROPIC OSCILLATOR IN A NONNORMAL BASIS

Before analyzing in a nonnormal basis the anisotropic oscillator whose ratio of frequencies is

$$(\omega_1/\omega_2) = (k_2/k_1), \tag{2.1}$$

where $k_1$, $k_2$ are two relatively prime integers, we continue the discussion of Ref. 1 to obtain information on action and angle variables.

As was shown previously[1] the canonical transformations that map the anisotropic oscillator in to an isotropic one affect independently the coordinates and momenta $(x_1, p_1)$ and $(x_2, p_2)$ of the two degrees of freedom. We can therefore suppress the index $i = 1, 2$ associated with them and discuss the problem of a particle in a one-dimensional oscillator whose mass is unity and whose frequency[1] is $k^{-1}$ with $k$ integer, i. e., the Hamiltonian

$$H = \tfrac{1}{2}(p^2 + k^{-2}x^2). \tag{2.2}$$

The action and angle variables for this problem[10] are

$$J = kH = \tfrac{1}{2}(kp^2 + k^{-1}x^2), \qquad w = -\text{ang}\tan(kp/x), \tag{2.3}$$

which clearly are canonically conjugate as the Poisson bracket is unity, i. e.,

$$\{w, J\} = \frac{\partial w}{\partial x}\frac{\partial J}{\partial p} - \frac{\partial w}{\partial p}\frac{\partial J}{\partial x} = 1. \tag{2.4}$$

We differ from the standard literature[10] by suppressing a factor $2\pi$ in $J$ and a corresponding factor $(2\pi)^{-1}$ in $w$.

By introducing the usual definition of creation and annihilation variables

$$\eta = (1/\sqrt{2})(k^{-1/2}x - ik^{1/2}p), \quad \xi = (1/\sqrt{2})(k^{-1/2}x + ik^{1/2}p), \tag{2.5}$$

the action-angle variables take the form

$$J = \eta\xi, \tag{2.6a}$$

$$w = -(i/2)\ln(\eta/\xi), \tag{2.6b}$$

and, inverting this last relation, we obtain

$$\eta = J^{1/2}\exp(iw), \tag{2.7a}$$

$$\xi = J^{1/2}\exp(-iw). \tag{2.7b}$$

We now consider the canonical transformation of Ref. 1 that maps an oscillator of frequency $k^{-1}$ into an oscillator of unit frequency. In terms of the annihilation and creation variables it takes the form[1]

$$\eta' = k^{-1/2}\eta^{(1+k)/2}\xi^{(1-k)/2}, \tag{2.8a}$$

$$\xi' = k^{-1/2}\eta^{(1-k)/2}\xi^{(1+k)/2}. \tag{2.8b}$$

The new action-angle variables given by

$$J' = \eta'\xi', \tag{2.9a}$$

$$w' = -(i/2)\ln(\eta'/\xi'), \tag{2.9b}$$

are related to the old ones of (2. 6) in the simple fashion

$$J' = k^{-1}J, \tag{2.10a}$$

$$w' = kw. \tag{2.10b}$$

The canonical transformation (2. 8), when written in the approrpiate quantum form, [1] provides the creation and annihilation operators that act on the subsets of normalized states

$$|n\rangle_\lambda \equiv [nk+\lambda)!]^{-1/2}\eta^{nk+\lambda}|0\rangle, \tag{2.11}$$

where $|0\rangle$ is the ground state of the oscillator and $\lambda$, which characterizes the subset, [1] takes the values $\lambda = 0, 1, 2, \cdots, k-1$.

As shown in Ref. 1 the operator form of $\eta'$, $\xi'$ corresponding to (2. 8), when acting on the states $|n\rangle_\lambda$, gives

$$\eta'|n\rangle_\lambda = (n+1)^{1/2}|n+1\rangle_\lambda, \tag{2.12a}$$

$$\xi'|n\rangle_\lambda = n^{1/2}|n-1\rangle_\lambda. \tag{2.12b}$$

Instead of the normalized set of states (2. 11) we could have taken the nonnormalized one defined by

$$|n)_\lambda \equiv (n!)^{-1/2}\eta^{nk+\lambda}|0\rangle, \tag{2.13}$$

which we denote by a round ket to distinguish them from the angular ket (2. 11). For these states it is immediately clear that we could take as creation operator

$$\eta'' = \eta^k \tag{2.14a}$$

since

$$\eta'' |n\rangle_\lambda = (n+1)^{1/2} |n+1\rangle_\lambda. \tag{2.15}$$

What is the corresponding annihilation operator $\xi''$? From the commutation rules $[\xi'', \eta''] = 1$, $[\xi, \eta] = 1$ it suggests itself that we put $\xi'' = \partial/\partial\eta''$, $\xi = \partial/\partial\eta$, and thus

$$\xi'' = \frac{\partial}{\partial\eta''} = \left(\frac{d\eta''}{d\eta}\right)^{-1} \frac{\partial}{\partial\eta} = k^{-1}\eta^{1-k}\xi. \tag{2.14b}$$

We easily check that the classical Poisson bracket of $\eta''$ and $\xi''$ is $\{\eta'', \xi''\} = i$ and thus $\xi''$ is the annihilation variable[1] corresponding to $\eta''$. The quantum mechanical operator[1] has a different form for each subset of nonnormalized states (2.13) characterized by $\lambda$. In fact, by a similar reasoning to the one carried out in Ref. 1 for the normalized states, we see that the quantum mechanical operator corresponding to (2.14b) has the form

$$\xi'' = k^{-1}\eta^{-k}(\eta\xi - \lambda), \tag{2.16}$$

as for each subset $\lambda = 0, 1, \cdots, k-1$ we have

$$\xi'' |n\rangle_\lambda = (n)^{1/2} |n-1\rangle_\lambda. \tag{2.17}$$

The analysis of the last few paragraphs indicates that it is quite simple to find the canonical transformation (2.14) that provides, when we pass to the quantum picture, the raising and lowering operators that act on the subsets of nonnormalized states (2.13). There is, though, one serious drawback. While $\eta'$, $\xi'$ are conjugate to each other in the complex variable sense, i.e., from (2.8) we see that

$$\xi' = \eta'^{*}, \tag{2.18}$$

this is *not* the case for $\eta''$, $\xi''$. Thus we have for the nonnormalized set of states (2.13) a canonical transformation (2.14) that is no longer real,[1] i.e., had we defined $x''$, $p''$ by the relations

$$\eta'' = (1/\sqrt{2})\,(x'' - ip''), \tag{2.19a}$$

$$\xi'' = (1/\sqrt{2})\,(x'' + ip''), \tag{2.19b}$$

corresponding to an oscillator of frequency 1, $x''$ and $p''$ would not be real functions of $x, p$ though their Poisson bracket is still 1 and thus they are canonically conjugate.

This puzzle appeared already in the sector problem of Ref. 2 where when acting on a nonorthonormal basis we were also led to a complex canonical tranformation.

We can solve the puzzle in the present one-dimensional oscillator problem in a simple way. Rather than defining $x''$, $p''$ through the relations (2.19), we go first to the action and angle variable associated with $\eta''$, $\xi''$, i.e.,

$$J'' = \eta''\xi'', \tag{2.20a}$$

$$w'' = -(i/2)\ln(\eta''/\xi''). \tag{2.20b}$$

From (2.14a, b) we see that

$$J'' = k^{-1}\eta\xi = k^{-1}J, \tag{2.21a}$$

$$\begin{aligned} w'' &= -(i/2)\ln[k(\eta/\xi)^k(\eta\xi)^{k-1}] \\ &= kw - (i/2)\ln(kJ^{k-1}). \end{aligned} \tag{2.21b}$$

As the original action-angle variables $J$, $w$ are real, we see that $J''$ is real while $w''$ is complex. But the imaginary part of $w''$ is a function of $J$ alone and thus its Poisson bracket with $J''$, which is proportional to $J$, is then zero. Thus while $w''$ is canonically conjugate to $J''$, it is also clear that if we just take the real part of $w''$, i.e., $kw$, this would also be an angle variable corresponding to the action variable $J''$. But from (2.10) we see that

$$J'' = J', \tag{2.22a}$$

$$\mathrm{Re}\,w'' = w', \tag{2.22b}$$

and thus, if we had restricted ourselves only to the real part of $w''$, we would have the action and angle variables associated with an orthonormal set of states from which, by a formula similar to (2.7), we would have obtained $\eta'$, $\xi'$ which are complex conjugate to each other. Finally, by defining $x'$, $p'$ by an expression of the type (2.19), they would be *real* functions of $x, p$ that are canonically conjugate and represent the canonical transformation that maps the oscillator of frequency $k^{-1}$ into an oscillator of unit frequency.

We have thus shown how to derive the real canonical transformations required by our problem by first obtaining the complex ones which are associated with a nonnormalized basis.

We will consider also another procedure, related to the quantum picture, of deriving $\eta'$, $\xi'$ once we know $\eta''$, $\xi''$. As we mentioned in (2.15) $\eta''$, when acting on the nonnormalized subsets of states $|n)_\lambda$ behaves as a creation operator. What happens when we apply $\eta''$ to the normalized set of states $|\lambda\rangle_\lambda$? From (2.11) and (2.14a) we see that

$$\eta'' |n\rangle_\lambda = \{[(n+1)k+\lambda]!/(nk+\lambda)!\}^{1/2} |n+1\rangle_\lambda. \tag{2.23}$$

The operator $\eta\xi$ when applied to the state (2.23) gives

$$\begin{aligned} (\eta\xi)\eta'' |n\rangle_\lambda &= \eta\xi\{[(nk+\lambda)!]^{-1/2}\eta^{[(n+1)k+\lambda]}|0\rangle\} \\ &= [(n+1)k+\lambda]\eta'' |n\rangle_\lambda, \end{aligned} \tag{2.24}$$

and thus from (2.23) we can write

$$\begin{aligned} &[(\eta\xi)(\eta\xi-1)\cdots(\eta\xi-k+1)]^{-1/2}k^{-1/2}(\eta\xi-\lambda)^{1/2}\eta^k|n\rangle_\lambda \\ &\quad = (n+1)^{1/2} |n+1\rangle_\lambda, \end{aligned} \tag{2.25}$$

where the operator in the left-hand side acting on $|n\rangle_\lambda$ is exactly the raising operator (4.2a) of Ref. 1. To pass now to the classical picture, we use the correspondence principle. As $\eta\xi$ is the number operator, we consider only cases in which $\eta\xi$ are large so that all the integers in the operator on the left-hand side of (2.25) can be disregarded as compared to $\eta\xi$. We have then that the operator becomes the dynamical variable

$$k^{-1/2}(\eta\xi)^{-(k-1)/2}\eta^k = k^{-1/2}\eta^{(1+k)/2}\xi^{(1-k)/2}, \tag{2.26}$$

which coincides with $\eta'$ of (2.8a). For $\xi'$ we just need to consider the relation $\xi' = \eta'^{*}$ to obtain (2.8b).

Thus, by applying to the normalized basis the raising operator of the nonnormalized one, we can obtain by the above reasoning the raising operator of the normalized basis. Using then the correspondence principle, we can obtain the classical canonical transformation that maps the harmonic oscillator of frequency $k^{-1}$ into another one of frequency 1.

Returning now to the two-dimensional anisotropic oscillator whose ratio of frequencies is rational, all we have to do is to add to the variables appearing in the equations $x$, $p$, $\eta$, $\xi$, etc. an index $i=1,2$ that corresponds to the two degrees of freedom. In this way we have the normalized and nonnormalized states

$$|n_1 n_2\rangle_{\lambda_1\lambda_2} = |n_1\rangle_{\lambda_1} |n_2\rangle_{\lambda_2}, \tag{2.27a}$$

$$|n_1 n_2)_{\lambda_1\lambda_2} = |n_1)_{\lambda_1} |n_2)_{\lambda_2}. \tag{2.27b}$$

The generators of the $U(2)$ group that connect all states in these two basis are respectively

$$\eta'_i \xi'_j, \quad i,j=1,2, \tag{2.28a}$$

$$\eta''_i \xi''_j, \tag{2.28b}$$

where $\eta'_i$, $\xi'_j$ are the quantum mechanical operators given in Ref. 1 while $\eta''_i$ and $\xi''_j$ are given respectively by (2.14a) and (2.16), where an index $i$ or $j$ has to be added to all variables.

If by the generators of the Lie algebra of the group responsible for accidental degeneracy we understand those operators that connect all states of the same energy, it is irrelevant whether we work in a normalized, or more generally orthonormalized basis, or if it does not have this property so long as it is complete. But when we look at the problem in the classical picture, we would like to have real canonical transformations whose effect on the orbits in phase space we understand. Thus it is convenient that if we originally obtained our results on a nonorthonormalized basis because it turned out to be simpler, we also derive later the results for the orthonormalized one for which the canonical transformation is real.

In Ref. 2 we discussed completely the symmetry group of canonical transformations for the problem of an isotropic oscillator in a sector of an angle $\pi/q$ with $q$ integer, for a nonorthonormalized but complete set of states. In the next section we shall discuss the same problem when the basis is orthonormalized and in the process establish a procedure which seems applicable to all problems whose spectrum has the form (1.1).

## 3. THE PROBLEM OF AN OSCILLATOR IN A SECTOR WHEN THE BASIS IS ORTHONORMAL

As in Ref. 2, we shall express the orthonormal and nonorthonormal states in terms of powers of creation operators acting on the ground state. As shown there, these operators take the form

$$\eta_\pm = (1/\sqrt{2})(x_\pm - ip_\pm), \tag{3.1}$$

where

$$x_\pm = (1/\sqrt{2})(x_1 \pm ix_2), \tag{3.2a}$$

$$p_\pm = (1/\sqrt{2})(p_1 \pm ip_2). \tag{3.2b}$$

The orthonormal states are then given by the angular ket[2]

$$|nN\rangle = \{n![n+q(N+1)]!\}^{-1/2}\, 2^{-1/2}(\eta_+\eta_-)^n[\eta_+^{q(N+1)} - \eta_-^{q(N+1)}]\,|0\rangle, \tag{3.3}$$

while the nonorthonormal will be designated by the round ket[2]

$$|nN) = [n!N!]^{-1/2}(\eta_+\eta_-)^n(\eta_+{}^q + \eta_-{}^q)^N(\eta_+{}^q - \eta_-{}^q)\,|0\rangle. \tag{3.4}$$

These states were taken from Ref. 2, but for convenience in the later developments we replaced $\nu_1$, $\nu_2$ appearing there by $n, N$.

If we define now the operators

$$a^+ \equiv \eta_+\eta_-, \tag{3.5a}$$

$$A^+ \equiv \eta_+{}^q + \eta_-{}^q, \tag{3.5b}$$

we immediately see from (3.4) that

$$a^+|nN) = (n+1)^{1/2}|n+1\ N), \tag{3.6a}$$

$$A^+|nN) = (N+1)^{1/2}|n\ N+1), \tag{3.6b}$$

and thus $a^+, A^+$ are the creation operators for the nonorthonormal basis.

As discussed in Ref. 2 (where we denoted $a^+$ by $\eta_1$ and $A^+$ by $\eta_2$), the annihilation operators for the nonorthonormal basis are *not* the Hermitian conjugate of the creation operators. Designating these annihilation operators by $\tilde{a}$, $\tilde{A}$ we can symbolically[2] express them as

$$\tilde{a} = \frac{\partial}{\partial a^+}, \quad \tilde{A} = \frac{\partial}{\partial A^+}, \tag{3.7}$$

and thus, using (3.5), we obtained[2]

$$\tilde{a} = (\eta_+^{q-1}\xi_- - \eta_-^{q-1}\xi_+)(\eta_+^q - \eta_-^q)^{-1}, \tag{3.8a}$$

$$\tilde{A} = (\eta_+\xi_+ - \eta_-\xi_-)q^{-1}(\eta_+^q - \eta_-^q)^{-1}, \tag{3.8b}$$

where we employed the symbolic relation

$$\frac{\partial}{\partial\eta_\pm} = \xi_\pm = \frac{1}{\sqrt{2}}(x_\mp + ip_\mp). \tag{3.9}$$

From (3.4) and (3.8) we immediately see that

$$\tilde{a}\,|nN) = n^{1/2}|n-1\ N), \tag{3.10a}$$

$$\tilde{A}\,|nN) = N^{1/2}|n\ N-1). \tag{3.10b}$$

Thus creation and annihilation operators for the nonorthonormal basis can be derived straightforwardly. We propose now to obtain the corresponding operators for the orthonormal basis which we shall designate by a caret above the symbol, i.e., $\hat{a}^+, \hat{A}^+, \hat{a}, \hat{A}$.

We begin by applying to the orthonormal basis the Hamiltonian and angular momentum of a particle in a two-dimensional isotropic harmonic oscillator potential. The operator form of these observables is given by

$$H = \eta_+\xi_+ + \eta_-\xi_-, \tag{3.11}$$

$$M = \eta_+\xi_+ - \eta_-\xi_-. \tag{3.12}$$

The orthonormal state (3.3) is an eigenstate of $H$ with eigenvalue

$$2n + q(N+1). \tag{3.13}$$

On the other hand it is *not* an eigenstate of $M$ because it satisfies the boundary conditions that require that the wave function vanishes at $\varphi = 0$, $\pi/q$. The classical angular momentum is given at the boundaries by the product of the radius vector by the component of the velocity normal to the boundary. After collision this component changes sign but not magnitude so that $M$ goes into $-M$. Thus classically $M^2$ would be an integral of motion of the problem and therefore we expect that $|nN\rangle$ of (3.3) should be an eigenstate of $M^2$. This we can check directly and find that its eigenvalue is given by

$$q^2(N+1)^2. \tag{3.14a}$$

We now introduce an operator which we call absolute value of the angular momentum and denote by $|M|$. It is defined by the fact that its eigenfunctions are still given by (3.3) but its eigenvalue is the positive square root of (3.14), i.e.,

$$q(N+1). \tag{3.14b}$$

Symbolically this means that

$$|M| = [M^2]^{1/2}. \tag{3.15}$$

In the classical picture this implies that

$$|M| = \begin{cases} \eta_+\xi_+ - \eta_-\xi_- & \text{when } M > 0, \quad (3.16a) \\ \eta_-\xi_- - \eta_+\xi_+ & \text{when } M < 0. \quad (3.16b) \end{cases}$$

We now proceed to apply to the state $|nN\rangle$ of (3.3) the operators $a^+$, $A^+$ of (3.5). We begin by noting that

$$a^+|nN\rangle = \left(\frac{(n+1)![(n+1)+q(N+1)]!}{n![n+q(N+1)]!}\right)^{1/2} |n+1, N\rangle$$
$$= [(n+1)+q(N+1)]^{1/2}(n+1)^{1/2}|n+1, N\rangle. \tag{3.17}$$

Thus $[(n+1)+q(N+1)]^{-1/2}a^+$ acts as a creation operator on the orthonormal basis with respect to the index $n$. Making now use of the spectra (3.13), (3.14b) of the operators $H$, $|M|$, we can write $\hat{a}^+$ as

$$\hat{a}^+ = [\tfrac{1}{2}(H+|M|)]^{-1/2}a^+, \tag{3.18}$$

where from (3.17) we immediately see that

$$\hat{a}^+|nN\rangle = (n+1)^{1/2}|n+1, N\rangle. \tag{3.19}$$

From the discussion of the previous section we see that as the basis $|nN\rangle$ is orthonormal the corresponding annihilation operator is just the Hermitian conjugate of (3.19), i.e.,

$$\hat{a} = a[\tfrac{1}{2}(H+|M|)]^{-1/2}, \tag{3.20a}$$

$$a = (a^+)^+ = \xi_+\xi_-. \tag{3.20b}$$

It remains now only to determine $\hat{A}^+$ as $\hat{A}$ is again just its Hermitian conjugate. We start by applying the operator $A^+$ of (3.5b) to the state (3.3) which gives

$$A^+|nN\rangle = \mu_{nN}|n\,N+1\rangle + \nu_{nN}|n+q, N-1\rangle, \tag{3.21}$$

where $\mu_{nN}$, $\nu_{nN}$ are two constants given by

$$\mu_{nN} = \{[n+q(N+2)]!/[n+q(N+1)]!\}^{1/2},$$
$$\nu_{nN} = [(n+q)!/n!]^{1/2}. \tag{3.22}$$

Comparing (3.22) with (3.17), we see that now the application of the operator $A^+$ to $|nN\rangle$ does not give a single state, as happened when we applied $a^+$, but a linear combination of two of them corresponding to the same energy. Thus we would like to have another operator that gives again a linear combination of the type (3.21) which we could use together with $A^+$ to obtain an operator that when acting on $|nN\rangle$ gives only a state $|n\,N+1\rangle$. This is very easy to find as $|nN\rangle$ is an eigenstate of $M^2$ and thus the commutator

$$[M^2, A^+] \tag{3.23}$$

when applied to $|nN\rangle$ gives a result similar to (3.21) with other values of $\mu_{nN}$, $\nu_{nN}$. Combining then $A^+$ with this commutator, we easily see that

$$\{q^2(2N+1)A^+ + [M^2, A^+]\}|nN\rangle$$
$$= 4q^2(N+1)\{[n+q(N+2)]!/[n+q(N+1)]!\}^{1/2}|n\,N+1\rangle. \tag{3.24}$$

If the multiplicative factor on the right-hand side of (3.24), divided by $(N+1)^{1/2}$, is passed to the left-hand side, and if we make use of the eigenvalues of $H$, $|M|$ given by (3.13), (3.14b), we obtain for the creation operator $\hat{A}^+$ in the orthonormal basis the explicit form

$$\hat{A}^+ = \tfrac{1}{4}q^{-3/2}$$
$$\times [2^{-q}(H+|M|)(H+|M|-2)\cdots(H+|M|-2q+2)]^{-1/2}$$
$$\times(|M|-q)^{-1/2}\{qA^+(2|M|-q) + [M^2, A^+]\}. \tag{3.25}$$

The corresponding annihilation operator $\hat{A}$ is obtained by taking the Hermitian conjugate of $\hat{A}^+$. As $H$, $|M|$, $M^2$ are Hermitian, $\hat{A}$ is given by (3.25) when we invert the order of the factors and replace $A^+$ by

$$A = \xi_+^q + \xi_-^q. \tag{3.26}$$

We have thus obtained explicitly the raising and lowering operators in the orthonormal basis. They are considerably more complicated than the corresponding operators in the nonorthonormal basis, but have the advantage that when we pass to the classical limit they will give rise to *real* rather than *complex* canonical transformations.

As discussed in the previous section, the classical limit can be obtained with the help of the correspondence principle, which in the present case implies that the eigenvalues of the operators $H$, $|M|$ are much larger than 1. This implies that the operators of the type $H+|M|+\gamma$, where $\gamma$ is any integer reduce to $H+|M|$. Furthermore, from (3.16) we see that in the classical limit

$$\tfrac{1}{2}(H+|M|) \to \begin{cases} \eta_+\xi_+ & \text{when } M > 0, \\ \eta_-\xi_- & \text{when } M < 0. \end{cases} \tag{3.27}$$

We note also that the commutator

$$[M^2, A^+] = M[M, A^+] + [M, A^+]M = q[M(\eta_+^q - \eta_-^q) + (\eta_+^q - \eta_-^q)M]. \tag{3.28}$$

In the classical limit $M$ and $\eta_+^q - \eta_-^q$ commute and the commutator reduces to twice the first term of the right-hand side of (3.28).

Taking all the previous observations into account, we have the following expressions for the $\hat{a}^+, \hat{a}, \hat{A}^+, \hat{A}$ in the classical limit. When the angular momentum $M > 0$, then

$$\hat{a}^+ = (\eta_+\xi_+)^{-1/2}\eta_+\eta_-, \tag{3.29a}$$

$$\hat{a} = (\eta_+\xi_+)^{-1/2}\xi_+\xi_-, \tag{3.29b}$$

$$\hat{A}^+ = q^{-1/2}(\eta_+\xi_+ - \eta_-\xi_-)^{1/2}(\eta_+\xi_+)^{-q/2}\eta_+^q, \tag{3.29c}$$

$$\hat{A} = q^{-1/2}(\eta_+\xi_+ - \eta_-\xi_-)^{1/2}(\eta_+\xi_+)^{-q/2}\xi_+^q. \tag{3.29d}$$

When the angular momentum $M < 0$, then

$$\hat{a}^+ = (\eta_-\xi_-)^{-1/2}\eta_+\eta_-, \tag{3.30a}$$

$$\hat{a} = (\eta_-\xi_-)^{-1/2}\xi_+\xi_-, \tag{3.30b}$$

$$\hat{A}^+ = q^{-1/2}(\eta_-\xi_- - \eta_+\xi_+)^{1/2}(\eta_-\xi_-)^{-q/2}\eta_-^q, \tag{3.30c}$$

$$\hat{A} = q^{-1/2}(\eta_-\xi_- - \eta_+\xi_+)^{1/2}(\eta_-\xi_-)^{-q/2}\xi_-^q. \tag{3.30d}$$

All symbols are interpreted as classical observables defined in terms of the $\eta_+, \eta_-, \xi_+, \xi_-$ given by (3.1), (3.9).

As indicated in Refs. 1, 2 the Poisson brackets of any two observables $F$, $G$ can be written as

$$\{F, G\} = i\left(\frac{\partial F}{\partial \eta_+}\frac{\partial G}{\partial \xi_+} - \frac{\partial F}{\partial \xi_+}\frac{\partial G}{\partial \eta_+}\right) + i\left(\frac{\partial F}{\partial \eta_-}\frac{\partial G}{\partial \xi_-} - \frac{\partial F}{\partial \xi_-}\frac{\partial G}{\partial \eta_-}\right), \tag{3.31}$$

and thus for both $M > 0$ and $M < 0$ we check from (3.29), (3.30) that

$$\begin{aligned} &\{\hat{a}^+, -i\hat{a}\} = \{\hat{A}^+, -i\hat{A}\} = 1, \\ &\{\hat{a}^+, \hat{A}^+\} = \{\hat{a}^+, \hat{A}\} = \{\hat{a}, \hat{A}^+\} = \{\hat{a}, \hat{A}\} = 0, \end{aligned} \tag{3.32}$$

so that $\hat{A}^+, \hat{a}^+, -i\hat{A}, -i\hat{a}$ can be considered as canonically conjugate variables in four-dimensional phase space.

We furthermore see from (3.29) that for $M > 0$

$$\hat{a}^+\hat{a} = \eta_-\xi_-, \tag{3.33a}$$

$$\hat{A}^+\hat{A} = q^{-1}(\eta_+\xi_+ - \eta_-\xi_-) \tag{3.33b}$$

while from (3.30) we have that for $M < 0$

$$\hat{a}^+\hat{a} = \eta_+\xi_+, \tag{3.34a}$$

$$\hat{A}^+\hat{A} = q^{-1}(\eta_-\xi_- - \eta_+\xi_+). \tag{3.34b}$$

Thus in both cases the Hamiltonian (3.11) is given by

$$H = 2\hat{a}^+\hat{a} + q\hat{A}^+\hat{A}, \tag{3.35}$$

which corresponds to an anisotropic oscillator whose ratio of frequencies is $q/2$.

Expressing now $\hat{a}^+, \hat{a}, \hat{A}^+, \hat{A}$ in terms of coordinate and momenta $x, p, X, P$ through the usual relations[1] for frequencies 2 and $q$, i.e.,

$$\begin{aligned} &\hat{a}^+ = (1/\sqrt{2})(\sqrt{2}\,x - i\,p/\sqrt{2}), \quad \hat{a} = (1/\sqrt{2})(\sqrt{2}\,x + i\,p/\sqrt{2}), \\ &\hat{A}^+ = (1/\sqrt{2})(\sqrt{q}\,X - i\,P/\sqrt{q}), \quad \hat{A} = (1/\sqrt{2})(\sqrt{q}X + i\,P/\sqrt{q}), \end{aligned} \tag{3.36}$$

we obtain from (3.29) or (3.30) and (3.1), (3.2), (3.9) the canonical transformation that connects them with $x_1, p_1, x_2, p_2$. Thus the problem of an isotropic oscillator in a sector of angle $\pi/q$ is mapped by this transformation on the anisotropic oscillator whose Hamiltonian is

$$H = \tfrac{1}{2}(p^2 + 4x^2) + \tfrac{1}{2}(P^2 + q^2X^2). \tag{3.37}$$

It remains to see whether the Hamiltonian (3.37) is not subject to any constraints as was the case for the original isotropic oscillator due to the barriers of the sector. To see that the motion takes place now on the full configuration space $(x, X)$ and the momenta $(p, P)$ are continuous throughout this motion, we analyze in detail the behavior of $\hat{a}^+, \hat{A}^+$ as function of time for the case $q = 3$. The motion within the sector of the original $(x_1x_2)$ plane is illustrated in Fig. 1 of Ref. 2. As indicated there $\eta_\pm$, $\xi_\pm$ are given in terms of the polar coordinates and momenta $r, \varphi, p_r, p_\varphi$ by

$$\eta_\pm = \tfrac{1}{2}[\exp(\pm i\varphi)](r \pm r^{-1}p_\varphi - ip_r), \tag{3.38a}$$

$$\xi_\pm = \eta_\pm^* = \tfrac{1}{2}[\exp(\mp i\varphi)](r \pm r^{-1}p_\varphi + ip_r). \tag{3.38b}$$

Furthermore, as we assumed in the original configuration space $(x_1x_2)$ an isotropic oscillator of frequency unity, we have

$$\eta_\pm = (\eta_{\pm 0})\exp(it), \tag{3.39a}$$

$$\xi_\pm = (\xi_{\pm 0})\exp(-it), \tag{3.39b}$$

where $(\eta_{\pm 0})$, $(\xi_{\pm 0})$ are the initial values of these variables given in terms of $r_0, \varphi_0, p_{r0}, p_{\varphi 0}$ by expressions similar to (3.38).

We proceed now to analyze the evolution in time of $\eta_\pm$ due to the presence of the barriers of the sector. The motion for $q = 3$ is given by the heavy lines of Fig. 1 in Ref. 2, and we see that in general there will be six branches to the trajectory. We can start with any of them initiating the motion at the barrier $\varphi_0 = 0$ with $p_{\varphi_0} > 0$ as indicated by the arrow marked 1 in Fig. 1 of the present paper. The initial conditions are then

$$r_0 = \alpha > 0, \quad \varphi_0 = 0, \quad p_{r0} = \beta, \quad p_{\varphi 0} = \gamma > 0. \tag{3.40}$$

Designating with an upper index 1 the $\eta_\pm$ of this first branch, we have

$$\eta_\pm^{(1)} = \tfrac{1}{2}(\alpha \pm \alpha^{-1}\gamma - i\beta)\exp(it). \tag{3.41}$$

After collision at the barrier $\varphi = (\pi/3)$ the motion takes places in the second branch which, using the method of images, can be thought of as starting with the initial conditions

$$r_0 = \alpha, \quad \varphi_0 = 2\pi/3, \quad p_{r0} = \beta, \quad P_{\varphi 0} = -\gamma. \tag{3.42}$$

Using these values, we can then write $\eta_\pm^{(2)}$ associated with the second branch. When the motion reaches $\varphi = 0$, we can again use the method of images to find what are the initial conditions for the third branch and thus get $\eta_\pm^{(3)}$. The different starting points with the direction of the angular momentum are marked in Fig. 1 of the present paper by the arrows numbered 1 to 6. We then obtain for the six branches

$$\eta_\pm^{(\kappa)} = \tfrac{1}{2}\exp(\pm i\varphi_{0\kappa})(\alpha \mp (-1)^\kappa\alpha^{-1}\gamma - i\beta)\exp(it), \tag{3.43}$$

where $\kappa = 1, 2, 3, 4, 5, 6$ and

$$\begin{aligned} &\varphi_{01} = 0, \quad \varphi_{02} = 2\pi/3, \quad \varphi_{03} = 4\pi/3, \\ &\varphi_{04} = 4\pi/3, \quad \varphi_{05} = 2\pi/3, \quad \varphi_{06} = 0. \end{aligned} \tag{3.44}$$

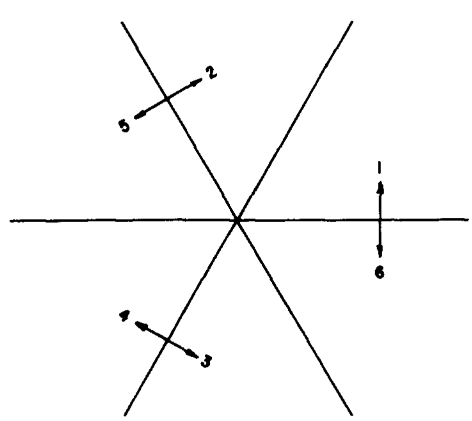



FIG. 1. The problem of the sector of angle $\pi/3$. The arrows indicate the direction of the velocity normal to the wall for each of the six branches of the trajectory when we use the method of images. Thus for branches 1, 3, 5 the angular momentum is positive and for branches 2, 4, 6 it is negative.

We see that $\eta_\pm^{(\kappa)}$ (and thus also $\xi_\pm^{(\kappa)}$ which is its conjugate) suffers discontinuities as we go from branch $\kappa$ to branch $\kappa+1$ at the point of collision. But this is not reflected in $\hat{a}^+, \hat{A}^+$. The latter change form at the barriers of the sector where $p_{\varphi 0}$ and thus $M$ changes sign, as indicated in (3.29), (3.30). Thus from (3.43) we obtain that on all six branches of the original motion in the sector $\hat{a}^+, \hat{A}^+$ are given by

$$\hat{a}^+ = \tfrac{1}{2}[(\alpha - i\beta)^2 - \alpha^{-2}\gamma^2]\exp(i2t)/[(\alpha+\alpha^{-1}\gamma)^2+\beta^2]^{1/2}, \tag{3.45a}$$

$$\hat{A}^+ = q^{-1/2}[(\alpha+\alpha^{-1}\gamma)^2+\beta^2]^{-q/2}\gamma^{1/2}(\alpha+\alpha^{-1}\gamma - i\beta)^q\exp(iqt). \tag{3.45b}$$

We have now written the result for arbitrary integer $q$ rather than only $q=3$ as the extension of the arguments is obvious.

As seen in Fig. 1 of Ref. 2 the time required to cover all branches of the orbit in the sector of configuration space $(x_1x_2)$ is the same as the one needed to traverse the ellipse in the full plane without barriers. As our frequency is unity this time is $t=2\pi$. After covering then all branches in the space $(x_1x_2)$, we have that $\hat{a}^+, \hat{A}^+$, and thus also $x, X, p, P$ given by (3.36), return to their original values. Thus we have proved that the canonical transformation relating $x_1, x_2, p_1, p_2$ to $x, X, p, P$ maps the isotropic oscillator in a sector on an anisotropic oscillator with ratio of frequencies $q/2$, over the full new configuration space $x, X$.

The remaining steps required now to find the group of canonical transformations responsible for accidental degeneracy for the problem of the sector have already been discussed in Ref. 2. The Hamiltonian (3.35) can also be written as

$$H = 2q[q^{-1}\hat{a}^+\hat{a} + \tfrac{1}{2}\hat{A}^+\hat{A}], \qquad \text{if } q \text{ is odd}, \tag{3.46a}$$

$$H = q[(q/2)^{-1}\hat{a}^+\hat{a} + \hat{A}^+\hat{A}] \qquad \text{if } q \text{ is even}. \tag{3.46b}$$

We can thus make a further transformation of the type discussed in the previous section, and also in Ref. 1, to map the problem on an isotropic oscillator in the full plane. The symmetry group of this last oscillator is $U(2)$, and it is responsible for accidental degeneracy for this problem. We could then write the symmetry group responsible for accidental degeneracy of the oscillator in a sector of angle $\pi/q$ as

$$\mathcal{T}^{-1}\begin{bmatrix} \tfrac{1}{2}(U+U^*) & (i/2)(U-U^*) \\ -(i/2)(U-U^*) & \tfrac{1}{2}(U+U^*) \end{bmatrix}\mathcal{T}, \tag{3.47}$$

where $U$ is the two dimensional unitary matrix and $\mathcal{T}$ is the nonlinear canonical transformation relating $x_1, x_2, p_1, p_2$ to the coordinates and momenta associated to an isotropic oscillator in the full plane as discussed in the previous paragraphs.

We now proceed to apply a sequence of similar steps to finding the group responsible for accidental degeneracy in the Calogero problem for the case of three particles. The procedure of this and the following section indicates that the method will work for all cases in which the spectrum has the form (1.1) as we discussed in the Introduction.

## 4. CANONICAL TRANSFORMATIONS AND THE CALOGERO PROBLEM

As discussed by Calogero, [5] the Hamiltonian of his problem for the three-particle case in one dimension is

$$\tilde{H} = \tilde{H}_0 + V,$$

$$\tilde{H}_0 = -\tfrac{1}{2}\sum_{i=1}^{3}\frac{\partial^2}{\partial x_i^2} + \tfrac{1}{6}[(x_1-x_2)^2 + (x_1-x_3)^2 + (x_2-x_3)^2],$$

$$V = g\left[\frac{1}{(x_1-x_2)^2} + \frac{1}{(x_2-x_3)^2} + \frac{1}{(x_1-x_3)^2}\right]. \tag{4.1}$$

Introducing the Jacobi coordinates in polar form, i.e.,

$$\begin{aligned} r\sin\varphi &= (1/\sqrt{2})(x_1-x_2), \\ r\cos\varphi &= (1/\sqrt{6})(x_1+x_2-2x_3), \\ X &= (1/\sqrt{3})(x_1+x_2+x_3), \end{aligned} \tag{4.2}$$

we have that the intrinsic Hamiltonian $H$, from which the center of mass motion has been eliminated, takes the form

$$H = H_0 + V, \tag{4.3a}$$

$$H_0 = -\frac{1}{2}\left(\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2} - r^2\right) - 1, \tag{4.3b}$$

$$V = \frac{9g}{2r^2\sin^2 3\varphi}, \tag{4.3c}$$

where for convenience in later notation we subtracted 1 from $H_0$ so that it becomes identical to (3.11) and use was made of the relation[5]

$$\sin^{-2}\varphi + \sin^{-2}(\varphi + 2\pi/3) + \sin^{-2}(\varphi + 4\pi/3) = 9\sin^{-2}(3\varphi). \tag{4.4}$$

We now proceed, as in the case of the sector problem, to discuss orthonormal and nonorthonormal eigenstates of the Hamiltonian $H$. The former have been derived by Calogero and we shall denote them by an angular ket $|nN\rangle$ whose explicit form is[5]

$$|nN\rangle = (-1)^n 2^\tau \Gamma(\tau)\left(\frac{3(N!)(n!)(N+\tau)}{\pi\,\Gamma(N+2\tau)\Gamma(n+3N+3\tau+1)}\right)^{1/2}$$
$$\times r^{3N+3\tau}\exp[-(r^2/2)]L_n^{3N+3\tau}(r^2)(\sin 3\varphi)^\tau C_N^\tau(\cos 3\varphi), \tag{4.5a}$$

where

$$\tau = \tfrac{1}{2} + (\tfrac{1}{4} + g)^{1/2}, \tag{4.5b}$$

and $L_n^m(r^2)$, $C_N^\tau(\cos 3\varphi)$ are respectively Laguerre and Gegenbauer polynomials of the arguments indicated. The kets $|nN\rangle$ are normalized for $\varphi$ in the interval $0 \leqslant \varphi \leqslant (\pi/3)$ as the repulsive potential $V$ prevents the particles from going outside these limiting values of the angle $\varphi$. The phase factor $(-1)^n$ is introduced to establish a complete paralelism with the sector problem as will be seen below.

The ket (4.5a) is an eigenstate of the Hamiltonian $H$ of (4.3) and of the angular operator[5]

$$M^2 = p_\varphi^2 + (9g/\sin^2 3\varphi), \quad \text{where } p_\varphi = \frac{1}{i}\frac{\partial}{\partial\varphi}. \tag{4.6}$$

The corresponding eigenvalues are[5]

$$H|nN\rangle = (2n+3N+3\tau)|nN\rangle, \tag{4.7a}$$

$$M^2|nN\rangle = 9(N+\tau)^2|nN\rangle. \tag{4.7b}$$

From (4.7a) we immediately see that the Calogero problem has the same accidental degeneracy as that of an anisotropic oscillator whose ratio of frequencies is 2/3.

As in the case of the sector, we define the operator $|M|$ as the one which when applied to $|nN\rangle$ has as eigenvalue the positive square root of that of $M^2$, i.e.,

$$|M|\,|nN\rangle = 3(N+\tau)\,|nN\rangle. \tag{4.8}$$

We now turn our attention to the nonorthonormal basis introduced by Perelomov[6] for the Calogero problem. He starts by considering two creation operators[6] which have simple commutation properties with the Hamiltonian (4.3). In this paper we shall denote the operators by $b^+$, $B^+$ and their commutators with $H$ and among themselves are[6]

$$[H, b^+] = 2b^+, \quad [H, B^+] = 3B^+, \quad [b^+, B^+] = 0. \tag{4.9}$$

We shall express these operators in terms of the $\eta_\pm$ of (3.38a), where $p_r, p_\varphi$ are replaced respectively by

$$p_r = -i\frac{\partial}{\partial r}, \quad p_\varphi = -i\frac{\partial}{\partial\varphi}. \tag{4.10}$$

From Perelomov's paper[6] we have (suppressing irrelevant multiplicative constants) that

$$b^+ = \eta_+\eta_- - 9g(4r^2\sin^2 3\varphi)^{-1}, \tag{4.11}$$

$$B^+ = \eta_+^3 + \eta_-^3 + \frac{27}{4}\frac{g}{r^2\sin^2 3\varphi} \times[(\cos 3\varphi)(r - ip_r) + (i/3r)(\sin 3\varphi)p_\varphi]. \tag{4.12}$$

The Hamiltonian $H$ of (4.3) can in turn be written in terms of $\eta_+, \eta_-$ as

$$H = \eta_+\xi_+ + \eta_-\xi_- + 9g(2r^2\sin^2 3\varphi)^{-1}. \tag{4.13}$$

It is particularly easy to check the commutation relations (4.9) if we express them as Poisson brackets and calculate them classically using (4.11), (4.12), (4.13) and (3.38a). It is interesting to note that $b^+, B^+$ reduce to $a^+, A^+$ of the sector problem for $q=3$ when $g=0$.

With the help of $b^+, B^+$ it becomes now possible to obtain a complete nonorthonormal basis for the Calogero problem.[6] We shall designate these states by the round ket $|nN)$ defined by

$$|nN) = (n!N!)^{-1/2}(b^+)^n(B^+)^N|00\rangle, \tag{4.14}$$

where $|00\rangle$ is the ground state of the Calogero problem given by (4.5) when $n=N=0$. From (4.9) it is clear that these states correspond to the same eigenvalue of the energy (4.7a) as the orthonormal states $|nN\rangle$. Furthermore, it is obvious that

$$b^+|nN) = (n+1)^{1/2}|n+1\ N),$$
$$B^+|nN) = (N+1)^{1/2}|n\ N+1). \tag{4.15}$$

As in the previous sections we expect that the annihilation operators for the nonorthonormal basis are *not* the Hermitian conjugates of $b^+, B^+$. We shall not proceed to derive them but rather go directly to the operators, which we shall denote by $\hat{b}^+, \hat{B}^+$, which are the raising ones for the orthonormal basis. The procedure we shall follow will be a carbon copy of the one employed for the sector problem.

We start by applying $b^+, B^+$ to the orthonormal basis $|nN\rangle$. Perelomov[6] has shown that

$$b^+|nN\rangle = \lambda_{nN}|n+1\ N\rangle, \tag{4.16}$$

$$B^+|nN\rangle = \mu_{nN}|n\ N+1\rangle + \nu_{nN}|n+3, N-1\rangle, \tag{4.17}$$

where

$$\lambda_{nN} = [(n+1+3N+3\tau)(n+1)]^{1/2}, \tag{4.18a}$$

$$\mu_{nN} = \left(\frac{\Gamma(n+3N+3\tau+4)}{\Gamma(n+3N+3\tau+1)}\right)^{1/2}\left(\frac{(N+1)(N+2\tau)}{(N+\tau)(N+\tau+1)}\right)^{1/2}, \tag{4.18b}$$

$$\nu_{nN} = \left(\frac{N(N+2\tau-1)(n+3)(n+2)(n+1)}{(N+\tau)(N+\tau-1)}\right)^{1/2}. \tag{4.18c}$$

By using the operators $H$, $|M|$ whose eigenvalues are given in (4.7a), (4.8), we immediately see from (4.18a) that the operator

$$\hat{b}^+ = [\tfrac{1}{2}(H+|M|)]^{-1/2}\,b^+ \tag{4.19}$$

has the property

$$\hat{b}^+|nN\rangle = (n+1)^{1/2}|n+1\ N\rangle. \tag{4.20}$$

For $\hat{B}^+$ we require, as in the case of the sector problem, another operator which when combined with it would allow us to eliminate the ket $|n+3, N-1\rangle$. It is clear that the commutator $[M^2, B^+]$ when applied to $|nN\rangle$ gives a linear combination of states similar to (4.17), i.e.,

$$[M^2, B^+]|nN\rangle = \mu'_{nN}|n\ N+1\rangle + \nu'_{nN}|n+3, N-1\rangle, \tag{4.21}$$

where

$$\mu'_{nN} = 9[2(N+\tau)+1]\mu_{nN}, \tag{4.22a}$$

$$\nu'_{nN} = 9[-2(N+\tau)+1]\nu_{nN}. \tag{4.22b}$$

Combining then the operators $B^+$, $[M^2, B^+]$, we get

$$\{[18(N+\tau)-9]B^+ + [M^2, B^+]\}|nN\rangle = 36(N+\tau)\mu_{nN}|n\ N+1\rangle. \tag{4.23}$$

Finally making use again of the eigenvalues (4.7a), (4.8) of the operators $H$ and $|M|$ we can write the raising operator $\hat{B}^+$ for the orthonormal basis as

$$\hat{B}^+ = (3\sqrt{6})^{-1}|M|^{1/2}[(|M|-3)(|M|+3\tau-3)(H+|M|) \times(H+|M|-2)(H+|M|-4)]^{-1/2} \times\{3B^+(2|M|-3) + [M^2, B^+]\}, \tag{4.24}$$

which clearly has the property that

$$\hat{B}^+|nN\rangle = (N+1)^{1/2}|n\ N+1\rangle. \tag{4.25}$$

The annihilation operators $\hat{b}, \hat{B}$ are then the Hermitian conjugates of the creation ones and their effect on the kets $|nN\rangle$ are

$$\hat{b}|nN\rangle = n^{1/2}|n-1\ N\rangle, \quad \hat{B}|nN\rangle = N^{1/2}|n\ N-1\rangle. \tag{4.26}$$

Thus the operator

$$2\hat{b}^+\hat{b} + 3\hat{B}^+\hat{B} + 3\tau \tag{4.27}$$

is diagonal in the basis $|nN\rangle$ and its eigenvalue are given by (4.7a), so that it is a representation of the Hamiltonian of the Calogero problem in terms of the new creation and annihilation operators.

The operators $\hat{b}^+, \hat{B}^+$ reduce to those of the sector problem for $q=3$ when the potential $V$ goes to zero. This is seen immediately as $b^+$, $B^+$ reduce to $a^+, A^+$; $H$, $|M|$ become those of an isotropic oscillator and, when $g=0$, $\tau=1$.

As in the sector problem we are interested in the classical limit so that we can obtain the canonical transformation that maps the Calogero problem on an anisotropic oscillator [as indicated by the expression (4.27) of the Hamiltonian] whose ratio of frequencies is 2/3. Again we use the correspondence principle in which the quantum numbers associated with $H$, $|M|$ are large. We can then disregard integers or $\tau$ as compared with the eigenvalues of $H$, $|M|$ and thus in the classical limit we have

$$\hat{b}^+ = [\tfrac{1}{2}(H+|M|)]^{-1/2}\, b^+, \tag{4.28a}$$

$$\hat{B}^+ = (3\sqrt{6})^{-1}(H+|M|)^{-3/2}|M|^{-1/2}(6B^+|M| + i\{M^2, B^+\}). \tag{4.28b}$$

As usual in these cases we have replaced the commutator $[F,G]$ by the Poisson bracket $i\{F,G\}$, which is allowed quantum mechanically and gives an expression that has meaning also in the classical limit.

If we want to calculate explicitly $\hat{b}^+, \hat{B}^+$ it is convenient to express $b^+, B^+, H, M^2$, in terms of $\eta_+, \eta_-, \xi_+, \xi_-$. From (3.38) we note that

$$\begin{aligned}
p_\varphi &= \eta_+\xi_+ - \eta_-\xi_-,\\
p_r &= i[(\eta_+ + \xi_-)(\eta_- + \xi_+)]^{-1/2}(\eta_+\eta_- - \xi_+\xi_-)\\
r &= [(\eta_+ + \xi_-)(\eta_- + \xi_+)]^{1/2},\\
\exp(i\varphi) &= [(\eta_+ + \xi_-)/(\eta_- + \xi_+)]^{1/2}
\end{aligned} \tag{4.29}$$

and thus we obtain

$$b^+ = \eta_+\eta_- + 9g\,\frac{(\eta_+ + \xi_-)^2(\eta_- + \xi_+)^2}{[(\eta_+ + \xi_-)^3 - (\eta_- + \xi_+)^3]^2} \tag{4.30a}$$

$$\begin{aligned}
B^+ = (\eta_+^3 + \eta_-^3) - &\frac{27g}{2[(\eta_+ + \xi_-)^3 - (\eta_- + \xi_+)^3]^2}\\
&\times\{[(\eta_+ + \xi_-)^3 + (\eta_- + \xi_+)^3][2\eta_+\eta_- + \eta_+\xi_+ + \eta_-\xi_-]\\
&+ \tfrac{1}{3}[(\eta_+ + \xi_-)^3 - (\eta_- + \xi_+)^3](\eta_+\xi_+ - \eta_-\xi_-)\},
\end{aligned} \tag{4.30b}$$

$$H = (\eta_+\xi_+ + \eta_-\xi_-) - 18g\,\frac{(\eta_+ + \xi_-)^2(\eta_- + \xi_+)^2}{[(\eta_+ + \xi_-)^3 - (\eta_- + \xi_+)^3]^2}\quad, \tag{4.30c}$$

$$M^2 = (\eta_+\xi_+ - \eta_-\xi_-)^2 - 36g\,\frac{(\eta_+ + \xi_-)^3(\eta_- + \xi_+)^3}{[(\eta_+ + \xi_-)^3 - (\eta_- + \xi_+)^3]^2}. \tag{4.30d}$$

The Poisson bracket appearing in $\hat{B}^+$ can be calculated using (3.31).

We have then the explicit canonical transformation that takes us from $\eta_+, \eta_-, \xi_+, \xi_-$ into $\hat{b}^+, \hat{B}^+, \hat{b}, \hat{B}$ and maps the classical Calogero problem into an anisotropic oscillator whose ratio of frequencies is 2/3.

As discussed in the last part of the previous section, the anisotropic oscillator can in turn be mapped into an isotropic one, and thus the group of canonical transformation responsible for accidental degeneracy in the Calogero problem is given again by a realization of the $U(2)$ group of the form (3.47). The canonical transformation $\mathcal{T}$ appearing in (3.47) is now the one that takes us from the relative Jacobi coordinates and their corresponding momenta to $\eta_+, \eta_-, \xi_+, \xi_-$, from there to $\hat{b}^+, \hat{B}^+, \hat{b}, \hat{B}$ as indicated above, then to the creation and annihilation operators of the isotropic oscillator as discussed in Sec. 2, and finally to coordinate and momenta operators defined in terms of the latter by relations similar to (2.19).

The canonical transformation $\mathcal{T}$ is a very complicated one and it seems difficult to visualize how we could have determined it from a purely classical reasoning. By using quantum mechanics and both the orthonormal basis of Calogero[5] and the nonorthonormal one of Perelomov[6] we were able to find it, though the relations (4.28), (4.30), show the complexity of its explicit form.

## 5. CONCLUSION

Turning now to the general two-dimensional problem whose spectrum has the form (1.1), we can say that the group responsible for its accidental degeneracy is a realization of the $U(2)$ group of the form (3.47). The steps required to derive this symmetry group were outlined in Sec. 1.

It is interesting to note that the generators of the Lie algebra of the $U(2)$ symmetry group of the isotropic oscillator in the full plane, i.e., the $\eta_i'\xi_j'$ of (2.28a), are integrals of motion of this problem. If we have a canonical transformation $\mathcal{T}$ that maps any problem whose spectrum is given by (1.1) into an isotropic oscillator in the full plane, we can use it to express $\eta_i'\xi_j'$, $i,j=1,2$, in terms of the coordinates and momenta of our original problem and thus obtain interesting integrals of motion for it.

We could extend our class of problems to $n$-dimensional configuration space if we assume that the spectrum is some linear combination with integer coefficients of the quantum numbers $\nu_1, \nu_2 \cdots \nu_n$. A problem of this type is given by the $n+1$ particle Calogero problem.[5,6] A simpler case appears when we consider $n$ particles in a one-dimensional harmonic oscillator potential, but restricted to obey Fermi statistics. In the latter case both an orthonormal and nonorthonormal[6] basis can be obtained in terms of polynomial functions in the creation operators $\eta_1, \eta_2 \cdots \eta_n$, acting on the antisymmetric states of lowest energy, and a discussion similar to the one presented in the previous sections can be carried out for this problem. The group of canonical transformations would again be given by (3.47) but with the $U$ appearing there referring to the $n$-dimensional unitary group.

We have given a general procedure for deriving the group of canonical transformations responsible for accidental degeneracy in a wide class of problems. We think that it provides a framework for a general understanding of these problems, which, except for some recent attempts,[11–13] have been attacked by the independent study of each case.

## ACKNOWLEDGMENTS

We are indebted to C. Boyer, E. Chacón, J. Flores, P.A. Mello, W. Miller Jr., and K. B. Wolf for helpful discussions. One of the authors (M. Moshinsky) would like to express his thanks to the Centre de Recherches Mathématiques, Université de Montréal, to which he was invited as an Aisenstadt professor.

*Permanent address: Instituto de Física, Universidad de México. Member of the Instituto Nacional de Energia Nuclear and El Colegio Nacional.

†Supported in part by the National Research Council of Canada and the Ministere de l'Education du Québec.

[1]J. D. Louck, M. Moshinsky, and K. B. Wolf, J. Math. Phys. **14**, 692 (1973).

[2]J. D. Louck, M. Moshinsky, and K. B. Wolf, J. Math. Phys. **14**, 696 (1973).

[3]V. A. Fock, Z. Phys. **98**, 145 (1935); V. Bargmann, Z. Phys. **99**, 576 (1936). For a review see M. J. Englefield, *Group Theory and the Coulomb Problem* (Wiley, New York, 1972).

[4]J. M. Jauch and E. L. Hill, Phys. Rev. **57**, 641 (1940). For a review see M. Moshinsky, *The Harmonic Oscillator in Modern Physics: From Atoms to Quarks* (Gordon & Breach, New York, 1969). Yu. N. Demkov, Zh. Eksp. Teor. Fiz. **44**, 2007 (1963) [Sov. Phys. JETP **17**, 1349 (1963)]; F. Duimio and G. Zambotti, Nuovo Cimento A **43**, 1203 (1966); A. Cisneros and H. V. McIntosh, J. Math. Phys. **11**, 870 (1970).

[5]F. Calogero, J. Math. Phys. **12**, 2191 (1969).

[6]A. M. Perelomov, Teor. Mat. Fiz. **6**, 364 (1971) [Theor. Math. Phys. **6**, 263 (1971)].

[7]P. Winternitz, Ya. A. Smorodinskii, M. Uhlir, and I. Fris, Yad. Fiz. **4**, 625 (1966) [Sov. J. Nucl. Phys. **4**, 444 (1967)]; I. Fris, V. Mandrosov, Ya. A. Smorodinskii, M. Uhlir, and P. Winternitz, Phys. Lett. **16**, 354 (1965).

[8]A. A. Makarov, Ya. A. Smorodinskii, Kh. V. Valiev, and P. Winternitz, Nuovo Cimento A **52**, 1061 (1967).

[9]P. Kramer, M. Moshinsky, and T. Seligman, "Complex Extensions of Canonical Transformations and Quantum Mechanics," in *Group Theory and Applications*, *Vol.* 3, edited by E. Loebl (Academic, New York, to be published, 1974).

[10]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1951), pp. 288—94.

[11]E. Onofri and M. Pauri, J. Math. Phys. **13**, 533 (1972); **14**, 1106 (1973).

[12]J. M. Sourian, *Structures des systemes dynamiques* (Dunod, Paris, 1970).

[13]A. Simoni, F. Zaccaria, and B. Vitale, Nuovo Cimento A **51**, 448 (1967); G. Caratin, G. Marmo, A. Simoni, B. Vitale, and F. Zaccaria, Preprint, Naples, 1974.

# On the exact solutions of the Wick-rotated fermion–antifermion Bethe–Salpeter equation

D. zum Winkel

*Deutsches Elektronen-Synchrotron DESY, Hamburg, West Germany*
(Received 22 July 1974)

Exact solutions of the fermion-antifermion Bethe–Salpeter equation with harmonic kernels are given for bound states with zero mass. The algebraic structure of these BS amplitudes is analyzed.

## 1. INTRODUCTION

In this paper we consider the fermion—antifermion Bethe—Salpeter equation in the Wick-rotated form for boundstate mass $M=0$.[1] The interaction kernel is specialized to harmonic forces.

The interest in this form of the BS equation is connected with the description of the mesons as quark—antiquark boundstates.[2] Since one needs smooth kernels in order to obtain linear Regge trajectories for the meson mass spectrum, it is worth studying an approximation of these kernels by harmonic interaction. Under the dynamical assumption of heavy quarks ($M \ll 2m_{\text{quark}}$) the restriction to boundstate mass $M=0$ is a useful starting point.

Under these specifications the BS equation takes the form

$$(\gamma q - im)\chi(q)(\gamma q - im) = (\mathcal{R}\chi)(q),$$
$$\mathcal{R} = \sum_{i=S,V,T,A,P} K^i \mathcal{P}^i, \tag{1}$$
$$K^i = \alpha^i - \beta^i \Box_q.$$

The $\mathcal{P}^i$ are projection operators on the Dirac matrices $1$, $\gamma_\mu$, $\sigma_{\mu\nu}$, $\gamma_5\gamma_\mu$, $\gamma_5$ (denoted by $S, V, T, A, P$).

## 2. EXACT SOLUTIONS

The BS equation for $M=0$ decomposes into three Dirac sectors $S+V$, $T+A$, $P$. This fact is used in the ansatz

$$\chi(q) = \chi_0 - \frac{i}{2m}\{\gamma q, \chi_0\} = \chi_0 + \chi_1, \tag{2}$$

where $\chi_1$ is of type $V, S, A, T, -$, if $\chi_0$ is of type $S, V, T, A, P$. Because of

$$(\gamma q - im)(\chi_0 + \chi_1)(\gamma q - im) = -(q^2 + m^2)(\chi_0 - \chi_1), \tag{3}$$

insertion of the ansatz (2) into Eq. (1) leads to two scalar equations

$$(-q^2 - m^2 - K^0)\chi_0 = 0,$$
$$(-q^2 - m^2 + K^1)\chi_1 = 0. \tag{4a}$$

The spin dependence appears only in the coupling of $\chi_0$ and $\chi_1$:

$$\chi_1 = -\frac{i}{2m}\{\gamma q, \chi_0\}. \tag{4b}$$

(Up to this point no special momentum dependence of the kernel is assumed.)

In the case of harmonic potentials Eq. (4a) are four-dimensional oscillator equations. Because of the opposite signs of the kernels in (4a) and because of the coupling (4b) we must demand $\beta^0 = -\beta^1 = \beta$. (In the following we make the substitution $q \to q/\beta^{1/4}$). $\chi(q)$ is a solution of the BS equation, if $\chi_0$ and $\chi_1$ satisfy the same oscillator equation. If $\chi_0$ is an arbitrary oscillator solution, then $\{\gamma q, \chi_0\}$ is, in general, not an eigensolution belonging to a certain eigenvalue. This is most easily seen by introducing the well-known creation and annihilation operators

$$a_\mu^\pm = \frac{1}{\sqrt{2}}(x_\mu \mp ip_\mu) = \frac{1}{\sqrt{2}}\left(x_\mu \mp \frac{\partial}{\partial x_\mu}\right)$$

with the commutation relations

$$[a_\mu^-, a_\nu^+] = \delta_{\mu\nu}, \quad [a_\mu^-, a_\nu^-] = 0 \quad [a_\mu^+, a_\nu^+] = 0. \tag{5}$$

By writing

$$\chi_1 = -\frac{i}{2\sqrt{2}m}\{\gamma_\mu, (a_\mu^+ + a_\mu^-)\chi_0\},$$

it is seen that $\chi_1$ is also an oscillator eigenfunction, if

$$\{\gamma_\mu, q_\mu\chi_0\} = \frac{1}{\sqrt{2}}\{\gamma_\mu, a_\mu^+\chi_0\}$$

TABLE I. The complete solutions of the BS equation are given by Eq. (2). $\bar{X}$ denotes an arbitrary scalar oscillator eigenfunction Also, the numbers $C_\nu$ can be chosen freely within the assumptions we made. By specializing $\bar{X}$ and $C_\nu$, one can arrange that the solutions have certain transformation properties.

| Dirac sector | Conditions on $\chi_0$ | | |
|---|---|---|---|
| | $\{\gamma_\lambda, a_\lambda^- \chi_0\} = 0$ | $\{\gamma_\lambda, a_\lambda^+ \chi_0\} = 0$ | $\{\gamma_\lambda, a_\lambda^\pm \chi_0\} = 0$ |
| $S+V$ | $\chi_0 = \gamma_\mu (a_\nu^- a_\nu^- a_\mu^+ - a_\mu^- a_\nu^- a_\nu^+)\bar{X}$ | $\chi_0 = \gamma_\mu (a_\nu^+ a_\nu^+ a_\mu^- - a_\mu^+ a_\nu^+ a_\nu^-)\bar{X}$ | $\chi_0 = \epsilon^{\mu\nu\rho\sigma}\gamma_\mu a^+{}_\nu a_\rho^- C_\sigma \bar{X}$ |
| $T+A$ | $\chi_0 = \gamma_5\gamma_\nu a_\nu^- \bar{X}$,<br>$\chi_0 = \sigma_{\mu\nu} a_\mu^- C_\nu \bar{X}$ | $\chi_0 = \gamma_5\gamma_\nu a_\nu^+ \bar{X}$,<br>$\chi_0 = \sigma_{\mu\nu} a_\mu^+ C_\nu \bar{X}$ | $\chi_0 = \sigma_{\mu\nu} a_\mu^+ a_\nu^- \bar{X}$ |
| $P$ | | | $\chi_0 = \gamma_5 \bar{X}$ |

TABLE II. Exact solutions of the fermion—antifermion BS equation for harmonic interaction.

| Dirac sector | Solutions of the BS equation | Sector | $(j^+j^-)^{\Pi' C'}$ | Eigenvalues $\alpha^0$ $\alpha^1$ |
|---|---|---|---|---|
| | $\left(\gamma_\mu - \frac{i}{m\sqrt{2}} a^+_\mu\right)(a^-_\nu a^-_\nu a^+_\mu - a^-_\mu a^-_\nu a^+_\nu)\bar{\chi}$, | | | $\alpha^V = -m^2 - 2(N+1)\sqrt{\beta^V}$, $\alpha^S = m^2 + 2(N+2)\sqrt{\beta^V}$, |
| | | $I$ | $\left(\frac{n}{2}\frac{n}{2}\right)^{++}$ | $\beta^S = -\beta^V$ |
| $S \oplus V$ | $\left(\gamma_\mu - \frac{i}{m\sqrt{2}} a^-_\mu\right)(a^+_\nu a^+_\nu a^-_\mu - a^+_\mu a^+_\nu a^-_\nu)\bar{\chi}$ | | | $\alpha^V = -m^2 - 2(N+3)\sqrt{\beta^V}$, $\alpha^S = m^2 + 2(N+2)\sqrt{\beta^V}$ |
| | $\epsilon^{\mu\nu\rho\sigma}\gamma_\mu a^+_\nu a^-_\rho c_\sigma \bar{\chi}$ | $V^a: c^{(a)}_\sigma = \epsilon^{4\sigma\alpha\beta} a^+_\alpha a^-_\beta$<br>$V^b: c^{(b)}_\sigma = \delta_{4\sigma}$ | $\left(\frac{n+1}{2}\ \frac{n-1}{2}\right)^{0-}$ | $\alpha^V = -m^2 - 2(N+2)\sqrt{\beta^V}$ |
| | $\left(\gamma_5\gamma_\nu a^-_\nu + \frac{1}{m\sqrt{2}}\gamma_5\sigma_{\mu\nu} a^+_\mu a^-_\nu\right)\bar{\chi}$, | | | $\alpha^A = -m^2 - 2(N+1)\sqrt{\beta^A}$, $\alpha^T = m^2 + 2(N+2)\sqrt{\beta^A}$, |
| | | IV | $\left(\frac{n}{2}\frac{n}{2}\right)^{--}$ | $\beta^T = -\beta^A$ |
| | $\left(\gamma_5\gamma_\nu a^+_\nu + \frac{1}{m\sqrt{2}}\gamma_5\sigma_{\mu\nu} a^-_\mu a^+_\nu\right)\bar{\chi}$ | | | $\alpha^A = -m^2 - 2(N+3)\sqrt{\beta^A}$, $\alpha^T = m^2 + 2(N+2)\sqrt{\beta^A}$ |
| $T \oplus A$ | $\left(\sigma_{\mu\nu} a^-_\nu - \frac{1}{m\sqrt{2}}\gamma_5\gamma_\alpha \epsilon^{\alpha\beta\mu\nu} a^+_\beta a^-_\nu\right)$<br>$\times \epsilon^{\mu\rho\sigma\kappa} a^+_\rho a^-_\sigma d_\kappa \bar{\chi}$ | $\mathrm{VI}^a: d^{(a)}_\kappa = \epsilon^{4\kappa\gamma\delta} a^+_\gamma a^-_\delta$ | | $\alpha^T = -m^2 - 2(N+1)\sqrt{\beta^T}$, $\alpha^A = m^2 + 2(N+2)\sqrt{\beta^T}$, |
| | | | $\left(\frac{n+1}{2}\ \frac{n-1}{2}\right)^{0+}$ | $\beta^A = -\beta^T$ |
| | $\left(\sigma_{\mu\nu} a^+_\nu - \frac{1}{m\sqrt{2}}\gamma_5\gamma_\alpha \epsilon^{\alpha\beta\mu\nu} a^-_\beta a^+_\nu\right)$<br>$\times \epsilon^{\mu\rho\sigma\kappa} a^+_\rho a^-_\sigma d_\kappa \bar{\chi}$ | $\mathrm{VI}^b: d^{(b)}_\kappa = \delta_{4\kappa}$ | | $\alpha^T = -m^2 - 2(N+3)\sqrt{\beta^T}$, $\alpha^A = m^2 + 2(N+2)\sqrt{\beta^T}$ |
| | $\sigma_{\mu\nu} a^+_\mu a^-_\nu \bar{\chi}$ | II | $\left(\frac{n}{2}\frac{n}{2}\right)^{+-}$ | $\alpha^T = -m^2 - 2(N+2)\sqrt{\beta^T}$ |
| $P$ | $\gamma_5\bar{\chi}$ | III | $\left(\frac{n}{2}\frac{n}{2}\right)^{-+}$ | $\alpha^P = -m^2 - 2(N+2)\sqrt{\beta^P}$ $(N = n + 2r)$ |

or

$$= \frac{1}{\sqrt{2}}\{\gamma_\mu, a^-_\mu \chi_0\} \tag{6}$$

or

$$= 0.$$

In Table I the oscillator solutions $\chi_0$ satisfying these conditions are listed for the three Dirac sectors.

## 3. $O^{\Pi C}(4)$-SYMMETRY

The BS Eq. (1) is invariant under the transformations of the group $O(4)$ $(x'_\mu = O^\nu_\mu x_\nu)$ extended by 3-space reflections $\Pi$ and charge conjugation $C$.[3] In a group theoretical analysis one takes advantage of the fact that $O(4)$ is isomorphic to the direct product $SU(2) \otimes SU(2)$:

$$O^\nu_\mu = O^\nu_\mu(R_+, R_+), \quad R_\pm \in SU(2).$$

Therefore, the irreducible representations of $O(4)$ are characterized by two angular momenta $(j^+, j^-)$. Because of

$$\Pi(R_+, R_-)\,\Pi = (R_-, R_+), \quad C(R_+, R_-)\,C^{-1} = (R_+, R_-), \quad C\Pi C^{-1} = \Pi,$$

the irreducible representations of $O^{\Pi C}(4)$ are further characterized by an inner parity $\Pi' = \pm 1$, if $j^+ = j^-$, and by an inner charge conjugation $C' = \pm 1$.

The BS amplitude may be expanded in a base, in which the representation of $O^{\Pi C}(4)$ decomposes into irreducible ones. A complete reduction leads to six classes (sectors) of irreducible representations. By this the BS equation decomposes into six systems of hyperradial equations.[4,5]

We now specialize the solutions given in Table I in order to achieve their transformation according to these irreducible representations $(j^+, j^-)^{\Pi' C'}$. This is done by choosing $\bar{\chi}$ as oscillator eigenfunctions, which are also eigenfunctions of $(l^+)^2 + (l^-)^2$, of $l^2$ $(\mathbf{l} = \mathbf{l}^+ + \mathbf{l}^-)$, and $l_3$:

$$\bar{\chi}_{\substack{n,r\\ l,l_3}} = |q|^n\, L^{n+1}_r(q^2) \exp(-q^2/2)\, \mathcal{Y}^n_{l l_3}(\hat{q}). \tag{7}$$

These functions belong to the representation $(l^+, l^-)^{\Pi' C'} = (\frac{1}{2}n, \frac{1}{2}n)^{++}$. The energy eigenvalue is $\bar{\alpha} = n + 2r + 2$.

The parameters $C_\nu$ in Table I appear in sectors with $j^+ \neq j^-$. These representations have no inner parity $\Pi'$. We fix the $C_\nu$ in such a way, that the solutions $\chi(q)$ are eigenfunctions of the parity operator $\Pi$ with the eigenvalue $(-1)^j$ or $(-1)^{j+1}$ (denoted by "a" and "b", respectively).

By this procedure we find that the solutions given in Table I correspond in a unique way to the six sectors of the BS equation[6] and to the solutions obtained by solving the hyperradial equations.[5] The result is listed in Table II.

## ACKNOWLEDGMENT

I want to thank M. Krammer and H. Joos for their hints and help in doing this work.

[1]H. A. Bethe and E. E. Salpeter, Phys. Rev. **84**, 1232 (1951); N. Nakanishi, Prog. Theor. Phys. Suppl. **43**, 1 (1969).

[2]For references, see M. Böhm, H. Joos, and M. Krammer, Act. Phys. Austr. Suppl. **XI**, 3 (1973).

[3]S. Mandelstam, Proc. Roy. Soc. A **237**, 496 (1956).

[4]R. F. Keam, J. Math. Phys. **9**, 1462 (1968); **10**, 594 (1969); **11**, 394 (1970); **12**, 515 (1971).

[5]M. Böhm, H. Joos, and M. Krammer, Nucl. Phys. B **51**, 397 (1973); M. Krammer, DESY-Internal Report T-73/1 (1973).

[6]For details, see D. zum Winkel, Diplomarbeit, Hamburg, 1974.

# Four examples of the inverse method as a canonical transformation

D. W. McLaughlin*

*Department of Mathematics, Arizona University, Tucson, Arizona 85721*
(Received 8 August 1974)

The Toda lattice, the nonlinear Schrödinger equation, the sine–Gordon equation, and the Korteweg–de Vries equation are four nonlinear equations of physical importance which have recently been solved by the inverse method. For these examples, this method of solution is interpreted as a canonical transformation from the initial Hamiltonian dynamics to an "action-angle" form. This canonical structure clarifies the independence of an infinite number of constants of the motion and indicates the special nature of the solution by the inverse method.

Recently a large class of nonlinear dispersive wave equations has been integrated by the "inverse method." These equations describe a wide variety of physical models ranging from water waves to quantum optics. A review of this "inverse method," together with descriptions of its many physical applications, may be found in Ref. 1. The latest results on the method are contained in Refs. 2, 15, 16.

This method is based upon the association of a linear eigenvalue problem to the nonlinear wave equation. Given the nonlinear wave at time $t = 0$, this association maps the nonlinear wave dynamics into spectral data for the linear problem. The time evolution of this spectral data is easily computed. Then, at time $t$, the nonlinear wave is constructed from the spectral data at time $t$ by a Gel'fand–Levitan equation.

Clearly, the key to this method is the linear eigenvalue problem. Thus, it is of considerable interest to discover relations between this linear problem and the nonlinear physics, or the nonlinear wave equation. Such relations indicate the generality of the method and its applicability to other physical situations.

In two cases[3] we have shown that the appropriate linear problem actually describes the micro (quantum) physics of the medium which supports the nonlinear wave. In at least these cases, the linear problem has a direct physical interpretation.

TABLE I. The main results.

| | Toda lattice | Nonlin. Schröd. | Sine–Gordon | Korteweg–deVries |
|---|---|---|---|---|
| 1. Nonlinear dynamics | $\ddot{Q}_n = \exp(Q_{n-1} - Q_n) - \exp(Q_n - Q_{n+1})$ | $\hat{u} = iu_{xx} + i\boldsymbol{\kappa} u^2 u^*$ | $\hat{u}_x = \sin(u) \Big\vert \begin{array}{l} v = u_x \\ \dot{v} = \sin(u) \end{array}$ | $\tilde{u} = 6uu_x - u_{xxx}$ |
| 2. Hamiltonians | $H = \sum_{n=-\infty}^{\infty} \frac{P_n^2}{2} + [\exp(Q_{n-1} - Q_n) - 1]$ | $H = i\int_{-\infty}^{\infty}[P_x Q_x - (\boldsymbol{\kappa}/2)P^2 Q^2]\, dx$ | $H = \int_{-\infty}^{\infty}[\cos(\int_{-\infty}^{x} v(x')\, dx') - 1]\, dx$ | $H = \int_{-\infty}^{\infty}[u^3 + u_x/2]\, dx$ |
| 3. Canonical eq. from $H$ | $\dot{Q}_n = \frac{\partial H}{\partial P_n} = P_n$<br>$\dot{P}_n = -\frac{\partial H}{\partial Q_n} = e^{\Delta_n} - e^{\Delta_{n+1}}$,<br>$\Delta_n \equiv Q_{n-1} - Q_n$ | $\dot{Q} = \frac{\delta H}{\delta P} = -iQ_{xx} - i\boldsymbol{\kappa} PQ^2$<br>$\dot{P} = -\frac{\delta H}{\delta Q} = iP_{xx} = i\boldsymbol{\kappa} P^2 Q$ | $\dot{v} = \frac{\partial}{\partial x}\frac{\delta H}{\delta v}$<br>$= \frac{\partial}{\partial x}\int_{\infty}^{x} dx' \sin\left(\int_{-\infty}^{x'} v(x'')\, dx''\right)$ | $\dot{u} = \frac{\partial}{\partial x}\frac{\delta H}{\delta u}$<br>$= \partial_x [3u^2 - u_{xx}]$ |
| 4. Hamiltonian —"action angle" | $K = \sum_{j=1}^{N} + \frac{p_j^2}{2} + \int_0^{2\pi} 2\sin(\varphi) p(\varphi)\, d\varphi$ | $K = -\frac{2i}{3\boldsymbol{\kappa}}\sum_{j=1}^{N}(\bar{p}_j^3 - p_j^3)$<br>$+\int_{-\infty}^{\infty} 4\xi^2 p(\xi)\, d\xi$ | $K = i\sum_{j=1}^{N}(e^{-\bar{p}_j} - e^{-p_j})$<br>$-\frac{1}{2}\int_{-\infty}^{\infty}\frac{1}{\xi}p(\xi)\, d\xi$ | $K = -\frac{32}{5}\sum_{j=1}^{N} p_j^{5/2}$<br>$+\int_{-\infty}^{\infty} 8\xi^3 p(\xi)\, d\xi$ |
| 5. Canonical eq. "action angle" "$p \equiv 0$" | $\dot{q}_j = \frac{\partial K}{\partial p_j} = +p_j$<br>$\dot{q}(\varphi) = \frac{\delta K}{\delta p} = 2\sin(\varphi)$ | $\dot{q}_j = \frac{\partial K}{\partial p_j} = \frac{2i}{\boldsymbol{\kappa}}p_j^2$<br>$\dot{\bar{q}}_j = \frac{\partial K}{\partial \bar{p}_j} = -\frac{2i}{\boldsymbol{\kappa}}\bar{p}_j^2$<br>$\dot{q}(\xi) = \frac{\delta K}{\delta p} = 4\xi^2$ | $\dot{q}_j = \frac{\partial K}{\partial p_j} = +ie^{-p_j}$<br>$\dot{\bar{q}}_j = \frac{\partial K}{\partial \bar{p}_j} = -ie^{-\bar{p}_j}$<br>$\dot{q}(\xi) = \frac{\delta K}{\delta p} = -\frac{1}{2\xi}$ | $\dot{q}_j = \frac{\partial K}{\partial p_j} = -16p_j^{3/2}$<br>$\dot{q}(\xi) = \frac{\delta H}{\delta p} = 8\xi^3$ |
| 6. Canonical maps | $p(\varphi) \equiv -\cos 2\varphi/4\pi \sin\varphi)\ln\{1 + \vert b\vert^2\}$<br>$q(\varphi) \equiv \arg b(e^{i\varphi})$<br>$p_j = [(1/\zeta_j)^2 - \zeta_j^2]^{1/2}$<br>$q_j = -2[(1+\zeta_j^2)/(1-\zeta_j^2)]^{1/2}\ln c_j$ | $p(\xi) = -2i/\boldsymbol{\kappa}\pi)\ln\{1 - \vert b\vert^2\}$<br>$q(\xi) = \arg b(\xi)$<br>$p_j = 2i\zeta_j \quad \bar{p}_j = -2i\zeta_j^*$<br>$q_j = -(2/\boldsymbol{\kappa})\ln c_j$<br>$\bar{q}_j = -(2/\boldsymbol{\kappa})\ln c_j^*$ | $p(\xi) = (-1/\pi\xi)\ln\{1 - \vert b\vert^2\}$<br>$q(\xi) = \arg b(\xi)$<br>$p_j = \ln\zeta_j \quad \bar{p}_j = \ln\zeta_j^*$<br>$q_j = -\ln c_j^2 \quad \bar{q}_j = -\ln c_j^{*2}$ | $p(\xi) = (\xi/\pi)\ln\{1 + \vert b\vert^2\}$<br>$q(\xi) = \arg b(\xi)$<br>$p_j = -\zeta_j^2$<br>$q_j = 2\ln[ic_j a^{\dagger}(\zeta_j)]$ |

Here we take a different approach and display the canonical structure of the inverse method. We consider four equations of particular physical interest. Each of these nonlinear equations is written as a Hamiltonian system. The linear problem is then interpreted as a canonical transformation which maps this Hamiltonian system into the spectral data viewed as a Hamiltonian system of "action-angle" type.[4] Such "action-angle" Hamiltonians do not depend explicitly upon the $q$'s. As such, the canonical equations of motion for the spectral data are trivially integrable. These yield $p$'s which are constants of the motion and $q$'s which vary linearly with time.

This canonical structure provides a clear interpretation of the infinite number of conservation laws which these nonlinear equations possess. In addition, calling upon our experience with action angle transformations in classical mechanics, we realize that such systems are quite special. This provides some feeling for the sensitivity of the inverse method to changes in the nonlinear wave equation. Finally, these canonical structures would seem of interest as models for quantized field theories.[17]

Our results are displayed in Table I. The Korteweg–de Vries (KdV) column was obtained by Zakharov and Faddeev[5] in a fundamental paper. Our arguments for the other three equations closely follow those of Ref. 5. However, it is now clear that both the Toda lattice and the nonlinear Schödinger equation provide a better model than KdV for a study of the canonical structure of the inverse method because the computations involved are more explicit. Also, it now seems clear that the inverse method can quite generally be interpreted as a canonical map from the nonlinear dynamics to a Hamiltonian system of action angle type. Currently, Flaschka and Newell[6,16] are investigating this generality with regard to the Clarkson inverse formalism.[2]

For the rest of this note we go through Table I indicating the type of computations and discussing several distinctions between the various equations. In line 1 (of Table I) we display the nonlinear dynamics. In lines 2 and 3 this dynamics is expressed in canonical form. Notice that in the sine-Gordon and KdV cases, explicit $p$'s and $q$'s are not identified. Such an identification can be made at the expense of embedding the structure in a larger space[1,6,7]; however, this is not necessary.[5,8] It

TABLE II. The linear problems.

| | Toda lattice | Nonlin. Schröd. | Sine—Gordon | Korteweg—de Vries |
|---|---|---|---|---|
| 1. Linear prob. "$L\Phi=\lambda\Phi$" | $(L\Phi)(n)\equiv\alpha_{n-1}\Phi(n-1)$ $+\alpha_n\Phi(n+1)+\beta_n\Phi(n)$,<br>$\alpha_n\equiv\frac{1}{2}\exp(\Delta_n/2)$<br>$\beta_n\equiv-\frac{1}{2}P_{n-1}$, $\Delta_n\equiv Q_{n-1}-Q_n$ | $\Phi=\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}$<br>$L\Phi\equiv\begin{pmatrix}\partial_x & -iu/2\\ +\frac{iu^*}{2} & -\partial_x\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}$ | $\Phi=\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}$<br>$L\Phi\equiv\begin{pmatrix}\partial_x & v/2\\ v/2 & -\partial_x\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}$ | $L\Phi=-\partial_{xx}\Phi+u\Phi$ |
| 2. Spectrum of $L$ | $2\lambda=\zeta+\zeta^{-1}$ $\zeta=\xi+i\eta$ | $\lambda=-i\zeta$ $\zeta=\xi+i\eta$ | $\lambda=-i\zeta$ $\zeta=\xi+i\eta$ | $\lambda=\zeta^2$ $\zeta=\xi+i\eta$ |
| Point | $\zeta_j\in(-1,1)$ $(j=1,2,\ldots N)$ | $\zeta_j$ $(j=1,2,\ldots,N)$ | $\zeta_j$ $(j=1,2,\ldots,N)$ | $\zeta_j$ $(j=1,2,\ldots,N)$ |
| Continuous | $\zeta=e^{i\varphi}$ $\forall$ $\phi\in[0,2\pi)$ | $\zeta=\xi$ $\forall$ $\xi\in(-\infty,\infty)$ | $\zeta=\xi$ $\forall$ $\xi\in(-\infty,\infty)$ | $\zeta=\xi$ $\forall$ $\xi\in(-\infty,\infty)$ |
| 3. Scattering bound. cond. | $\psi(n,\zeta)\approx\zeta^n$ as $n\to+\infty$ | $\psi(x,\xi)\approx\begin{pmatrix}0\\1\end{pmatrix}e^{i\xi x}$ as $x\to+\infty$ | $\psi(x,\xi)\approx\begin{pmatrix}0\\1\end{pmatrix}e^{i\xi x}$ as $x\to+\infty$ | $\psi(x,\xi)\approx e^{i\xi x}$ as $x\to+\infty$ |
| "Jost Solutions" | $\phi(n,\zeta)\approx\zeta^{-n}$ as $n\to-\infty$ | $\phi(x,\xi)\approx\begin{pmatrix}1\\0\end{pmatrix}e^{-i\xi x}$ as $x\to-\infty$ | $\phi(x,\xi)\approx\begin{pmatrix}1\\0\end{pmatrix}e^{-i\xi x}$ as $x\to-\alpha$ | $\phi(x,\xi)\approx e^{-i\xi x}$ as $x\to-\infty$ |
| 4. Def. of the coefficients $(a,b,c)$ | $\phi(n,\zeta)=b(\phi)\psi(n,\zeta)$ $+a(\phi)\psi(n,\zeta^{-1})$, $\zeta=e^{i\phi}$ | $\phi(x,\xi)=a(\xi)\bar{\psi}(x,\xi)$ $+b(\xi)\psi(x,\xi)$<br>$\bar{\psi}\equiv\begin{pmatrix}\bar{\psi}_2^*\\ -4_1^*\end{pmatrix}$ | $\phi(x,\xi)=a(\xi)\bar{\psi}(x,\xi)$ $+b(\xi)\psi(x,\xi)$<br>$\bar{\psi}=\begin{pmatrix}\psi_2^*\\ -\psi_1^*\end{pmatrix}$ | $\phi(x,\xi)=a(\xi)\psi^*$ $-b^*(\xi)\psi(x,\xi)$ |
| | $c_j$ denotes a normalization of the $j$th eigenvector | $c_j$ denotes a normalization of the $j$th eigenvector | $c_j$ denotes a normalization of the $j$th eigenvector | $c_j$ denotes a normalization of the $j$th eigenvector |
| 5. Scattering data | $\{\zeta_j\ \ c_j\ \ b(\phi)\}$ | $\{\zeta_j\ \ c_j\ \ b(\xi)\}$ | $\{\zeta_j\ \ c_j\ \ b(\xi)\}$ | $\{\zeta_j\ \ c_j\ \ b(\xi)\}$ |
| 6. The evolution of the scattering data | $\dot{b}=(\zeta-\zeta^{-1})b$<br>$\dot{\zeta}_j=0$<br>$\dot{c}_j=\frac{1}{2}(\zeta_j-\zeta_j^{-1})c_j$ | $\dot{b}=4i\xi^2 b$<br>$\dot{\zeta}_j=0$<br>$\dot{c}_j=4i\zeta_j^2c_j$ | $\dot{b}=-\frac{i}{2\xi}b$<br>$\dot{\zeta}_j=0$<br>$\dot{c}_j=-(i/2\zeta_j)c_j$ | $\dot{b}=8\zeta_j^3C_j$<br>$\dot{\zeta}_j=0$<br>$\dot{c}_j=8\zeta^3C_j$ |
| 7. A basic identity | $\ln a(\zeta)=\sum_{j=1}^N\ln\frac{\zeta_j(\zeta_j-\zeta)}{1-\zeta\zeta_j}$<br>$+\frac{i}{\pi}\int_0^{2\pi}d\phi\frac{e^{i\varphi}}{e^{i\varphi}-\zeta}$<br>$\times\ln\lvert a(e^{i\phi})\rvert$ | $\ln a(\zeta)=\sum_{j=1}^N\ln\left(\frac{\zeta-\zeta_j}{\zeta-\zeta_j^*}\right)$<br>$-\frac{1}{\pi i}\int_{-\infty}^{\infty}\frac{\ln\lvert a(\xi')\rvert}{\zeta-\xi'}d\xi'$ | $\ln a(\zeta)=\sum_{j=1}^N\ln\left(\frac{\zeta-\zeta_j}{\zeta-\zeta_j^*}\right)$<br>$-\frac{1}{\pi i}\int_{-\infty}^{\infty}\frac{\ln\lvert a(\xi')\rvert}{\zeta-\xi'}d\xi'$ | $\ln a(\zeta)=\sum_{j=1}^N\ln\left(\frac{\zeta-\zeta_j}{\zeta-\zeta_j^*}\right)$<br>$-\frac{1}{\pi i}\int_{-\infty}^{\infty}\frac{\ln\lvert a(\xi')\rvert}{\zeta-\xi'}d\xi'$ |
| 8. "$\mathcal{H}=\mathcal{K}$" | $\ln a(\zeta)\approx\sum_{j=0}^{\infty}\zeta^jA_j$, $\zeta\to0$<br>$A_2=-\mathcal{K}=-\mathcal{H}$ | $\ln a(\zeta)\approx\sum_{j=1}^{\infty}\zeta^{-j}A_j$, $\zeta\to\infty$<br>$A_3=-(\mathfrak{X}/16)\mathcal{K}=$ $-(\mathfrak{X}/16)\mathcal{H}$ | $\ln a(\zeta)\approx\sum_{j=1}^{\infty}\zeta^jA_j$, $\zeta\to0$<br>$A_1=-i\mathcal{K}=-i\mathcal{H}$ | $\ln a(\zeta)\approx\sum_{j=1}^{\infty}\zeta^{-j}A_j$, $\zeta\to\infty$<br>$A_5=-(16i)^{-1}\mathcal{K}=-(16i)^{-1}\mathcal{H}$ |

should be remarked that for the Clarkson formalism, this embedding seems natural.[6,16] Finally notice the Hamiltonian is not the standard one, $h$, used to describe the sine-Gordon dynamics in laboratory coordinates $(X, T)$,

$$h=\int_{-\infty}^{\infty}[\tfrac{1}{2}(Q_T^2+Q_X^2)+(\cos Q-1)]\,dx$$
$$=\int_{-\infty}^{\infty}[\tfrac{1}{2}(P^2+Q_X^2)+(\cos Q-1)]\,dX. \tag{1}$$

In lines 4 and 5 of Table I, we display the action angle Hamiltonian dynamics. The canonical maps connecting the two Hamiltonian systems are described by the linear eigenvalue problems which are summarized in Table II. This table is self-explanatory. Its entrees were obtained by Gardner, Greene, Kruskal, and Muira[9] for KdV, by Zakharov and Shabat[10] for nonlinear Schrödinger, by Ablowitz, Kaup, Newell, and Segur[11] for sine—Gordon, and by Flaschka[12,13] for the Toda lattice. Notice in particular that the scattering data listed in line 5 is defined through the Jost solutions of line 3. The evolution of this scattering data is summarized in line 6.

The existence of a one-to-one map between the scattering data and the nonlinear wave dynamics follows from Gel'fand—Levitan theory. In addition, the identifications made in line 6 of Table I show that the evolution of the scattering data (line 6, Table II) can be represented through the action angle Hamiltonian systems listed in lines 3 and 4 of Table I. Thus, the proof that these maps are canonical transformations rests on the verification that $H$ is mapped into $K$.

In all four cases, this is established by expansions

$$\ln a(\zeta)\approx\sum_{i=-\infty}^{\infty}A_i\zeta^{-i}. \tag{2}$$

The coefficients $A_i$ can be represented in two ways. One way begins from the identity displayed in line 7 of Table II, which follows from Cauchy's theorem and the identity $|a|^2+|b|^2=1$. The main results of this computation are displayed in line 8. In particular, the action angle Hamiltonian $K$ is always one of these coefficients. Alternatively, the coefficients $A_i$ can be expressed in terms of the nonlinear waves. In all cases, $H=K$.

The explicit calculations for $A_i$ in terms of the nonlinear waves differs slightly between the sine—Gordon case and the other three examples. For the sine—Gordon case, $H=K=iA_{-1}$. The negative index indicates that we need $\ln a(\zeta)$ near $\zeta=0$. This is reminiscent of scattering length computation in quantum mechanics.[14] Following closely the type of arguments used in quantum texts, we obtain the identity

$$\frac{d}{d\zeta}\ln a(\zeta)=-i\int_{-\infty}^{\infty}[\Phi_1\psi_2+\Phi_2\psi_1-a]\,dx. \tag{3}$$

For the sine—Gordon case, it is easy to compute the Jost solutions at $\zeta=0$,

$$\psi=\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}=\begin{pmatrix}\sin\int_{-\infty}^{x}\frac{v(x')}{2}\,dx'\\ \cos\int_{-\infty}^{x}\frac{v(x')}{2}\,dx'\end{pmatrix},$$
$$\varphi=\begin{pmatrix}\varphi_1\\ \varphi_2\end{pmatrix}=\begin{pmatrix}\cos\left(\int_{-\infty}^{x}\frac{v(x')}{2}\,dx'\right)\\ -\sin\left(\int_{-\infty}^{x}\frac{v(x')}{2}\,dx'\right)\end{pmatrix}. \tag{4}$$

Inserting (4) into (3) yields $A_{-1}$ and explicitly shows that $iA_{-1}=H$. For the other three cases, $H$ is obtained from an expansion of $\ln a$ by the same expansion procedure used in Ref. 5. This distinction can be traced to the pole in the dispersion relation for the linearized sine—Gordon equation. In fact, the analytic structure of this dispersion relation seems to determine which coefficients in the expansion of $\ln a(\zeta)$ yield $H$.[6]

For the sine—Gordon and KdV cases one must also check the invariance of the simpletic form. For KdV this is done in Ref. 5. Presumably the sine—Gordon case would follow analogously. This check is not necessary for the Toda lattice and nonlinear Schrödinger cases since there explicit $p$'s and $q$'s have been identified.

Canonical structure provides a clear description of the infinite number of constants of the motion. Clearly the action angle momentum $(p(\xi), p_j)$ constitute "$n$" constants of the motion in our "$2n$"-dimensional phase space. On the other hand, notice that $a(\zeta)$ is constant in time; thus, by (2), the coefficients $A_i$ are also constant in time. In fact, these $A_i$ represent the celebrated infinite number of conservation laws (polynomial conservation laws for KdV.) Specifying these $A_i$ fixes $\ln a(\zeta)$ by (2), which in turn fixes $(p(\xi), p_i)$.

Given an infinite number of conservation laws, the exact meaning of their independence is always a problem. Here, the canonical structure clearly displays the independence of the infinite set of constants of the motion $\{A_i\}$. Each of these can be specified arbitrarily and independently. Each such specification fixes the action angle momenta $(p(\xi), p_i)$ and thus fixes a class of solutions. [To uniquely select one solution from this class, one must fix $(q(\xi), q_i)$ initially.] Thus, fixing specific values for these constants of the motion $A_i$ which are obtained from the "infinite number of conservation laws" fixes a class of solutions. Changing one value $A_i$ changes this class. A detailed discussion of the points covered in the last two paragraphs will appear in Ref. 18.

Actually, for the KdV and sine—Gordon case, only one half of the $\{A_i\}$ can be arbitrarily specified, every other one being identically zero. Recall that in these two cases, the $p$'s and $q$'s for the nonlinear wave could only be identified at the expense of an embedding. The wave dynamics described by KdV (or by sine—Gordon) lives on a constraint in this larger space. In "scattering-data space" this constraint takes the form $A_{2j} \equiv 0$, $j =$ integer.

## ACKNOWLEDGMENTS

It is my pleasure to acknowledge many beneficial conversations with J. Corones, H. Flaschka, and A. Newell during the course of this work.

*Supported in part by NSF Grant GP37627, Iowa State University, Ames, Iowa, and in part by Office of Naval Research under Contract No. N00014-67-A-0467-0027 at the Appl. Math. Summer Institute.

[1]A.C. Scott, F.Y.F. Chu, and D.W. McLaughlin, Proc. IEEE 61, 1443 (1973).
[2]M. Ablowitz, K. Kaup, A. Newell, and H. Segar, to appear, Studies in Appl. Math.
[3]D. McLaughlin and J. Corones, Phys. Rev. A, Nov.-Dec. 1975.
[4]H. Goldstein, *Classical Mechanics* (Addison-Wesley, London, 1950), pp. 285—307.
[5]V.E. Zakharov and L.D. Faddeev, Funct. Anal. Appl. 5, 280 (1971).
[6]H. Flaschka and A. Newell, private communication.
[7]C.S. Gardner, J. Math. Phys. 12, 1548 (1971).
[8]V.I. Arnold and A. Avez, *Ergodic Problems in Mechanics* (Benjamin, New York, 1968), pp. 1—6.
[9]C.S. Gardner, J.M. Greene, M.D. Kruskal, and R.M. Miura, Phys. Rev. Lett. 19, 1095 (1967).
[10]V.E. Zakharov and A.B. Shabat, Zh. Eksp. Teor. Fiz. 61, 118 (1971) [Sov. Phys. JETP 34, 62 (1962)].
[11]M. Ablowitz, D. Kaup, A. Newell, and H. Sequr, Phys. Rev. Lett. 31, 125 (1973).
[12]H. Flaschka, Prog. Theor. Phys. 51, 703 (1974).
[13]H. Flaschka, Phys. Rev. B 9, 1924 (1974).
[14]A. Messiah, *Quantum Mechanics*, Vol. I (North-Holland, Amsterdam, 1964), pp. 406—408.
[15]V.E. Zakharov and A.B. Shabet, Funct. Anal. Appl. 8, No. 3, 43 (in Russian) (1974).
[16]H. Flaschka and A.C. Newell, to appear in *Proceedings of 1974 Battelle Summer Rencontres*.
[17]R. Dashen, B. Hasslacher, and A. Neveu, three preprints, Inst. for Adv. Study, Princeton (1974).
[18]D. McLaughlin, to appear Proceedings N.S.F. Workshop, Vanderbuilt Univ., Sept. (1974).

# A stochastic Gaussian beam. II

D. W. McLaughlin*

*Department of Mathematics, University of Arizona, Tucson, Arizona 85721*
(Received 19 August 1974)

The propagation of a Gaussian beam in a strongly focusing medium is considered. The medium is subject to random deformations of the beam axis. The average intensity and the intensity fluctuations on the beam axis and the mean population remaining in the fundamental mode are computed when the random inhomogeneities are weak and the distance between the source and observation points is large. All results for random axis deformations are compared to those obtained earlier for random width perturbations. The mean intensity off the beam axis and the mean population transfer into higher modes are also discussed.

## 1. INTRODUCTION

Consider the initial value problem for the Schrödinger equation,

$$2i\partial_t\psi=\Delta\psi-[\mathbf{y}-\epsilon\boldsymbol{\eta}(t,\omega)]^2\psi,\quad 0<\epsilon\ll 1,$$
$$\psi(t=0,\mathbf{y})=(1/\sqrt{\pi})\exp(-\mathbf{y}^2/2). \tag{1.1}$$

Here $\mathbf{y}\equiv(y_1,y_2)$, $\boldsymbol{\eta}=(\eta_1,\eta_2)$, and $\mathbf{y}^2=y_1^2+y_2^2$. Physically, $\psi$ represents an electromagnetic beam traveling in a focusing medium subject to random imperfections. The $t$ axis denotes the beam axis. In (1.1) the parabolic approximation to the reduced wave equation has been employed.[1] In this model the fundamental mode of the beam is initially excited.

The stochastic perturbation $\eta(t,\omega)$ models random deformations of the beam axis. It would correspond to misalignment problems in a lens system. In Ref. 1, we have studied a stochastic perturbation which describes random changes of the focusing strength, or "width of the lens"

$$2i\partial_t\psi=\Delta\psi-[1-\epsilon\beta(t,\omega)]\mathbf{y}^2\psi,$$
$$0<\epsilon\ll 1,\quad \beta \text{ a scalar process},\tag{1.2}$$
$$\psi(t=0,\mathbf{y})=(1/\sqrt{\pi})\exp(-\mathbf{y}^2/2).$$

One purpose of this note is to compare the effects of random axis deformations, as modeled by (1.1), with those of random width changes, as modeled by (1.2).

The process $\eta(t,\omega)$ is taken to be mean zero and stationary,

$$E\{\eta(t,\cdot)\}=0,\quad E\{\eta_i(t,\cdot)\eta_j(s,\cdot)\}=R_{ij}(t-s). \tag{1.3}$$

In addition the usual mixing conditions are assumed, in particular,

$$R_{ij}(s)\to 0 \text{ as } s\to\infty \text{ sufficiently rapidly}. \tag{1.4}$$

The goal of this note is to compute the expected value of certain functions of the random field $\psi$ [as defined by (1.1)] in the limit of weak stochastic perturbations ($0<\epsilon\ll 1$), uniformly through large beam distance ($t=O(\tau/\epsilon^2)$, $\tau$ finite). All results for random axis deformations are compared with those under random width changes. It should be noted that in all cases rather explicit formula are obtained.

In particular, we compute the expected value of the intensity,

$$E\{|\psi(t,\mathbf{y})|^2\}, \tag{1.5a}$$

the fluctuation of the intensity,

$$E\{[|\psi(t,\omega)|^2-E\{|\psi(t,\omega)|^2\}]^2\}; \tag{1.5b}$$

and that portion of the beam which remains in the fundamental mode,

$$E\{|\langle h_{00}|\psi(t)\rangle|^2\}=E\{|\int d^2y(1/\sqrt{\pi})\exp(-\mathbf{y}^2/2)\psi(t,\mathbf{y})|^2\}. \tag{1.5c}$$

In addition, the expected modal transfer into the $(p-q)^{\text{th}}$ mode $h_{pq}$ is studied,

$$E\{|\int d^2y\, h_{pq}(\mathbf{y})\psi(t,\mathbf{y})|^2\}. \tag{1.5d}$$

Marcuse[2] has studied an initial-boundary value problem closely related to (1.1)—(1.2). His problem is more realistic than that studied here; consequently, his approach (a modal analysis) is harder to justify with estimates of accuracy. We view the present work, as well as the work in Refs. 1 and 3, as complementary to studies such as Ref. 2. By using an idealized initial model, we are able to obtain explicit results within that model and to give formal error estimates within the model.

Finally, we mention that (1.1) and (1.2) could also be interpreted as quantum mechanical harmonic oscillators in which the ground (lowest energy) state is excited at $t=0$. The two types of randomness described would (i) shake the equilibrium position of the oscillator, and (ii) shake the "spring coefficient." The problem is to compute the manner in which each type of randomness distributes the population from the lowest energy state into higher ones.

In Sec. 2, we describe the key representation. In Sec. 3, we discuss the average intensity and its fluctuation. In Sec. 4, we describe the modal transfer.

## 2. THE KEY REPRESENTATION

For the moment, fix a realization of the process $\eta$, and hence of the random field $\psi$ as defined by (1.1). Motivated by geometrical optics, we seek $\psi$ of the form

$$\psi(t,\mathbf{y})=A(t)\exp[iS(t,\mathbf{y})]. \tag{2.1}$$

A short computation based upon this ansatz verifies that $\psi$ can be written as

$$\psi(t,\mathbf{y})=\frac{1}{\sqrt{\pi}}\exp(it)\exp(-\mathbf{y}^2/2)\exp[ie^{it}\Delta_j(t)y^j+i\delta(t)], \tag{2.2}$$

where $\Delta(t,\omega)$ and $\delta(t,\omega)$ are defined by

$$\partial_t \Delta = -\epsilon \exp(-it)\eta, \quad \Delta(t=0)=0, \tag{2.3a}$$

$$\partial_t \delta = \tfrac{1}{2}\exp(2it)\Delta^2 + \tfrac{1}{2}\epsilon^2\eta^2, \quad \delta(t=0)=0. \tag{2.3b}$$

Integrating Eqs. (2.3) and inserting the result into (2.2) yields the basic representation

$$\psi(t,\mathbf{y}) = (1/\sqrt{\pi})\exp(it)\exp(-\mathbf{y}^2/2)\exp[\epsilon A_j(t)y^j + \epsilon^2 B(t)], \tag{2.4}$$

where $A_j = A_j^R + iA_j^I$, $B = B^R + iB^I$, and

$$A_j^R \equiv \int_0^t \sin(t-\sigma)\eta_j(\sigma)d\sigma, \tag{2.5a}$$

$$A_j^I \equiv -\int_0^t \cos(t-\sigma)\eta_j(\sigma)d\sigma, \tag{2.5b}$$

$$B^R \equiv -\tfrac{1}{2}\int_0^t d\sigma \int_0^\sigma d\sigma_1 \int_0^\sigma d\sigma_2 \times \sin(2\sigma-\sigma_1-\sigma_2)\eta_k(\sigma_1)\eta^k(\sigma_2), \tag{2.5c}$$

$$B^I \equiv \tfrac{1}{2}\int_0^t d\sigma(\eta^2(\sigma) + \int_0^\sigma d\sigma_1 \int_0^\sigma d\sigma_2 \times \cos(2\sigma-\sigma_1-\sigma_2)\eta_k(\sigma_1)\eta_k(\sigma_2)). \tag{2.5d}$$

All computations in this note are based upon the key representation (2.4)—(2.5). For convenience, we list those quantities of primary interest (1.5a,b,c,d), in terms of $\mathbf{A}^R$, $\mathbf{A}^I$, $B^R$, and $B^I$:

$$I(t,\mathbf{y}) \equiv |\psi(t,\mathbf{y})|^2 = (1/\pi)e^{-\mathbf{y}^2}\exp[\epsilon 2A_j^R(t)y^j + \epsilon^2 2B^R(t)], \tag{2.6a}$$

$$J_{pq}(t) \equiv |\langle h_{pq}|\psi(t,\cdot)\rangle|^2 = \frac{\epsilon^{2(p+q)}|A_1|^{2p}|A_2|^{2q}}{2^{(p+q)}p!\,q!}\exp(\epsilon^2\{2B^R(t) + [\mathbf{A}^2(t) + \mathbf{A}^{*2}(t)]/4\}),$$
$$p,q = 0,1,2,\cdots. \tag{2.6b}$$

$J_{pq}(t)$, Eq. (2.6b), denotes the modal transfer function from the fundamental $(h_{00})$ mode at $t=0$ to the $pq$th mode $(h_{pq})$ at $t=t$. To obtain (2.6b) requires some computation. First realize that the mode $h_{pq}$ is defined by[1]

$$h_{pq}(\mathbf{y}) \equiv (\pi 2^p p!\,2^q q!)^{-1/2}H_p(y_1)H_q(y_2)\exp(-\mathbf{y}^2/2), \tag{2.7a}$$

where $H_p(y_1)$ denotes the $p$th Hermite polynomial.[4] Inserting $h_{pq}$ and $\psi$ as represented by (2.4) into (1.5d) and using the two identities,[4]

$$\int_{-\infty}^{\infty} dx\, H_p(x+y)e^{-x^2} = 2^{-p/2} = \sum_{k=0}^{[p]}\binom{p}{2k}H_{p-2k}(\sqrt{2}y)\left(\sqrt{\pi}\,\frac{(2k)!}{k!}\right), \tag{2.7b}$$

$$\sum_{k=0}^{[p]}[k!\,(p-2k)!]^{-1}H_{p-2k}(y) = \frac{(2y)^p}{p!}, \tag{2.7c}$$

yields (2.6b). In (2.7b,c), $[p]\equiv p$ if $p$ is even and is $(p-1)$ if $p$ is odd; $\binom{p}{q}$ denotes the binomial coefficient.

In the rest of this paper, we compute expected values of (2.6a,b) in the limit as $\epsilon \to 0$ uniformly in $t$ through $t = O(\tau/\epsilon^2)$, $\tau$ finite. Frequently such computations are carried out by a perturbation theory using some form of Khashminskii's diffusion theorem. This technique is used in Ref. 1 for the case of random width perturbations. Here key representation (2.4)—(2.5) is so explicit that such techniques are not needed. In comparing with Ref. 1, notice there that Eqs. (2.3a,b) are replaced by a variable coefficient linear differential equation whose solution is not known in closed form. This ordinary differential equation is analyzed by Khashminskii's theorem. Here we merely integrate (2.3a,b).

## 3. THE INTENSITY AND ITS FLUCTUATION

In this section, we compute the expected value of the intensity and its fluctuation and compare these results for random axis deformations with the analogous results for random width changes. The parabolic approximation (1.1) to the reduced wave equation is most accurate on the beam axis $(\mathbf{y}\equiv 0)$.[5] There the intensity is given by

$$I(t,\mathbf{y}\equiv 0) = (1/\pi)\exp[2\epsilon^2 B^R(t,\omega)]. \tag{3.1}$$

For finite, fixed $t$, the expected value of the on-axis intensity remains constant,

$$E\{I(t,\mathbf{y}\equiv 0)\} = 1/\pi + O(\epsilon^2), \quad t \text{ fixed}. \tag{3.2}$$

However, as $t$ increases, this expected value will suffer large distance effects. To compute these effects, consider the definition of $B^R(t,\omega)$, Eq. (2.5c),

$$B^R(t,\omega) \equiv -\tfrac{1}{2}\int_0^t d\sigma \int_0^\sigma d\sigma_1 \int_0^\sigma d\sigma_2 \times \sin(2\sigma-\sigma_1-\sigma_2)\eta_k(\sigma_1)\eta^k(\sigma_2). \tag{3.3}$$

Interchanging integrals and evaluating the integral over $\sigma$ first yields

$$B^R(t,\omega) = \tfrac{1}{4}\int_0^t d\sigma_1 \int_0^t d\sigma_2 \times[\cos(2t-\sigma_1-\sigma_2) - \cos(\sigma_1-\sigma_2)]\eta_k(\sigma_1)\eta^k(\sigma_2). \tag{3.3'}$$

We are interested in the behavior of (3.3) for large $t$; therefore, we consider the time average

$$\overline{B^R} \equiv \lim_{t\to\infty}\frac{1}{4t}\int_0^t d\sigma_1 \int_0^t d\sigma_2 \times[\cos(2t-\sigma_1-\sigma_2) - \cos(\sigma_1-\sigma_2)]\eta_k(\sigma_1)\eta^k(\sigma_2). \tag{3.4}$$

To compute this limit, we assume that the process $\eta$ belongs to the class of stochastic processes for which this limit is the same for almost all realizations.[6] For this class, the limit $\overline{B^R}$ is equal to its mean value. Thus,

$$\overline{B^R} = E\{\overline{B^R}\} = \lim_{t\to\infty}\frac{1}{4t}\int_0^t d\sigma_1 \int_0^t d\sigma_2 \times[\cos(2t-\sigma_1-\sigma_2) - \cos(\sigma_1-\sigma_2)]E\{\eta_k(\sigma_1)\eta^k(\sigma_2)\}. \tag{3.5}$$

Using the stationarity of the process $\eta$, we compute from (3.5)

$$\overline{B^R} = -\tfrac{1}{2}\int_{-\infty}^{\infty} d\sigma \cos\sigma\{\mathrm{tr}[R(\sigma)]\},$$
$$R_{jk}(\sigma) \equiv E\{\eta_j(\sigma+s)\eta_k(s)\}. \tag{3.6}$$

Assumption (1.4) guarantees that this distance average is finite.

Returning to (3.1), we use (3.5) to calculate the large distance behavior of $I(t,0)$ in the limit $\epsilon\to 0$, $t\to\infty$, $\epsilon^2 t = \tau$ finite,

$$I(t,0) \approx (1/\pi)\exp(-2\epsilon^2 t|\overline{B^R}|), \quad t \gg 1, \tag{3.7}$$

and, since $\overline{B^R}$ is independent of realization, we obtain

$$E\{I(t,0)\} \approx (1/\pi)\exp(-2\epsilon^2 t|\overline{B^R}|) + O(\epsilon),$$

RANDOM AXIS CASE. (3.8A)

Expression (3.8) is uniformly valid through $t = O(\tau/\epsilon^2)$. Certainly for finite, fixed $t$ it reduces to (3.2).

Result (3.8) is a main result in this section. It states that weak random $[0(\epsilon)]$ axis deformations cause an $O(1)$

exponential decay in the distance scale $\tau=\epsilon^2 t$ of the on-axis intensity, with decay constant $2|\overline{B^R}|$. In Ref. 1 we found that weak random $[O(\epsilon)]$ width changes left the intensity unperturbed

$$E\{I(t,0)\}\approx 1/\pi+O(\epsilon). \tag{3.8W}$$

uniformly through $t=O(\tau/\epsilon^2)$, $\tau$ finite.

Thus, random axis deformations induce a stronger effect on the intensity than do random width variations.

Experimentally misalignment causes more problems than width uncertainties.[7] In Ref. 5, we have shown that misalignment causes larger errors in the parabolic approximation to the reduced wave equation than random width variations. Here we see that even within the framework of the parabolic approximation, misalignment causes a stronger effect on the intensity than does random width variations.

Using exactly the same procedure, we compute the fluctuation in the on-axis intensity,

$$E\{[I(t,0)-E\{I(t,0)\}]^2\}=0+O(\epsilon),\quad \text{RANDOM AXIS CASE}, \tag{3.9A}$$

uniformly through $t=O(\tau/\epsilon^2)$, $\tau$ finite. At first glance this result is somewhat surprising. However, it is easily understood after noticing that the random function $I(t,0)$ depends upon the input stochastic process $\eta$ only through "averages" such as

$$\epsilon^2 t\frac{1}{t}\int_0^t d\sigma_1\int_0^t d\sigma_2\eta_k(\sigma_1)\eta^k(\sigma_2)\cos(\sigma_1-\sigma_2).$$

Thus, ergodicity will force the limit ($\epsilon\to 0$, $t\to\infty$, $\epsilon^2 t=\tau$ finite) to be almost surely independent of realization. This in turn forces results such as (3.9A).

In any case, result (3.9A) is certainly different from the random width case, where, in Ref. [1], we found

$$E\{[I(t,0)-E\{I(t,0)\}]^2\}=(1/\pi^2)[\exp(2\epsilon^2\gamma t)-1]+O(\epsilon),$$
$$\text{RANDOM WIDTH CASE}, \tag{3.9W}$$

uniformly through $t=O(\tau/\epsilon^2)$, $\tau$ finite. Here

$$\gamma\equiv\int_0^\infty E\beta(s+\sigma)\beta(\sigma)\cos 2\sigma\,d\sigma.$$

Off axis ($y\neq 0$) the situation is more complicated. Not only is the parabolic approximation less accurate,[5] but calculations within the parabolic approximation seem more difficult. For example, consider computing the expected value of the intensity $I$,

$$I(t,\mathbf{y})\equiv(1/\pi)e^{-y^2}\exp[2\epsilon A_j^R(t,\omega)y^j+2\epsilon^2 B^R(t,\omega)]. \tag{3.10}$$

As we have seen, the term $\epsilon^2 B^R(t,\omega)\approx\epsilon^2 t\,\overline{B^R}$ as $t\to\infty$; and hence contributes to the large distance behavior of $E\{I(t,\mathbf{y})\}$. The term $\epsilon A_j^R(t,\omega)y^j$ in (3.10) also contributes to the large distance behavior, although its effect is harder to calculate. It is easy to see that it does not contribute in a distance of $O(1/\epsilon)$:

$$\begin{aligned}A_j^R(t,\omega)&\approx t\lim_{t\to\infty}\frac{1}{t}\int_0^t\sin(t-\sigma)\eta_j(\sigma)d\sigma\\&\equiv t\,\overline{A_j^R}=t\,E\{\overline{A_j^R}\}\\&=t\lim_{t\to\infty}\frac{1}{t}\int_0^t\sin(t-\sigma)\,E\{\eta_j(\sigma)\}\,d\sigma=0\end{aligned} \tag{3.11}$$

by the mean zero property of $\eta_j$. To see that this term does contribute in a distance of $O(1/\epsilon^2)$, we expand (3.10) as a power series in $\epsilon$:

$$\begin{aligned}I(t,\mathbf{y})&\approx(1/\pi)e^{-y^2}\exp[2\epsilon^2 B^R(t,\omega)]\\&\times[1+2\epsilon A_j^R(t,\omega)y^j+2\epsilon^2 A_j^R(t,\omega)A_k^R(t,\omega)y^jy^k+O(\epsilon^3)].\end{aligned} \tag{3.12}$$

As $t\to\infty$, (3.11) shows that the $O(\epsilon)$ term in (3.12) goes to zero. As for the $O(\epsilon^2)$ term:

$$\begin{aligned}&A_j^R(t,\omega)A_k^R(t,\omega)\\&\approx t\lim_{t\to\infty}\frac{1}{t}\int_0^t d\sigma_1\int_0^t d\sigma_2\sin(t-\sigma_1)\sin(t-\sigma_2)\eta_j(\sigma_1)\eta_k(\sigma_2)\\&=t\lim_{t\to\infty}\frac{1}{t}\int_0^t d\sigma_1\int_0^t d\sigma_2\sin(t-\sigma_1)\sin(t-\sigma_2)\\&\quad\times E\{\eta_j(\sigma_1)\eta_k(\sigma_2)\}\\&=t\lim_{t\to\infty}\frac{1}{2t}\int_0^t d\sigma_1\int_0^t d\sigma_2\\&\quad\times[\cos(\sigma_1-\sigma_2)-\cos(2t-\sigma_1-\sigma_2)]R_{jk}(\sigma_1-\sigma_2)\\&=t\int_{-\infty}^\infty R_{jk}(\sigma)\cos\sigma\,d\sigma,\quad t\gg 0.\end{aligned} \tag{3.13}$$

Here we have again used the property that averages over $t$ are almost surely independent of realization. Combining (3.12)–(3.13) yields

$$\begin{aligned}E\{I(t,\mathbf{y})\}&\approx(1/\pi)e^{-y^2}\exp(-2\epsilon^2 t|\overline{B^R}|)\\&\times[1+2\epsilon^2 t\gamma_{jk}y^jy^k+O((\epsilon A(t)y)^3)],\end{aligned} \tag{3.14}$$

where

$$\gamma_{jk}\equiv\int_{-\infty}^\infty R_{jk}(\sigma)\cos\sigma\,d\sigma.$$

Formula (3.14) again shows that random axis deformations lead to an exponential decaying factor in the intensity as $t$ increases. In addition, for any fixed $\epsilon^2 t>0$, (3.14) shows an increase in the intensity just off the beam axis. The random axis deformations cause a local spreading of the beam on the $\tau=\epsilon^2 t$ distance scale.

One is tempted to try to correct the secular behavior displayed in (3.14) by some "two-timing" scheme such as discussed in Ref. 6. However, scaling the distance $t$ leads to results such as

$$\begin{aligned}E\{I(t,\mathbf{y})\}&\approx(1/\pi)\exp(-2\epsilon^2 t|\overline{B^R}|)\\&\times\exp[-(\delta_{jk}-2\epsilon^2 t\gamma_{jk})y^jy^k].\end{aligned}$$

Comparing this result with $I(t,y)$, Eq. (3.10), clearly indicates that it cannot be much more accurate than (3.14).

## 4. MODAL TRANSFER FUNCTION

In this section we compute the expected value of the modal transfer function $J_{pq}$, with particular emphasis upon that portion of the beam which remains in the fundamental mode, $J_{00}$. From (2.6b) we obtain

$$\begin{aligned}J_{00}(t)&=|\langle h_{00}|\psi(t)\rangle|^2\\&=\exp(\epsilon^2\{2B^R(t,\omega)+[\mathbf{A}^2(t,\omega)+\mathbf{A}^{2*}(t,\omega)]/4\}).\end{aligned} \tag{4.1}$$

Thus, we must compute the long distance behavior of $\mathrm{Re}[\mathbf{A}^2(t,\omega)]$. From (2.5a,b) we obtain

$$\mathbf{A}(t,\omega)=i\int_0^t e^{-i(t-\sigma)}\eta(\sigma)d\sigma,$$

$$\mathbf{A}^2(t,\omega)=-\int_0^t d\sigma_1\int_0^t d\sigma_2 \exp[-i(2t-\sigma_1-\sigma_2)]\eta_k(\sigma_1)\eta^k(\sigma_2).$$

Thus, $\mathrm{Re}[\mathbf{A}^2(t,\omega)]$ is given by

$$\mathrm{Re}[\mathbf{A}^2(t,\omega)]=(-)\int_0^t d\sigma_1\int_0^t d\sigma_2 \cos(2t-\sigma_1-\sigma_2)\eta_k(\sigma_1)\eta^k(\sigma_2)$$

As $t\to\infty$, this term is $O(1)$ as the following computation shows:

$$\begin{aligned}\mathrm{Re}[A^2(t,\omega)]&=-\int_0^t d\sigma_1\int_0^t d\sigma_2\cos(2t-\sigma_1-\sigma_2)\eta_k(\sigma_1)\eta^k(\sigma_2)\\&=-\int_0^t d\sigma_1\int_0^t d\sigma_2\cos(2t-\sigma_1-\sigma_2)\,\mathrm{tr}\,R(\sigma_1-\sigma_2)\\&=-\tfrac{1}{2}\int_0^t d\sigma_-\,\mathrm{tr}R(\sigma_-)\Big(\int_{\sigma_-}^{2t-\sigma_-}d\sigma_+\cos(2t-\sigma_+)\\&\quad+\int_{-\sigma_-}^{2t+\sigma_-}d\sigma_+\cos(2t-\sigma_+)\Big)\\&=-\sin 2t\int_0^t d\sigma\,\mathrm{tr}R(\sigma)\cos\sigma\\&=-\sin 2t\int_0^\infty d\sigma\,\mathrm{tr}R(\sigma)\cos\sigma\quad\text{as}\quad t\to\infty.\end{aligned}$$

In this computation, we have used the fact that averages over $t$ are almost surely independent of realization, stationarity, and property (1.4). Thus, we obtain

$$\mathrm{Re}[A^2(t,\omega)]\approx-\sin 2t\,\mathrm{tr}(\gamma_{jk})\quad\text{as}\quad t\to\infty,\tag{4.2}$$

where $\gamma_{jk}$ is defined by (3.14). Equation (4.2), together with the long distance behavior of $B^R(t,\omega)$ already calculated, yields

$$E\{J_{00}(t)\}\approx\exp[-2|\overline{B^R}|\epsilon^2 t-\tfrac{1}{2}\epsilon^2\,\mathrm{tr}(\gamma_{jk})\sin 2t]+O(\epsilon),$$

$$\text{RANDOM AXIS CASE,}\tag{4.3A}$$

uniformly in $t$ through $t=O(\tau/\epsilon^2)$, $\tau$ finite,

Result (4.3A) is the main result of this section. It shows that the population decays from the fundamental mode exponentially on the distance scale $\tau=\epsilon^2 t$ with decay constant

$$2|\overline{B^R}|=\int_0^\infty\cos\sigma\, E\{\eta_j(\sigma+s)\eta_j(s)\}.$$

Result (4.3A) should be compared with the random width case for which we found in Ref. 1

$$E\{J_{00}(t)\}\approx\frac{4\exp(-\gamma\epsilon^2 t/4)}{\sqrt{\pi}}\int_0^\infty\frac{\rho^2\exp(-\rho^2)}{\cosh\rho\sqrt{\gamma\epsilon^2 t}}\,d\rho+O(\epsilon),$$

$$\text{RANDOM WIDTH CASE,}\tag{4.3W}$$

uniformly through $t=0(\tau/\epsilon^2)$, $\tau$ finite, where

$$\gamma\equiv\tfrac{1}{2}\int_0^\infty\cos(2s)\,E\{\beta(s+\sigma)\beta(\sigma)\}ds.$$

We conclude with a few comments about the modal transfer from the fundamental mode at $t=0$ to the $pq$th mode at distance $t$,

$$\begin{aligned}J_{pq}(t)=&[\epsilon^{2(p+q)}(|A_1|^2)^p(|A_2|^2)^q/2^{(p+q)}p!q!]\\&\times\exp(\epsilon^2\{2B^R(t)+[\mathbf{A}^2(t)+\mathbf{A}^{*2}(t)]/4\})\end{aligned}\tag{4.4}$$

Notice that the over-all exponential decay factor is the same for all modes; however, the modes above the fundamental have an additional multiplicative factor. This factor forces the behavior

$$J_{pq}(0)=\begin{cases}1, & p=q=0,\\0, & \text{otherwise},\end{cases}$$

as it must. Also notice that, except for $(p=0,q=0)$, $(p=1,q=0)$, and $(p=0,q=1)$, the average value of $J_{pq}$ will depend upon higher moments of the process $\eta$. These will yield a behavior of the type

$$\begin{aligned}E\{J_{pq}(t)\}\approx\ & c_{pq}(\epsilon^2 t)^{p+q}\exp[-2|\overline{B^R}|\epsilon^2 t\\&-\tfrac{1}{2}\epsilon^2\,\mathrm{tr}(\gamma_{jk})\sin(2t)]+O(\epsilon),\quad\text{RANDOM AXIS CASE,}\end{aligned}\tag{4.5A}$$

uniformly through $t=O(\tau/\epsilon^2)$, $\tau$ finite,

where the constants $c_{pq}$ will depend upon higher moments of the process $\eta$.

Any given mode, once excited by the random process, decays exponentially with a decay rate which is common to all modes. The random disturbance initially excites the $(p-q)$th mode by transferring the population from the fundamental mode. This transfer process into the $(p-q)$th mode occurs as the power $(\epsilon^2 t)^{p+q}$. Thus, the maximum population for a given mode $p$-$q$ is achieved at a distance $\tau_{\max}=\epsilon^2 t_{\max}$ which is inversely proportional to the sum $(p+q)$. The population cascades away from the fundamental mode. Very similar results were obtained for the random width case in Ref. 3, although the dependence on higher moments did not appear there.

Finally, we remark that these results are actually more general than would appear. They do not depend in any essential way on the "ergodic-like" hypothesis on $\eta$ which we used for purposes of calculation. We could have used arguments of the Khasminskii type to see this. However, representation (2.4)–(2.5) is so explicit that the above arguments seemed the most direct means of computation.


## ACKNOWLEDGMENTS

It is my pleasure to acknowledge a conversation with Dr. J. Arnaud who told me of the "misalignment problem." Also, I had several beneficial conversations with Dr. J.B. Keller and Dr. G. Papanicolaou while these computations were in progress.

*Supported in part by NSF Grant GP37627, Iowa State University, Ames, Iowa, and in part by the Office of Naval Research under contract number N00014-67-A-0467-0027 at the Applied Mathematics Summer Institute.



[1]G.C. Papanicolaou, D. McLaughlin, and R. Burridge, J. Math. Phys. 14, 84 (1973).
[2]D. Marcuse, Bell Syst. Tech. J. 53, 195 (1974).
[3]D. McLaughlin and S. Nelson, manuscript in preparation.
[4]W. Magnus, F. Oberhetinger, and R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966).
[5]D. McLaughlin, manuscript in preparation.
[6]G.C. Papanicolaou and J.B. Keller, SIAM J. Appl. Math. 21, 287 (1971).
[7]J.A. Arnaud, private communication.

# Lorentz covariance of the Yukawa$_2$ quantum field theory

Oliver A. McBryan *

*Department of Mathematics, University of Toronto, Toronto, Canada M5S 1A1*[†]

Yong Moon Park

*Mathematical Science Department, IBM, Thomas J. Watson Research Center, Yorktown Heights, New York 10598*

(Received 10 May 1974)

We prove Lorentz covariance of the field algebras in the Yukawa$_2$ quantum field theory, thus completing the verification of the Haag–Kastler axioms for this model. We also study the energy–momentum tensor of the theory. In particular, we prove an estimate of the form $P(f)^2 \le \text{const}(H(g)+I)^2$ dominating local momenta by local Hamiltonians. Most of the results are a consequence of this estimate. These methods apply also to the $(\phi^4)_2$ model.

## 1. INTRODUCTION AND RESULTS

The Yukawa$_2$ quantum field theory is described by fields $\phi(x,t)$, $\psi(x,t)$ representing a boson and a fermion, respectively. The inhomogeneous Lorentz group in two space—time dimensions is the three-parameter group

$$\{a,\Lambda_\beta\}(x,t)=(a_1+x\cosh\beta+t\sinh\beta,\ a_0+t\cosh\beta+x\sinh\beta).$$

Lorentz covariance of the theory requires that corresponding to each transformation $\{a,\Lambda_\beta\}$ there is a unitary operator $U(a,\beta)$ such that

$$U(a,\beta)\phi(h)U(a,\beta)^{-1}=\phi(\{a,\Lambda_\beta\}h), \tag{1.1}$$

$$U(a,\beta)\psi(h)U(a,\beta)^{-1}=S(\Lambda_\beta)\psi(\{a,\Lambda_\beta\}h). \tag{1.2}$$

Here $\phi(h)=\int dx\,dt\,h(x,t)\phi(x,t)$, $\psi(h)=\int dx\,dt\,h(x,t)\psi(x,t)$, $h\in\mathcal{S}(R^2)$, and $\{a,\Lambda_\beta\}h$ is defined by $\{a,\Lambda_\beta\}h(x,t)=h(\{a,\Lambda_\beta\}^{-1}(x,t))$. $S(\Lambda_\beta)$ is the $2\times2$ matrix

$$S(\Lambda_\beta)=e^{\beta\gamma_5/2},\quad \gamma_5=\begin{bmatrix}1&0\\0&-1\end{bmatrix}.$$

In this paper we show how to construct such a unitary operator $U(a,\beta)$ for fields localized in a bounded region $B$ of space—time.

*Theorem* 1.1: For each Lorentz rotation $\{a,\Lambda_\beta\}$ and bounded region $B$ of space—time, there is a unitary operator $U(a,\beta)$ satisfying (1.1) and (1.2) whenever $\operatorname{supp}h\subset B$.

Lorentz covariance of the infinite volume field algebras now follows immediately, as in Ref. 1 for the $\phi_2^4$ theory. This completes the verification of the Haag—Kastler axioms for the Yukawa$_2$ model.

The Eqs. (1.1) and (1.2) are valid as identities for self-adjoint operators and for bounded operators, respectively. $U(a,\beta)$ may be chosen to be strongly continuous in $a,\beta$ on any closed interval.

For notation and estimates for the Yukawa$_2$ model see Refs. 2—6. We start with the time-zero symmetric energy—momentum tensor $T_{\mu\nu}=T_{\nu\mu}$:

$$T_{00}(x)=T_0(x)+g(x)T_I(x)+g^2(x)T_C(x),$$

$$T_{01}(x)=P(x),$$

where $g(\cdot)\geq 0\in C_0^\infty(R)$ and

$$T_0(x)=\tfrac12:\pi^2+(\nabla\phi)^2+m^2\phi^2:-\tfrac12:\bar\psi\gamma^1\nabla\psi-\nabla\bar\psi\gamma^1\psi+2M\bar\psi\psi:,$$

$$T_I(x)=:\lambda\bar\psi\psi\phi:$$

$$T_C(x)=-E-\frac{\delta m^2}{2}:\phi^2:,$$

$$P(x)=-\tfrac12:\pi\nabla\phi+\nabla\phi\pi:-\tfrac12:\bar\psi\gamma^0\nabla\psi-\nabla\bar\psi\gamma^0\psi:. \tag{1.3}$$

Here $E$ and $\delta m^2$ are the infinite vacuum energy density and boson mass renormalization counter terms. For real $f$, $f=1$ or $f\in\mathcal{S}(R)$, we define

$$T(f)=\int dx f(x)T_{00}(x),\quad P(f)=\int dx f(x)T_{01}(x),$$

$$M(f)=T(xf)\quad (f\neq1). \tag{1.4}$$

$T(f)$ is given by a limit as a momentum cutoff $\kappa$ is removed, see Sec. 2. In particular, $T(1)$ is the Yukawa$_2$ Hamiltonian with a space-cutoff $g(\cdot)$:

$$H\equiv H(g)=T(1)=H_0+H_I(g)+C(g),$$

while $P\equiv P(1)$ is the global momentum. We will prove, as one expects on formal grounds, that $T(f)$, $P(f)$, and $M(f)$ generate time translations, space translations, and Lorentz boosts, respectively, for observables which are suitably localized. Our proof depends on estimates proved in Sec. 2:

$$\pm T(f)\leq \text{const}(H+I),\quad P(f)^2\leq\text{const}(H+I)^2,$$

$$[iH,T(f)]=P(f'), \tag{1.5}$$

as well as the known positivity and self-adjointness of $H$. We do not require positivity of the Lorentzian $M(f)$. For simplicity we consider throughout this paper only $f(\cdot)$ for which $f=1$ or $f=x$ on $\{x:\text{dist}(x,\operatorname{supp}g)\leq1\}$. For more general functions $f$, see Ref. 7.

For a bounded region $B$ of space—time, we define its casual shadow on the $x$ axis by

$$I_B=[b_-,b_+],\quad b_\pm=\pm\sup\{\pm x+|t|:(x,t)\in B\}.$$

Corresponding to a Lorentz transformation $\{a,\Lambda_\beta\}$, we define

$$\{a,\Lambda_\beta\}B=\{(x',t'):(x',t')=\{a,\Lambda_\beta\}(x,t),(x,t)\in B\},$$

and for a given Lorentz transformation $\{a_0,\Lambda_{\beta_0}\}$ and bounded region $B_0$, we define

$$B=B(a_0,\beta_0)=\bigcup_{\substack{|a_i|\leq|a_{0,i}|\\|\beta|\leq|\beta_0|}}\{a,\Lambda_\beta\}B_0.$$

We now choose $g(\cdot)$ above so that $g(x)=1$ on $I_B$ and for the remainder of the paper $g(\cdot)$, $a_0$, $\beta_0$, and $B_0$ will be fixed. We will also need the algebras $\mathfrak{A}(B)$, $\mathfrak{A}_0(B)$ generated by bounded functions of fields localized in $B$ and of time-zero fields localized in $I_B$, respectively. Our principal result is:

*Theorem* 1.2: For $f\in \mathcal{S}(R)$ or $f=1$, $T(f)$, $P(f)$ define self-adjoint operators, any core for $H$ is a core for $T(f)$, any core for $H_0$ is a core for $P(f)$, $P^2(f) \leq c(f)^2(H+I)^2$, and $T(f_1+f_2)=T(f_1)+T(f_2)$, $P(f_1+f_2) = P(f_1)+P(f_2)$. Furthermore, $T(f)$ maps $D(H^2)$ into $D(H)$, $\exp[iT(f)t]$ leaves $D(H)$ invariant, and $[iH, T(f)]=P(f')$. The operator $H\exp[iT(f)t]$ is strongly continuous in $t$ on $D(H)$ with strong derivative $iH\exp[iT(f)t]T(f)$ on $D(H^2)$ and satisfies $\|H\exp[iT(f)t]\theta\| \leq \exp[c(f')]\|H\theta\|$, $\theta\in D(H)$. Similar statements apply for $M(f)$, $f\in \mathcal{S}(R)$.

*Theorem* 1.3: If $f=1$ on supp$g$, then the unitary operators

$$U(a,\beta)=\exp[ia_0T(f)]\exp[-ia_1P(f)]\exp[i\beta M(f)$$

generate the Lorentz transformations (1.1), (1.2) for fields localized in $B_0$ and for $|a_i|\leq|a_{0,i}|$, $|\beta|\leq|\beta_0|$.

The proof of Theorem 1.3 depends on the following theorem which, along with Theorem 1.2, is proved later in this section:

*Theorem* 1.4: If $f=1$ on supp$g$ and supp$h\subset B$, then as forms on $D(H)\times D(H)$:

$$[iM(f),\phi(h)]=-\phi(x\,\partial_t h+t\,\partial_x h), \qquad (1.6)$$

$$[iM(f),\psi(h)]=-\psi(x\,\partial_t h+t\,\partial_x h)+\tfrac{1}{2}\gamma_5\psi(h). \qquad (1.7)$$

Also $T(1-f)$ and $P(1-f)$ commute with $\mathfrak{A}_0(B)$.

*Proof of Theorem* 1.3: First we consider the case of pure Lorentz rotations. For $\theta\in D(H^2)$, $f=1$ on supp$g$, supp$h\subset B$, we define

$$F_1(\beta;x,t)=(\theta,\exp[iM(f)\beta]\phi(x,t)\exp[-iM(f)\beta]\theta),$$

$$F_2(\beta;x,t)=S^{-1}(\Lambda_\beta)(\theta,\exp[iM(f)\beta]\psi(x,t)\exp[-iM(f)\beta]\theta)$$

By Theorem 1.2, $F_i(\beta;x,t)$ and $(\partial F_i/\partial\beta)(\beta;x,t)$ are defined and continuous in $\beta,x,t$—note that $\phi(x,t)$, $\psi(x,t)$ are continuous in $x,t$ on $D(H)\times D(H)$. For arbitrary $h\in C_0^\infty(B)$ we define

$$F_i(\beta;h)=\int dx\,dt\,h(x,t)F_i(\beta;x,t),$$

and from Theorem 1.4 we have

$$\frac{d}{d\beta}F_i(\beta;h)=F_i(\beta;-(x\,\partial_t+t\,\partial_x)h).$$

By a result for partial differential equations,[1] it follows that for all $\beta,x,t$ with $|\beta|\leq|\beta_0|$ and $(x,t)\in B_0$

$$F_i(\beta;x,t)=F_i(0;\Lambda_\beta(x,t)).$$

By continuity, this proves (1.1) and (1.2) for Lorentz rotations on $D(H)\times D(H)$, even without smearing in $x,t$:

$$U(0,\beta)\phi(x,t)U(0,\beta)^{-1}=\phi(\Lambda_\beta(x,t)),$$

$$U(0,\beta)\psi(x,t)U(0,\beta)^{-1}=S(\Lambda_\beta)\psi(\Lambda_\beta(x,t)).$$

The extension to operator identities on smearing with $h\in C_0^\infty(B_0)$ is immediate, noting that $U(0,\beta)D(H)\subset D(H)$, that $D(H)$ is a core for $\phi(h)$, and that $\psi(h)$ is bounded.

The unitary operator $U'(a)=\exp(ia_0H)\exp(-ia_1P)$ is a generator of space–time translations for $\phi(\Lambda_\beta h)$, $\psi(\Lambda_\beta h)$ provided $|a_0|\leq|a_{0,0}|$, $|a_1|\leq|a_{0,1}|$, $|\beta|\leq|\beta_0|$, and supp$h\subset B_0$. Using the Trotter product formula for $H=T(f)+T(1-f)$, $P=P(f)+P(1-f)$, and Theorem 1.4, we conclude that $U(a)=\exp[ia_0T(f)]\exp[-ia_1P(f)]$ also generates translations, completing the proof of Theorem 1.3 and establishing (1.1) and (1.2).

Theorem 1.2 follows immediately from the estimates (1.5) and the following general result for forms $A$ relatively bounded by a positive self-adjoint operator $H$, which is proved in Ref. 8. We write $\dot{A}=[iH,A]$, $R=(H+I)^{-1}$, $C^\infty(H)=\cap_{m=0}^\infty D(H^m)$.

*Theorem* 1.5: Let $A$ be a symmetric form on $C^\infty(H)\times C^\infty(H)$ and suppose that as forms on $C^\infty(H)\times C^\infty(H)$:

(1) $\pm A\leq \mathrm{const}(H+I)^n$,

(2) $\dot{A}R$ is bounded.

Then $A$ defines a self-adjoint operator $A^-$, any core for $H^n$ is a core for $A^-$, $A^-$ maps $D(H^{n+1})$ into $D(H)$, and $\exp(iA^-t)$ leaves $D(H)$ invariant. The operator $H\exp(iA^-t)$ is strongly continuous on $D(H)$ with strong derivative $iH\exp(iA^-t)A$ on $D(H^{n+1})$ and satisfies

$$\|H\exp(iA^-t)\theta\|\leq\exp(c\,|t|)\|H\theta\|,\quad \theta\in D(H).$$

*Proof of Theorem* 1.4: Since $f=1$ on supp$g$ we have $T(1-f)=T_0(1-f)$. By Theorem 1.2, the domain $\mathcal{D}$ of vectors with finite numbers of particles and wavefunctions in Schwartz space is a core for $T_0(1-f)$ and $P(1-f)$. On $\mathcal{D}$, $\exp[i\phi_0(f_1)+i\pi_0(f_2)]$ has a power series expansion and by explicit computation we find that $T_0(1-f)$, $P(1-f)$ commute with $\phi_0(f_1)$, $\pi_0(f_2)$, $\psi_0(f_3)$, and $\bar{\psi}_0(f_4)$ on $\mathcal{D}\times\mathcal{D}$, provided supp$f_i\subset I_B$. A subscript zero denotes time-zero fields. Thus $T(1-f)$, $P(1-f)$ commute with $\mathfrak{A}_0(B)$.

To prove (1.6) we write $M\equiv M(f)$, and then for $\theta\in D(H^2)$

$$(\theta,[iM,\phi(h)]\theta)=\int dt(\theta(t),[iM(-t),\phi_0(h(\cdot,t))]\theta(t)), \qquad (1.8)$$

where $\theta(t)=\exp(-itH)\theta$, $M(t)=\exp(itH)M\exp(-itH)$. By Theorem 1.2, $M\chi\in D(H)$ for $\chi\in D(H^2)$ and

$$M(-t)\chi=M\chi-\int_0^t ds\exp(-isH)P((xf)')\exp(isH)\chi.$$

Substituting into (1.8), we obtain

$$\begin{aligned}(\theta,[iM,\phi(h)]\theta)&=\int dt(\theta(t),[iM,\phi_0(h(\cdot,t))]\theta(t))\\&\quad-\int dt\int_0^t ds(\theta(t-s),[iP((xf)'),\phi(h(\cdot,t),s)]\theta(t-s)).\end{aligned} \qquad (1.9)$$

In Theorem 2.12 of Sec. 2 we compute the commutator

$$[iM,\phi_0(f_1)]=\pi_0(xf_1),\quad \mathrm{supp}f_1\subset I_B, \qquad (1.10)$$

on $D(H)\times D(H)$. Thus the first term in (1.9) reduces to

$$\int dt(\theta(t),\pi_0(\cdot\,h(\cdot,t))\theta(t))=-(\theta,\phi(x\,\partial_t h)\theta),$$

using the boson equation of motion. In the second term we write $P((xf)')=P+P((xf)'-1)$. For $0\leq|s|\leq|t|$ the spectral projections of $\phi(h(\cdot,t),s)$ are contained in $\mathfrak{A}_0(B)$ and since $(xf)'=1$ on supp$g$, $\phi(h(\cdot,t),s)$ commutes with $P((xf)'-1)$. Noting that $P$ generates space translations, the second term in (1.9) reduces to

$$-\int dt\int_0^t ds(\theta(t-s),\phi(\partial\cdot h(\cdot,t),s)\theta(t-s))$$
$$=-(\theta,\phi(t\,\partial_x h)\theta)$$

since the integrand is actually independent of $s$. This completes the proof of (1.6), as forms on $D(H^2)\times D(H^2)$. Extension to $D(H)\times D(H)$ is immediate by continuity. The proof of (1.7) follows by similar methods, where now by Theorem 2.12, if $\operatorname{supp} f_1\subset I_B$

$$[iM,\psi_0(f_1)]$$
$$=\gamma_0\{\gamma^1\psi_0(x\,\partial_x f_1)+M\psi_0(xf_1)+\lambda(\psi_0\phi_0)(xf_1)+\tfrac{1}{2}\gamma^1\psi_0(f_1)\}. \tag{1.11}$$

Using the Fermi field equation of motion this reduces to

$$[iM,\psi_0(f_1)]=(\partial_t\psi)_{t=0}(xf_1)+\tfrac{1}{2}\gamma_5\psi_0(f_1).$$

## 2. MOMENTUM CUTOFFS

Section 2 deals with the momentum cutoff operators and their limits. In 2 A we define $T_\kappa(f)$, $P(f)$, and $H_\kappa$ and in 2 B we show that $T_\kappa(f)$ and $P(f)$ are relatively bounded by $H_\kappa$, uniformly in $\kappa$. In 2 C we consider the commutators $[iH_\kappa,T_\kappa(f)]$, $[iH_\kappa,P(f)]$, and we discuss the limiting operators $T(f)$, $M(f)$ in Sec. 2 D.

### A. Definition of $T_\kappa(f)$, $P(f)$

We introduce momentum cutoffs as in Ref. 2 by means of a cutoff function $\chi_\kappa(k,p_1,p_2)=\chi(k/\kappa)\chi(p_1/\kappa)\chi(p_2/\kappa)$, $\tilde{\chi}\in C_0^\infty$, $\chi(0)=1$. The quantities (1.3), (1.4) are then given as follows:

$$T_0(f)=T^b(f)+T^f(f),\quad P(f)=P^b(f)+P^f(f),$$
$$T^b(f)=\int dk_1dk_2[T_1(f;k_1,k_2)a^*(k_1)a(-k_2)$$
$$+T_2(f;k_1,k_2)(a^*(k_1)a^*(k_2)+a(-k_1)a(-k_2))],$$
$$T^f(f)=\int dp_1dp_2[T_1(f;p_1,p_2)(b^*(p_1)b(-p_2)$$
$$+b'^*(p_1)b'(-p_2))+T_2(f;p_1,p_2)(b^*(p_1)b'^*(p_2)$$
$$+b'(-p_1)b(-p_2))],$$
$$T_{I,\kappa}(f)=\int dk\,dp_1dp_2[w_\kappa^c(f;k,p_1,p_2)(b^*(p_1)b'^*(p_2)$$
$$+b'(-p_1)b(-p_2))+w_\kappa(f;k,p_1,p_2)(b^*(p_1)b(-p_2)$$
$$+b'^*(p_1)b'(-p_2))](a^*(k)+a(-k)),$$
$$T_{C,\kappa}(f)=-\int dx f(x)[E_\kappa+\tfrac{1}{2}\delta m_\kappa^2:\phi^2(x):].$$

$P^b(f)$, $P^f(f)$ are given by expressions similar to $T^b(f)$, $T^f(f)$ but with kernels $P_i(f;\cdot,\cdot)$, and in addition the coefficient of $aa$ or $b'b$ is negative. We will reserve the variables $k,l$ for bosons and $p,q$ for fermions and trust that no confusion will arise. The kernels and renormalization constants occurring above are given by

$$T_1(f;k_1,k_2)=\tilde{f}(k_1+k_2)(8\pi\mu_1\mu_2)^{-1/2}(\mu_1\mu_2-k_1k_2+m^2),$$
$$T_2(f;k_1,k_2)=\tilde{f}(k_1+k_2)(32\pi\mu_1\mu_2)^{-1/2}(-\mu_1\mu_2-k_1k_2+m^2),$$
$$T_1(f;p_1,p_2)=\tilde{f}(p_1+p_2)(\omega_2+\omega_1)w_1(p_1,p_2),$$
$$T_2(f;p_1,p_2)=\tilde{f}(p_1+p_2)(\omega_2-\omega_1)w_2(p_1,p_2),$$
$$P_1(f;k_1,k_2)=\tilde{f}(k_1+k_2)(8\pi\mu_1\mu_2)^{-1/2}(-\mu_1k_2+\mu_2k_1),$$
$$P_2(f;k_1,k_2)=\tilde{f}(k_1+k_2)(32\pi\mu_1\mu_2)^{-1/2}(-\mu_1k_2-\mu_2k_1),$$
$$P_1(f;p_1,p_2)=-\tilde{f}(p_1+p_2)(p_2-p_1)w_1(p_1,p_2),$$
$$P_2(f;p_1,p_2)=\tilde{f}(p_1+p_2)(p_2-p_1)w_2(p_1,p_2),$$
$$w_\kappa(f;k,p_1,p_2)=-\lambda(\pi\mu)^{-1/2}\tilde{f}(k+p_1+p_2)w_1(p_1,-p_2)$$
$$\times\chi_\kappa(k,p_1,p_2),$$
$$w_\kappa^c(f;k,p_1,p_2)=-\lambda(\pi\mu)^{-1/2}\tilde{f}(k+p_1+p_2)w_2(p_1,-p_2)$$
$$\times\chi_\kappa(k,p_1,p_2),$$
$$E_\kappa=-\frac{\lambda^2}{\pi}\int dp_1dp_2\mu(p_1+p_2)^{-1}(\mu(p_1+p_2)+\omega_1+\omega_2)^{-1}$$
$$\times|w_2(p_1,-p_2)\chi_\kappa(-p_1-p_2,p_1,p_2)|^2+\text{const}+o(1),$$
$$\delta m_\kappa^2=-\frac{\lambda^2}{2\pi}\int d\xi\,\omega(\xi)^{-1}\left|\chi_\kappa\left(0,\frac{\xi}{2},-\frac{\xi}{2}\right)\right|^2+\text{const}+o(1).$$

The constants in these terms are independent of $\chi_\kappa$ while the $o(1)$ terms vanish as $\kappa\to\infty$. The $w_i(p_1,p_2)$ are

$$w_1(p_1,p_2)=(32\pi\omega_1\omega_2)^{-1/2}(\nu(p_1)\nu(-p_2)+\nu(-p_1)\nu(p_2)),$$
$$w_2(p_1,p_2)=(32\pi\omega_1\omega_2)^{-1/2}(\nu(p_1)\nu(p_2)-\nu(-p_1)\nu(-p_2)),$$

and

$$\mu(k)=(k^2+m^2)^{1/2},\quad\omega(p)=(p^2+M^2)^{1/2},$$
$$\nu(p)=(\omega(p)+p)^{1/2}.$$

We note the following properties of the functions $w_i(p_1,p_2)$:

*Lemma* 2.1: $w_i(p_1,p_2)$ are symmetric in $p_1,p_2$, $w_1(p_1,p_2)>0$ and $|w_1(p_1,p_2)|^2+|w_2(p_1,p_2)|^2=(8\pi)^{-1}$. With $\xi=p_1-p_2$, $\eta=p_1+p_2$

$$w_1(p_1,p_2)\le\omega(\xi)\omega(\eta)^{-1},\quad|w_2(p_1,p_2)|\le\omega(\eta)^2\omega(\xi)^{-2},$$
$$|w_i(p_1,p_2')-w_i(p_1,p_2)|\le2|p_2'-p_2|(\omega_2+\omega_2')^{-1}.$$

It is easily checked that the quantities defined above give operators on $D(H_0)$, noting that $T_2(f;\cdot,\cdot)$ and $P_2(f;\cdot,\cdot)$ are in $L_2(R^2)$. The cutoff energy density is

$$T_\kappa(f)\equiv T_0(f)+T_{I,\kappa}(fg)+T_{C,\kappa}(fg^2)$$

and, in particular, we define the cutoff Hamiltonians $H_\kappa\equiv H(g,\kappa)\equiv T_\kappa(1)$. It can be shown that our choice of energy counterterm differs from that of Glimm and Jaffe[2] only by

$$|\|g\|_2^2E_\kappa-E_2(g,\kappa)|=\text{const}+o(1).$$

Choosing the constants in $E_\kappa$ so that inf spectrum $H_\kappa=0$, our $H_\kappa$ therefore agrees with theirs.

### B. $\pm T_\kappa(f)$, $\pm P(f)\le\text{const}(H_\kappa+I)$

We introduce an auxiliary lower cutoff $\rho$ as in Ref. 2 by making the change $\chi_\kappa(k,p_1,p_2)\to\chi_{\kappa,\rho}(k,p_1,p_2)$ where

$$\chi_{\kappa,\rho}(k,p_1,p_2)=\chi_\kappa(k,p_1,p_2)(1-\theta_\rho(p_1-p_2)).$$

Here $\theta_\rho(p)=1$, $|p|\le\rho$, $\theta_\rho(p)=0$, $|p|>\rho$. We denote the sum of the *finite* energy and mass renormalizations by $\Delta C_\kappa(f)$. For $f\in\mathcal{S}(R)$ of $f=1$ the corresponding operators $T_{\kappa,\rho}(f)$ satisfy:

*Lemma* 2.2: Let $\epsilon>0$, $\tau>0$. Then for sufficiently large $\kappa_0,\rho_0$ and $\kappa\ge\kappa_0$, $\rho\ge\rho_0$ there are constants uniform in $\kappa$ with

$$H_\kappa-H_{\kappa,\rho}+\Delta C_\kappa(g^2)\ge-\epsilon N_\tau-\text{const},$$
$$\pm\{T_\kappa(f)-T_{\kappa,\rho}(f)\}\le\text{const}(N_\tau+I).$$

*Proof*: The first estimate is Lemma 7.1.3 of Ref. 2 while the second follows by the method of Lemma 6.3.5 of Ref. 2. The conditions $\kappa \geq \kappa_0$, $\rho \geq \rho_0$ are required because of the finite mass renormalization.

For $\frac{1}{2} < \tau < 1$ we choose nonzero constants $c(\tau) > 0$, $\alpha(\tau)$, $\beta(\tau)$ which satisfy the inequality

$$\int dl \{ |\alpha T_1(f;k,l)| + |\beta P_1(f;k,l)| \} < \bar{\mu}(k) \equiv \mu(k) - c\mu(k)^\tau \tag{2.1}$$

and the corresponding inequality for fermion kernels. We denote the diagonal parts of $T_0(f)$, $P(f)$ by $T_1(f)$, $P_1(f)$. We now apply the dressing transformation of Ref. 2:

$$b(p,\epsilon) \to \hat{b}(p,\epsilon) = b(p,\epsilon) + [b(p,\epsilon), \Gamma W^C_{\kappa,\rho}(g)],$$

$$\Gamma W^C_{\kappa,\rho}(g) = \int dk\, dp_1\, dp_2 w^c_{\kappa,\rho}(g;k,p_1,p_2) b^*(p_1) b'^*(p_2) \times \left\{ \frac{a^*(k)}{\mu + \omega_1 + \omega_2} + \frac{a(-k)}{\omega_1 + \omega_2} \right\}$$

to $\overline{H}_0 \equiv H_0 - cN_\tau - \alpha T_1(f) - \beta P_1(f) \geq 0$, obtaining an expansion:

$$0 \leq \hat{\overline{H}}_0 = -cN_\tau - \beta P_1(f) + H_1 - \alpha H_1^T + H_2 - \alpha H_2^T - \beta(H_1^P + H_2^P). \tag{2.2}$$

$H_1$ contains terms resembling $H_{\kappa,\rho}$, $H_1^T$ contains similar terms resembling $T_{\kappa,\rho}(f) - T_2(f)$ while $H_2$, $H_2^T$, and $H_i^P$ include all other terms. For given $\epsilon > 0$ we prove:

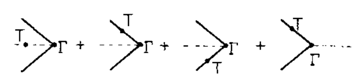
*Lemma* 2.3: For all $\kappa$ and for $\rho \geq \rho_1$ sufficiently large

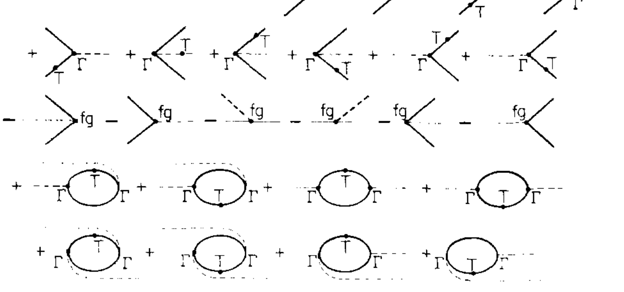
(i) $\pm(H_1 - H_{\kappa,\rho} + \Delta C_\kappa(g^2)) \leq \epsilon(N_\tau + I)$,

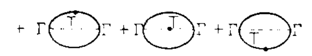
(ii) $\pm\{H_1^T - (T_{\kappa,\rho}(f) - T_2(f))\} \leq \text{const}(N_\tau + I)$,

(iii) $H_2 - \alpha H_2^T - \beta(H_1^P + H_2^P) \leq \epsilon(N_\tau + I)$.

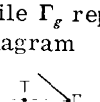
*Proof*: (i) is proved in Ref. 2. The proof of (ii) is similar. We find (for the diagrammatic conventions used, see Ref. 4)

$H_1^T - (T_{\kappa,\rho}(f) - T_2(f)) =$ 



 $- T_{C,\kappa,\rho}(fg^2)$. (2.3)

The vertices $T_f$, $f$ denote the kernels $T_1(f;\cdot,\cdot)$ and $w^c_{\kappa,\rho}(f;\cdot,\cdot,\cdot)$ while $\Gamma_g$ represents the kernels of $\Gamma W^C_{\kappa,\rho}(g)$. The diagram



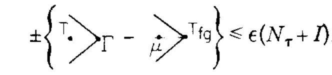
has an effective kernel $T_f * \Gamma_g$, where

$$(T_f * \Gamma_g)(k,p_1,p_2) = \int dl\, T_1(f;k,-l)\Gamma_g(l,p_1,p_2),$$

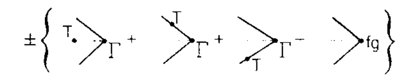
which we compare with $(\mu\Gamma_{fg})(k,p_1,p_2) \equiv \mu(k)\Gamma_{fg}(k,p_1,p_2)$. Using the properties of $w_i(p_1,p_2)$ from Lemma 2.1 and the properties of the cutoffs $\chi_\kappa$, we find

$$(T_f * \Gamma_g)(k,p_1,p_2) = (\mu\Gamma_{fg})(k,p_1,p_2) + A_1(k,p_1,p_2) \tag{2.4}$$

where for $\tau > 0$ and $\rho > \rho_1(\tau)$ sufficiently large, $\|\mu^{-\tau/2}A_1(k,p_1,p_2)\|_2 < \epsilon$. By an $N_\tau$ estimate it follows that for $\rho > \rho_1$

 $\leq \epsilon(N_\tau + I)$.

Similar estimates apply to the second and third terms in (2.3). Thus combining these estimates and noting that $(\mu + \omega_1 + \omega_2)\Gamma_{fg} = fg$, we have

 $\leq \epsilon(N_\tau + I)$.

In this way, we find that (2.4) and the corresponding estimate for convolution with a fermion momentum allows a complete cancellation of diagrams in (2.3), to within terms dominated by $\epsilon(N_\tau + I)$, plus the finite renormalization term $\Delta C_\kappa(fg^2)$ which is dominated by const$(N_\tau + I)$. This completes the proof of (ii).

We now introduce the operator $W$ on $L^2(R)$ with kernel

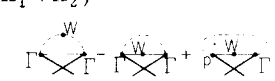
$$W(k,l) = \bar{\mu}(k)\delta(k+l) - \alpha T_1(f;k,l) - \beta P_1(f;k,l).$$

There is a corresponding definition for the fermion case. By (2.1)

$$0 \leq W \leq 2\bar{\mu} \quad \text{or} \quad 0 \leq W \leq 2\bar{\omega}, \tag{2.5}$$

where $\bar{\mu}$, $\bar{\omega}$ are multiplication by $\bar{\mu}(\cdot)$, $\bar{\omega}(\cdot)$. Then

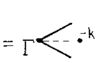
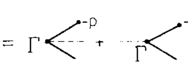
$H_2 - \alpha H_2^T - \beta(H_1^P + H_2^P)$

$= \sum_{i=1}^{9} Z_i -$ 

$- \int dp\, dq B(p)^* W(p,q) B(-q) + (b \to b')$

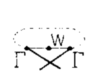
$+ \int dk\, dl A(k)^* W(k,l) A(-l) - \beta H_1^P$,

where

$A(k) =$  , $B(p) =$ 

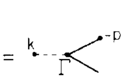
The $Z_i$ are given in Ref. 2 and dominated there by $\epsilon(N_\tau + I)$. $H_1^P$ contains diagrams identical to $H_1^T$ with $T$ replaced by $P$, see (2.3), but using the equivalent of (2.4) and combining terms in groups as for $H_1^T$ we find the resulting terms are dominated by $\epsilon(N_\tau + I)$, $\tau > \frac{1}{2}$, without cancellations being needed—a consequence of the approximate conservation of total momentum given by integration with $g(x)$ in $H(g,\kappa)$.

The first two diagrams and the term involving $B(p)$ are all negative and may thus be ignored. The remaining diagram has the form

 $= \int dk \int dp\, dq B_1(k,p)^* W(p,q) B_1(k,-q)$

$\leq 2 \int dk \int dp B_1(k,p)^* \bar{\omega}(p) B_1(k,p)$

where

$B_1(k,p) =$ 

and we used (2.5). Similarly,

$$\int dk\, dl A(k)^* W(k,l) A(-l) \leq 2 \int dk A(k)^* \bar{\mu}(k) A(k).$$

Elementary $N_\tau$ estimates give

$$\|A(k)(N_\tau+I)^{-1/2}\| \le \mathrm{const}\rho^{-\delta}\mu^{-1-\tau/2+2\delta},$$
$$\|B_1(k,p)(N_\tau+I)^{-1/2}\| \le \mathrm{const}\rho^{-\delta}\mu^{-1/2}\omega^{-1/2-\tau/2+2\delta} \times\mu(k+p)^{-1/2-\delta},$$

and thus both remaining terms are dominated by $\epsilon(N_\tau+I)$, $\tau>\frac{1}{2}$, for sufficiently large $\rho$. This proves (iii).

Taking $\epsilon<c/5$, $|\alpha|$ sufficiently small, $\rho\ge\max(\rho_0,\rho_1)$ and $\kappa\ge\kappa_0$, we obtain from (2.2) and Lemmas 2.2, 2.3

$$\pm\{\alpha(T_\kappa(f)-T_2(f))+\beta P_1(f)\}\le H_\kappa+\mathrm{const},$$

and since $T_2(f)$, $P_2(f)$ have $L_2$ kernels, we have, for $\kappa\ge\kappa_0$,

*Theorem* 2.4: For $f\in\mathcal{S}(R)$ or $f=1$, there is a constant independent of $\kappa$ with $\pm T_\kappa(f)$, $\pm P(f)\le\mathrm{const}(H_\kappa+I)$.

### C. Commutators with $H_\kappa$

For convenience, we will adopt the notation

$$:\phi\bar\psi\Gamma\psi:(A)=\int dy\,dy_1\,dy_2:\phi(y)\bar\psi(y_1)\Gamma\psi(y_2):A(y,y_1,y_2)$$

throughout this section. Thus $T_{I,\kappa}(f)=:\phi\bar\psi\psi:(B_\kappa(f))$,

$$B_\kappa(f;y,y_1,y_2)=\lambda(2\pi)^{-3/2}\int dx\,f(x)\tilde\chi_\kappa(x-y,x-y_1,x-y_2).$$

*Theorem* 2.5: As forms on $D(H_\kappa)\times D(H_\kappa)$, $f\in\mathcal{S}(R)$ or $f=1$,

$$[iH_\kappa,T_\kappa(f)]=P(f')+C_\kappa,\quad [iH_\kappa,P(f)]=\sum_{i=1}^{m}D_{i,\kappa},$$

where $\|R_\kappa C_\kappa R_\kappa\|\le\mathrm{const}\kappa^{-\epsilon}$ and for each $i$ either $\|R_\kappa D_{i,\kappa}R_\kappa^{1-\delta}\|\le\mathrm{const}$ or $\|R_\kappa^{1-\delta}D_{i,\kappa}R_\kappa\|\le\mathrm{const}$, for some $\delta>0$.

*Proof*: Elementary computations on $\mathcal{D}\times\mathcal{D}$ give

$$[iH_\kappa,T_\kappa(f)]=P(f')+:\pi\bar\psi\psi:(C_{1,\kappa}(f))+:\phi\bar\psi\gamma^0\gamma^1\psi:(C_{2,\kappa}(f)) +:\phi\bar\psi\gamma^0\psi:(C_{3,\kappa}(f))+:\phi\phi\bar\psi\gamma^0\psi:(C_{4,\kappa}(f)) +:\bar\psi\gamma^0\psi:(C_{5,\kappa}(f)),$$

$$[iH_\kappa,P(f)]=(T_0-m^2:\phi^2:+M:\bar\psi\psi:)(f')-:\phi\bar\psi\psi:(C_{6,\kappa}(f)) +\phi(C_{7,\kappa}(f))+\tfrac{1}{2}\delta m_\kappa^2:\phi^2:((fg^2)'),$$

where the kernels involved are

$$C_{1,\kappa}(f;y,y_1,y_2)=B_\kappa(fg;y,y_1,y_2)-f(y)B_\kappa(g;y,y_1,y_2),$$

$$C_{2,\kappa}(f;y,y_1,y_2)=\left(\frac{\partial}{\partial y_1}-\frac{\partial}{\partial y_2}\right)B_\kappa(fg;y,y_1,y_2) -\left(f(y_1)\frac{\partial}{\partial y_1}-f(y_2)\frac{\partial}{\partial y_2}+\tfrac{1}{2}f'(y_1)-\tfrac{1}{2}f'(y_2)\right)\times B_\kappa(g;y,y_1,y_2),$$

$$C_{3,\kappa}(f;y,y_1,y_2)=M(f(y_1)-f(y_2))B_\kappa(g;y,y_1,y_2),$$

$$C_{4,\kappa}(f;y,y',y_1,y_2)=\int dz[B_\kappa(g;y,y_1,z)B_\kappa(fg;y',z,y_2) -B_\kappa(fg;y',y_1,z)B_\kappa(g;y,z,y_2)],$$

$$C_{5,\kappa}(f;y_1,y_2)=\int dk(4\pi\mu)^{-1}\int dy\,dy'\exp[-ik(y-y')] \times C_{4,\kappa}(f;y,y',y_1,y_2),$$

$$C_{6,\kappa}(f;y,y_1,y_2)=\left(f(y)\frac{\partial}{\partial y}+f(y_1)\frac{\partial}{\partial y_1}+f(y_2)\frac{\partial}{\partial y_2}+f'(y) +\tfrac{1}{2}f'(y_1)+\tfrac{1}{2}f'(y_2)\right)B_\kappa(g;y,y_1,y_2),$$

$$C_{7,\kappa}(f;y)=M\int dp(2\pi\omega)^{-1}\int dy_1\,dy_2\exp[ip(y_1-y_2)] \times C_{8,\kappa}(f;y,y_1,y_2),$$

$$C_{8,\kappa}(f;y,y_1,y_2)=-\left(f(y_1)\frac{\partial}{\partial y_1}+f(y_2)\frac{\partial}{\partial y_2}+\tfrac{1}{2}f'(y_1)+\tfrac{1}{2}f'(y_2)\right) \times B_\kappa(g;y,y_1,y_2),$$

and these commutators extend to $D(H_\kappa)\times D(H_\kappa)$ by continuity. Denoting the term with kernel $C_{i,\kappa}(f;\cdot)$ by $C_{i,\kappa}$, we define $C_\kappa=\sum_{i=1}^5 C_{i,\kappa}$. In order to study $C_{i,\kappa}$, $i=1,\ldots,5$, we obtain estimates for the Fourier transforms

$$|\tilde C_{1,\kappa}(f;k,p_1,p_2)|\le\mathrm{const}\kappa^{-\delta}\mu^{-1+\alpha+\delta}(\omega_1+\omega_2)^{-\alpha}h(k+p_1+p_2),$$
$$|\tilde C_{2,\kappa}(f;k,p_1,p_2)|\le\mathrm{const}\kappa^{-\delta}(\omega_1+\omega_2)^{\delta}h(k+p_1+p_2),$$
$$|\tilde C_{3,\kappa}(f;k,p_1,p_2)|\le\mathrm{const}\kappa^{-\delta}\mu^{-1+\alpha+\beta+\delta}\omega_1^{-\alpha}\omega_2^{-\beta}h(k+p_1+p_2),$$
$$|\tilde C_{4,\kappa}(f;k,k',p_1,p_2)|\le\mathrm{const}\kappa^{-\delta}(\mu+\mu'+\omega_1+\omega_2)^{-1+\delta} \times h(k+k'+p_1+p_2),$$
$$|\tilde C_{5,\kappa}(f;p_1,p_2)|\le\mathrm{const}\kappa^{-\delta}\mu(p_1-p_2)^{3\delta}h(p_1+p_2),$$

where $h(\cdot)$ is of rapid decrease, $\alpha,\beta\ge0$, and $\alpha+\beta\le1-\delta$. The fermion components of each $C_{i,\kappa}$ are given by

$$:\bar\psi\Gamma_i\psi:(x,y)=-(2/\pi)^{1/2}\int dp_1\,dp_2\exp(-ip_1x-ip_2y) \times\{u_i(p_1,p_2)(b^*(p_1)b'^*(p_2)\pm b'(-p_1)b(-p_2)) +v_i(p_1,p_2)(b^*(p_1)b(-p_2)\pm b'^*(p_2)b(-p_1))\}.$$

Here $\Gamma_1=1$, $\Gamma_2=\gamma^0\gamma^1$, $\Gamma_3,\Gamma_4,\Gamma_5=-i\gamma^0$, the + signs occur for $i=1$ and

$$u_2(p_1,p_2)=v_1(p_1,-p_2)=-v_3(p_1,p_2)=w_1(p_1,p_2),$$
$$u_1(p_1,-p_2)=u_3(p_1,p_2)=v_2(p_1,p_2)=w_2(p_1,p_2).$$

$C_{1,\kappa}$, $C_{2,\kappa}$ resemble closely the interaction Hamiltonian and by a standard expansion[9] $\|R_\kappa C_{i,\kappa}R_\kappa\|\le\mathrm{const}\kappa^{-\epsilon}$, $i=1,2$. Renormalizations are not needed in either case. For $C_{1,\kappa}$ the minus sign in $a^*-a$, from $\pi(y)$, ensures cancellation of terms between the expansions for $b^*b'^*$ and $b'b$ which would otherwise require vacuum energy renormalization while the terms corresponding to mass diagrams are not divergent on account of the factor $\mu^\alpha(\omega_1+\omega_2)^{-\alpha}$ in $|\tilde C_{1,\kappa}|$. For $C_{2,\kappa}$ the relative sign difference of $b^*b'^*$ and $b'b$ ensures cancellations of all divergences between the corresponding expansions.

The terms $C_{3,\kappa}$, $C_{4,\kappa}$, and $C_{5,\kappa}$ with fermion parts $:\bar\psi\gamma^0\psi:$ resemble the charge density[4] and are well-behaved because of approximate total momentum conservation. Thus elementary $L_2$ and $L_1-L_\infty$ bounds and first and second order estimates[2-4] yield $\kappa^{-\epsilon}$ dominated norms in each case.

In order to discuss $[iH_\kappa,P(f)]$ we introduce

$$T_{I,\kappa}^R(f)=T_{I,\kappa}(f)+2T_{C,\kappa}(fg),$$
$$j_\kappa(f)=:\bar\psi\psi:(f)-N_\kappa\phi(fg),$$
$$N_\kappa=-\frac{\lambda}{2\pi}\int d\xi\,\omega(\xi)^{-1}\chi_\kappa(0,\xi/2,-\xi/2).$$

Then the commutator may be rewritten as

$$[iH_\kappa,P(f)]=(T_\kappa+Mj_\kappa-m^2:\phi^2:)(f')-T_{I,\kappa}^R(fg') +\phi(C_{7,\kappa}(f)+MN_\kappa f'g)-:\phi\bar\psi\psi:(C_{6,\kappa}(f)-B_\kappa(fg')).$$

We decompose $T^R_{I,\kappa}(f)$ into $T^{R,+}_{I,\kappa}+T^0_{I,\kappa}+T^{R,-}_{I,\kappa}$, where $T^{R,+}_{I,\kappa}=T^+_{I,\kappa}(f)+T_{C,\kappa}(fg)$, $T^{R,-}_{I,\kappa}=(T^{R,+}_{I,\kappa})^*$ and +, 0, or − denotes the component creating 2, 0, or −2 fermions, respectively.

*Lemma* 2.6: For $\alpha>\frac{1}{2}$, each of $R^\alpha_\kappa j_\kappa(f)R^\alpha_\kappa$, $R_\kappa T^{R,+}_{I,\kappa}R^\alpha_\kappa$, $R^{1/2}_\kappa T^0_{I,\kappa}R^{1/2}_\kappa$, $R^\alpha_\kappa T^{R,-}_{I,\kappa}R_\kappa$ is uniformly bounded in $\kappa$.

*Proof*: The estimates for the diagonal components of $j_\kappa$ and $T^R_{I,\kappa}$ use $N_\tau$ estimates. For the fermion creation and annihilation terms we use a standard expansion[9] which exhibits renormalization cancellations explicitly.

Returning to Theorem 2.5 and using Theorem 2.4, we see that only the $\phi$ and $\phi\bar{\psi}\psi$ terms remain to be estimated. For the kernel of $\phi$ we find

$$|\tilde{C}_{7,\kappa}(f)(k)+MN_\kappa(f'g)^\sim(k)|\leq \text{const}\kappa^{-\delta}\tilde{h}(k).$$

Thus

$$\|R^{1/2}_\kappa\phi(C_{7,\kappa}(f)+MN_\kappa(f'g))R^{1/2}_\kappa\|\leq \text{const}\kappa^{-\delta},$$

and a similar treatment to that for $T^R_{I,\kappa}$ gives

$$\|R_\kappa:\bar{\psi}\psi\phi\overset{+}{:}(C_9)R^\alpha_\kappa\|+\|R^{1/2}_\kappa:\bar{\psi}\psi\phi\overset{0}{:}(C_9)R^{1/2}_\kappa\|$$
$$+\|R^\alpha_\kappa:\bar{\psi}\psi\phi\overset{-}{:}(C_9)R_\kappa\|\leq \text{const}\kappa^{-\delta}$$

where $C_9(\cdot)=C_{6,\kappa}(f;\cdot)-B_\kappa(fg';\cdot)$.

### D. The ultraviolet limit

In this section we remove the momentum cutoff and show that the corresponding densities $T(f), P(f)$ satisfy (1.5). We will need two results for forms $A$ defined on $C^\infty(H)\times C^\infty(H)$, $H=H^*\geq 0$. For the proofs of these results see Ref. 8. We write $R(\lambda)=(H+\lambda+I)^{-1}$, $R=R(0)$, $\dot{A}=[iH,A]$, and we suppose that as forms:

$$\dot{A}=\sum_{i=1}^N B_i.$$

For positive $n, m, \delta$ we define

$$\alpha(r)=\min(r, n/2+1-\delta),\quad \beta(r)=r+n/2-m-\delta.$$

*Theorem* 2.7: Suppose $R^{1/2}AR^{1/2}$ is bounded. Then $AR$ is bounded provided for each $i$ there is a $\delta_i>0$ with either $RB_iR^{1-\delta_i}$ or $R^{1-\delta_i}B_iR$ bounded. Also $AR^{1+\delta}$ is bounded, $\delta>0$, if $RB_iR$ is bounded, all $i$.

*Theorem* 2.8: Suppose either $R^{n/2}AR^{n/2}$ or $AR^n$ is bounded, $A$ symmetric. Then $A$ defines an operator, essentially self-adjoint on any core for $H^n$ provided there are $\mu_i$, $0\leq\mu_i\leq 1$, such that $R^{\mu_i}B_iR^{1-\mu_i}$ are bounded.

As an immediate consequence of Theorems 2.4, 2.5, 2.7, 2.8, we have for $f\in\mathcal{S}(R)$ or $f=1$, and $\alpha>1$,

*Theorem* 2.9: $T_\kappa(f)R^\alpha_\kappa$ and $P(f)R_\kappa$ are uniformly bounded; $P(f)$ is essentially self-adjoint on cores for $H_0$.

*Theorem* 2.10: $R^\alpha_\kappa T_\kappa(f)R^\alpha_\kappa$ and $T_\kappa(f)R^{2\alpha}_\kappa$ converge in norm as $\kappa\to\infty$, $\alpha>1$; $P(f)R_\kappa$ converges weakly to $P(f)R$.

*Proof*: With the notation $\delta A=A_{\kappa_2}-A_{\kappa_1}$ we have

$$\delta(R^\alpha T(f)R^\alpha)=\delta R^\alpha T_{\kappa_2}(f)R^\alpha_{\kappa_2}+R^\alpha_{\kappa_1}\delta T(f)R^\alpha_{\kappa_2}+R^\alpha_{\kappa_1}T_{\kappa_1}(f)\delta R^\alpha.$$

Thus since $\|\delta R^\alpha\|\leq \text{const}\kappa_0^{-\epsilon}$, $\kappa_0=\min(\kappa_1,\kappa_2)$,[9] and $T_\kappa(f)R^\alpha_\kappa$ is uniformly bounded, norm convergence of $R^\alpha_\kappa T_\kappa(f)R^\alpha_\kappa$ follows from the estimate

$$\|R_{\kappa_1}\delta T(f)R_{\kappa_2}\|\leq \text{const}\kappa_0^{-\epsilon},$$

which is proved by following the proof of resolvent convergence of the Hamiltonians $H_\kappa$ in Ref. 9. The estimates are all similar so we do not repeat them here.

Norm convergence of $T_\kappa(f)R^{2\alpha}_\kappa$ results from the identity

$$T_\kappa(f)R^{2\alpha}_\kappa=R^\alpha_\kappa T_\kappa(f)R^\alpha_\kappa-ic(2\alpha,2)\int d\lambda\,\lambda^{1-\alpha}\{R_\kappa(\lambda)\dot{T}_\kappa(f)R_\kappa(\lambda)^2$$
$$+R_\kappa(\lambda)^2\dot{T}_\kappa(f)R_\kappa(\lambda)\}R^\alpha_\kappa,$$

where $\dot{T}_\kappa(f)\equiv P(f')+C_\kappa$, with, by Theorem 2.5, $\|R_\kappa C_\kappa R_\kappa\|\leq\text{const}\kappa^{-\epsilon}$.

Weak convergence of $P(f)R_\kappa$ to $P(f)R$ is a consequence of the self-adjointness of $P(f)$ and the uniform boundedness of $P(f)R_\kappa$.

We now define the energy density $T(f)$ in the ultraviolet limit as a form on $D(H^\alpha)\times D(H^\alpha)$, $\alpha>1$:

$$T(f)=(H+I)^\alpha\cdot\lim_{\kappa\to\infty}R^\alpha_\kappa T_\kappa(f)R^\alpha_\kappa\cdot(H+I)^\alpha. \tag{2.6}$$

From Theorems 2.4, 2.5, 2.10 we conclude that

$$\pm T(f)\leq\text{const}(H+I),$$
$$\dot{T}(f)\equiv[iH,T(f)]=P(f'),$$

on $D(H^{1+\alpha})\times D(H^{1+\alpha})$, completing the proof of estimates (1.5) and of Theorem 1.2. Essential self-adjointness of $T(f)$ follows by Theorem 2.8 and we redefine $T(f)$ to be the self-adjoint closure. From the norm convergence of $T_\kappa(f)R^{2\alpha}_\kappa$ we can show that

$$T(f)=\text{strong graph}\lim T_\kappa(f),$$

and as a consequence we obtain strong resolvent convergence:

*Theorem* 2.11: $T_\kappa(f)$ converges to $T(f)$ in the strong resolvent sense.

There remains only the computation of the commutators of $M(f)=T(xf)$ with time zero fields:

*Theorem* 2.12: The Eq. (1.10) and (1.11) hold on $D(H)\times D(H)$.

*Proof*: The commutators of time-zero fields with $M_\kappa(f)$ are easily computed on $\mathcal{D}\times\mathcal{D}$ and by continuity on $D(H^\alpha_\kappa)\times D(H^\alpha_\kappa)$, $\alpha>2$. Strong convergence of $M_\kappa(f)R^\alpha_\kappa$ and of the other terms in these commutators gives (1.10) and (1.11) on $D(H^\alpha)\times D(H^\alpha)$, and by continuity we extend to $D(H)\times D(H)$.

*Research supported in part by the National Research Council of Canada.

†Present address: Department of Mathematics, Rockefeller University, New York, N.Y. 10021.

[1]J. Cannon and A. Jaffe, Commun. Math. Phys. 17, 261 (1970).

[2]J. Glimm and A. Jaffe, "Quantum Field Theory Models," in *Statistical Mechanics and Quantum Field Theory*, edited by G. Dewitt and R. Stora (Gordon and Breach, New York, 1971).

[3]J. Dimock, Ann. Phys. 72, 177 (1972).

[4]O. McBryan, "Vector currents in the Yukawa$_2$ quantum field theory," University of Toronto, preprint.

[5]O. McBryan, "The $N$-th order Renormalized Pull-Through Formula for the Yukawa$_2$ Quantum Field Theory," Rockefeller University, preprint.

[6]O. McBryan, "Higher Order Estimates for the Yukawa$_2$ Quantum Field Theory," Commun. Math. Phys., to appear.

[7]O. McBryan, "Lorentz Covariance and Local Generators for the Lorentz Group in the Yukawa$_2$ Model," University of Toronto, preprint.

[8]O. McBryan, "Self-adjointness of Relatively Bounded Quadratic Forms and Operators," J. Funct. Anal., to appear.

[9]J. Glimm and A. Jaffe, Ann. Phys. 60, 321 (1970).

[10]O. McBryan, Nuovo Cimento 18A, 654 (1973).

# Lattice Green's function for $B$-site lattice in spinel

Michiko Inoue

*RCA Research Laboratories, Inc., Tokyo International 100-31, Japan*
(Received 26 July 1974)

We show that the lattice Green's function at an arbitrary site of a $B$-site lattice in a spinel with nearest neighbor interactions can be expressed in terms of Green's functions for a face-centered cubic lattice. Results of numerical calculations of the $B$-site lattice Green's functions at the origin and over the third neighboring sites are presented.

A lattice Green's function for $B$-site lattice in spinel is a quantity of particular interest in studying problems such as the energy spectrum of impurities, defects in a $B$-site magnetic spinel, or hopping conduction in magnetite, etc. In this paper we show that the lattice Green's function of a $B$-site lattice at an arbitrary site with nearest neighbor interactions can be expressed in terms of Green's functions for face-centered cubic (f.c.c.) lattice which are given as a linear combination of products of the complete elliptic integrals of the first and second kinds.[1]

A spinel structure has an overall cubic symmetry. A unit cell accommodates two types of sites for cations, namely, the tetrahedral $A$-site and the octahedral $B$-site. The $B$-sites are composed of four interpenetrating f.c.c. sublattices in such a way that each site of a sublattice is surrounded by six nearest neighbor sites which belong to three other sublattices, as illustrated in Fig. 1.

We consider the case of a $B$-site ferromagnet in spinel with nearest neighbor isotropic exchange coupling in the presence of the external magnetic field along the $z$ axis. The Hamiltonian of the system is given by

$$H=-J\sum_{\mathbf{j}_m}\sum_{\rho_{mn}}\mathbf{S}_{\mathbf{j}_m}\cdot\mathbf{S}_{\mathbf{j}_m+\rho_{mn}}-g\mu_B\sum_{\mathbf{j}_m}S^z_{\mathbf{j}_m}H_0, \tag{1}$$

where $\mathbf{j}_m$ is a lattice vector at the $j$th site of sublattice $m$, $\rho_{mn}$ is a vector joining the adjacent sites of sublattices $m$ and $n$, and $J>0$.

We shall calculate the lattice Green's function using Zubarev's technique of double-time Green's functions,[2] as defined by

$$G(t-t')\equiv\langle\langle a(t);b(t')\rangle\rangle=-i\theta(t-t')\langle[a(t),b(t')]\rangle, \tag{2}$$

where $a$ and $b$ are physical operators, $\langle\ \rangle$ means taking a statistical average over a grand canonical ensemble, and

$$\theta(t)=\begin{cases}1, & t>0\\ 0, & t<0.\end{cases}$$

In the following we shall restrict our attention to the one-spin Green's function at $T=0\,^\circ$K. The $\omega$ Fourier component of the one-spin Green's function

$$g_{\mathbf{j}_m\mathbf{1}_{m'}}(\omega)\equiv\langle\langle S^+_{\mathbf{j}_m};S^-_{\mathbf{1}_{m'}}\rangle\rangle_\omega$$
$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}dt\,\langle\langle S^+_{\mathbf{j}_m}(t);S^-_{\mathbf{1}_{m'}}(t')\rangle\rangle\exp[i\omega(t-t')], \tag{3}$$

satisfies the equation of motion,

$$\hbar\omega g_{\mathbf{j}_m\mathbf{1}_{m'}}(\omega)=\frac{S^z_{\mathbf{j}_m}}{\pi}\delta_{\mathbf{j}_m\mathbf{1}_{m'}}+2J[\langle\langle S^+_{\mathbf{j}_m}\sum_{\rho_{mn}}S^z_{\mathbf{j}_m+\rho_{mn}};S^-_{\mathbf{1}_{m'}}\rangle\rangle_\omega$$
$$-\langle\langle S^z_{\mathbf{j}_m}\sum_{\rho_{mn}}S^+_{\mathbf{j}_m+\rho_{mn}};S^-_{\mathbf{1}_{m'}}\rangle\rangle_\omega]+g\mu_BH_0g_{\mathbf{j}_m\mathbf{1}_{m'}}(\omega). \tag{4}$$

At $T=0\,^\circ$K, the system is in the ground state where all the spins are completely aligned along the external field, the following decoupling is valid:

$$\langle\langle S^+_{\mathbf{j}_m}\sum_{\rho_{mn}}S^z_{\mathbf{j}_m+\rho_{mn}};S^-_{\mathbf{1}_{m'}}\rangle\rangle_\omega=ZS\langle\langle S^+_{\mathbf{j}_m};S^-_{\mathbf{1}_{m'}}\rangle\rangle_\omega,$$

and

$$\langle\langle S^z_{\mathbf{j}_m}\sum_{\rho_{mn}}S^+_{\mathbf{j}_m+\rho_{mn}};S^-_{\mathbf{1}_{m'}}\rangle\rangle_\omega=S\sum_{\rho_{mn}}\langle\langle S^+_{\mathbf{j}_m+\rho_{mn}};S^-_{\mathbf{1}_{m'}}\rangle\rangle_\omega. \tag{5}$$

These reduce Eq. (4) to a closed form

$$[\hbar\omega-2ZJS-g\mu_BH_0]g_{\mathbf{j}_m\mathbf{1}_{m'}}(\omega)+2JS\sum_{\rho_{mn}}g_{\mathbf{j}_m+\rho_{mn}\mathbf{1}_{m'}}(\omega)$$
$$=\frac{S}{\pi}\delta_{\mathbf{j}_m\mathbf{1}_{m'}}, \tag{6}$$

where $Z$ is the number of the nearest neighbors and $S$ is the magnitude of spin.

Introducing the spatial Fourier transform of $g_{\mathbf{j}_m\mathbf{1}_{m'}}(\omega)$,

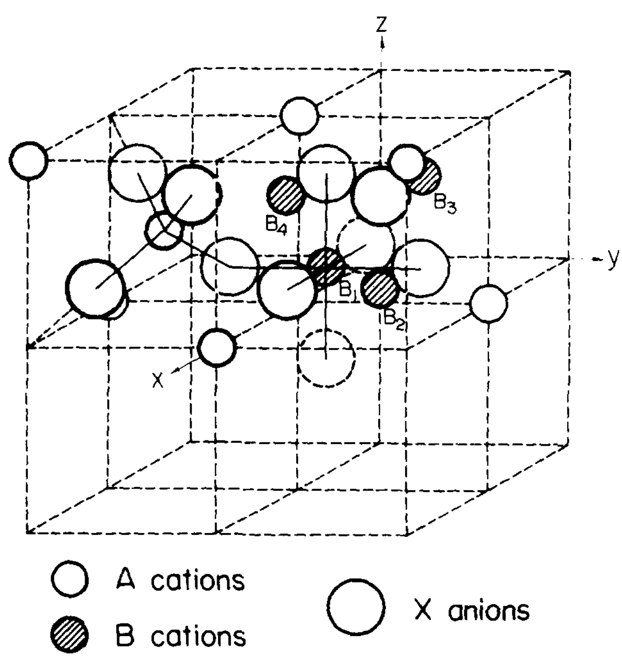



FIG. 1. Spinel crystal structure.

$$g_{j_m l_{m'}}(\omega)=\frac{1}{N}\sum_{\mathbf{k}}\sum_{\mathbf{k'}}g_{mm'}(\mathbf{k},\mathbf{k}',\omega)\exp[i(\mathbf{k}\cdot\mathbf{j}_m-\mathbf{k}'\cdot\mathbf{l}_{m}')], \tag{7}$$

we rewrite Eq. (6)

$$[\hbar\omega-2ZJS-g\mu_B H_0]g_{mm'}(\mathbf{k},\mathbf{k}',\omega)$$
$$+2JS\sum_{n\neq m}\gamma_{mn}(\mathbf{k})g_{nm'}(\mathbf{k},\mathbf{k}',\omega)=\frac{S}{\pi}\delta_{mm'}\delta_{\mathbf{kk'}}, \tag{8}$$

with

$$\gamma_{mn}(\mathbf{k})=2\cos\mathbf{k}\cdot\rho_{mn}. \tag{9}$$

In matrix notation Eq. (8) can be written as

$$D(\mathbf{k},\omega)g(\mathbf{k},\mathbf{k}',\omega)=\frac{S}{\pi}I\delta_{\mathbf{kk'}}, \tag{10}$$

where

$$D(k,\omega)=\begin{bmatrix}\alpha & \beta\cos\mathbf{k}\cdot\rho_{12} & \beta\cos\mathbf{k}\cdot\rho_{13} & \beta\cos\mathbf{k}\cdot\rho_{14}\\ \beta\cos\mathbf{k}\cdot\rho_{21} & \alpha & \beta\cos\mathbf{k}\cdot\rho_{23} & \beta\cos\mathbf{k}\cdot\rho_{24}\\ \beta\cos\mathbf{k}\cdot\rho_{31} & \beta\cos\mathbf{k}\cdot\rho_{32} & \alpha & \beta\cos\mathbf{k}\cdot\rho_{34}\\ \beta\cos\mathbf{k}\cdot\rho_{41} & \beta\cos\mathbf{k}\cdot\rho_{42} & \beta\cos\mathbf{k}\cdot\rho_{43} & \alpha\end{bmatrix} \tag{11}$$

with

$$\alpha=\hbar\omega-2ZJS-g\mu_B H_0,$$
$$\beta=4JS, \tag{11'}$$

and $g(\mathbf{k},\mathbf{k}',\omega)$ is a $4\times4$ matrix having elements $g_{mm'}(\mathbf{k},\mathbf{k}',\omega)$ for $m,m'=1,2,3,4,$ and $I$ is a unit matrix. From (10) we get the Green's function matrix

$$g(\mathbf{k},\mathbf{k},\omega)=\frac{S}{\pi}D^{-1}(\mathbf{k},\omega)\delta_{\mathbf{k},\mathbf{k}'},$$

and with (8), the Green's function

$$g_{j_m l_{m'}}(\omega)=\frac{S}{N\pi}\sum_{\mathbf{k}}[D^{-1}(\mathbf{k},\omega)]_{mm'}\exp[ik(j_m-l_m)]. \tag{12}$$

The inverse of the matrix $D(\mathbf{k},\omega)$ can be written as

$$[D(\mathbf{k},\omega)^{-1}]_{mn}=d_{mn}(\mathbf{k},\omega)/|D(\mathbf{k},\omega)|, \tag{13}$$

where $d_{mn}(\mathbf{k},\omega)$ is a cofactor of $D_{mn}(\mathbf{k},\omega)$. For spinel structures it is necessary to determine two of the cofactors independently since those remainings can be obtained from the two cofactors by proper symmetry operations as shown by the following relations:

$$d_{mn}(\mathbf{k},\omega)=d_{nm}(\mathbf{k},\omega),$$
$$d_{11}(k_x,k_y,k_z,\omega)=d_{22}(k_x,k_y,-k_z,\omega)=d_{33}(-k_x,k_y,k_z,\omega)$$
$$=d_{44}(k_x,-k_y,k_z,\omega),$$
$$d_{12}(k_x,k_y,k_z,\omega)=d_{13}(k_z,k_y,k_x,\omega)=d_{14}(k_x,k_z,k_y,\omega)$$
$$=d_{23}(k_x,-k_z,-k_y,\omega)=d_{24}(-k_z,k_y,-k_x,\omega)$$
$$=d_{34}(-k_y,-k_x,k_z,\omega). \tag{14}$$

The determinant $|D(\mathbf{k},\omega)|$ and the two cofactors $d_{11}(\mathbf{k},\omega)$ and $d_{12}(\mathbf{k},\omega)$ are calculated to be

$$|D(\mathbf{k},\omega)|=(\alpha-\beta)^2\left(E-\cos\frac{ck_x}{2}\cos\frac{ck_y}{2}-\cos\frac{ck_y}{2}\cos\frac{ck_z}{2}-\cos\frac{ck_z}{2}\cos\frac{ck_x}{2}\right), \tag{15a}$$

$$d_{11}(\mathbf{k},\omega)=(\alpha-\beta)\left[\alpha^2+\alpha\beta-\frac{\beta^2}{2}\left(1+\cos\frac{c(k_x-k_y)}{2}+\cos\frac{c(k_y-k_z)}{2}+\cos\frac{c(k_z-k_x)}{2}\right)\right], \tag{15b}$$

and

$$d_{12}(\mathbf{k},\omega)=-\beta(\alpha-\beta)\left(\alpha\cos\frac{c}{4}(k_x+k_y)-\beta\cos\frac{c}{4}(k_x-k_y)\cos\frac{ck_z}{2}\right), \tag{15c}$$

with

$$E=[(\alpha+\beta)^2/\beta^2]-1 \tag{15d}$$

and $c$, the lattice constant of the spinel structure. The two cofactors $d_{11}(\mathbf{k},\omega)$ and $d_{12}(\mathbf{k},\omega)$ give rise to two distinct lattice Green's functions, namely, intrasublattice and intersublattice functions. Here we note that a site on a $B$-sublattice in spinel is carried to any other site of the same or different sublattice by a proper combination of a translation of $l\hat{x}+m\hat{y}+n\hat{z}$ and that of $\rho_{12}=\frac{1}{2}(\hat{x}+\hat{y})$, where $l+m+n$ is zero or an even integer, and $\hat{x}$, $\hat{y}$, and $\hat{z}$ are the primitive translation vectors for a f.c.c. lattice with lattice constant $c$. Then, after some calculations with (12), (13), and (14), we obtain the general formula

$$g(l,m,n)=\frac{1}{8\pi JN}\frac{1}{(\epsilon-\Delta-4)}\{[(\epsilon-\Delta-3)(\epsilon-\Delta-2)-\tfrac{1}{2}]D(l,m,n)-\tfrac{1}{4}[D(l+1,m-1,n)+D(l-1,m+1,n)$$
$$+D(l,m+1,n-1)+D(l,m-1,n+1)+D(l-1,m,n+1)+D(l+1,m,n-1)]\}, \tag{16a}$$

for the intrasublattice Green's function, and

$$g(l+\tfrac{1}{2},m+\tfrac{1}{2},n)=-\frac{1}{4\pi JN}\frac{1}{(\epsilon-\Delta-4)}[2(\epsilon-\Delta-3)D(l,m,n)+(2\epsilon-2\Delta-7)D(l+1,m+1,n)-2D(l+1,m,n-1)$$
$$-D(l,m-1,n-1)], \tag{16b}$$

for the intersublattice Green's function of the $B$-site lattice. Here $D(l,m,n)$ is defined by

$$D(l,m,n)=\frac{1}{\pi^3}\int_0^\pi\int_0^\pi\int_0^\pi dx\,dy\,dz\times\frac{\cos lx\cos my\cos nz}{E-\cos x\cos y-\cos y\cos z-\cos z\cos x}, \tag{17}$$

and $E$, $\epsilon$, and $\Delta$ are dimensionless units given by

$$E=(\epsilon-\Delta-3)(\epsilon-\Delta-1),$$
$$\epsilon=\hbar\omega/\beta, \tag{18}$$
$$\Delta=g\mu_B H_0/\beta.$$

For the range of values $\epsilon>4+\Delta$ and $\epsilon<\Delta$, $D(l,m,n)$ and, hence, $g$'s are real, while for $4+\Delta>\epsilon>\Delta$, they are complex numbers and $\epsilon$ should be replaced by $\epsilon-is$ where $s$ is a positive infinitesimal number. Then, by use of the relation

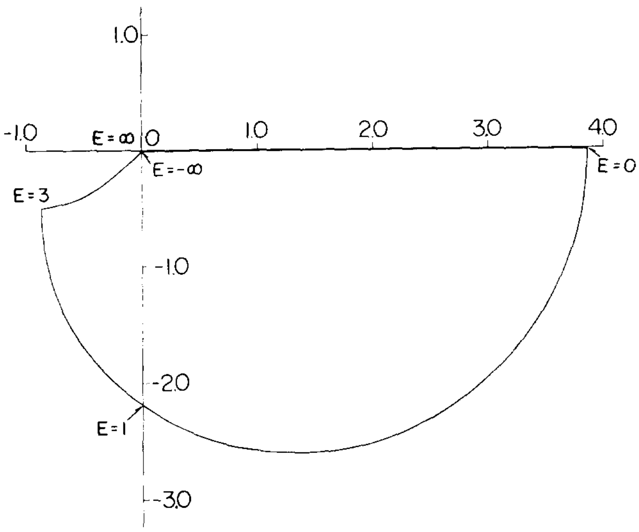



FIG. 2. The values of $q_{1+}$ in the complex plane.

$$\lim_{s\to 0^+}\frac{1}{\epsilon \mp is-\epsilon_0}=P\frac{1}{\epsilon-\epsilon_0}\pm i\pi\delta(\epsilon-\epsilon_0),$$

the denominator of the integrand in (17) becomes

$$\frac{1}{E-is'-\cos x\cos y-\cos y\cos z-\cos z\cos x},$$

where $s'>0$ for $\epsilon-2-\Delta>0$ and $s'<0$ for $\epsilon-2-\Delta<0$. This enables us to express $D(l,m,n)$ in terms of the f.c.c. lattice Green's function $G(l,m,n)$ and its complex conjugate $G(l,m,n)$ as follows;

$$\begin{aligned} D(l,m,n)&=G(l,m,n)\quad \text{for } \epsilon-2-\Delta>0\\ &=\bar{G}(l,m,n)\quad \text{for } \epsilon-2-\Delta<0. \end{aligned} \tag{19}$$

We shall explicitly evaluate the $B$-site lattice Green's functions at the origin and over the three neighboring sites. From (16) and (17) we have

$$g_0\equiv g(0,0,0)=\frac{1}{8\pi JN}\left(\frac{1}{\epsilon-is-4-\Delta}+(\epsilon-3-\Delta)D(0,0,0)\right),$$

$$g_1\equiv g(\tfrac12,\tfrac12,0)=\frac{1}{24\pi JN}\left(\frac{\epsilon-5-\Delta}{\epsilon-is-4-\Delta}-(\epsilon-3-\Delta)\times(\epsilon-2-\Delta)D(0,0,0)\right),$$

$$\begin{aligned} g_2\equiv g(1,\tfrac12,\tfrac12)=\frac{1}{16\pi JN}\Big(&[\tfrac13(\epsilon-1-\Delta)(\epsilon-3-\Delta)^2(\epsilon-5-\Delta)\\ &+\tfrac34]D(0,0,0)+\tfrac14[2D(2,0,0)-D(2,2,0)]\\ &-\frac{(\epsilon-3-\Delta)(\epsilon-5-\Delta)}{3}\Big)\frac{1}{\epsilon-is-4-\Delta}, \end{aligned}$$

and

$$\begin{aligned} g_3\equiv g(1,1,0)=\frac{1}{8\pi JN}\Big[\Big(&\tfrac13(\epsilon-1-\Delta)(\epsilon-2-\Delta)(\epsilon-3-\Delta)\\ &+\frac{1}{4(\epsilon-is-4-\Delta)}\Big)D(0,0,0)-\frac{1}{4(\epsilon-is-4-\Delta)}\\ &\times[2D(2,0,0)-D(2,2,0)]-\frac{\epsilon-2-\Delta}{3}\Big]. \end{aligned}\tag{20}$$

We note that for evaluating the above four $g$'s for $-\infty<\epsilon<\infty$, we need to calculate the three f.c.c. lattice Green's functions $G(0,0,0)$, $G(2,0,0)$, and $G(2,2,0)$ for the range of $-1\leq E<\infty$. For the case of $E>3$, these f.c.c. Green's function can be evaluated from Eqs. (3.18a), (3.18b), (3.11), and (3.16) in Ref. 1 as a function of the complete elliptic integrals of the first and second kinds, $K(k_\pm)$ and $E(k_\pm)$ with the real moduli $k_\pm$. For practical calculations of $G$'s inside the energy band, $3>E>-1$, it is convenient to use the expressions of $K(k_\pm)$ and $E(k_\pm)$ transformed for $E<-1$ (See Ref. 1), leading to

$$G(0,0,0)=\frac{4}{\pi^2(1+E)}\,\frac{1}{a_+a_-}K(q_+)K(q_-),$$

$$\begin{aligned} G(2,0,0)=\frac{4}{\pi^2(1+E)^3(1-a_+^2)(1-a_-^2)a_+a_-}&[K(q_+)K(q_-)\\ &+E(q_+)E(q_-)-K(q_+)E(q_-)-K(q_-)E(q_+)], \end{aligned}$$

and

$$\begin{aligned} G(2,2,0)=(-\tfrac23E^2+\tfrac43E+1)&G(0,0,0)+2G(2,0,0)-\tfrac43(1+E)\\ +\frac{8(1+E)}{\pi^2}\Big(&\frac{1}{a_+a_-}K(q_+)K(q_-)+2a_+a_-E(q_+)E(q_-)\\ &-\frac{a_-}{a_+}K(q_+)E(q_-)-\frac{a_+}{a_-}K(q_-)E(q_+)\Big), \end{aligned}\tag{21}$$

where

$$a_\pm=\frac{1}{2^{1/2}}\left(1+\frac{(1-E)(-1-E)^{1/2}(3-E)^{1/2}}{(-1-E)^2}\mp\frac{4(-E)^{1/2}(-1-E)^{1/2}}{(-1-E)^2}\right)^{1/2},$$

and

$$q_\pm=2[(-E)^{1/2}\mp(-1-E)^{1/2}]/[1-E+(3-E)^{1/2}(-1-E)^{1/2}]. \tag{21'}$$

In (21′) the square root should be defined by

$$\begin{aligned} (x-E)^{1/2}&=(x-E)^{1/2}\quad \text{for } x-E\geq 0\\ &=i(E-x)^{1/2}\quad \text{for } x-E<0. \end{aligned}$$

Now we shall apply the procedure of analytical continuation[3] to $K(q_\pm)$ and $E(q_\pm)$, both of which are multivalued functions of complex variables $q_\pm$ with branch lines joining $+1$ and $+\infty$, and $-1$ and $-\infty$ along the real axis. Let's specify $K(q_\pm)$ and $E(q_\pm)$, the principal Riemann branches on the first sheet, and use superscript II to refer to branches on the next sheet connected from region $\operatorname{Re} q_\pm>0$, $\operatorname{Im} q_\pm<0$ on the first sheet to region $\operatorname{Re} q_\pm>0$, $\operatorname{Im} q_\pm>0$ across the branch cut between $+1$ and $+\infty$. The branches II of $K^{(\mathrm{II})}(q_\pm)$ and $E^{(\mathrm{II})}(q_\pm)$ can be expressed in terms of the principal branches,

$$K^{(\mathrm{II})}(q_\pm)=K(q_\pm)-2iK'(q_\pm)$$

and

$$E^{(\mathrm{II})}(q_\pm)=E(q_\pm)-2i[K'(q_\pm)-E'(q_\pm)]. \tag{22}$$

For the special case of $0\geq E\geq -1$, it is possible to transform $K(q_\pm)$ and $E(q_\pm)$ with complex moduli $q_\pm$ into functions with real moduli as shown below. We first perform the transformation of $K(q_\pm)$ and $E(q_\pm)$ in the analytic region, $E<-1$, according to

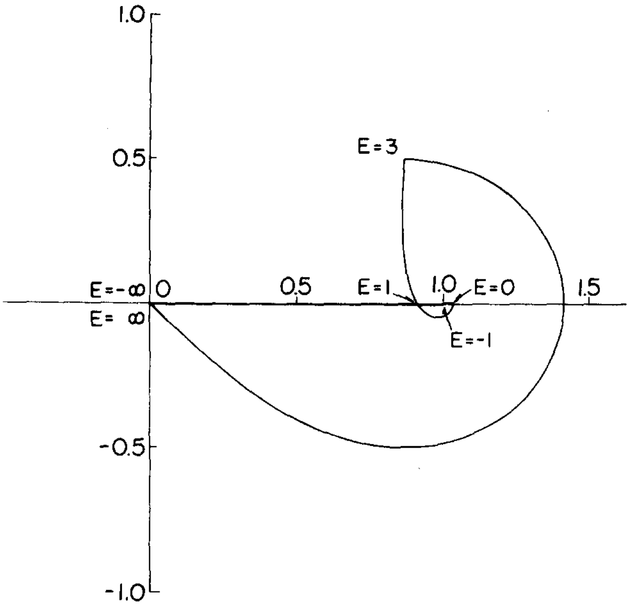



FIG. 3. The values of $q_{1-}$ in the complex plane.

$$K(q_\pm)=\frac{1}{1+q_\pm}K(q_{1_\pm}),$$

and

$$E(q_\pm)=\tfrac{1}{2}[(1+q_\pm)E(q_{1_\pm})+(1-q_\pm)K(q_{1_\pm})], \tag{23}$$

where

$$q_{1_\pm}\equiv\frac{2\sqrt{q_\pm}}{1+q_\pm}=2[2+(-E)^{1/2}(1-E)\mp(1+E)(3-E)^{1/2}]^{-1/2}. \tag{23'}$$

The values of $q_{1_\pm}$ for $-\infty<E<\infty$ are sketched in Figs. 2

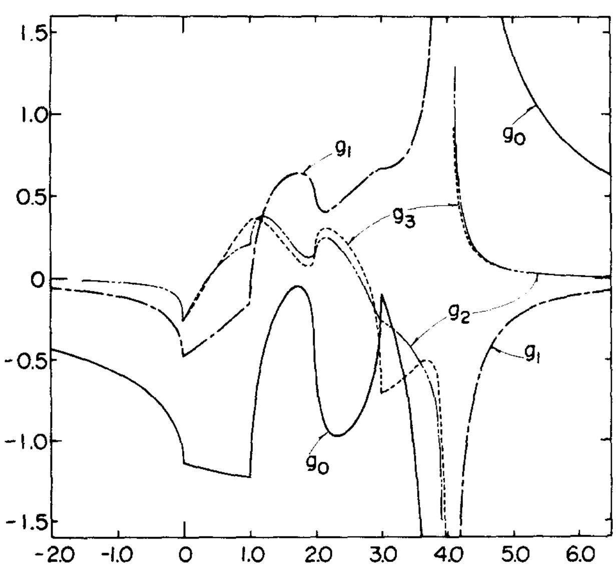



FIG. 4. The real parts of Green's function for $B$-site lattice in spinel. Equations (20), $g_0$, $g_1$, $g_2$ and $g_3$ are multiplied by $8\pi JN$.

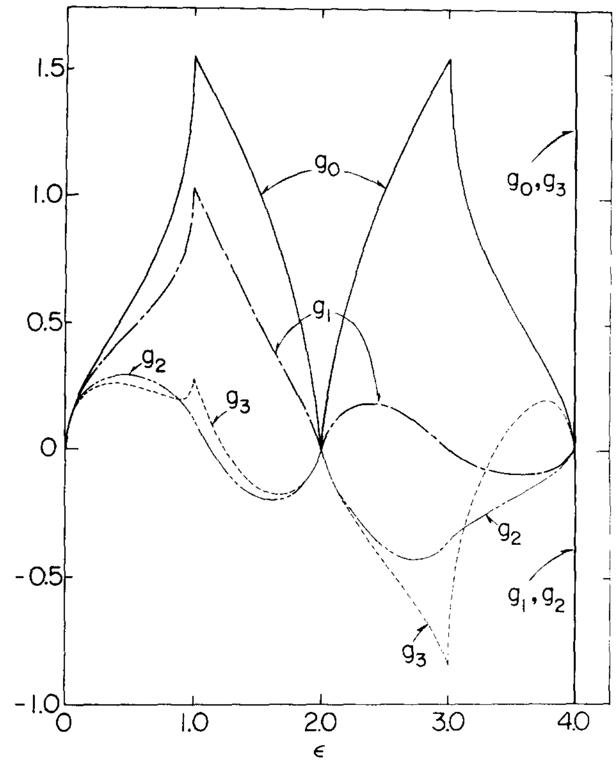



FIG. 5. The imaginary parts of Green's function for $B$-site lattice in spinel. Equations (20) are multiplied by $8\pi JN$.

and 3. When the value of modulus is greater than unity we can use the transformation

$$K(q_1)=\frac{1}{q_1}\left[K\left(\frac{1}{q_1}\right)\pm iK'\left(\frac{1}{q_1}\right)\right],$$

and

$$E(q_1)=q_1\left[E\left(\frac{1}{q_1}\right)\mp iE\left(\frac{1}{q_1}\right)-\frac{(q_1^2-1)}{q_1^2}K\left(\frac{1}{q_1}\right)\pm i\frac{1}{q_1^2}K'\left(\frac{1}{q_1}\right)\right], \tag{24}$$

where the upper sign refers to the case $\mathrm{Im}\,(1/q_{1_\pm})^2>0$ and the lower signs to $\mathrm{Im}\,(1/q_{1_\pm})^2<0$. Equations (22), (23), and (24) lead to the following expressions for $K(q_\pm)$ and $E(q_\pm)$ in terms of $K(1/q_{1_\pm})$ and $E(1/q_{1_\pm})$ with real moduli $1/q_{1_\pm}$:

$$K(q_+)=\frac{1+i|q_{1+}'|}{2q_+}\left[K\left(\frac{1}{q_{1+}}\right)-iK'\left(\frac{1}{q_{1+}}\right)\right],$$

$$K(q_-)=\frac{1-i|q_{1-}'|}{2q_{1-}}\left[K\left(\frac{1}{q_{1-}}\right)-i3K'\left(\frac{1}{q_{1-}}\right)\right],$$

$$E(q_+)=\frac{q_{1+}}{1+i|q_{1+}'|}\left[E\left(\frac{1}{q_{1+}}\right)+iE'\left(\frac{1}{q_{1+}}\right)\right]+\frac{i}{q_{1+}}\left[|q_{1+}'|K\left(\frac{1}{q_{1+}}\right)-K'\left(\frac{1}{q_{1+}}\right)\right],$$

and

$$E(q_-)=\frac{q_{1-}}{1-i|q_{1-}'|}\left[E\left(\frac{1}{q_{1-}}\right)+i3E'\left(\frac{1}{q_{1-}}\right)\right]-\frac{i}{q_{1-}}\left[|q_{1-}'|K\left(\frac{1}{q_{1-}}\right)+3K'\left(\frac{1}{q_{1-}}\right)\right],$$

with

$$q_{1\pm}' = \sqrt{1 - q_{1\pm}^2} \tag{25}$$

For $3 > E > 0$ and applying analytical continuation using (22), $K(q_\pm)$ and $E(q_\pm)$ are directly computed by extending the method of arithmetic-geometric mean from real to complex variables.

The results of numerical calculations of the four Green's functions of $B$-site lattice Eq. (20) for $\Delta = 0$ are plotted in Figs. 4 and 5 where the Green's functions are multiplied by a factor $8\pi JN$. The imaginary part of $g(0,0,0)$ corresponds to the density of states which consist of one acoustic branch for $0 < \epsilon < 2$ and three optical branches for $2 < \epsilon \leq 4$. Two of the optical branches have a $\delta$-function -type spectrum at $\epsilon = 4$ for the present system.


[1]M. Inoue, J. Math. Phys. 15, 704 (1974).
[2]D. N. Zubarev, Sov. Phys. Usp. 3, 320 (1960).
[3]T. Morita and T. Horiguchi, J. Math. Phys. 12, 986 (1970).

# Fields and radiation due to a charge incident on a conducting plane

**Neal J. Maresca and Richard L. Liboff**

*Department of Physics, School of Electrical Engineering and School of Applied and Engineering Physics, Cornell University, Ithaca, New York 14850*
(Received 11 May 1973)

The exact fields of a uniformly moving charged particle which passes through a point hole at $t = 0$ in a grounded, infinitely conducting plane are obtained. Calculation of the fields for all time $(-\infty \leq t \leq +\infty)$ reveals a singular spherical pulse emanating from the origin at $t = 0$ which destroys fields within its sphere in the left half-space, and generates them within its sphere in the right half-space. Further calculation exposes a reversible exchange of reactive field energy between the left and right half-spaces, together with a dissipative radiation loss. A review of techniques which employ the integral $\int \mathbf{E} \cdot \mathbf{j} dV$ to obtain radiation energy is given. For the problem at hand $(-\infty \leq t \leq +\infty)$ it is found that this integral represents radiation alone while in the restricted problem $(-\infty \leq t \leq 0)$, the integral contributes both reactive and radiative (resistive) energy with reactive gain exceeding radiative loss by the factor $\beta^2$.

## I. INTRODUCTION AND SUMMARY OF RESULTS

When a uniformly moving charged particle makes a transition between two media of different optical properties, it radiates. This transition radiation was first investigated by Frank and Ginzburg,[1] and subsequently by Garibyan,[2] and others. It has been pointed out by Ott and Shmoys,[3] however, that most of the analysis is directed towards the evaluation of radiation profiles, intensity and polarization, whereas little attention is given to the explicit nature of the fields. In this paper, the exact relativistic fields of a uniformly moving charged particle normally incident on an infinitely conducting plane are obtained.

The analysis consists of a straightforward solution of Maxwell's equations. The magnetic field is first resolved into time derivatives of its Cartesian components, which are then expressed as Fourier integrals. Evaluation of the Fourier coefficients reduces the problem to one of integration, and subsequent deformations of the integration contour in the $\omega$ plane yield the solution in all regions of space and time.

The fields are found to be simple in form, although highly singular at the reflected and transmitted wavefronts. Before impact, and in various regions after impact, the solution conforms to that of an image picture from which the causal nature of the fields is clearly evident. Inspection of either the starting equations or the final solution reveals an interesting "symmetry" to the problem, namely, the field (either electric or magnetic) on any $z$ plane added to the field on the mirror $-z$ plane equals the sum of the actual and image fields on the original $z$ plane.

Knowledge of the explicit fields allows calculation of both the spectral and angular distributions of radiation. The resulting expressions are in agreement with those originally obtained by Frank and Ginzburg.[1] The total transition radiation energy, however, is found to diverge. This is a consequence of the apparent infinite accelerations accompanying the instantaneous annihilation of the actual-image charge pair at the origin $(z, t) = 0$, as seen by an observer in the left half-space. In like manner an observer in the right half-space detects a singular pulse owing to the instantaneous creation of a charge pair at $(z, t) = 0$. In an attempt to remove these singularities, we cast the total radiated energy into a form amenable to approximation by computing the total work done by the particle. This provides an expression for the distribution of radiation with respect to transverse wavenumber. Following Liboff,[4] an approximate cutoff is introduced into the divergent integral by placing a hole in the plane, and the resulting expression for the finite radiation energy is found to agree with that of Dnestrovskii and Kostomorov[5] in the ultrarelativistic limit.

The expressions for the spectral and wavenumber distributions of radiation are confirmed by a calculation of the irreversible work done by the particle while in the left half-space. Such a method[6] relies on a decomposition of the electric field into advanced and retarded parts, and an identification of the "resistive" (radiative) and "reactive" (inertial) contributions to the work.

An interesting effect is seen in the restricted $(-\infty \leq t \leq 0)$ problem. Namely, it is found that energy gained by the particle from stored field energy exceeds energy lost to radiation by the factor $\beta^2$.

As formulated, the problem considers the charge as approaching, striking and passing through the plane. It can be shown, however, that the solution includes the following as special cases: a charge approaching and stopping at an infinitely conducting plane; a charge starting from and leaving such a plane; and the Bremsstrahlung problems of pair annihilation and pair creation. The only restriction is that in all cases the charges move with constant velocity.

## II. ANALYSIS

### A. The problem

Let a point charge $+e$ move with constant velocity $v$ from $(z, t) = -\infty$ to $(z, t) = +\infty$ striking an infinitely conducting plane at $(z, t) = 0$. The problem is illustrated in Fig. 1.

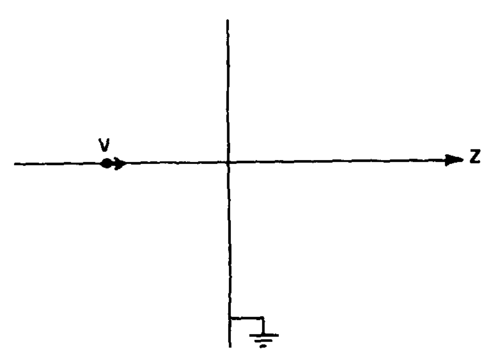



FIG. 1. Experimental configuration, $t < 0$.

The fields are given by Maxwell's equations[7]

$$\nabla \times \mathbf{E}(\mathbf{r},t) = -\mu \frac{\partial \mathbf{H}(\mathbf{r},t)}{\partial t},$$

$$\nabla \times \mathbf{H}(\mathbf{r},t) = \epsilon \frac{\partial \mathbf{E}\ \mathbf{r},t}{\partial t} + \mathbf{j}(\mathbf{r},t),$$

where $\mathbf{j} = \hat{\mathbf{k}} ev\,\delta(z - vt)\,\delta(x)\,\delta(y)$. We take the values of the parameters $\mu$ and $\epsilon$ to be those of free space in order not to complicate the analysis with the presence of Cerenkov radiation. The consideration of more general expressions for these parameters is reserved for a future analysis.

Inasmuch as the current is confined to the $z$ direction, the magnetic field will be totally azimuthal and may be resolved into its $x$ and $y$ components as follows:

$$\mathbf{H}(\mathbf{r},t) = \frac{\partial}{\partial t}\,(F(\mathbf{r},t)\,\hat{\mathbf{i}} + G(\mathbf{r},t)\,\hat{\mathbf{j}}).$$

Written in this form, the functions $F$ and $G$ are found to satisfy the equations

$$\frac{\partial}{\partial t}\left(-\nabla^2 + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) F(\mathbf{r},t) = +e\delta(x)\,\delta\left(t - \frac{z}{v}\right)\frac{d}{dy}\,\delta(y), \tag{1a}$$

$$\frac{\partial}{\partial t}\left(-\nabla^2 + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) G(\mathbf{r},t) = -e\delta(y)\,\delta\left(t - \frac{z}{v}\right)\frac{d}{dx}\,\delta(x), \tag{1b}$$

where $c$ is the velocity of light in vacuum. We concentrate on obtaining $F(\mathbf{r},t)$.

Introducing a triple Fourier integral in the form

$$F(\mathbf{r},t) = \frac{1}{(2\pi)^3}\int_{-\infty}^{+\infty} e^{i(\eta x + \xi y - \omega t)}\,\tilde{F}(\eta\xi z,\omega)\,d\eta d\xi d\omega, \tag{2}$$

reduces Eq. (1a) to an ordinary differential equation for the Fourier coefficient $\tilde{F}$,

$$\left(\frac{d^2}{dz^2} + \alpha^2\right)\tilde{F}(\eta\xi z,\omega) = \frac{e\xi}{\omega}\exp\left(i\frac{z}{v}\,\omega\right),$$

where $\alpha^2 = (\omega^2/c^2) - (\eta^2 + \xi^2)$.

The solution is easily found to be

$$\tilde{F}(\eta\xi z,\omega) = \frac{e\xi}{\omega}\,\frac{\exp[i(z/v)\omega]}{\alpha^2 - (\omega^2/v^2)} + A(\eta\xi\omega)\,e^{i\alpha|z|}.$$

The solution is the sum of the particular and homogeneous solutions, represented by the first and second terms, respectively. The unknown homogeneous coefficient $A$ is determined by requiring the tangential electric field to vanish at $z = 0$, appropriate to an infinitely conducting plane. In terms of $\tilde{F}$, this condition becomes $d\tilde{F}/dz = 0$ at $z = 0$ which yields

$$A(\eta\xi\omega) = \mathrm{sgn}(z)\,\frac{e\xi}{v}\,\frac{1}{\alpha[\alpha^2 - (\omega^2/v^2)]}.$$

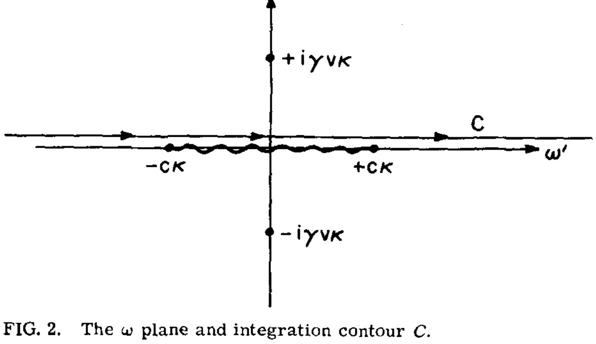



FIG. 2. The $\omega$ plane and integration contour $C$.

The Fourier coefficient $\tilde{F}$ is now determined, and may be substituted into the Fourier integral given by (2) to obtain

$$\begin{aligned} H_x(\mathbf{r},t) &= \frac{1}{(2\pi)^3}\frac{\partial}{\partial t}\int_{-\infty}^{+\infty}\frac{e\xi}{\omega}\,\frac{\exp[i(z/v)\omega]}{\alpha^2 - (\omega^2/v^2)}\,e^{i(\eta x + \xi y - \omega t)}\,d\eta d\xi d\omega \\ &\quad - \frac{\mathrm{sqn}(z)}{(2\pi)^3}\frac{\partial}{\partial t}\int_{-\infty}^{+\infty}\frac{e\xi}{v}\,\frac{e^{i\alpha|z|}}{\alpha[\alpha^2 - (\omega^2/v^2)]} \\ &\quad \times e^{i(\eta x + \xi y - \omega t)}\,d\eta d\xi d\omega. \end{aligned} \tag{3}$$

The first term of this expression is easily evaluated (Appendix 1) and found to be the $x$ component of the magnetic field of a charge $e$ moving with constant velocity $v$ in free space. This field is denoted by $H_x^a(\mathbf{r},t)$, and given by

$$\begin{aligned} H_x^a(\mathbf{r},t) &= \frac{(-e)}{4\pi}\,\gamma v\,\sin\phi\int_0^\infty J_1(\kappa\rho)\,\kappa e^{-\gamma\kappa|z - vt|}\,d\kappa \\ &= \frac{(-e)}{4\pi}\,\gamma v\,\sin\phi\,\frac{\rho}{[\rho^2 + \gamma^2(z - vt)^2]^{3/2}}, \end{aligned}$$

where we have introduced the polar coordinates $\begin{smallmatrix} x=\rho\cos\phi \\ y=\rho\sin\phi \end{smallmatrix}$. This "actual" field, so denoted since it is associated with the moving charge, is to be distinguished from the "image" field associated with the moving image at $z = -vt$, and given by $H_x^i(\rho,z,t) = H_x^a(\rho,-z,t)$. We also introduce the "previous" fields

$$H_x^p(\mathbf{r},t) = \begin{cases} H_x^i + H_x^a, & z < 0 \\ 0, & z > 0 \end{cases}$$

which, as will be shown below, are the fields seen by an observer at $(\mathbf{r},t)$ previous to $t = r/c$, where $r$ is the distance from the point of impact.

Equation (3) may now be written as

$$H_x(\mathbf{r},t) = H_x^a(\mathbf{r},t) - \frac{\mathrm{sgn}(z)}{(2\pi)^3}\frac{\partial}{\partial t}\int_{-\infty}^{+\infty}\frac{e\xi}{v}\,\frac{e^{i(\eta x + \xi y + \alpha|z| - \omega t)}}{\alpha[\alpha^2 - (\omega^2/v^2)]} \times d\eta d\xi d\omega. \tag{4}$$

We denote by $\alpha'$ the real part of $\alpha$, and by $\alpha''$ the imaginary part. In order to prevent the magnetic field from diverging as $|z| \to \infty$, we require $\alpha'' > 0$ everywhere on the integration path. Also, to insure left-moving waves for $z < 0$ and right-moving waves for $z > 0$ (the radiation condition) we further require $\alpha' < 0$ for $\omega < 0$, and $\alpha' > 0$ for $\omega > 0$. These conditions on $\alpha'$ and $\alpha''$ specify the integration path in the $\omega$ plane. Introducing the coordinates and integration variables

$$x = \rho\cos\phi, \qquad \eta = \kappa\cos\phi',$$
$$y = \rho\sin\phi, \qquad \xi = \kappa\sin\phi',$$

(4) becomes

$$H_x(\mathbf{r},t) = H_x^a - \mathrm{sgn}(z)\,\frac{ei}{(2\pi)^2}\,(-v\gamma^2)\,\sin\phi\,\frac{\partial}{\partial t}\int_0^\infty J_1(\kappa\rho)\kappa^2 \times \int_C \frac{e^{i(\alpha|z| - \omega t)}}{\alpha[\omega^2 + (\gamma v\kappa)^2]}\,d\omega d\kappa \tag{5}$$

where $\alpha = [(\omega^2/c^2) - \kappa]^{1/2}$.

Fig. 2 shows the placement of the poles, the branch cut, and the integration contour $C$ in the $\omega$ plane.

### B. Contour deformations

The solution to the problem now lies in the evaluation of the integral expression contained in (5). In this section we eliminate the $\omega$ integration by deforming the contour into the upper and lower planes. We note that as $\omega \to \infty$,

$$e^{i(\alpha|z|-\omega t)} \to e^{i\omega[(|z|/c)-t]}.$$

Thus the convergence of the $\omega$ integration depends on $|z| \gtrless ct$.

For $|z| > ct$, the integration converges at infinity in the upper half plane and we may lift the contour to $\omega'' = +\infty$, as shown in Fig. 3. The only contribution is that due to the pole at $\omega = +i\gamma v\kappa$, and is found to be

$$\begin{aligned} H_x(\mathbf{r}, t) &= H_x^a - \mathrm{sgn}(z)\frac{(-e)}{4\pi}\gamma v \sin\phi \int_0^\infty J_1(\kappa\rho)\kappa e^{-\gamma\kappa(|z|-vt)} d\kappa \\ &= H_x^a + \begin{cases} +H_x^i, & z < 0 \\ -H_x^a, & z > 0 \end{cases}, \\ &= H^p(\mathbf{r}, t), \quad |z| > vt. \end{aligned}$$

Although the $\kappa$ integration converges for $|z| > vt$, this does not contradict the initial $|z| > ct$ requirement, but replaces it. It must be kept in mind that (5) contains a double integration and therefore any restrictions for convergence on the $\omega$ integration will be altered, in the final result, by the presence of the $\kappa$ integration. In this case we find that whereas the $\omega$ integration converges for $|z| > ct$, the entire double integration is valid for $|z| > vt$.

For $|z| < ct$, the $\omega$ integral converges in the lower half place. Before deforming the contour, however, we first apply the Fourier convolution theorem to obtain

$$\int_{-\infty}^{+\infty} \frac{e^{i(\alpha|z|-\omega t)}}{\alpha[\omega^2 + (\gamma v\kappa)^2]} d\omega \equiv \int_{-\infty}^{+\infty} f(u) g(t-u)\, du,$$

where

$$f(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{e^{i(\alpha|z|-\omega u)}}{\alpha} d\omega,$$

$$g(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{e^{-i\omega u}}{\omega^2 + (\gamma v\kappa)^2} d\omega.$$

This is convenient since $f(u)$ now contains no poles, and $g(u)$ no branches. The integral $g(t-u)$ is easily found to be

$$g(t-u) = \frac{1}{\gamma v\kappa}\sqrt{\pi/2}\; e^{-\gamma v\kappa|t-u|}.$$

For $f(u)$, the contour may be deformed into the lower plane provided $u > |z|/c$. This encircles the branch cut, as shown in Fig. 4, and the entire integral is found to be a representation of a zero-order Bessel function,[8]

$$f(u) = -(2\pi i c/\sqrt{2\pi})\, J_0[\kappa c\sqrt{u^2 - (|z|^2/c^2)}], \qquad u > |z|/c.$$

Collecting results for $|z| < ct$ and substitution into (5), after the change of variable $u' = u/c$, yields

$$\begin{aligned} H_x(\mathbf{r}, t) &= H_x^a - \mathrm{sgn}(z)\frac{(-e)}{4\pi}\gamma \sin\phi \frac{\partial}{\partial t}\int_0^\infty J_1(\kappa\rho)\kappa \\ &\quad \times \Big(\int_{|z|}^{ct} J_0(\kappa\sqrt{u'^2 - |z|^2})\, e^{-\gamma v\kappa[t-(u'/c)]} du' \\ &\quad + \int_{ct}^{\infty} J_0(\kappa\sqrt{u'^2 - |z|^2})\, e^{+\gamma v\kappa[t-(u'/c)]} du'\Big) d\kappa \\ &= H_x^a - \mathrm{sgn}(z)\frac{(-e)}{4\pi}\gamma \sin\phi \frac{\partial}{\partial t}\int_0^\infty J_1(\kappa\rho)\kappa (I_{|z|}^{ct} + I_{ct}^{\infty}) d\kappa, \end{aligned} \tag{6}$$

where this last equation defines the integrals $I_{|z|}^{ct}$ and $I_{ct}^{\infty}$.

### C. Evaluation of $I_{|z|}^{ct}$ and $I_{ct}^{\infty}$

To evaluate the integrals occurring in the integrand of (6), we will expand the integrands, perform the indicated integrations, then resum. For this purpose we make use of the addition theorem for Bessel functions[9]

$$J_\nu(\tilde{\omega}) = \left(\frac{Z - ze^{+i\phi}}{Z - ze^{-i\phi}}\right)^{\nu/2} \sum_{n=-\infty}^{+\infty} J_{\nu+n}(Z) J_n(z) e^{in\phi},$$

where $\tilde{\omega} = (Z^2 + z^2 - 2Zz\cos\phi)^{1/2}$.

Thus,

$$J_0(\kappa\sqrt{u'^2 - |z|^2}) = 2\sum_{n=0}^{\infty}{}' (-1)^n J_{2n}(i\kappa|z|) J_{2n}(\kappa u'),$$

where the prime on the summation weights the $n = 0$ term with a factor $\frac{1}{2}$. By further introducing the representation

$$J_n(z) = \frac{i^{-n}}{2\pi}\int_{-\pi}^{+\pi} e^{i(z\cos\theta + n\theta)} d\theta,$$

we find

$$I_{|z|}^{ct} = \frac{e^{-\gamma v\kappa t}}{2\pi}\, 2\sum_{n=0}^{\infty}{}' J_{2n}(i\kappa|z|)\int_{-\pi}^{+\pi} d\theta\, e^{2in\theta}\int_{|z|}^{ct} \times e^{u'(i\kappa\cos\theta + \gamma\beta\kappa)} du',$$

$$I_{ct}^{\infty} = \frac{e^{+\gamma v\kappa t}}{2\pi}\, 2\sum_{n=0}^{\infty}{}' J_{2n}(i\kappa|z|)\int_{-\pi}^{+\pi} d\theta\, e^{2in\theta}\int_{ct}^{\infty} \times e^{u'(i\kappa\cos\theta - \gamma\beta\kappa)} du'.$$

The $u'$ integrations are elementary, producing factors $e^{i|z|\kappa\cos\theta}$ and $e^{ict\kappa\cos\theta}$ which we expand in the form

$$e^{iz\cos\theta} = \sum_{m=-\infty}^{+\infty} i^m J_m(z) e^{im\theta} = 2\sum_{m=0}^{\infty}{}' i^m J_m(z)\cos m\theta.$$

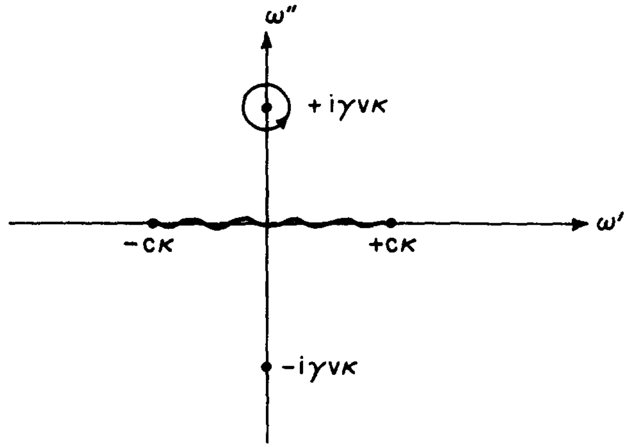



FIG. 3. Deformation of the integration contour into the upper half plane for $|z| > ct$.

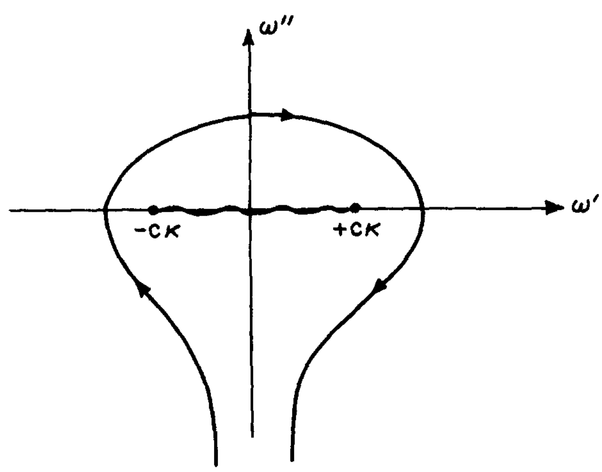



FIG. 4. Deformation of the integration contour of $f(u)$ into the lower half plane for $u > |z|/c$. The integrand of $f(u)$ contains no poles.

The remaining $\theta$ integrations are performed by first writing

$$\cos A \cos B = \tfrac{1}{2}[\cos(A+B) + \cos(A-B)]$$

and noting[10] that

$$\int_0^\pi \frac{\cos n\theta}{1 + a\cos\theta}\, d\theta = \frac{\pi}{\sqrt{1-a^2}}\left(\frac{\sqrt{1-a^2}-1}{a}\right)^n, \qquad a^2 < 1, \quad n > -1.$$

This last step removes all integrations and yields

$$I^{ct}_{|z|} = \frac{2}{\gamma\kappa}\sum_{n=0}^{\infty'}(-1)^n\left(\sum_{m=0}^{2n-1'}(-1)^m[\gamma(1-\beta)]^{2n-m} + \sum_{m=2n}^{\infty'}[\gamma(1-\beta)]^{m-2n} + \sum_{m=0}^{\infty'}[\gamma(1-\beta)]^{m+2n}\right)$$
$$\times [J_m(\kappa ct) - e^{\gamma\kappa(|z|\beta - vt)}J_m(\kappa|z|)]J_{2n}(i\kappa|z|),$$

$$I^{\infty}_{ct} = \frac{2}{\gamma\kappa}\sum_{n=0}^{\infty'}(-1)^n\left(\sum_{m=0}^{2n-1'}[\gamma(1-\beta)]^{2n-m} + \sum_{m=2n}^{\infty'}(-1)^m[\gamma(1-\beta)]^{m-2n} + \sum_{m=0}^{\infty'}(-1)^m[\gamma(1-\beta)]^{m+2n}\right)$$
$$\times J_m(\kappa ct)J_{2n}(i\kappa|z|).$$

Comparison of these two expressions indicates that except for the second term in the brackets occurring in $I^{ct}_{|z|}$, the sum $I^{ct}_{|z|} + I^{\infty}_{ct}$ will contain only even terms in $m$ due to the placement of the $(-1)^m$ factors. Thus with $\zeta \equiv \gamma^2(1-\beta)^2$ the sum becomes

$$I^{ct}_{|z|} + I^{\infty}_{c}$$
$$= -\frac{2}{\gamma\kappa}e^{\gamma\kappa(|z|\beta - vt)}\sum_{n=0}^{\infty'}(-1)^n$$
$$\times\left(\sum_{m=0}^{2n-1'}(-1)^m[\gamma(1-\beta)]^{2n-m} + \sum_{m=2n}^{\infty'}[\gamma(1-\beta)]^{m-2n} + \sum_{m=0}^{\infty'}[\gamma(1-\beta)]^{2n+m}\right)$$
$$\times J_{2n}(i\kappa|z|)J_m(\kappa ct) + \frac{4}{\gamma\kappa}\sum_{n=0}^{\infty'}(-1)^n$$
$$\times\left(\sum_{m=0}^{n-1'}\zeta^{n-m} + \sum_{m=n}^{\infty'}\zeta^{m-n} + \sum_{m=0}^{\infty'}\zeta^{m+n}\right)J_{2n}(i\kappa|z|)J_{2m}(\kappa ct)$$
$$\equiv \textstyle\sum_1 + \sum_2,$$

where $\sum_1$ consists of the first three terms, and $\sum_2$ of the second three. It is not difficult to evaluate $\sum_1$ (Appendix 2) and we find its contribution to the field is given by

$$H_x(\text{due to } \textstyle\sum_1) = \begin{cases} +H^i_x, & z < 0 \\ -H^a_x, & z > 0\end{cases}, \qquad |z| < vt.$$

We recall from Sec. B above that these same fields were found for $|z| > vt$. We see, then, that they exist everywhere in space.

To simplify $\sum_2$, we rewrite it as follows:

$$\textstyle\sum_2 = \displaystyle\frac{4}{\gamma\kappa}\sum_{n=0}^{\infty'}(-1)^n\left(\sum_{m=0}^{n-1'}\zeta^{n-m} + \sum_{m=n}^{\infty'}\zeta^{m-n} + \sum_{m=0}^{\infty'}\zeta^{m+n}\right)$$
$$\times J_{2n}(i\kappa|z|)J_{2m}(\kappa ct)$$
$$= \frac{4}{\gamma\kappa}\sum_{\alpha=0}^{\infty'}\zeta^\alpha\sum_{l=0}^{\infty'}(-1)^l[J_{2\alpha-2l}(\kappa ct) + J_{2\alpha+2l}(\kappa ct)]J_{2l}(i\kappa|z|)$$
$$= \frac{4}{\gamma\kappa}\sum_{\alpha=0}^{\infty'}\zeta^\alpha\sum_{l=-\infty}^{+\infty}(-1)^l J_{2\alpha+2l}(\kappa ct)J_{2l}(i\kappa|z|).$$

The $l$ summation in this last expression can be computed exactly with the help of a modified form of the Bessel function addition theorem,[9] and we obtain

$$\textstyle\sum_2 = \displaystyle\frac{2}{\gamma\kappa}\sum_{\alpha=0}^{\infty'}\zeta^\alpha(\lambda^\alpha + \lambda^{-\alpha})J_{2\alpha}(\kappa\sqrt{\tau^2 - |z|^2}),$$

where $\lambda = (\tau + |z|)/(\tau - |z|)$, $\tau = ct$, $\zeta = \gamma^2(1-\beta)^2$, and the sum converges for $|z| < \tau$.

Collecting all results, including those of Section B above, we find that the magnetic field in all regions of space and time is given by

$$H_x(\mathbf{r}, t) = H^p_x - \operatorname{sgn}(z)\frac{(-e)c}{2\pi}\sin\phi\frac{\partial}{\partial\tau}$$
$$\times\sum_{\alpha=0}^{\infty'}\zeta^\alpha(\lambda^\alpha + \lambda^{-\alpha})\int_0^\infty J_1(\kappa\rho)J_{2\alpha}(\kappa\sqrt{\tau^2 - |z|^2})\,d\kappa$$
$$\equiv H^p_x - \operatorname{sgn}(z)\frac{(-e)c}{2\pi}\sin\phi\,\Lambda. \tag{7}$$

### D. Reduction of $\Lambda$

In this section we calculate the contribution to the field due to $\Lambda$ as defined in (7),

$$\Lambda = \frac{\partial}{\partial\tau}\sum_{\alpha=0}^{\infty'}\zeta^\alpha(\lambda^\alpha - \lambda^{-\alpha})\int_0^\infty J_1(\kappa\rho)J_{2\alpha}(\kappa\sqrt{\tau^2 - |z|^2})\,d\kappa.$$

The integral occurring in this expression is known and given by[9,10]

$$\int_0^\infty J_1(\kappa\rho)J_{2\alpha+2}(\kappa\sqrt{\tau^2 - |z|^2})\,d\kappa$$
$$= \begin{cases} \dfrac{\rho}{\tau^2 - |z|^2}P^{(1,0)}_\alpha\left(1 - \dfrac{2\rho^2}{\tau^2 - |z|^2}\right), & r < ct, \\[2ex] \dfrac{(-1)^\alpha}{2\rho}, & r = ct, \\[2ex] 0, & r > ct, \end{cases}$$

where $P^{(1,0)}_\alpha(x)$ are the Jacobi polynomials[10] generated as follows:

$$\frac{2}{R'(1 - z + R')} = \sum_{\alpha=0}^\infty z^\alpha P^{(1,0)}_\alpha(x), \qquad R' = \sqrt{1 - 2xz + z^2}, \quad |z| < 1.$$

The conditions $r \gtrless \tau$ have been written in place of the conditions $P \gtrless \sqrt{\tau^2 - |z|^2}$, and the solution separates into three cases.

#### 1. *Case 1: $r > ct$*

Here we see that $\Lambda = 0$, and consequently by (7)

$$H_x(\mathbf{r}, t) = H^p_x(\mathbf{r}, t),$$

so that for this, the simplest of the three cases, fields reduce to the previous fields, $H^p_x$.

#### 2. *Case 2: $r < ct$*

In this case we must evaluate

$$\Lambda = \frac{\partial}{\partial\tau}\left[\frac{\rho}{\tau^2 - |z|^2}\sum_{\alpha=0}^\infty (z')^{\alpha+1}P^{(1,0)}_\alpha\left(1 - \frac{2\rho^2}{\tau^2 - |z|^2}\right)\right]$$
$$+ \text{ same with } z' \to z''$$

where $z' = \zeta\lambda = \gamma^2(1-\beta)^2(\tau + |z|/\tau - |z|)$ and $z'' = \zeta\lambda^{-1}$.

The sums are evaluated with the help of the generating formula, and after some algebra we find

$$H_x(\mathbf{r}, t) = H^p_x + \begin{cases} -H^a_x - H^i_x, & z < 0, \\ +H^a_x + H^i_x, & z > 0, \end{cases}$$
$$= \begin{cases} 0, & z < 0, \\ H^a_x + H^i_x, & z > 0. \end{cases}$$

The results for the cases $r \gtrless ct$ are depicted in Fig. 5. For $t < 0$, the image picture dominates as a $z < 0$ observer sees the fields $H_x = H_x^a + H_x^i$, whereas a $z > 0$ observer, being shielded from the left by an infinitely conducting plane, sees no fields. For $t > 0$, a hemispherical pulse propagating at the speed of light destroys all fields behind it for $z < 0$, and only after a time $t = r/c$ is a $z < 0$ observer aware that the image picture is no longer valid. Similarly, a $z > 0$ observer must await the information that a charge has passed into his space, after which he finds the image picture in force. It is interesting to consider the charge as having effected a hole in the plane through which the fields in the left space leak symmetrically into the right space.

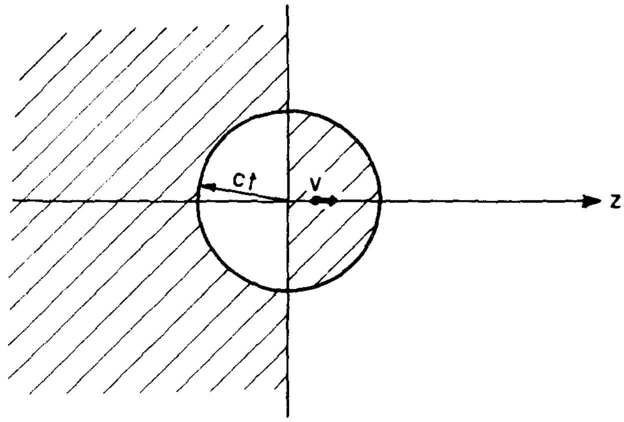



FIG. 5. The solution for $t > 0$. Shaded regions represent $\mathbf{H} = \mathbf{H}^a + \mathbf{H}^i$. Unshaded regions have zero field.

### 3. Case 3: $r = ct$

In this case the following relations will prove useful[8]:

(i) $J_{\alpha-1}(x) = \dfrac{\alpha}{x} J_\alpha(x) + J'_\alpha(x)$,

(ii) $\int_0^\infty J_1(\kappa\rho) J_1(\kappa\sqrt{\tau^2 - |z|^2})\, \kappa d\kappa = \dfrac{1}{\rho}\, \delta(\rho - \sqrt{\tau^2 - |z|^2})$,

(iii)
$$\int_0^\infty J_1(\kappa\rho) J_{2\alpha+1}(\kappa\sqrt{\tau^2 - |z|^2})\, \kappa d\kappa = \frac{1}{\rho} \sum_{l=1}^{\alpha} (-1)^{l+\alpha} \times 4l \int_0^\infty J_1(\kappa\rho)\, J_{2l}(\kappa\sqrt{\tau^2 - |z|^2})\, d\kappa + \frac{(-1)^\alpha}{\rho}\, \delta(\rho - \sqrt{\tau^2 - |z|^2}).$$

Relation (iii) is obtained by repeated application of the recurrence relations for Bessel functions.

We now carry out the time derivative operation in $\Lambda$ and employ (i) above to obtain

$$\Lambda = \frac{2|z|}{\tau^2 - |z|^2} \sum_{\alpha=0}^{\infty\,\prime} \alpha\zeta^\alpha(\lambda^{-\alpha} - \lambda^\alpha) \int_0^\infty J_1(\kappa\rho) J_{2\alpha}(\kappa\sqrt{\ }) d\kappa + \sum_{\alpha=0}^{\infty\,\prime} \zeta^\alpha(\lambda^{-\alpha} + \lambda^\alpha) \int_0^\infty J_1(\kappa\rho) J_{2\alpha-1}(\kappa\sqrt{\ }) \frac{\kappa\tau}{\sqrt{\ }}\, d\kappa - \frac{2\tau}{\tau^2 - |z|^2} \sum_{\alpha=0}^{\infty\,\prime} \alpha\zeta^\alpha(\lambda^{-\alpha} + \lambda^\alpha) \int_0^\infty J_1(\kappa\rho) J_{2\alpha}(\kappa\sqrt{\ }) d\kappa \equiv \Lambda_1 + \Lambda_2 + \Lambda_3. \tag{8}$$

In $\Lambda_2$, we separate the $\alpha = 0, 1$ terms explicitly and make use of relations (ii) and (iii) to obtain

$$\begin{aligned}
\frac{\rho^2}{\tau} \Lambda_2 &= \delta(\rho - \sqrt{\tau^2 - |z|^2}) \times \left(-1 + \zeta(\lambda + \lambda^{-1}) + \sum_{\alpha=1}^{\infty} (-\zeta)^{\alpha+1}(\lambda^{\alpha+1} + \lambda^{-\alpha-1})\right) \\
&\quad + \sum_{\alpha=1}^{\infty} \zeta^{\alpha+1}(\lambda^{\alpha+1} + \lambda^{-\alpha-1}) \sum_{l=1}^{\alpha} (-1)^{l+\alpha} \times 4l \int_0^\infty J_1(\kappa\rho) J_{2l}(\kappa\sqrt{\ }) d\kappa \\
&= -\frac{1}{2} \frac{\beta \sin^2\theta}{1 - \beta^2\cos^2\theta}\, \delta(\rho - \sqrt{\tau^2 - |z|^2}) \\
&\quad + 4 \sum_{\alpha=0}^{\infty} \zeta^{\alpha+2}(\lambda^{\alpha+1} + \lambda^{-\alpha-1}) \sum_{l=0}^{\alpha} (-1)^{l+\alpha}(l+1) \times \int_0^\infty J_1(\kappa\rho) J_{2l+2}(\kappa\sqrt{\ }) d\kappa,
\end{aligned}$$

where we have summed all terms involving the $\delta$-function, made the substitutions $|z| = r\cos\theta$, $\rho = r\sin\theta$, $r = \tau$, and slightly rearranged the summation on the remaining integral expression. We note that

$$\delta(\rho - \sqrt{\tau^2 - |z|^2}) = \delta(r\sin\theta - \sqrt{\tau^2 - r^2\cos^2\theta}) = \frac{\delta(r - \tau)}{|\tau/-\sqrt{\tau^2 - r^2\cos^2\theta}|_{\tau=r}} = \sin\theta\,\delta(r - \tau)$$

Returning these results to (8) and eliminating the vanishing $\alpha = 0$ terms in $\Lambda_1$ and $\Lambda_2$, yields

$$\begin{aligned}
\Lambda &= -\frac{1}{2} \frac{\beta\sin\theta}{1 - \beta^2\cos^2\theta} \frac{\delta(r - ct)}{r} \\
&\quad + \frac{2|z|}{\tau^2 - |z|^2} \sum_{\alpha=0}^{\infty} (\alpha+1)\zeta^{\alpha+1}(\lambda^{-\alpha-1} - \lambda^{\alpha+1}) \int_0^\infty J_1(\kappa\rho) J_{2\alpha+2}(\kappa\rho) d\kappa \\
&\quad + \frac{4\tau}{\tau^2 - |z|^2} \sum_{\alpha=0}^{\infty} \zeta^{\alpha+2}(\lambda^{-\alpha-1} + \lambda^{\alpha+1}) \sum_{l=0}^{\alpha} (-1)^{l+\alpha} \times (l+1) \int_0^\infty J_1(\kappa\rho) J_{2l+2}(\kappa\rho) d\kappa - \frac{2\tau}{\tau^2 - |z|^2} \\
&\quad \times \sum_{\alpha=0}^{\infty} (\alpha+1)\zeta^{\alpha+1}(\lambda^{-\alpha-1} + \lambda^{\alpha+1}) \int_0^\infty J_1(\kappa\rho) J_{2\alpha+2}(\kappa\rho) d\kappa.
\end{aligned}$$

From above, each integral is known and we find, aside from the $\delta$-function,

$$\Lambda = \frac{1}{\rho(\tau^2 - |z|^2)} \Bigg((|z| - \tau) \sum_{\alpha=0}^{\infty} (-\zeta/\lambda)^{\alpha+1}(\alpha+1) - (|z| + \tau) \sum_{\alpha=0}^{\infty} (-\zeta\lambda)^\alpha(\alpha+1) + \tau \sum_{\alpha=0} (-1)^\alpha \times [(\zeta\lambda)^{\alpha+2} + (\zeta/\lambda)^{\alpha+2}](\alpha+1)(\alpha+2)\Bigg).$$

These sums are easily evaluated, and writing $|z| = r\cos\theta$, $\rho = r\sin\theta$, $r = \tau$ yields, for the magnetic field at $r = ct$,

$$\begin{aligned}
H_x(\mathbf{r}, t) &= H_x^p - \operatorname{sgn}(z)\sin\phi\left[-\frac{1}{2}\frac{(-e)}{4\pi}\frac{v\sin\theta}{(\gamma r)^2}\left(\frac{1}{(1+\beta\cos\theta)^3} + \frac{1}{(1-\beta\cos\theta)^3}\right) - \frac{(-e)}{4\pi} v \frac{\sin\theta}{1 - \beta^2\cos^2\theta}\frac{\delta(r - ct)}{r}\right] \\
&= H_x^p - \operatorname{sgn}(z)\left(-\frac{1}{2}(H_x^a + H_x^i) + \frac{ev}{2\pi}\sin\phi \times \frac{\sin\theta}{1 - \beta^2\cos^2\theta}\frac{\delta(r - ct)}{r}\right).
\end{aligned}$$

The calculation for $H_y$ is similar to that for $H_x$ and the total field is found to be given by

$$\mathbf{H}(\mathbf{r}, t) = \hat{\phi}\, H(\mathbf{r}, t),$$

where

$$\begin{aligned}
H(\mathbf{r}, t) &= \begin{cases} H^a + H^i, & z < 0, \\ 0, & z > 0, \end{cases} \quad r > ct, \\
&= \begin{cases} 0, & z < 0, \\ H^a + H^i, & z > 0, \end{cases} \quad r < ct, \\
&= \begin{cases} \dfrac{1}{2}(H^a + H^i) - \dfrac{ev}{2\pi}\dfrac{\sin\theta}{1 - \beta^2\cos^2\theta}\dfrac{\delta(r - ct)}{r}, & z < 0, \\ \dfrac{1}{2}(H^a + H^i) + \dfrac{ev}{2\pi}\dfrac{\sin\theta}{1 - \beta^2\cos^2\theta}\dfrac{\delta(r - ct)}{r}, & z > 0, \end{cases} \quad r = ct.
\end{aligned} \tag{9}$$

Inspection of the solution reveals the interesting "symmetry"

$$\mathbf{H}(\rho, z, t) + \mathbf{H}(\rho, -z, t) = \mathbf{H}^a(\mathbf{r}, t) + \mathbf{H}^i(\mathbf{r}, t)$$

which is also apparent from (4). It is noted that this symmetric sum (in $z$) does not include the singular radiation pulse. A similar condition also holds for the electric field (replacing $\mathbf{H}$ by $\mathbf{D}$), associated with (9), which we now compute in two ways.

First, we recall that the field $H^a$ is due to a charge $+e$ in vacuum moving uniformly from $z = -\infty$ to $z = +\infty$, and that $H^i$ is identical to $H^a$ except that both the charge and direction of motion are reversed. These fields, then, can also be found by appropriate Lorentz transformations on the Coulomb fields of charges at rest in vacuum;[4] a technique by which we easily find the electric fields associated with $H^a + H^i$ to be

$$D_\rho = \frac{e}{4\pi}\gamma\rho\left(\frac{1}{[\rho^2 + \gamma^2(z - vt)^2]^{3/2}} - \frac{1}{[\rho^2 + \gamma^2(z + vt)^2]^{3/2}}\right)$$
$$= D^a_\rho + D^i_\rho,$$
$$D_z = \frac{e}{4\pi}\gamma\left(\frac{z - vt}{[\rho^2 + \gamma^2(z - vt)^2]^{3/2}} - \frac{z + vt}{[\rho^2 + \gamma^2(z + vt)^2]^{3/2}}\right)$$
$$= D^a_z + D^i_z.$$

Second, the electric radiation fields are obtained directly from Maxwell's equations using the known magnetic radiation fields, and are given by

$$D_\rho = \frac{e\beta}{2\pi}\frac{\sin\theta\cos\theta}{1 - \beta^2\cos^2\theta}\frac{\delta(r - ct)}{r}\,\mathrm{sgn}(z),$$
$$D_z = -\frac{e\beta}{2\pi}\frac{\sin^2\theta}{1 - \beta^2\cos^2\theta}\frac{\delta(r - ct)}{r}\,\mathrm{sgn}(z).$$

The radiation fields, therefore, are found to be

$$\mathbf{H}(\mathbf{r}, t) = \hat{\boldsymbol{\phi}}\,\frac{ev}{2\pi}\frac{\sin\theta}{1 - \beta^2\cos^2\theta}\frac{\delta(r - ct)}{r}\,\mathrm{sgn}(z)$$
$$\mathbf{D}(\mathbf{r}, t) = (\hat{\boldsymbol{\rho}}\cos\theta - \hat{\mathbf{k}}\sin\theta)\,\frac{e\beta}{2\pi}\frac{\sin\theta}{1 - \beta^2\cos^2\theta}\frac{\delta(r - ct)}{r}\,\mathrm{sgn}(z). \tag{10}$$

## III. TRANSITION ENERGY

The fields given by (10) are seen to be normal to each other and to the direction of propagation. The electric radiation field is linearly polarized and lies in the plane determined by the ray and the particle trajectory. Such linear polarization is characteristic of transition radiation. The results obtained above agree with the asymptotic fields previously reported by Garibyan,[2] who considers a charge incident on a dielectric half-space. The nature of the transition radiation may be investigated in one of three ways due to Poynting's theorem,[11]

$$\oint_A \mathbf{S}\cdot d\mathbf{A} = -\int_v \mathbf{j}\cdot\mathbf{E}\,dV - \frac{1}{2}\frac{\partial}{\partial t}\int_v (E^2 + B^2)\,dV, \tag{11}$$

which expresses the power radiated in terms of the rate at which the current does work and the rate of storage of electromagnetic energy: One may integrate the lhs of (11) for all time to obtain expressions for both the angular and spectral distributions of radiation [method (a)]. The integration of the rhs [method (b)] provides the wavenumber distribution of radiated energy. Finally, one may compute the irreversible work done by the charge on the fields in the left half-space, due to the presence of the plane [method (c)]; a technique employed by Schwinger[6] in obtaining the radiation from an electron in arbitrary motion. In the following all three techniques are employed to examine the radiation due to the impact of the particle on the plane.

(a) From the lhs of (11) the energy radiated into the solid angle $d\Omega = \sin\theta\,d\theta\,d\phi$ is

$$\frac{d\mathcal{E}_r}{d\Omega} = r^2\int_{-\infty}^{+\infty}(\mathbf{S}\cdot\hat{r})\,dt = \frac{r^2}{c\epsilon_0}\int_{-\infty}^{+\infty}|\mathbf{H}(\mathbf{r}, t)|^2\,dt, \tag{12}$$

where $\mathbf{S}$ is the Poynting vector associated with the radiation fields (10),

$$\mathbf{S} = \mathbf{E}\times\mathbf{H} = \frac{\hat{\mathbf{r}}}{c\epsilon_0}\,|\mathbf{H}(\mathbf{r}, t)|^2.$$

To obtain the spectral distribution of radiation, we introduce the Fourier representation

$$\mathbf{H}(\mathbf{r}, \omega) = \int_{-\infty}^{+\infty} e^{i\omega t}\,\mathbf{H}(\mathbf{r}, t)\,dt = \hat{\phi}\,\frac{e\beta}{2\pi}\frac{\sin\theta}{1 - \beta^2\cos^2\theta}\frac{e^{i\omega(r/c)}}{r},$$

allowing (12) to be written in the form

$$\frac{d\mathcal{E}_r}{d\Omega d\omega} = \frac{r^2}{2\pi c\epsilon_0}\,|\mathbf{H}(\mathbf{r}, \omega)|^2 = \frac{1}{\pi c\epsilon_0}\left(\frac{e\beta}{2\pi}\right)^2\frac{\sin^2\theta}{(1 - \beta^2\cos^2\theta)^2}. \tag{13}$$

In the ultrarelativistic case, this radiation profile is strongly peaked in the directions $\theta \sim mc^2/E \sim \gamma^{-1}$ and $\theta \sim \pi - \gamma^{-1}$. In the nonrelativistic case the pattern is equivalent to that due to the sudden creation of a dipole at $z = 0$. Performing the $d\Omega$ integration in (13) over $2\pi$ solid angle yields the total energy radiated per unit frequency into the left half-space

$$\frac{d\mathcal{E}_r}{d\omega} = \frac{1}{c\epsilon_0}\left(\frac{e\beta}{2\pi}\right)^2\left(\frac{1}{2}\,\frac{1 + \beta^2}{\beta^3}\ln\frac{1 + \beta}{1 + \beta} - \frac{1}{\beta^2}\right). \tag{14}$$

This is in agreement with the formula of Frank and Ginzburg.[1]

(b) We now integrate the rhs of (11) from $t = -\infty$ to $t = +\infty$. The $(E^2 + B^2)$ term vanishes, as its value at $+t$ and $-t$ is the same, and we find

$$\mathcal{E}_r = -\int_{-\infty}^{+\infty}\mathbf{j}\cdot\mathbf{E}\,dV\,dt. \tag{15}$$

Since the charge does work only against the field due to the plane, we write (3) in the form

$$\mathbf{H}_x = \mathbf{H}_x^{(1)} + \mathbf{H}_x^{(2)},$$

where $\mathbf{H}_x^{(1)}$ represents the field of a free charge, and $\mathbf{H}_x^{(2)}$ arises from the presence of the plane. Thus,

$$\mathbf{H}^{(2)} = \frac{\partial}{\partial t}\left(\hat{\mathbf{i}}F^{(2)} + \hat{\mathbf{j}}G^{(2)}\right)$$

and by Maxwell's equations we find the $z$ component of the associated electric field to be

$$\epsilon_0 E_z^{(2)} = \frac{\partial}{\partial x}G^{(2)} - \frac{\partial}{\partial y}F^{(2)}$$
$$= \mathrm{sgn}(z)\,\frac{ie}{(2\pi)^3}\int_{-\infty}^{+\infty}\frac{(\eta^2 + \xi^2)}{v\alpha[\alpha^2 - (\omega^2/v^2)]}$$
$$\times\, e^{i\alpha|z|}e^{i(\eta x + \xi y - \omega t)}\,d\xi d\eta d\omega. \tag{16}$$

With this result and $j = ev\delta(z - vt)\,\delta(x)\,\delta(y)$, Eq. (15) becomes

$$\mathcal{E}_r = \frac{ie^2}{(2\pi)^3\epsilon_0}\int_{-\infty}^{+\infty}\frac{(\eta^2 + \xi^2)}{\alpha[\alpha^2 - (\omega^2/v^2)]}\,d\xi d\eta d\omega\int_{-\infty}^{+\infty}dt\,e^{-i\omega t}$$
$$\times\left(\int_{-\infty}^{0}e^{-i\alpha z}\,dz - \int_0^{\infty}e^{+i\alpha z}\,dz\right)\delta(z - vt).$$

Introducing the change of variable

$$x' = z - \xi, \quad 2z = y' + x', \quad \xi = vt, \quad y' = z + \xi, \quad 2\xi = y' - x', \tag{17}$$

which carries a Jacobian $|J| = \frac{1}{2}$, we find

$$\mathcal{E}_r = 2\frac{e^2v^2}{(2\pi)^3\epsilon_0}\int_{-\infty}^{+\infty}\frac{(\eta^2+\xi^2)}{(\alpha^2v^2-\omega^2)^2}\frac{\omega}{\alpha}\,d\xi d\eta d\omega.$$

Changing to polar coordinates and carrying out the $\omega$ integration with $\omega = \kappa c \sin\zeta$ yields, upon disregarding imaginary quantities,

$$\frac{d\mathcal{E}_r}{d\kappa} = \frac{e^2}{4\pi\epsilon_0}\gamma\beta^2. \tag{18}$$

This represents the energy radiated per unit transverse wavenumber into both the left and right half-spaces. The total energy, however, is seen to diverge; a result also apparent from (14). This divergence is generated by the passage of the charge through a hole of zero radius, causing an apparent infinite deceleration when the charge meets its image (to a $z < 0$ observer). Such an acceleration will produce a singular pulse, as is evident in the radiation fields of (10). In an attempt to render the radiated energy finite, we approximate the total transition energy in the case that the charge passes through a small hole of radius $r_0$ in the plane by cutting off the integration in (18) at $1/r_0$. Thus,

$$\mathcal{E}_r \simeq \frac{e^2}{4\pi\epsilon_0}\gamma\beta^2\int_0^{1/r_0}d\kappa = \frac{e^2}{4\pi\epsilon_0 r_0}\gamma\beta^2. \tag{19}$$

If $E$ is the kinetic energy of the charge then in the ultrarelativistic case the radiation loss becomes

$$\mathcal{E}_r = \frac{e^2}{4\pi\epsilon_0 r_0}\frac{E}{m_0c^2}$$

in agreement with Dnestrovskii and Kostomarov.[5] Using (19) we may also establish the relative decrease in particle energy upon transition to be

$$\frac{E'}{E} \simeq 1 - \frac{r_e}{r_0}\frac{\gamma\beta^2}{\gamma-1},$$

where $E'$ is the final kinetic energy of the particle, $r_e$ the classical electron radius, and $E = (\gamma-1)m_0c^2$ the initial particle energy.

(c) Finally, we calculate the work done $\mathcal{E}_T$ by the charge while in the left half-space. This requires an integration of (11) from $t = -\infty$ to $t = 0$,

$$\mathcal{E}_T = -\int_{-\infty}^0 dt\int \mathbf{j}\cdot\mathbf{E}dV. \tag{20}$$

However, the $(E^2 + B^2)$ term of (11) no longer vanishes as in (b), and (20) must contain contributions from both radiation and stored (reactive) energy. To separate these effects we follow Schwinger[6] and write the time dependence of (16) in the form

$$e^{-i\omega t} = \tfrac{1}{2}(e^{-i\omega t} + e^{+i\omega t}) + \tfrac{1}{2}(e^{-i\omega t} - e^{+i\omega t}), \tag{21}$$

thereby effecting a decomposition of $E_z^{(2)}$ into the sum and difference of retarded and advanced fields. The first "reactive" term of (21) causes the power to change sign under time reversal, and is, therefore, to be associated with stored field energy. The second "resistive" term gives the radiated power, and we find

$$\mathcal{E}_r = \frac{ic^2}{(2\pi)^3\epsilon_0}\int_{-\infty}^0 dz\,\delta(z-vt)\int_{-\infty}^0 dt\int_{-\infty}^{+\infty}\frac{(\eta^2+\xi^2)}{\alpha[\alpha^2-(\omega^2/v^2)]}\times e^{-i\alpha z}\,\frac{e^{-i\omega t}-e^{+i\omega t}}{2}\,d\xi d\eta d\omega.$$

Employing the change of variable (17), we have

$$\mathcal{E}_r = \frac{e^2v^2}{(2\pi)^3\epsilon_0}\int_{-\infty}^{+\infty}\frac{(\eta^2+\xi^2)}{(\alpha^2v^2-\omega^2)^2}\frac{\omega}{\alpha}\,d\xi d\eta d\omega \tag{22}$$

from which there results

$$\frac{d\mathcal{E}_r}{d\omega} = \frac{1}{c\epsilon_0}\left(\frac{e\beta}{2\pi}\right)^2\int_0^{\pi/2}\frac{\sin^3\zeta d\zeta}{(1-\beta^2\cos^2\zeta)^2} = \frac{1}{c\epsilon_0}\left(\frac{e\beta}{2\pi}\right)^2 \times\left(\frac{1}{2}\,\frac{1+\beta^2}{\beta^3}\ln\frac{1+\beta}{1-\beta} - \frac{1}{\beta^2}\right), \tag{23a}$$

$$\frac{d\mathcal{E}_r}{d\kappa} = \left(\frac{ev}{2\pi}\right)^2\frac{\kappa^3}{\epsilon_0}\int_{-\infty}^{+\infty}\frac{\omega d\omega}{\alpha(\alpha^2v^2-\omega^2)^2} = \frac{e^2}{8\pi\epsilon_0}\gamma\beta^2. \tag{23b}$$

Relations (23a) and (23b), which represent the energy radiated into the left half-space, are obtained from (22) after the respective substitutions $\kappa = \omega/c\sin\zeta$ and $\omega = c\kappa\sin\zeta$, and the retention of only real quantities.

It is not difficult to show that the energy radiated into each half-space is the same so that (23) is seen to agree with the results of (a) and (b) above. In a similar way, the stored energy is found to be

$$\frac{d\mathcal{E}_s}{d\kappa} = -\left(\frac{ev}{2\pi}\right)^2\frac{v\kappa^3}{\epsilon_0}\int_{-\infty}^{+\infty}\frac{d\omega}{(\alpha^2v^2-\omega^2)^2} = -\frac{e^2}{8\pi\epsilon_0}\gamma. \tag{24}$$

The fact that this is negative indicates that the energy stored in $\mathbf{E}^{(2)}, \mathbf{B}^{(2)}$ is being removed by the charge. A calculation for $(z,t) > 0$ shows that this term changes sign, whereas the radiation term does not. Thus, the radiation energy is irretrievably lost while the energy stored in $\mathbf{E}^{(2)}, \mathbf{B}^{(2)}$ is reversibly transferred to the charge (see Fig. 5). The transfer of energy involved in this process is analogous to that which occurs in an initially charge LRC circuit which is closed for one cycle during which the stored energy in $C$ is transferred to the stored energy in $L$ and dissipated in $R$. In the limit that $R/LC \to 0$, with $R$ small but finite, complete energy transfer requires an infinite interval.

The total work done by the charge while in the left half-space is then

$$\frac{d\mathcal{E}_T}{d\kappa} = \frac{d\mathcal{E}_r}{d\kappa} + \frac{d\mathcal{E}_s}{d\kappa} = -\frac{e^2}{8\pi\epsilon_0}\frac{1}{\gamma}. \tag{25}$$

This result is easily confirmed by using in (20) the explicit form of $E_z^{(2)}$ which, for $(z,t) < 0$, is given by

$$E_z^{(2)} = \frac{e}{4\pi\epsilon_0}\frac{|z+vt|}{[\rho^2+\gamma^2(z+vt)^2]^{3/2}} = \frac{e}{4\pi\epsilon_0}\int_0^\infty J_0(\kappa\rho)\kappa e^{\gamma\kappa(z+vt)}d\kappa.$$

Thus,

$$\mathcal{E}_T = -\int_{-\infty}^0 dt\int jE_z^{(2)}dV = -\frac{e^2}{8\pi\epsilon_0}\frac{1}{\gamma}\int_0^\infty d\kappa \tag{26}$$

agreeing with (25). Expression (26) has also been found by Liboff,[4] who considers the work done by a charge moving along the axis of a semi-infinite cylindrical cavity, then taking the limit as the radius passes to infinity. We note that this expression includes both radiation and stored energy.

In conclusion, we note that expression (19) is subject to experimental investigation. This result represents the total energy lost to radiation by a charge of arbitrary velocity passing through a hole of radius $r_0$ in a plate. As shown above, this loss leads to a fractional decrease in particle energy $(E'/E)$ upon transition,

$$\frac{E'}{E} \simeq 1 - \frac{r_e}{r_0}\frac{\gamma\beta^2}{\gamma-1}$$

where $r_e$ is the classical electron radius. One also notes in this regard that to obtain this radiation loss it is necessary to follow the particle along its entire trajectory. Over the left half of the trajectory, (i.e., the restricted problem) particle energy both diminishes by radiation and increases at the expense of a loss in the "stored" energy in the fields. To within the order of validity of this calculation,[12] one finds that the energy gained from the fields exceeds the energy lost to radiation by the factor $\beta^2$ [compare (23b) and (24)]. The net gain, given by (25), may be expressed as

$$\mathcal{E}_T = \mathcal{E}_0/\gamma,$$

where $\mathcal{E}_0$ is the point charge self energy

$$\mathcal{E}_0 = \frac{e^2}{8\pi\epsilon_0}\int_0^\infty d\kappa.$$

In the limit $v \to c$, $\mathcal{E}_T$ approaches zero. This is a manifestation of the fact that at this speed, stored energy gained is balanced by radiation lost.[13] That is, a measurement of particle energy at the surface of the plate would, in this idealization, reveal no increase of initial particle kinetic energy. Over the entire trajectory, on the other hand, particle energy change is due only to radiation loss and by (23) is

$$\mathcal{E}_r = \gamma\beta^2\mathcal{E}_0.$$

In the ultrarelativistic limit ($v \to c$) this loss increases[12] as $\gamma$.

## IV. EQUIVALENT PROBLEMS

Finally, we consider four additional problems. First, suppose the charge stops at the plane. In this case, the current is given by

$$\mathbf{j}(\mathbf{r}, t) = ev\delta(z - vt)\,\delta(x)\,\delta(y)\,\theta(-t)\,\hat{\mathbf{k}},$$

where $\theta(z)$ is the unit step function

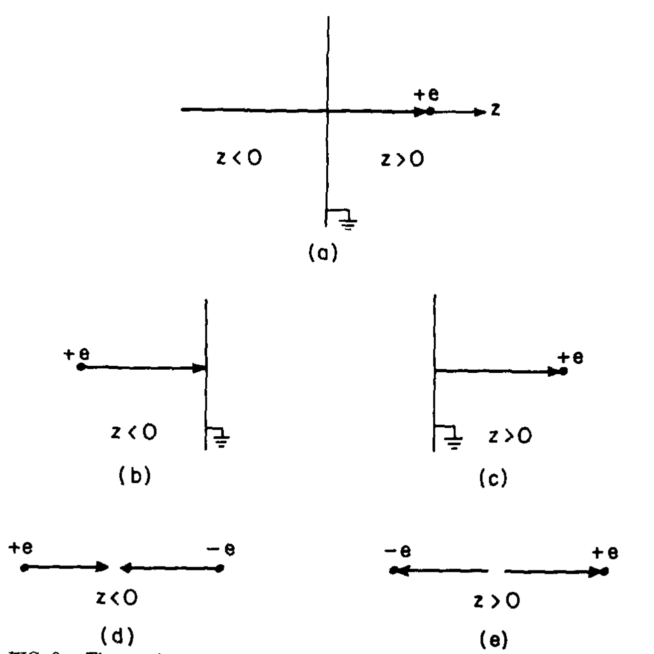



FIG. 6. The equivalent problems. For a $z < 0$ observer the original problem (a) is equivalent to: (b), a charge $+e$ moving from $(z, t) = -\infty$ and stopping at the plane: and (d), two opposite charges approaching each other and disappearing at $(z, t) = 0$. For a $z > 0$ observer (a) is equivalent to: (c), a charge $+e$ starting from the plane at $(z, t) = 0$ and moving to $(z, t) = +\infty$: and (e), two opposite charges appearing at $(z, t) = 0$ and separating from one another.

$$\theta(z) = \begin{cases} 1, & z > 0, \\ 0, & z < 0. \end{cases}$$

In order to proceed from (1), the Fourier integral of $\mathbf{j}(\mathbf{r}, t)$ must be computed. For this purpose we introduce the representation

$$\theta(z) = \lim_{\epsilon \to 0} \frac{i}{2\pi}\int_{-\infty}^{+\infty} \frac{e^{i\alpha z}}{\alpha + i\epsilon}$$

to obtain

$$j(\mathbf{r}, \omega) = \lim_{\epsilon \to 0} \frac{i}{2\pi}\int_{-\infty}^{+\infty} e^{i\omega t}\,\delta(z - vt)\int_{-\infty}^{+\infty} \frac{e^{i\alpha t}}{\alpha + i\epsilon}\,d\alpha dt$$
$$= \begin{cases} 0, & z > 0, \\ \dfrac{1}{v}\exp[i(z/v)\,\omega], & z < 0, \end{cases}$$

where the irrelevant spatial dependence of $j(\mathbf{r}, t)$ has been suppressed. For $z < 0$, this problem is identical to that considered in the paper, wherein $j(z, \omega) = (1/v)\exp[i(z/v)\omega]$.

Similarly, for $z > 0$ the problem of the particle passing through the plane is found to be equivalent to that of a charge starting at the plane and moving uniformly to infinity.

Thirdly, we consider the annihilation of two opposite charges moving uniformly towards $z = 0$ in the absence of the plane. Here,

$$\mathbf{j}(\mathbf{r}, t) = ev\,\delta(x)\,\delta(y)[\delta(z - vt) + \delta(z + vt)]\,\theta(-t)\,\hat{\mathbf{k}}$$

and we find

$$j(z, \omega) = \frac{1}{v}\exp[-i(|z|/v)\omega].$$

Again, for $z < 0$ this problem is equivalent to that treated in the main text of the paper. In that analysis we introduced the plane by requiring the tangential electric field to vanish at $z = 0$. In the present annihilation problem, however, the tangential field naturally vanishes at $z = 0$, and the restriction is therefore of no consequence.

Finally, we note that for $z > 0$ the main solution yields the fields of the creation of two opposite charges at $z = 0$ which separate uniformly to $z = \pm\infty$. In all cases, all observers see zero total charge.

In summary, then the problem of a uniformly moving charge passing through an infinitely conducting plane contains the solution to the following problems: (1) For $z < 0$, the problem of a charge stopping at the plane and giving rise to the Bremsstrahlung fields of pair annihilation (see Fig. 6b, d). (2) For $z > 0$, the problem of a charge starting at the plane and giving rise to the Bremsstrahlung fields of pair creation (see Fig. 6c, e).

It should be noted that these fields represent pure transition radiation. In the case of a plane with structure, Bremsstrahlung radiation arising from collisions within the plane must be distinguished from these transition fields.

## ACKNOWLEDGMENT

This research was supported in part by the Physics Branch of the Office of Naval Research under Contract N0014-67-A-0077-0015.

## APPENDIX A: INVERSION OF THE PARTICULAR SOLUTION

From (3), we wish to evaluate

$$H_x^a(\mathbf{r}, t) = \frac{1}{(2\pi)^3}\frac{\partial}{\partial t}\int_{-\infty}^{+\infty} \frac{e\xi}{\omega}\,\frac{\exp[i(z/v)\omega]}{\alpha^2 - (\omega^2/v^2)}$$

$$\times e^{i(\eta x+\xi y-\omega t)}d\eta d\xi d\omega,$$

where $\alpha^2 = (\omega^2/c^2) - (\eta^2 + \xi^2)$.
Introducing the polar coordinates $\substack{x=\rho\,\cos\phi\\ y=\rho\,\sin\phi}$ and integration variables $\substack{\eta=\kappa\,\cos\phi'\\ \xi=\kappa\,\sin\phi'}$ yields

$$H_x^a = \frac{-e}{(2\pi)^2}(v^2\gamma^2) \times \sin\phi \int_0^\infty J_1(\kappa\rho)\,\kappa^2 \int_{-\infty}^{+\infty} \frac{e^{i\omega(z/v-t)}}{\omega^2+(\gamma v\kappa)^2}\,d\omega d\kappa.$$

Simple poles occur at $\omega = \pm i\gamma v\kappa$ in the $\omega$ plane. For $z > vt$ the contour is completed in the upper plane, and for $z < vt$ in the lower. In both cases the results are the same yielding

$$H_x^a(\mathbf{r},t) = \frac{(-e)}{4\pi}\gamma v \sin\phi \int_0^\infty J_1(\kappa\rho)\,\kappa e^{-\gamma\kappa|z-vt|}d\kappa = \frac{(-e)}{4\pi}\gamma v \frac{\rho}{[\rho^2+\gamma^2(z-vt)^2]^{3/2}}.$$

## APPENDIX B: EVALUATION OF $\Sigma_1$

In this section we evaluate the sum

$$\sum\nolimits_1 = \frac{-2}{\gamma\kappa} e^{\gamma\kappa(|z|\beta-vt)} \sum_{n=0}^{\infty}{}' (-1)^n \times\left(\sum_{m=0}^{2n-1'} (-1)^n[\gamma(1-\beta)]^{2n-m} + \sum_{m=2n}^{\infty'}[\gamma(1-\beta)]^{m-2n} + \sum_{m=0}^{\infty'}[\gamma(1-\beta)]^{m+2n}\right) J_{2n}(i\kappa|z|)J_m(\kappa ct)$$

which appears in Sec. B. To do this, we consider the integral

$$\sum\nolimits_1' = \int_{|z|}^\infty J_0(\kappa\sqrt{u'^2-|z|^2})\,e^{-\gamma v\kappa(t+u'/c)}du'.$$

By an analysis similar to that for $I_{|z|}^{ct}$ and $I_{ct}^\infty$, we find the expansion

$$\sum\nolimits_1' = \frac{+2}{\gamma\kappa} e^{-\gamma\kappa(|z|\beta+vt)} \sum_{n=0}^{\infty}{}' (-1)^n \times\left(\sum_{m=0}^{2n-1'} [\gamma(1-\beta)]^{2n-m} + \sum_{m=2n}^{\infty'}(-1)^m[\gamma(1-\beta)]^{m-2n} + \sum_{m=0}^{\infty'}(-1)^m[\gamma(1-\beta)]^{m+2n}\right) J_{2n}(i\kappa|z|)J_m(\kappa|z|).$$

However, the integral $\sum_1'$ can also be obtained exactly,[10] and is found to be

$$\sum\nolimits_1' = \frac{1}{\kappa\gamma}\, e^{-\gamma v\kappa t}e^{-|z|\kappa\gamma}.$$

We now suppose both $\sum_1$ and $\sum_1'$ to be functions of $\gamma$, and by inspection find

$$\sum\nolimits_1(\gamma) = e^{-2\gamma v\kappa t}\sum\nolimits_1'(-\gamma) = e^{-2\gamma v\kappa t}\left(-\frac{1}{\kappa\gamma}e^{+\gamma v\kappa t}e^{+|z|\kappa\gamma}\right) = -\frac{1}{\kappa\gamma}e^{-\gamma\kappa(vt-|z|)}.$$

Returning to Eq. (6), then, we have

$$\begin{aligned} H_x(\text{due to } \textstyle\sum_1) &= H_x^a - \operatorname{sgn}(z)\frac{(-e)}{4\pi}\gamma\sin\phi\frac{\partial}{\partial t}\int_0^\infty J_1(\kappa\rho)\,\kappa\textstyle\sum_1(\gamma)\,d\kappa \\ &= H_x^a - \operatorname{sgn}(z)\frac{(-e)}{4\pi}\gamma v\sin\phi\int_0^\infty J_1(\kappa\rho)\,\kappa e^{-\gamma\kappa(vt-|z|)}d\kappa \\ &= H_x^a + \begin{cases} +H_x^i, & z<0, \\ -H_x^a, & z>0, \end{cases} \qquad |z| < vt. \end{aligned}$$


[1] I. Frank and V. Ginzburg, J. Phys. 9, 353 (1945)
[2] G. M. Garibyan, Sov. Phys. JETP 6, 1079 (1958). A list of references on transition radiation is to be found in the summary article by I. Frank, Uspekhi 8, 729 (1966).
[3] E. Ott and J. Shmoys, Quart. J. Appl. Math. 25, 377 (1968).
[4] R. L. Liboff, J. Math. Physics 11, 1296 (1970); 13, 272, (1972); 13, 1828 (1972).
[5] Y. N. Dnestrovskii and D. P. Kostomorov, Sov. Phys. Dokl. 4, 158 (1959). See also, O. A. Kolpakov and V. I. Kotov, Sov. Phys., Tech. Phys. 9, 1072 (1965).
[6] J. Schwinger, Phys. Rev. 75, 1912 (1949).
[7] To change to Gaussian units, replace **E**, **H** and $e$ with $\mathbf{E}/\sqrt{4\pi\epsilon_0}$, $\mathbf{H}/\sqrt{4\pi\epsilon_0}$, and $e\sqrt{4\pi\epsilon_0}$, respectively.
[8] P. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[9] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Macmillan, New York, 1945).
[10] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965).
[11] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1967).
[12] We note that the analysis assumes a constant particle velocity in which case the result for $\mathcal{E}_r$ is valid so long as it is small compared to particle kinetic energy, $(\gamma-1)m_0c^2$.
[13] Alternatively, we may argue that at $\beta \approx 1$ the particle does not detect its image while in the left half-space (see Ref. 4).

# Some solutions of the Einstein field equations for a rotating perfect fluid

Andrzej Krasiński*

*Institute of Theoretical Physics, Warsaw University, Warsaw, Poland*

(Received 4 September 1973)

The equations of isentropic rotational motion of a perfect fluid are investigated with use of Darboux' theorem. It is shown that, together with the equation of continuity, they guarantee the existence of four scalar functions on space-time, which constitute a dynamically distinguished set of coordinates. It is assumed that in this coordinate system the metric tensor is constant along the lines tangent to velocity and vorticity fields. Under these assumptions a complete set of solutions of the field equations with $T_{ij} = (\epsilon + p)u_i u_j - pg_{ij}$ is found. They divide into three families, first of which contains six types of new solutions with nonzero pressure. The second family contains only the Gödel's solution, and the third one, only the Lanczos' solution. Symmetry groups, exterior metrics, type of conformal curvature, geometrical and physical properties of the new solutions are investigated. A short review of other models of rotating matter is given.

## INTRODUCTION

It was not long after the creation of the general relativity theory that people tried to construct a solution of the Einstein field equations for rotating matter. The problem was interesting both from theoretical and observational point of view because nobody knew how to describe the rotational motion in the formalism of general relativity while many stars and galaxies exhibited visible rotation. Today even the possibility of rotation of the universe in the large is admitted.[1]

However, for quite a long time models of rotating matter were constructed under very special assumptions. The authors either used the method of "slow rotation" approximation (first paper by J. Lense and H. Thirring[2] in 1918) or assumed the energy-momentum tensor corresponding to dust (K. Lanczos[3] in 1924 and many others). It was not till 1967 that M. Trümper[4] clearly stated the problem of searching for solutions with pressure different from zero, but he has just written down the field equations and stopped after arriving at some general statements. There were a few papers whose authors went further but they left the problem behind when the equations were simplified and nearly integrated (i.e., there remained only one or two equations to be solved). They gave at most special cases of solutions which were mathematically simple (e.g., J. Stewart and G. F. R. Ellis,[5] J. Wainwright.[6])

Until 1972, in fact, just two complete results were obtained—by H. D. Wahlquist[7] in 1968 and E. Herlt[8] in 1972. The aim of the present paper was to supply new metrics of this kind. I have used the method of description of the isentropic rotational motion of the perfect fluid introduced by J. Plebański.[9] Under the assumptions, which are clearly stated in Sec. 1, the field equations were completely integrated. The resulting metrics divide into three families, the first of which contains six types of new solutions with nonzero pressure. Each of the other families contains just one solution known before.

The first family solutions are investigated in detail. Their symmetry groups, exterior metrics, type of conformal curvature, geometrical and physical properties are established and discussed. A few special cases are investigated in more detail. I also give a short review of the solutions found by other authors.

Most of the material presented in Sec. 1 is taken from J. Plebański's paper.[9]

## 1. THE EQUATIONS OF MOTION AND DYNAMICALLY DISTINGUISHED COORDINATES

Throughout the paper we shall use the signature $(+---)$. The equations of motion of a perfect fluid have the form:

$$T^{\alpha\beta}{}_{;\beta} = 0, \tag{1.1}$$

where

$$T^{\alpha\beta} = (\epsilon + p)u^\alpha u^\beta - pg^{\alpha\beta}. \tag{1.2}$$

The quantity $(\epsilon + p)$ is called the enthalpy density. Let $\mathcal{H}$ denote the enthalpy per unit mass,

$$\mathcal{H} = (\epsilon + p)/\rho, \tag{1.3}$$

where $\rho$ is the density of the rest-mass. Independently of (1.1) the conservation of the total rest mass is postulated:

$$(\rho u^\alpha)_{;\alpha} = 0. \tag{1.4}$$

By virtue of (1.3) and (1.4) Eqs. (1.1) take the form

$$0 = T_\alpha{}^\beta{}_{;\beta} = \rho u^\beta (\mathcal{H} u_\alpha)_{;\beta} - p_{,\alpha}. \tag{1.5}$$

The enthalpy in phenomenological thermodynamics obeyed the following identity:

$$d\mathcal{H} = (1/\rho)\,dp + TdS. \tag{1.6}$$

This equation may be considered to be the definition of temperature and entropy in general relativity. Namely, only two of the state functions $(\mathcal{H}, \rho, p)$ can be independent. Therefore the form $(d\mathcal{H} - (1/\rho)\,dp)$ has an integrating factor which we denote by $1/T$ and its inverse we call the temperature. Then the form $(1/T)(d\mathcal{H} - 1/\rho)dp)$ is a total differential of a function $S$ which we call entropy.

With the help of (1.6) we get in (1.5)

$$\rho[u^\beta(\mathcal{H}u_\alpha)_{;\beta} - \mathcal{H}_{,\alpha} + TS_{,\alpha}] = 0. \tag{1.7}$$

Now the identities $u^\alpha u_\alpha = 1$ and $u^\beta u_{\beta;\alpha} = 0$ allow us to write (1.7) as

$$[(\mathcal{H}u_\alpha)_{,\beta} - (\mathcal{H}u_\beta)_{,\alpha}]u^\beta + TS_{,\alpha} = 0. \tag{1.8}$$

These are the equations of motion of a perfect fluid in a form equivalent to (1.1).

We shall confine ourselves to isentropic motions, where $S_{,\alpha} = 0$. Then (1.3) and (1.6) imply

$$d[(\epsilon + p)/\rho] = (dp)/\rho. \tag{1.9}$$

We see that $d\epsilon = [(\epsilon + p)/\rho]d\rho$ and so $\epsilon = \epsilon(\rho)$, $p = p(\rho)$; in other words, $\rho = \rho(p)$ and $\epsilon = \epsilon(p)$. Thus (1.9) is an ordinary differential equation, and we can integrate it to obtain

$$\epsilon + p = \rho c^2 \left(H_0 + \frac{1}{c^2} \int_0^p \frac{dp}{\rho(p)}\right) \tag{1.10}$$

where $H_0 = \text{const}$. If we assume that $\epsilon(p=0) = \rho(0)c^2$ then $H_0 = 1$. Let us denote

$$H \stackrel{\text{def}}{=} H_0 + \frac{1}{c^2} \int_0^p \frac{dp}{\rho(p)}. \tag{1.11}$$

Then Eqs. (1.8) with $S_{,\alpha} = 0$ take the form

$$[(Hu_\alpha)_{,\beta} - (Hu_\beta)_{,\alpha}]u^\beta = 0. \tag{1.12}$$

Now we recall two theorems which will be useful later. We give both of them in the special case of a four-dimensional manifold. Their general forms can be found in Refs. 10–12.

*Theorem 1 (Darboux):* Let $\omega$ be a differential form of the 1st order, then

(1) $(d\omega \wedge d\omega \neq 0) \Leftrightarrow$ (there exists the set of functions $\sigma, \tau, \xi, \eta$ such that $\omega = \sigma d\tau + \eta d\xi$);

(2) $(d\omega \wedge d\omega = 0$ but $\omega \wedge d\omega \neq 0) \Leftrightarrow (\sigma = 1$ above$)$;

(3) $(\omega \wedge d\omega = 0$ but $d\omega \neq 0) \Leftrightarrow (\xi = 1$ in (1)$)$;

(4) $(d\omega = 0) \Leftrightarrow (\sigma = \xi = 1$ in (1)$)$.

Its proof is given in Ref. 10.

For an antisymmetric tensor $F_{\alpha\beta}$ the following form can be defined:

$$Pf(F_{\alpha\beta}) = \tfrac{1}{8}\, \epsilon^{\alpha\beta\gamma\delta} F_{\alpha\beta} F_{\gamma\delta} \tag{1.13}$$

where $\epsilon^{\alpha\beta\gamma\delta}$ is the Levi–Civita symbol. We have

*Theorem 2:*

$$[Pf(F_{\alpha\beta})]^2 = \det(F_{\alpha\beta}).$$

The proof can be found in Refs. 11 and 12.

Now let $F_{\alpha\beta} \stackrel{\text{def}}{=} (Hu_\alpha)_{,\beta} - (Hu_\beta)_{,\alpha}$. We see from (1.12) that $\det(F_{\alpha\beta}) = 0$ and so from Theorem 2 $Pf(F_{\alpha\beta}) = 0$ which means that $F_{[\alpha\beta} F_{\gamma\delta]} = 0$.

Let us define $\omega = Hu_\alpha dx^\alpha$. Then $F_{\alpha\beta}dx^\alpha \wedge dx^\beta = -2d\omega$, and so $d\omega \wedge d\omega = 0$.

Now Theorem 1 implies that there exist functions $\tau, \xi, \eta$ such that $\omega = d\tau + \eta d\xi$, i.e.,

$$Hu_\alpha = \tau_{,\alpha} + \eta\xi_{,\alpha}, \tag{1.14}$$

$$F_{\alpha\beta} = \xi_{,\alpha}\eta_{,\beta} - \xi_{,\beta}\eta_{,\alpha}. \tag{1.15}$$

This representation of $Hu_\alpha$ is introduced and discussed in more detail in Ref. 9.

When $F_{\alpha\beta} = 0$ we call the motion irrotational. When $F_{\alpha\beta} \neq 0$ we call it rotational. To distinguish rotational and irrotational motions we can use as well the vorticity vector $w^\alpha$:

$$w^\alpha = -(-g)^{-1/2} \epsilon^{\alpha\beta\gamma\delta} u_\beta u_{\gamma,\delta}. \tag{1.16}$$

In the local inertial frame at a point $p$ [where $u^\alpha = \delta^\alpha_0$, $g_{\alpha\beta}(p) = \mathrm{diag}(+1, -1, -1, -1)$] the vector $w^\alpha$ has the components $w^\alpha = (0, -(1/c)\mathbf{W})$ where $\mathbf{W} = \mathrm{rot}\mathbf{v}$, $\mathbf{v}$ – the Newtonian velocity vector. Thus the differentiation between rotational and irrotational motions based on $w^\alpha$ agrees with that in Newtonian physics. Moreover, we have

*Theorem 3:*

$$(F_{\alpha\beta} = 0) \Leftrightarrow (w^\alpha = 0).$$

Therefore, this differentiation agrees with that based on $F_{\alpha\beta}$, too. Consequently, we can consider $F_{\alpha\beta}$ to be the angular velocity tensor. But there is a definition of the angular velocity tensor, given by J. Ehlers[13,14]

$$\Omega_{\alpha\beta} = u_{[\alpha;\beta]} - u_{[\alpha;|\rho|}u^\rho u_{\beta]} \tag{1.17}$$

With the help of the equations of motion (1.12) it is easy to show that

$$F_{\alpha\beta} = 2H\Omega_{\alpha\beta}, \tag{1.18}$$

so our definition of rotational motion agrees with that of Ehlers.

From now on we shall deal with rotating matter only, so we assume

$$F_{\alpha\beta} \neq 0. \tag{1.19}$$

It means that all the three functions in (1.14) have linearly independent gradients. Equation (1.12) implies that $u^\alpha\xi_{,\alpha} = u^\alpha\eta_{,\alpha} = 0$. This, together with the equation of continuity $[(-g)^{1/2}\rho u^\alpha]_{,\alpha} = 0$, allows us to define the fourth function $\zeta$ in the following way:

$$(-g)^{1/2}\rho u^\alpha = \epsilon^{\alpha\beta\gamma\delta}\xi_{,\beta}\eta_{,\gamma}\zeta_{,\delta}. \tag{1.20}$$

(For the details see again.[9]) By contraction of (1.14) and (1.20) we get

$$g = -\rho^{-2}H^{-2}\left(\frac{\partial(\tau, \xi, \eta, \zeta)}{\partial(x^0, x^1, x^2, x^3)}\right)^2. \tag{1.21}$$

If (1.14) and (1.20) are assumed, then the equations of motion and continuity are just identities.

Of course we can use the functions $(\tau, \xi, \eta, \zeta)$ as new coordinates. If we do, then (1.14), (1.20), and (1.21) reduce to

$$u^\alpha = H\delta^\alpha_0, \tag{1.22}$$

$$u_\alpha = H^{-1}\delta^0_\alpha + x^2 H^{-1}\delta^1_\alpha, \tag{1.23}$$

$$g = -\rho^{-2}H^{-2}. \tag{1.24}$$

We also have

$$w^\alpha = \rho H^{-1}\delta^\alpha_3 \tag{1.25}$$

and, since $u_\alpha = g_{\alpha\rho}u^\rho$,

$$\begin{aligned} g_{00} &= H^{-2}, \\ g_{01} &= x^2 H^{-2}, \\ g_{02} &= g_{03} = 0. \end{aligned} \tag{1.26}$$

The functions $(\tau, \xi, \eta, \zeta)$ are not unique. The coordinate transformations preserving the properties (1.22)–(1.26) are of the form:

$$x^0 = x^{0\prime} - S(x^{1\prime}, x^{2\prime}),$$
$$x^1 = F(x^{1\prime}, x^{2\prime}),$$
$$x^2 = G(x^{1\prime}, x^{2\prime}), \qquad (1.27)$$
$$x^3 = x^{3\prime} + T(x^{1\prime}, x^{2\prime}),$$

where $T$ is completely arbitrary, while $F$ and $G$ must obey the equation

$$F_{,1'}G_{,2'} - F_{,2'}G_{,1'} = 1. \qquad (1.28)$$

$S$ is fixed by the equations

$$GF_{,1'} - x^{2\prime} = S_{,1'}, \qquad (1.29)$$
$$GF_{,2'} = S_{,2'}.$$

We see that one of the functions $F$ and $G$ is arbitrary and once it is fixed, the other is given by (1.28). Therefore, together with $T$ we have two arbitrary functions in (1.27). Notice that all functions in (1.27) depend only on two variables $x^1$ and $x^2$.

Now the idea arises: If the whole metric tensor also depends only on $x^1$ and $x^2$, then the transformations (1.27) may allow us to simplify the metric further. So we assume

$$\left(\frac{\partial}{\partial x^0}\right)g_{\alpha\beta} = \left(\frac{\partial}{\partial x^3}\right)g_{\alpha\beta} = 0. \qquad (1.30)$$

This condition is covariant with the transformations (1.27). As a consequence of (1.22) and (1.25) it can be written as

$$\partial_u g_{\alpha\beta} = \partial_w g_{\alpha\beta} = 0, \qquad (1.31)$$

where $\partial_u = u^\alpha(\partial/\partial x^\alpha)$, $\partial_w = w^\alpha(\partial/\partial x^\alpha)$.

These two assumptions are sufficient to integrate the Einstein field equations for the metric fulfilling (1.24) and (1.26) to the very end. No additional simplifying assumptions are made here. We shall explain the geometrical meaning of the assumptions (1.31) later. Notice that the first of (1.31) means that $u^\alpha$ is colinear with a timelike Killing vector, so the expansion and shear of the velocity field vanish.

## 2. FIRST INTEGRALS OF THE FIELD EQUATIONS AND CLASSIFICATION OF THE SOLUTIONS

Since there are two arbitrary functions in (1.27), we can expect that it will be possible to make two more components of the metric tensor equal to 0. It is really the case. If we choose $F$, $G$, and $T$ so that the equations

$$g^{22}F_{,1'}F_{,2'} - g^{12}(F_{,1'}G_{,2'} + F_{,2'}G_{,1'}) + g^{11}G_{,1'}G_{,2'} = 0 \qquad (2.1)$$

and

$$T_{,1'} = -(g_{13}/g_{33})F_{,1'} - (g_{23}/g_{33})G_{,1'} \qquad (2.2)$$

hold, then in the new coordinates $(x^{0\prime}, x^{1\prime}, x^{2\prime}, x^{3\prime})$ we have, in addition to (1.24) and (1.26),

$$g_{12} = g_{13} = 0 \qquad (2.3)$$

The set of Eqs. (1.28)–(2.1) makes sense no matter what $g_{\alpha\beta}$ is. Equation (2.2) makes sense because. Theorem 3, (1.19), and (1.25) imply that $g_{33} \neq 0$.

Substituting (2.3) in (2.1) and (2.2), we get a new set of equations which determines the transformations (1.27) preserving all the properties (1.24), (1.26), and (2.3).

From now on there is no arbitrary function in (1.27).

It is time to use the field equations. If the right-hand side of the equations

$$R^\alpha{}_\beta = (\kappa/c^2)(T^\alpha{}_\beta - \tfrac{1}{2}\delta^\alpha{}_\beta T) + \Lambda\delta^\alpha{}_\beta, \qquad \kappa = 8\pi k/c^2, \qquad (2.4)$$

is given by (1.2), (1.11), and (1.10), then it must be $R^0{}_3 = R^1{}_3 = 0$. These two equations when integrated yield the result

$$g_{23} = K(x^2)g_{33}, \qquad (2.5)$$

where $K$ is an arbitrary function of one variable. Now we can verify that the coordinate transformation

$$x^0 = x^{0\prime} + x^{1\prime}x^{2\prime},$$
$$x^1 = x^{2\prime}, \qquad x^2 = -x^{1\prime}, \qquad (2.6)$$
$$x^3 = x^{3\prime} - \int K(x^2)\,dx^2$$

fulfills all Eqs. (1.28), (1.29), (2.1), (2.2), and yields, in addition,

$$g_{23} = 0. \qquad (2.7)$$

In the new coordinates it is easier to compute the Ricci tensor. From the equations $R^1{}_0 = R^2{}_0 = 0$, we easily find that

$$g_{33} = G\rho^{-1}H^3, \qquad G = \text{const} < 0. \qquad (2.8)$$

We classify the solutions into three families:

Family I in which

$$\rho_{,\alpha} \neq 0, \qquad p \neq 0. \qquad (2.9)$$

Family II in which $\rho_{,\alpha} = 0$ and consequently

$$H_{,\alpha} = p_{,\alpha} = 0. \qquad (2.10)$$

Family III in which

$$p = 0. \qquad (2.11)$$

This classification is invariant. We are going to discuss each family separately.

## 3. THE FIRST FAMILY OF SOLUTIONS

Using the complete set of the field equations one can prove that by a suitable choice of coordinate system we obtain

$$\rho = \rho(x^2), \qquad \text{and so} \qquad H = H(x^2), \qquad p = p(x^2). \qquad (3.1)$$

Then the field equations reduce to the set of ordinary differential equations, and after integration they yield

$$ds^2 = H^{-2}(dx^0)^2 + 2x^2H^{-2}dx^0dx^1 + [(x^2)^2 - W/G]H^{-2}(dx^1)^2 + (W\rho H)^{-1}(dx^2)^2 + G\rho^{-1}H^3(dx^3)^2, \qquad (3.2)$$

where

$$W = (G + \kappa(x^2)^2 + Bx^2 + E, \qquad B, E = \text{const}, \qquad (3.3)$$

$$\rho = D\frac{H^5}{W}\exp\left(\int \frac{Gx^2}{W}\,dx^2\right), \qquad D = \text{const} < 0, \qquad (3.4)$$

$$H = |Mu_1 + Nu_2|^{1/3}, \qquad M, N = \text{const} \qquad (3.5)$$

$u_1$ and $u_2$ are the linearly independent solutions of the equation

$$u_{,22} - \frac{W_{,2} - Gx^2}{W} u_{,2} + \frac{3}{4}\left(-\frac{W_{,22}}{W} + \frac{W_{,2}^2}{W^2} - \frac{Gx^2 W_{,2}}{W^2} + \frac{G}{W}\right) u = 0. \quad (3.6)$$

The pressure $p$ is given by the formula resulting from (1.11):

$$p = c^2 \int \rho H_{,2} dx^2 + p_0. \quad (3.7)$$

Whenever an inequality for a constant appears above or below, it results from two conditions:

(1) $\rho, p, H > 0$.

(2) The signature of the metric is $(+ - - -)$.

The absolute value in (3.5) is needed to assure that $H > 0$.

The solutions of the first family divide into six types according as to whether $W$ has two complex roots, two real roots, one real root or degenerates to a polynomial of a lower degree.

It is clear from (3.4) that when the sign of $W$ is not the same for all values of $x^2$, then $\rho$ may be positive only in some range of values of $x^2$. The boundaries of this range (i.e., the roots of $W$) are singular points of $\rho$, and outside of this range $\rho$ would be negative. In such a situation we have to find some exterior metric and match it to (3.2) so that the complete space–time has no singularities. This is done in Sec. 7. In the formulas given below an auxiliary constant $a \overset{\text{def}}{=} G/(G + \kappa)$ is occasionally used.

### Type I

$$W = (G + \kappa)(x^2 - b)(x^2 - c'), \qquad c' = b^*, \qquad a > 1. \quad (3.8)$$

$$u_1 = u + u^*, \qquad u_2 = -i(u - u^*). \quad (3.9)$$

$$u = \left(\frac{x^2 - b}{c' - K}\right)^{\beta}\left(\frac{x^2 - c'}{b - K}\right)^{\gamma} \times F\left(\alpha + \beta + \gamma, \alpha' + \beta + \gamma, 1 + \beta - \beta', \frac{x^2 - b}{c' - b}\right), \quad (3.10)$$

$K = \text{const} = K^*$, $F(., ., ., .)$—the hypergeometric function.

$$\left.\begin{matrix}\alpha\\ \alpha'\end{matrix}\right\} = \tfrac{1}{2}[a - 3 \pm (a^2 - 3a + 3)^{1/2}]. \quad (3.11)$$

$$\left.\begin{matrix}\beta\\ \beta'\end{matrix}\right\} = \frac{1}{2(b - c')}\{-(a - 2)b - 2c' \mp [a^2 b^2 + (b - c')(b - c' - ab)]^{1/2}\}. \quad (3.12)$$

$$\left.\begin{matrix}\gamma\\ \gamma'\end{matrix}\right\} = \frac{1}{2(b - c')}\{2b + (a - 2)c' \pm [a^2 c'^2 + (b - c')(b - c' + ac')]^{1/2}\} = \left\{\begin{matrix}\beta^*\\ \beta'^*\end{matrix}\right.. \quad (3.13)$$

### Type II

$$W = (G + \kappa)(x^2 - b)(x^2 - c'), \qquad b \text{ and } c' \text{ real}, \qquad b < c'. \quad (3.14)$$

$u_1$ is given by (3.10), $u_2$ is the standard second linearly independent solution.[15-17] The formulas for $\alpha, \alpha', \beta, \beta', \gamma, \gamma'$ are identical with (3.11)–(3.13). This time no analogue of the equations $\gamma = \beta^*$ and $\gamma' = \beta'^*$ holds.

The sign of $W$ is not constant. For $x^2 = b$ and $x^2 = c'$ the solution has singularities. When $a < 0$, the density of matter is positive in the region $b < x^2 < c'$; when $a > 1$ it is positive in the nonconnected region $x^2 < b$ and $x^2 > c'$.

### Type III

$$W = (G + \kappa)(x^2 - b)^2, \qquad b \neq 0, \qquad a > 1. \quad (3.15)$$

$$u_i = (x^2 - b)^{q_i} F\left(\tfrac{3}{2} - q_i, 4 - a - 2q_i, \frac{ab}{x^2 - b}\right), \qquad i = 1, 2, \quad (3.16)$$

$F(., ., .)$—the confluent hypergeometric function.

$$\left.\begin{matrix}q_1\\ q_2\end{matrix}\right\} = \tfrac{1}{2}[3 - a \pm (a^2 - 3a + 3)^{1/2}]. \quad (3.17)$$

$W$ has a constant sign, but $x^2 = b$ is a singular point of the solution.

### Type IV

$$W = (G + \kappa)(x^2)^2, \qquad a > 1. \quad (3.18)$$

$$u_i = (x^2)^{q_i}, \qquad i = 1, 2, \qquad q_i \text{ given by (3.17)}. \quad (3.19)$$

Again $W$ has a constant sign, but $x^2 = 0$ is a singular point.

### Type V

$$W = Bx^2 + E. \quad (3.20)$$

Here the coordinates can be chosen so that $B = \kappa$. We denote $E = \kappa E_0$ and we get

$$u_i = [\exp(x^2 + E_0)](-x^2 - E_0)^{q_i} F(q_i + E_0 - 1, 2q_i + E_0 - 1, -x^2 - E_0), \qquad i = 1, 2, \quad (3.21)$$

$$\left.\begin{matrix}q_1\\ q_2\end{matrix}\right\} = \tfrac{1}{2}[2 - E_0 \pm (E_0^2 - E_0 + 1)^{1/2}]. \quad (3.22)$$

The density of matter is positive in the region $x^2 < -E_0$.

### Type VI

$$W = E = \text{const} < 0, \quad (3.21)$$

$$u_1 = F(\tfrac{3}{8}, \tfrac{1}{2}, (\kappa/2E)(x^2)^2), \qquad u_2 = x^2 F(\tfrac{7}{8}, \tfrac{3}{2}, (\kappa/2E)(x^2)^2). \quad (3.22)$$

## 4. THE SECOND FAMILY OF SOLUTIONS

Here the field equations reduce to one partial differential equation. Again it can be shown that by a suitable choice of coordinates the metric can be made dependent only on one variable $x^2$. Then the solution appears to be unique (exact to coordinate transformations):

$$ds^2 = H^{-2}(dx^0 + x^2 dx^1)^2 - \tfrac{1}{2} H^{-2}(x^2)^2(dx^1)^2 - [\kappa\rho H(x^2)^2]^{-1}(dx^2)^2 - 2\kappa\rho^{-1}H^3(dx^3)^2, \quad (4.1)$$

where

$$\rho, p > 0 \quad \text{are arbitrary constants}, \quad (4.2)$$

$$H = 1 + (p/c^2\rho) = \text{const}, \quad (4.3)$$

$\Lambda = \tfrac{1}{2}\kappa[\rho - (p/c^2)]$ is the cosmological constant. (4.4)

The metric (4. 1) was found for the first time by H. M. Raval and P. C. Vaidya[18] and is a generalization to the case of constant but nonzero pressure of the well-known solution of Gödel[19] [if $p = 0$ then (4. 1) is precisely the Gödel's metric]. It is the limiting case $a = 2$, $M = 0$ of the Type IV solution from the first family.

## 5. THE THIRD FAMILY OF SOLUTIONS

One sees that when $p = 0$ (and consequently $\epsilon = \rho c^2$), then the equations of motion (1. 1) can be written in the form $(u_{\alpha,\beta} - u_{\beta,\alpha})u^\beta = 0$, and thus are the special case $H = 1$ of Eq. (1. 12). Therefore all the formulas up to (2. 8) hold for dust if $p = 0$ and $H = 1$ is substituted there. This time again one verifies that such coordinates exist, in which $\rho = \rho(x^2)$. The solution is unique:

$$ds^2 = (dx^0)^2 + 2x^2dx^0dx^1 + x^2(x^2+1)(dx^1)^2 + \frac{e^{x^2}}{\kappa a x^2}(dx^2)^2 - \frac{\kappa}{a}e^{x^2}(dx^3)^2, \quad (5.1)$$

where

$$\rho = ae^{-x^2}, \quad a = \text{const} > 0. \quad (5.2)$$

The metric has the proper signature in the region $x^2 < 0$. Here necessarily $\Lambda = 0$. This metric was found by K. Lanczos[3] in 1924 and was the first exact solution with rotating matter in the history of relativity. It was rediscovered next by W. J. van Stockum[20] in 1937 and J. P. Wright[21] in 1965. In fact, Lanczos and Wright also found the generalization of (5. 1) to the case $\Lambda \neq 0$, but this generalization does not fulfill the second of (1. 30).

Equation (5. 1) is the limiting case $E_0 = N = 1$, $M = \frac{3}{2}$ of the Type V solution from the first family (represented in slightly different coordinates, related to those of Type V by the transformation $x^0 = x^{0\prime} + x^1$, $x^2 = x^{2\prime} - 1$).

## 6. SYMMETRIES OF THE SOLUTIONS

I will not investigate the second and third family of solutions as they have been considered by many other authors.[3,6,18-22] The symmetry group for all the types of the first family solutions consists of the following transformations:

$$\begin{aligned} x^0 &= x^{0\prime} + t_0, \\ x^1 &= x^{1\prime} + t_1, \\ x^2 &= x^{2\prime}, \\ x^3 &= x^{3\prime} + t_3 \end{aligned} \quad (6.1)$$

with $t_0, t_1, t_3 = \text{const}$.

Thus it is 3-parametric Abelian group with the Killing vectors $k^\mu_{(i)} = \delta^\mu_i$, $i = 0, 1, 3$. It acts simply transitively on the timelike hypersurfaces $x^2 = \text{const}$. Such groups were classified by Bianchi into nine types.[23] (In fact, Bianchi classification is usually applied to groups acting on spacelike hypersurfaces, but no specific signature of the metric on the hypersurface is assumed and therefore such a classification is true for timelike homogeneous hypersurfaces, too). Since the group of transformations (6. 1) is Abelian, it is of Bianchi Type I, and the hypersurfaces $x^2 = \text{const}$ are flat. Notice that the group (6. 1) is completely characterized by four statements:

(1) There exist three commuting Killing vectors $\underset{(0)}{k^\mu}, \underset{(1)}{k^\mu}, \underset{(3)}{k^\mu}$ whose integral lines are the coordinate lines $(x^0, x^1, x^3)$. The $x^0$ line is timelike.

(2) The $x^2$ line is orthogonal to all the three $(x^0, x^1, x^3)$ lines.

(3) $g_{\mu\nu} \underset{(0)}{k^\mu} \underset{(3)}{k^\nu} = g_{\mu\nu} \underset{(1)}{k^\mu} \underset{(3)}{k^\nu} = 0$.

(4) $g_{\mu\nu} \underset{(0)}{k^\mu} \underset{(1)}{k^\nu} \neq 0$.

## 7. EXTERIOR SOLUTIONS

It is reasonable to look for exterior solutions having the same symmetry group as the interior ones to which they are to be matched. Taking Statements (1)–(4) above as axioms, we arrive at the metric form

$$ds^2 = (\alpha dx^0 + \beta dx^1)^2 - (\gamma dx^1)^2 - (\delta dx^2)^2 - (\epsilon dx^3)^2, \quad (7.1)$$

where $\alpha, \beta, \gamma, \delta, \epsilon$ are functions of one variable $x^2$. Two cases must be considered separately: $(\beta/\alpha)_{,2} = 0$ and $(\beta/\alpha)_{,2} \neq 0$.

In the first case the metric (7. 1) is static. The only nonflat solution of the empty space field equations (with $\Lambda = 0$) is then

$$\begin{aligned} ds^2 = {}& A_0^2(x^2)^{2a}(dx^0 + Sdx^1)^2 - A_1^2(x^2)^{-2(a-1)}(dx^1)^2 \\ & - A_2^2(x^2)^{2a(a-1)}(dx^2)^2 - A_3^2(x^2)^{2a(a-1)}(dx^3)^2, \end{aligned} \quad (7.2)$$

where $A_0, \ldots, A_3, S, a = \text{const}$. If $a = 0$ or $a = 1$, then (7. 2) is flat.

In the second case (7. 1) is stationary, nonstatic, and $(\beta/\alpha)$ can be taken as a new coordinate $x^2$. Then (7. 1) becomes closely analogous to (3. 2). The solutions of the empty space Einstein equations with the $\Lambda$ term divide into four types and are given by the formulas

$$\begin{aligned} ds^2 = {}& f^{-2}(dx^0)^2 + 2x^2f^{-2}dx^0dx^1 + [(x^2)^2 - V]f^{-2}(dx^1)^2 \\ & - \frac{J^2}{sf^6}\exp\left(-\int \frac{x^2}{V}dx^2\right)(dx^2)^2 \\ & - \frac{V}{sf^2}\exp\left(-\int \frac{x^2}{V}dx^2\right)(dx^3)^2, \end{aligned} \quad (7.3)$$

where $J^2 \neq 0 < s$ are constants and

$$V = (x^2)^2 + px^2 + q, \quad p, q = \text{const}, \quad (7.4)$$

$$f = (Pv_1 + Qv_2)^{1/3}, \quad P, Q = \text{const}. \quad (7.5)$$

$v_1$ and $v_2$ are two linearly independent solutions of the equation

$$v_{,22} - V^{-1}(V_{,2} - x^2)v_{,2} + \frac{3}{4}\left(-\frac{V_{,22}}{V} + \frac{V_{,2}^2}{V^2} - \frac{x^2V_{,2}}{V^2} + \frac{1}{V}\right)v = 0. \quad (7.6)$$

Now compare (7. 3)–(7. 6) with (3. 2)–(3. 6) and note the similarity.

### Type A

$$V = (x^2 - p_0)(x^2 - q_0), \quad q_0 = p_0^*. \quad (7.7)$$

$$\begin{aligned} v_1 &= \left(\frac{x^2 - p_0}{q_0 - L}\right)^\mu \left(\frac{x^2 - q_0}{p_0 - L}\right)^\nu, \\ v_2 &= \left(\frac{x^2 - p_0}{q_0 - L}\right)^{\mu'} \left(\frac{x^2 - q_0}{p_0 - L}\right)^{\nu'}, \quad L = L^* = \text{const}. \end{aligned} \quad (7.8)$$

$$\left.\begin{matrix}\mu \\ \mu'\end{matrix}\right\} = \frac{1}{2(p_0 - q_0)}[p_0 - 2q_0 \pm (p_0^2 - p_0q_0 + q_0^2)^{1/2}]. \quad (7.9)$$

$$\left.\begin{matrix}\nu \\ \nu'\end{matrix}\right\} = \frac{1}{2(p_0 - q_0)}[2p_0 - q_0 \mp (p_0^2 - p_0q_0 + q_0^2)^{1/2}] = \left\{\begin{matrix}\mu^* \\ \mu'^*\end{matrix}\right.. \quad (7.10)$$

$$\Lambda = (s/3J^2)PQ(p_0^2 - p_0q_0 + q_0^2)(p_0 - L)^k(q_0 - L)^l, \tag{7.11}$$

where $k = -q_0/(p_0 - q_0)$, $l = p_0/(p_0 - q_0)$.

### Type B

$$V = (x^2 - p_0)(x^2 - q_0), \quad p_0 \text{ and } q_0 \text{ real}, \quad p_0 < q_0. \tag{7.12}$$

The formulas for $v_1, v_2, \mu, \mu', \nu, \nu'$ are identical with (7.8)–(7.10), but now $\mu^* = \mu$, $\mu'^* = \mu'$. For $x^2 = p_0$ and $x^2 = q_0$ the metric has singularities, and it has the proper signature in the nonconnected region $x^2 < p_0$ and $x^2 > q_0$. The cosmological constant is given by (7.11).

### Type C

$$V = (x^2 - p_0)^2, \quad p_0 \neq 0, \tag{7.13}$$

$$v_1 = |x^2 - p_0|^{3/2}, \quad v_2 = v_1 \exp[p_0/(x^2 - p_0)]. \tag{7.14}$$

$$\Lambda = (s/3J^2)PQp_0^2. \tag{7.15}$$

The signature is proper for all values of $x^2$, but $x^2 = p_0$ is a singular point.

### Type D

$$V = (x^2)^2, \tag{7.16}$$

$$v_1 = |x^2|^{3/2}, \quad v_2 = |x^2|^{1/2}, \tag{7.17}$$

$$\Lambda = -(s/3J^2)Q^2. \tag{7.18}$$

Again the signature is proper everywhere, but $x^2 = 0$ is a singular point.

Now we have to say how these solutions are matched to the interior ones. The solutions of Types I, III, IV, and VI can have the exterior metric of Type A only. For the solutions of Types II and V the exterior metric is of Type A if the joining point $x^2 = r_0$ is at a distance from the singular point greater than some critical value. Otherwise the exterior metric is of Type B (or C), but both singularities of Type B metric (or the singular point of Type C metric) appear outside of matter.

It is interesting that all the four types of stationary exterior solutions can be obtained from the first family solutions by a formal substitution $\kappa = 8\pi k/c^2 \to 0$. Then Type I reduces to Type A, II reduces to $B$, III to C and IV to D. For obvious reasons the Types V and VI have no such analogs. The static metric (7.2) was discovered by E. Kasner[24] in 1925. Some cylindrically symmetric empty space solutions were considered by T. Lewis[25] in 1932. Kasner's solution (7.2) was one of them, but also there appeared Type A metric in the case $Q = 0$. The Type C metric in the case $P = 0$ is contained in Lewis' class,[25] but it is not given explicitly there. Finally, in the case $\Lambda = 0$ all the metrics from the present section are of the form given by Dautcourt, Papapetrou, and Treder.[26,27]

It should be emphasized that these references are rather accidental. The empty space metrics play only an auxiliary role in my paper, so I did not carry out any systematic search in the literature. In particular, I do not guarantee that the generalization of the stationary metrics to the case $\Lambda \neq 0$ is a new result. The generalization of (7.2) to the case $\Lambda \neq 0$ is unexpectedly very involved, so I do not present it here.

## 8. THE TYPE OF CONFORMAL CURVATURE

A special solution of Type IV from the first family is of Petrov type II. It is the metric

$$\begin{aligned} ds^2 = {} & N^{-2/3}(x^2)^{1-\sqrt{2}}[(dx^0)^2 + 2x^2dx^0dx^1 \\ & + 2(\sqrt{2}-1)(x^2)^2(dx^1)^2] + DN^{-2}(x^2)^{-5\sqrt{2}}(dx^2)^2 \\ & + (4DN^{2/3})^{-1}\kappa^2(x^2)^{-3\sqrt{2}}(dx^3)^2. \end{aligned} \tag{8.1}$$

All the other first-family solutions are of Petrov type I (general).

## 9. GEOMETRY OF THE SPACE-TIME

We have noticed in Sec. 6 that the hypersurfaces $x^2 =$ const are flat. Therefore, they can be embedded into the Minkowski space, i.e., they can be realized as some hypersurfaces $x^2 =$ const in the Minkowski space. It appears that this may be done only in four ways. The surfaces $x^0 =$ const, $x^2 =$ const can have the following geometry:

(1) Euclidean plane,

(2) surface of a cylinder with $x^3$ as the azimuthal angle and the $x^1$ line as the generator,

(3) surface of a cylinder with $x^1$ as the azimuthal angle and the $x^3$ line as the generator, parametrized by an observer at rest,

(4) the same surface as in (3), parametrized by an observer rotating about the axis of symmetry.

Since the velocity field is given by (1.22) and (1.23), we see that the particles of the fluid move inside the $x^2 = C$ hypersurface and follow the $x^1$ lines. Moreover, we know from (1.25) that the vorticity vector is tangent to $x^3$ lines, and we can compute quite easily the acceleration vector $\dot{u}_\alpha = H^{-1}H_{,2}\delta^2_\alpha$, which is tangent to $x^2$ lines.

This is enough to decide which case listed above is realized in our space–times of the first family. In cases (1) and (2) the acceleration, if present, is tangent to $x^1$ lines because the streamlines are straight. In cases (3) and (4) the acceleration has the direction of the radial line $x^2$, just as in our metrics. We decide that case (4) is a better model of our space–time since we do not expect that in the presence of rotating matter an observer at absolute rest would exist. It means that our space–time, when realized nonrelativistically, consists of co-axial cylinders rotating around an axis of symmetry with different angular velocities. All the physical quantities are constant on the surface of a fixed cylinder, but they vary from one cylinder to the other. The $x^2$-lines are geodesics orthogonal to the cylinders, $x^1$ lines are azimuthal circles, and $x^3$ lines are generators. Now we see that the second of the assumptions (1.30) meant just homogeneity in the direction of generators. We did not assume axial symmetry, but it resulted from the field equations.

## 10. PHYSICAL PROPERTIES OF THE SOLUTIONS

The velocity field has no shear or expansion, but it has rotation and acceleration. Rotation produces no red shift. According to Ehlers' formula[13] the red shift is equal to

$$(d\lambda)/\lambda = -\dot{u}_\alpha \delta_\perp x^\alpha, \tag{10.1}$$

where $\dot{u}_\alpha = H^{-1}H_{,2}\delta^2{}_\alpha$, and

$$\delta_\perp x^\alpha = (\delta^\alpha{}_\beta - u^\alpha u_\beta)\delta x^\beta, \tag{10.2}$$

$\delta x^\beta$ being the infinitesimal vector pointing from the observer to the particle sending light signals to him. The

TABLE I. Models of rotating matter.

| | | | | |
|---|---|---|---|---|
| Lanczos[3] 1924<br>van Stockum[20] 1937<br>Wright[21] 1965<br>*Ellis[22] 1967<br>*Krasiński 1973 | Gödel[19] 1949<br>*Wright[21] 1965<br>*Ozsváth[30] 1965<br>*Raval-Vaidya[18] 1966<br>*Ellis[22] 1967<br>*Wainwright[6] 1970<br>*Ozsváth[31] 1970<br>Bray[29] 1972<br>*Krasiński 1973 | Ozsváth-Schücking[32] 1962<br>*Ozsváth[30] 1965 | Ozsváth[30] 1965 | |
| Maitra[33] 1966 | Raval-Vaidya[18] 1966<br>Stewart-Ellis[5] 1968<br>Wainwright[6] 1970<br>*Krasiński 1973 | Ellis[22] 1967<br>*Wainwright[6] 1970 | Stewart-Ellis[5] 1968<br>*Wainwright[6] 1970 | |
| Wahlquist[7] 1968 | Wainwright[6] 1970 | Ozsváth[31] 1970 | Herlt[8] 1972 | Krasiński 1973 |

red shift given by (10. 1) is strongly anisotropic and thus rather not realistic. However, it is not obvious that red shift computed with respect to distant sources of light would also have such a strong anisotropy.

It is interesting that $\rho = \rho(x^2)$ and $p = p(x^2)$, and so we have an equation of state $\rho = \rho(p)$ given in a parametric way, which resulted from the field equations. It might be unexpected as one usually considers an equation of state to be independent of the field equations. But if the metric tensor depends only on one variable, the equation of state is always determined by the field equations.

Now look at (3. 4), (3. 5), (3. 7), and (3. 8)–(3. 13). There are six independent arbitrary constants—$D, M, N, G, b, c'$ entering the equation of state $\rho = \rho(p)$. In fact, this is a large class of equations of state.

One may expect simpler results when the parameters of the hypergeometric function in (3. 10) are such that $F(.\,,\,.\,.\,,\,.\,,\,.)$ degenerates to a polynomial. J. Plebański[28] even suggested that then it would be possible to obtain the equation of state in the form of the van der Waals isotherms.

This question has not been investigated.

In Type IV solutions if $M = 0$ or $N = 0$ then $\rho$ and $p$ obey the polytrope type equation of state $p \cdot \rho^{-\gamma} = \text{const}$, with $[5a - 6 + \epsilon(a^2 - 3a + 3)^{1/2}]\gamma = 6(a-1)$, where $\epsilon = +1$ corresponds to $N = 0$ and $\epsilon = -1$ corresponds to $M = 0$. The condition $a > 1$ implies that $\gamma < 0$, $1 < \gamma < \frac{4}{3}$ or $\gamma > \frac{3}{2}$.

## 11. A SURVEY OF MODELS OF ROTATING PERFECT FLUID OR DUST

This survey is made in the form of a table. Each "cell" of the table represents one solution obtained by different authors. A star preceding author's name means that he knew his predecessors and did not expect to be the first inventor of the solution. There is no star at Bray's name in the "Gödel's cell" because his solutions, when the electromagnetic field is absent, reduce precisely to the metric of Gödel, but this fact was not indicated in Bray's paper.[29]

For each of the solutions the coordinates $(\tau, \xi, \eta, \zeta)$ from (1. 14), (1. 20), and (1. 21) can be introduced, but this might be a subject of another paper.

No approximate solutions are taken into account.

A large expansion of this article is currently submitted to Acta Physica Polonica.

## ACKNOWLEDGMENTS

I am indebted to Professor Jerzy Plebański for pointing out this subject to me as well as for constant encouragement and valuable advice while writing this paper. It owes much to his remarks and patient discussions with me. I deeply appreciate the kind help of Dr. Marek Demiański in tracing the references. My special thanks to Dr. Leonid Grishchuk from the Sternberg Astronomical Institute in Moscow for only one but very profitable discussion and to Dr. Bogdan Mielnik for some very useful remarks during the seminars.

*Present address: Institute of Astronomy, Polish Academy of Sciences, Warsaw, Poland.


[1]A. M. Wolfe, Astrophys. J. **159**, L61 (1970).
[2]J. Lense and H. Thirring, Physik. Z. **19**, 156 (1918).
[3]K. Lanczos, Z. Physik. **21**, 73 (1924).
[4]M. Trümper, Z. Naturforsch. **22a**, 1347 (1967).
[5]J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).
[6]J. Wainwright, Commun. Math. Phys. **17**, 42 (1970).
[7]H. D. Wahlquist, Phys. Rev. **172**, 1291 (1968).
[8]E. Herlt, Wiss. Z. Friedrich-Schiller-Universität Jena, Math. Naturwiss. Reihe **21**, 19 (1972).
[9]J. Plebański, *Lectures on Non-Linear Electrodynamics* (Nordita, Copenhagen, 1970), pp. 107-115 and 130-141.
[10]S. Sternberg, *Lectures on Differential Geometry* (Prentice-Hall, Englewood Cliffs, N. J., 1964), p. 141.
[11]N. Bourbaki, *Éléments de mathématique. Premiere partie: Les structures fondamentales de l'Analyse. Livre II: Algèbre.* (Herman,Paris, 1948).
[12]S. Lang, *Algebra* (Addison-Wesley, Reading, Mass., 1965), p.372.
[13]J. Ehlers, Abhandl. Math. Naturwiss. Kl. Akad. Wiss. Lit. Mainz **11**, 791 (1961).
[14]J. Ehlers, in *Recent Developments in General Relativity* (Pergamon, New York, 1962), p.201.
[15]I. M. Ryżyk and I. S. Gradsztejn, *Tablice całek, sum, szeregów i iloczynów* (Państwowe Wydawnictwo Naukowe, Warsaw, 1964), p. 436 (in Polish).
[16]V. I. Smirnov, *Kurs wyšsej matematiki* (Gosudarstvennoje Izdatel'stvo Techniko-Teoretičeskoj Literatury, Moscow, 1950), Vol. 3, part II, pp. 369-376 (in Russian).
[17]P. M. Morse and H. Feschbach, *Methods of Theoretical Physics* (McGraw–Hill, New York, 1953), pp. 537-539 and 541-543.
[18]H. M. Raval and P. C. Vaidya, Ann. Inst. Poincaré A **4**, 21 (1966).
[19]K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[20]W. J. van Stockum, Proc. Roy. Soc. Edinburgh **57**, 135 (1937).
[21]J. P. Wright, J. Math. Phys. **6**, 103 (1965).
[22]G. F. R. Ellis, J. Math. Phys. **8**, 1171 (1967).
[23]V. A. Belinskii, I. M. Khalatnikov, and E. M. Lifschitz, Adv. Phys. **19**, 525 (1970).
[24]E. Kasner, Trans. Am. Math. Soc. **27**, 249 (1925).
[25]T. Lewis, Proc. Roy. Soc. London A **136**, 176 (1932).
[26]G. Dautcourt, A. Papapetrou, and H. Treder, Ann. Physik **9**, 330 (1962).
[27]I thank Dr. E. Kreisel from Potsdam, GDR, for the information about the reference above.
[28]Private communication.
[29]M. Bray, Compt. Rend. Acad. Sci. (Paris) A **274**, 809 (1972).
[30]I. Ozsváth, J. Math. Phys. **6**, 590 (1965).
[31]I. Ozsváth, J. Math. Phys. **11**, 2871 (1970).
[32]I. Ozsváth and E. Schücking, Nature **193**, 1168 (1962); Ann. Phys. **55**, 166 (1969).
[33]S. C. Maitra, J. Math. Phys. **7**, 1025 (1966).

# Preliminary results in the Banach space formulation of master equation and subdynamics theory in quantum statistics

**M. A. Alberti, P. Cotta-Ramusino,* and G. Ramella**

*Istituto di Fisica, Università di Milano, 20133 Milano, Italy*
(Received 4 June 1973)

The Liouville–von Neumann equation and a useful decomposition of the resolvent of the generator of the time evolution operators are obtained in the formulation of quantum statistics by means of the pair of Banach spaces ($\tau c$, $\mathbb{B}$), where $\tau c$ is the space of the "trace-class" operators and $\mathbb{B}$ is the space of all bounded operators, defined on a Hilbert space.

## 1. INTRODUCTION

The theory of "independent subdynamics," developed in Refs. 1-3, provides a new method to extract the macroscopic level from the quantum mechanical description of a macroscopic system.

It provides also a possible way to realize the "embedding" of the new axiomatic description of macroscopic systems, recently proposed by Ludwig,[4] into the quantum theory of $N$-body systems.

A similar but simpler problem is the "embedding" of unstable nonrelativistic particles into a Galilean field theory, which has been rigorously treated in Ref. 5.

An important step of the theory of indipendent subdynamics is to set up a decomposition of the resolvent of the generator of the time evolution operators by means of a suitable projection (see, e.g. Refs. 6-10).

So far such a decomposition has been found only in Hilbert spaces, as in the case

(a). of a pure states description as in Ref. 5; then the generator of the time evolution operators is the Hamiltonian operator; and

(b). of a statistical description in the Hilbert space of the Hilbert-Schmidt operators, called the Liouville space, as in Refs. 2 and 6; then the generator of the time evolution operators is the Liouville–von Neumann operator.

However the "Hilbert space" formulation of quantum statistics is not satisfactory, since one must restrict observables to be Hilbert–Schmidt operators.

In the physically interesting applications of quantum statistics one deals with the mean values $\langle A\rangle_t$ of observables $A$ which are bounded operators; the mean values are given by the expression

$$\langle A\rangle_t = \mathrm{tr}\,(AW(t)) \quad [W(t) \text{ is the statistical operator}] \tag{1.1}$$

Then a mathematical structure in which statistical descriptions of physical systems can be formalized in a more satisfactory way is a pair of Banach spaces which are in duality by the bilinear form (1.1).

Precisely we shall choose the pair of spaces $(\tau c, \mathbb{B})$ where $\tau c$ is the Banach space of all the "trace-class" operators, that is of all the operators $A$ defined on $\mathfrak{H}$, the "states-Hilbert space" of the given system, such that $\mathrm{tr}\,((A^+A)^{1/2} < \infty$ and $\mathbb{B}$ [or $\mathcal{B}(\mathfrak{H})$] is the dual space of $\tau c$, that is the Banach space of all bounded operators on $\mathfrak{H}$ (see Ref. 4).

In such a formalism the space of observables is much "wider" than the Liouville space, and on the other side the space of the statistical operators is as " small" as possible, since the linear manifold spanned by the statistical operators is exactly $\tau c$ (see Appendix).

Moreover, the choice of the pair of spaces $(\tau c, \mathbb{B})$ is suggested by the axiomatic foundations of quantum mechanic proposed by Ludwig in Ref. 11.

In this paper we use the pair of spaces $(\tau c, \mathbb{B})$ to rewrite the Liouville–von Neumann equation (Sec. 2) and to find the decomposition of the resolvent of the generator of the time evolution operators (Sec. 3).

## 2. LIOUVILLE-VON NEUMANN EQUATION IN $\tau c$

Let $\mathfrak{H}$ be the Hilbert space of the states of the given quantum system, $H = \int_{-\infty}^{+\infty} \lambda dE_\lambda$ the Hamiltonian, $U(t) =: \exp(-iHt) = \int_{-\infty}^{+\infty} \exp(-i\lambda t) dE_\lambda$ the time evolution operator.

It is well known that the family of operators $U(t)$ is a strongly continuous group of unitary operators on $\mathfrak{H}$.

LIST OF SYMBOLS[a]

(A) The symbol $=:$ stands for "defined as:"

(B) Let $X$ be a Banach space:

| | |
|---|---|
| $X^+$ | is the dual space of $X$. |
| $\langle f,g\rangle\, f \in X^+,\ \ g \in X$ | is the functional $f$ applied to $g$ |
| $\mathcal{B}(x)$ | is the algebra of all linear bounded operators on $X$ itself. |
| $\mathcal{C}(X)$ | is the set of closed operators from $X$ to itself. |
| $S < X$ | denotes that $S$ is a subspace (that is a closed linear manifold) of $X$. |
| $S^\perp\ (S \subset X)$ | is the set of all $f \in X^+$ such that $\langle f,g\rangle = 0\ \ \forall g \in S$. |

(C) let $A$ be a linear operator defined in a Banach space:

| | |
|---|---|
| $\mathfrak{D}(A)$ | is the domain of $A$. |
| $\mathcal{R}(A)$ | is the range of $A$. |
| $A^+$ | is the adjoint operator of $A$, if it exists. |
| $\bar{A}$ | is the closure of $A$, if it exists. |
| $\mathcal{G}(A)$ | is the graph of $A$. |
| $\mathrm{nul}(A)$ | is the dimension of the null space of $A$. |

(D) Let $\mathfrak{H}$ be the Hilbert space of the system and $A, B$ two operators defined on $\mathfrak{H}$:

| | |
|---|---|
| $(f,g)\ \ f,g \in \mathfrak{H}$ | is the inner product in $\mathfrak{H}$. |
| $((A,B))$ | stands for $\mathrm{tr}\,(A^+B)$. |
| $\lVert A\rVert$ | is the usual norm of $A$. |
| $\lVert A\rVert_2$ | stands for $(\mathrm{tr}\,(A^+A))^{1/2} = \sqrt{((A,A))}$. |
| $\lVert A\rVert_1$ | stands for $\mathrm{tr}\,((A^+A)^{1/2})$. |
| $\mathbb{B}, \mathcal{B}(\mathfrak{H})$ | is the algebra of all bounded operators on $\mathfrak{H}$. Its norm is $\lVert\cdot\rVert$. |
| $\mathcal{C}$ | is the Banach algebra of all completely continuous operators on $\mathfrak{H}$. Its norm is $\lVert\cdot\rVert$. |
| $\sigma c$ | is the Hilbert space of all Hilbert-Schmidt operators. Its inner product is $((\cdot,\cdot))$. |
| $\tau c$ | is the Banach space of all the "trace-class" operators. Its norm is $\lVert\cdot\rVert_1$. |

c.o.n.s. stands for "complete orthonormal system."

[a] For other notations see Lemmas A 1-A 5 in Appendix.

Let us now consider the "corresponding" time evolution operator in the $\tau c$ space:

$$\mathcal{U}(t)\colon \tau c \to \tau c$$
$$\mathcal{U}(t)A =: U(t)AU^+(t), \qquad \forall t, \forall A \in \tau c \tag{2.1}$$

(we recall that the product of any bounded operator with any operator in $\tau c$ is an operator in $\tau c$).

From the definition of $\mathcal{U}(t)$ it follows that:

1. $\mathcal{U}(t)$ is a group of operators on $\tau c$,
2. $\forall t$, $\mathcal{U}(t)$ is an isometric operator,
3. $\forall t$, $\mathcal{R}[\mathcal{U}(t)] = \tau c$,
4. $\mathcal{U}(t)$ is weakly continuous at $t = 0$.

Let us prove these statements:

1. It is immediate.

2. Let $\{f_n\}$ and $\{g_n\}$ be any two c.o.n.s. in $\mathfrak{H}$; since $\forall t$, $U(t)$ is an unitary operator on $\mathfrak{H}$; the mapping $\{f_n\} \to \{U^+(t)f_n\}$ is a bijective mapping of the set of all c.o.n.s. in $\mathfrak{H}$ onto itself. Then we have: $\sum |(U(t)WU^+(t)f_n, g_n)| = \sum |(WU^+(t)f_n, U^+(t)g_n)| \leq \|W\|_1$. That is $\forall$ c.o.n.s. $\{f_n\}$ and $\{g_n\}$, $\sum |(U(t)WU^+(t)f_n, g_n)|$ is convergent for every fixed $t$ and we have

$$\|\mathcal{U}(t)W\|_1 = \|U(t)WU^+(t)\|_1 = \underset{\{f_n\},\{g_n\}}{\text{l.u.b.}} \left(\textstyle\sum_n |(U(t)WU^+(t)f_n, g_n)|\right) = \underset{\{U_n\},\{V_n\}}{\text{l.u.b.}} \left(\textstyle\sum |(WU_n, V_n)|\right) = \|W\|_1,$$

where the l.u.b. extends over the set of all c.o.n.s. in $\mathfrak{H}$ (Ref. 12, 1.2.4).

3. It is immediate.

4. To show the weak continuity of $\mathcal{U}(t)$ at $t = 0$, we have to prove that $\forall A \in \mathcal{B}$, $\forall W \in \tau C$

$$\operatorname{tr}\,(\mathcal{U}(t)WA - WA) \to 0 \quad \text{for } t \to 0$$

Indeed, $\forall W \in \tau c$ and $A \in \mathcal{B}$ let us choose a c.o.n.s. in $\mathfrak{H}$, such that $\sum_1^\infty {}_n \|W^+f_n\| < \infty$.

Of such a c.o.n.s. there exists at least one (Ref. 12, 1.2.8). Then $\forall \epsilon > 0$, $\exists N1(\epsilon) > 0$ such that for $N \geq N1(\epsilon)$ we have

$$\sum_{N+1}^{\infty} {}_n \|W^+f_n\| \leq \epsilon/3\|A\|$$

and $\forall \epsilon > 0$, $\exists N_2(\epsilon) > 0$ such that for $N \geq N_2(\epsilon)$ we have

$$\left|\sum_{N+1}^{\infty} {}_n (WAf_n, f_n)\right| \leq \epsilon/3.$$

Let $\bar{N} = \max(N_1, N_2)$; $\forall \epsilon > 0 \; \exists \, \delta_\epsilon > 0$ such that $\forall n = 0, 1, \ldots, N$ and $\forall t$, $|t| < \delta_\epsilon$, we have

$$\|(WU^+(t)AU(t) - WA)f_n\| \leq \epsilon/3\bar{N}$$

since $U(t)$ is a strongly continuous operator at $t = 0$.

Then $\forall t$, $|t| < \delta_\epsilon$ we have

$$\begin{aligned}
|\operatorname{tr}(U(t)WU^+(t)A - WA)| &= |\operatorname{tr}(WU^+(t)AU(t) - WA)| \\
&\leq \Big|\sum_1^{\bar N}{}_n([WU^+(t)AU(t) - WA]f_n, f_n)\Big| + \Big|\sum_{\bar N+1}^{\infty}{}_n(WU^+(t)AU(t)f_n, f_n)\Big| + \Big|\sum_{\bar N+1}^{\infty}{}_n(WAf_n, f_n)\Big| \\
&\leq \sum_1^{\bar N}{}_n\|(WU^+(t)AU(t) - WA)f_n\| + \sum_{\bar N+1}^{\infty}{}_n\|U^+(t)AU(t)f_n\| \cdot \|W^+f_n\| + \frac{\epsilon}{3} \leq \tfrac{2}{3}\epsilon + \|A\| \sum_{\bar N+1}^{\infty}{}_n \|W^+f_n\| \leq \epsilon.
\end{aligned}$$

(We recall that the trace of an operator in $\tau c$ is independent of the choosen c.o.n.s.)

Now using the general theory of semigroups and the properties 1–4 above, it follows that the semigroup $\mathcal{U}(t)$, $t > 0$ verifies the following statements:

(A) $\mathcal{U}(t)$ is strongly continuous at $t = 0$, that is $\forall W \in \tau c \|\mathcal{U}(t)W - W\|_1 \to 0$ for $t \to 0^+$ (Ref. 13, Theorem 10.6.5.).

(B) $\mathcal{U}(t)$ is strongly continuous for $t \geq 0$ (Ref. 13, Theorem 10.5.5).

(C) Let $A_\eta =: \dfrac{\mathcal{U}(\eta) - \mathbb{I}_{\tau c}}{\eta}$ and

$$-i\,Ж =: \lim_{\eta \to 0+} A_\eta, \tag{2.2}$$

then Ж is a linear operator *densely defined in* $\tau c$, generally unbounded and $-i\,Ж$ is the generator of the semigroup $\mathcal{U}(t)$ (Ref. 13, Sec. 10.3).

(D) We have $\forall W \in \mathfrak{D}(Ж)$

$$s - \frac{d}{d\xi}\,\mathcal{U}(\xi)W = -i\,Ж\,\mathcal{U}(\xi)W = -i\,\mathcal{U}(\xi)\,Ж\,W. \tag{2.3}$$

(E) Ж *is a closed operator* (Ref. 13, Theorem 11.5.1).

(F) Every $z$ with $\operatorname{Im} z \neq 0$ belongs to the resolvent set of Ж.

*Proof:* first we have (Ref. 13, Sec. 11.5 and Theorem 12.3.1) every $\lambda$ with $\operatorname{Re}\lambda > 0$ belongs to the resolvent set of $-i$Ж and it holds:

$$\frac{1}{\lambda + i\,Ж} = \int_0^\infty e^{-\lambda x}\,\mathcal{U}(x)\,dx \tag{2.4}$$

$$\left\|\frac{1}{\lambda + i\,Ж}\right\| \leq \frac{1}{\operatorname{Re}\lambda}, \qquad \operatorname{Re}\lambda > 0. \tag{2.5}$$

Then, let us consider the semigroup $\mathcal{V}(t)$,

$$\mathcal{V}(t)\colon \tau c \to \tau c, \qquad \mathcal{V}(t)A =: U^+(t)AU(t), \qquad \forall A \in \tau c, \qquad t > 0.$$

The statements 1-4 and (A)-(F) obviously hold for the semigroup $\mathcal{V}(t)$ also.

Let $-i\,\tilde{Ж}$ be the generator of $\mathcal{V}(t)$.

Then

$$-i\,\tilde{Ж} =: s\text{–}\lim_{\eta \to 0+} \frac{\mathcal{V}(\eta) - \mathbb{I}_{\tau c}}{\eta} = s\text{–}\lim_{\eta \to 0+} \frac{\mathcal{U}(\eta)^{-1} - \mathbb{I}_{\tau c}}{\eta} = -\,s\text{–}\lim_{\eta \to 0+} \mathcal{U}(\eta)^{-1}\,\frac{\mathcal{U}(\eta) - \mathbb{I}_{\tau c}}{\eta}.$$

But $s\text{–}\lim_{\eta \to 0+} \mathcal{U}(\eta)^{-1} = \mathbb{I}_{\tau c}$, so we have $\tilde{Ж} = -\,Ж$, then every $\lambda$ with $\operatorname{Re}\lambda > 0$ belongs to the resolvent set ot $i\,Ж$ and (2.4), (2.5) still hold for the operator $i\,Ж$.

Finally, every $z$ with $\operatorname{Im} z \neq 0$ belongs to the resolvent set of Ж and we have

$$\left\|\frac{1}{Ж - z}\right\| \leq \frac{1}{|\operatorname{Im} z|}, \qquad \operatorname{Im} z \neq 0. \tag{2.6}$$

(G) $\forall A \in \tau c$, $\forall \epsilon > 0$ we have

$$\mathcal{U}(t)A = \lim_{\xi \to \infty} \frac{1}{2\pi i}\,\frac{1}{\xi} \int_0^\xi dw \int_{-w+i\epsilon}^{w+i\epsilon} dz\, \exp(-izt)\,\frac{1}{Ж - z}A \tag{2.7}$$

(Ref. 13, Sec. 11.7-E.7).

(H) $\forall A \in \mathfrak{D}(\mathcal{K})$, $\forall \epsilon > 0$ we have

$$\mathfrak{U}(t)A = \lim_{w \to \infty} \frac{1}{2\pi i} \int_{-w+i\epsilon}^{w+i\epsilon} dz \exp(-izt) \frac{1}{\mathcal{K} - z} A \tag{2.8}$$

(Ref. 13, Sec. 11.7-E. 6).

(I) $\forall C \in \mathfrak{D}(\mathcal{K})$ we have

$$\mathcal{K}C = \overline{[H, C]} \tag{2.9}$$

In fact, $\forall C \in \mathfrak{D}(\mathcal{K})$ and $\forall f \in \mathfrak{H}$ it holds that

$$\left\| \left( \frac{U(t)CU^+(t) - C}{t} + i\mathcal{K}C \right) f \right\| \leqslant \left\| \frac{U(t)CU^+(t) - C}{t} + i\mathcal{K}C \right\|_1 \|f\| \to 0 \quad \text{for } t \to 0^+$$

since $\forall f \in \mathfrak{D}(H)$, $\lim_{t \to 0^+} U(t)C[(U^+(t) - \mathbb{1})/t]f = iCHf$ [we recall that $-iH$ is the generator of the group $U(t)$], then

$$\lim_{t \to 0+} \frac{U(t) - \mathbb{1}_{\mathfrak{H}}}{t} Cf = \left( \lim_{t \to 0+} \frac{U(t)CU^+(t) - C}{t} f \right) - \left( \lim_{t \to 0+} U(t)C \frac{U^+(t) - \mathbb{1}_{\mathfrak{H}}}{t} f \right) = -i(\mathcal{K}C)f - iCHf.$$

Thus $Cf \in \mathfrak{D}(H)$ and it holds that

$$HCf = (\mathcal{K}C)f + CHf, \qquad \forall f \in \mathfrak{D}(H), \qquad \forall C \in \mathfrak{D}(\mathcal{K}).$$

Now $\mathcal{K}C \in \tau c$, then it is bounded and $\mathcal{K}C \supset [H, C]$; thus $[H, C]$ is a bounded operator, its domain is dense in $\tau c$ and $\mathcal{K}C = \overline{[H, C]}$.

(J) let $W(t) =: \mathfrak{U}(t)W(0)$, applying (2.3) and (2.9) to $W(t)$ one has the *Liouville-von Neumann Equation in the $\tau c$ Space*:

$$s\frac{d}{dt} W(t) = -i\mathcal{K}W(t) = -i\overline{[H, W(t)]} \tag{2.10}$$

(K) since $\mathfrak{D}(\mathcal{K})$ is dense in $\tau c$, $\mathcal{K}^+$ exists and it holds:

$$\left[\frac{1}{\mathcal{K} - z}\right]^+ = \frac{1}{\mathcal{K}^+ - z}$$

Then even the spectrum of $\mathcal{K}^+$ is contained in the real axis: $\mathcal{K}^+$ is obviously a closed operator on $\mathcal{B}$, but its domain is not necessarily dense in $\mathcal{B}$, since the $\tau c$ space is not reflexive.

(L) $\forall A \in \mathfrak{D}(\mathcal{K})$, we have

$$A \in \mathfrak{D}(\mathcal{K}^+) \text{ and } \mathcal{K}^+A = \mathcal{K}A. \tag{2.11}$$

In fact, $\forall A, B \in \mathfrak{D}(\mathcal{K})$ we have

$$\begin{aligned}\langle B, -i\mathcal{K}A\rangle &= \operatorname{tr}\left(B^+ \lim_{t \to 0+} \frac{\mathfrak{U}(t)A - A}{t}\right) = \lim_{t \to 0+} \operatorname{tr}\left(B^+ \frac{\mathfrak{U}(t)A - A}{t}\right) \\ &= \lim_{t \to 0+} \operatorname{tr}\left(\frac{B^+U(t)AU^+(t)}{t} - \frac{B^+A}{t}\right) \\ &= \lim_{t \to 0+} \operatorname{tr}\left(\boxed{\frac{U^+(t)B^+U(t) - B^+}{t}} A\right) \\ &= \lim_{t \to 0+} \operatorname{tr}\left[\left(\frac{U^+(t)BU(t) - B}{t}\right)^+ A\right] \\ &= \lim_{t \to 0+} \operatorname{tr}\left[\left(\frac{\mathfrak{V}(t)B - B}{t}\right)^+ A\right],\end{aligned}$$

since

$$\left\|\left|\frac{\mathfrak{V}(t)B - B}{t} - i\mathcal{K}B\right\| \leqslant \left\| \frac{\mathfrak{V}(t)B - B}{t} - i\mathcal{K}B\right\|_1 \to 0.$$

Then

$$\begin{aligned}\lim_{t \to 0+} \operatorname{tr}\left[\left(\frac{\mathfrak{V}(t)B - B}{t}\right)^+ A\right] &= \lim_{t \to 0} \left\langle \frac{\mathfrak{V}(t)B - B}{t}, A \right\rangle \\ &= \left\langle \lim_{t \to 0+} \frac{\mathfrak{V}(t)B - B}{t}, A \right\rangle = \langle i\mathcal{K}B, A\rangle.\end{aligned}$$

Hence it follows the statement.

## 3. DECOMPOSITION OF THE RESOLVENT $1/(\mathcal{K} - z)$

Let $\mathcal{P}$ be a projection operator defined on $\tau c$ with range $\mathcal{S}$.

Assume that $\mathcal{P}$ verifies the following properties:

I. 1 $\quad \forall A \in \tau c \quad \mathcal{P}^+A = \mathcal{P}A$

I. 2 $\quad \overline{\mathfrak{D}(\mathcal{K}\mathcal{P}) \cap \mathfrak{D}(\mathcal{K})} = \tau c$

Now let us consider the operator

$$\hat{\mathcal{K}} =: (\mathbb{1}_{\tau c} - \mathcal{P})\mathcal{K}(\mathbb{1}_{\tau c} - \mathcal{P}).$$

By I. 2 above, $\hat{\mathcal{K}}$ is densely defined. Moreover, let us suppose $\hat{\mathcal{K}}$ verifies

I. 3 $\quad \mathcal{R}(\hat{\mathcal{K}} - z) = \tau c, \qquad \operatorname{Im} z \neq 0.$

Afterward we consider the operator

$$\mathcal{K}' =: \mathcal{P}\mathcal{K} + \mathcal{K}\mathcal{P} + \hat{\mathcal{K}} - \mathcal{P}\mathcal{K}\mathcal{P}.$$

Obviously $\mathcal{K} \supset \mathcal{K}'$ and $\mathcal{K}'$ is densely defined, by I. 2. Furthermore, let us suppose

I. 4 $\quad \overline{\mathcal{R}(\mathcal{K}' - z)} = \tau c, \qquad \operatorname{Im} z \neq 0.$

We make some remarks on the assumptions I. 1-I. 4:

(1) If $\mathcal{K}\mathcal{P} \in \mathcal{B}(\tau c)$, assumptions I. 2 and I. 4 are an obvious consequence of equation $\mathcal{K} = \mathcal{K}'$.

(2) Assume the subspace $\mathcal{S}$ be finite dimensional. Then assumptions I. 2 and I. 4 may be changed with the assumption

I. 2′ $\quad \mathcal{S} \subset \mathfrak{D}(\mathcal{K})$.

In fact, if I. 2′holds $\mathcal{K}\mathcal{P} \in \mathcal{B}(\tau c)$ and by the former remark, I. 2 and I. 4 are verfied as well.

Vice versa if I. 2′ does not hold, we have $\mathcal{S} = \tilde{\mathcal{S}} \oplus \mathcal{T}$ where $\tilde{\mathcal{S}} =: \mathfrak{D}(\mathcal{K}) \cap \mathcal{S}$ and $\mathcal{T}$ is a suitable linear manifold different from $\{0\}$.

Obviously, $\mathcal{T} \cap \mathfrak{D}(\mathcal{K}) = \{0\}$ and so $\mathfrak{D}(\mathcal{K}\mathcal{P}) \cap \mathcal{T} = \{0\}$, hence $\overline{\mathfrak{D}(\mathcal{K}\mathcal{P})} \cap \mathcal{T} = \{0\}$. That is I. 2 does not hold.

Moreover, as regards assumption I. 1 we note that $\mathcal{S}$ is a subspace of the Hilbert space $\sigma c$ as well, and therefore a unique orthogonal projection $Q_{\mathcal{S}}$ of $\sigma c$ on $\mathcal{S}$ does exist. The restriction $\tilde{Q}_{\mathcal{S}}$ of $Q_{\mathcal{S}}$ to $\tau c$ is obviously a projection of $\tau c$ on $\mathcal{S}$, that verifies I. 1, and we may consider $\tilde{Q}_{\mathcal{S}}$ instead of $\mathcal{P}$.

In the following we shall give some properties of the operator $\hat{\mathcal{K}}$:

($\alpha$) $\hat{\mathcal{K}}$ is closable.

*Proof:* $\hat{Ж}$ and then $Ж(\mathbb{I}_{\tau c} - \mathcal{P})$ are densely defined, therefore $[Ж(\mathbb{I}_{\tau c} - \mathcal{P})]^+$ exists and it holds that

$$[Ж(\mathbb{I}_{\tau c} - \mathcal{P})]^+ \supset (\mathbb{I}_{\tau c} - \mathcal{P})^+ Ж^+.$$

Thus $(\mathbb{I}_{\tau c} - \mathcal{P})^+ Ж^+$ is closable and by (2.11) and I.1 the restriction to $\tau c$ of $(\mathbb{I}_{\tau c} - \mathcal{P})^+ Ж^+$ is $(\mathbb{I}_{\tau c} - \mathcal{P})\, Ж$. Hence by Lemma A.3 (in Appendix) $(\mathbb{I}_{\tau c} - \mathcal{P})\, Ж$ is also closable.

$\hat{Ж}$ is then closable because it is the product of a closable operator with a bounded operator.

($\beta$) $\forall A \in \mathfrak{D}(\hat{Ж})$ we have

$$\hat{Ж}^+ A = \hat{Ж} A \tag{3.1}$$

*Proof:* $\hat{Ж}^+ \supset (\mathbb{I}_{\tau c} - \mathcal{P})^+ Ж^+ (\mathbb{I}_{\tau c} - \mathcal{P})^+$

and

$$\mathfrak{D}(\hat{Ж}) = \{A \mid (\mathbb{I}_{\tau c} - \mathcal{P})A \in \mathfrak{D}(Ж)\}.$$

The statement follows immediately by (2.11) and I.1.

($\gamma$) $(\hat{Ж} - z)^{-1}$ exists $\forall z$, $\operatorname{Im} z \neq 0$.

*Proof:* Let $A \in \mathfrak{D}(\hat{Ж})$. Then

$$\begin{aligned}\|(\hat{Ж} - z)A\|_1^2 &\geq \|(\hat{Ж} - z)A\|_2^2 \\ &= \|(\hat{Ж} - \operatorname{Re} z)A\|_2^2 + |\operatorname{Im} z|^2 \|A\|_2^2 \\ &\geq |\operatorname{Im} z|^2 \|A\|_2^2 > 0\end{aligned}$$

when

$$A \neq 0, \quad \operatorname{Im} z \neq 0.$$

($\delta$) $(\bar{\hat{Ж}} - z)^{-1}$ exists $\forall z$, $\operatorname{Im} z \neq 0$.

*Proof:* Assume $A_n \in \mathfrak{D}(\hat{Ж})$. $\|A_n - A\|_1 \to 0$ and $\|(\hat{Ж} - z)A_n\|_1 \to 0$.

Let us suppose *ab absurdo* $A \neq 0$.[14]

Then

$$|\,\|A\|_2 - \|A_n\|_2| \leq \|A_n - A\|_2 \leq \|A_n - A\|_1 \to 0,$$

thus

$$\|A_n\|_2 \to \|A\|_2.$$

By ($\gamma$) we have $\|(\hat{Ж} - z)A_n\|_1 \geq |\operatorname{Im} z|\, \|A_n\|_2$.

Hence

$$\begin{aligned}0 = \lim_{n\to\infty} \|(\hat{Ж} - z)A_n\|_1 &\geq \lim_{n\to\infty} |\operatorname{Im} z|\, \|A_n\|_2 \\ &= |\operatorname{Im} z|\, \|A\|_2 \Rightarrow A = 0.\end{aligned}$$

($\epsilon$) $\forall z, \operatorname{Im} z \neq 0$,

$$(\bar{\hat{Ж}} - z)^{-1} \in \mathfrak{B}(\tau c) \tag{3.2}$$

*Proof:* From above it follows that

$$\mathfrak{R}(\bar{\hat{Ж}} - z) \supset \mathfrak{R}(\hat{Ж} - z) = \tau c.$$

Then applying the closed graph theorem, we have

$$(\bar{\hat{Ж}} - z)^{-1} \in \mathfrak{B}(\tau c), \qquad \forall z \text{ with } \operatorname{Im} z \neq 0.$$

Besides $(\bar{\hat{Ж}} - z)^{-1} \supset (\hat{Ж} - z)^{-1}$ and $\mathfrak{D}[(\hat{Ж} - z)^{-1}] = \tau c$ hence it follows the statement.

We remark that from $(\hat{Ж} - z)^{-1} \in \mathfrak{B}(\tau c)$ it follows that $\hat{Ж}$ is a closed operator.

Let us now define the following operators:

$$\mathfrak{N}_1(z) = \mathcal{P} Ж(\mathbb{I}_{\tau c} - \mathcal{P}) \frac{1}{\hat{Ж} - z} (\mathbb{I}_{\tau c} - \mathcal{P}), \qquad \operatorname{Im} z \neq 0, \tag{3.3}$$

$$\mathfrak{N}_2(z) = (\mathbb{I}_{\tau c} - \mathcal{P}) \frac{1}{\hat{Ж} - z} (\mathbb{I}_{\tau c} - \mathcal{P})\, Ж\, \mathcal{P}, \qquad \operatorname{Im} z \neq 0, \tag{3.4}$$

$$\mathfrak{M}(z) = \mathcal{P}\, Ж\, (\mathbb{I}_{\tau c} - \mathcal{P}) \frac{1}{\hat{Ж} - z} (\mathbb{I}_{\tau c} - \mathcal{P}) Ж \mathcal{P} - \mathcal{P} Ж \mathcal{P}, \qquad \operatorname{Im} z \neq 0. \tag{3.5}$$

Now we will prove the following properties for such operators:

1. $\mathfrak{N}_1(z) \in \mathfrak{B}(\tau c)$.

Indeed,

$$(\mathbb{I}_{\tau c} - \mathcal{P}) \frac{1}{\hat{Ж} - z} (\mathbb{I}_{\tau c} - \mathcal{P}) \in \mathfrak{B}(\tau c), \qquad Ж \in \mathfrak{C}(\tau c),$$

hence $Ж(\mathbb{I}_{\tau c} - \mathcal{P})(\hat{Ж} - z)^{-1}(\mathbb{I}_{\tau c} - \mathcal{P})$ is closed, but since its domain is $\tau c$, then it is a bounded operator as well.

Therefore also $\mathfrak{N}_1(z) \in \mathfrak{B}(\tau c)$.

2. $\mathfrak{M}(z)$ is a closable operator.

In fact, $\forall z\, \mathfrak{D}(\mathfrak{M}(z)) = \mathfrak{D}(Ж\mathcal{P})$ then $\mathfrak{M}(z)$ is densely defined.

So $\mathfrak{M}^+(z)$ exists and $[\mathfrak{M}^+(\bar{z})]_{\tau c} \supset \mathfrak{M}(z)$, by (2.11), (3.1) and I.1.

Hence by Lemma A3 (see Appendix) $\mathfrak{M}(z)$ is closable $\forall z$, $\operatorname{Im} z \neq 0$.

3. $\forall z$, $\operatorname{Im} z \neq 0$, we have

$$\mathfrak{N}_1(z) - \mathcal{P} = [z + \overline{\mathfrak{M}(z)}]\, \mathcal{P} \frac{1}{Ж - z}. \tag{3.6}$$

*Proof:* If $C \in \mathfrak{D}(Ж')$, one gets

$$(Ж - z)C = (\hat{Ж} - z)C + (\mathcal{P} Ж + Ж\mathcal{P} - \mathcal{P}Ж\mathcal{P})C.$$

Hence

$$\frac{1}{\hat{Ж} - z}(Ж - z)C = C + \frac{1}{\hat{Ж} - z}(\mathcal{P} Ж + Ж\mathcal{P} - \mathcal{P}Ж\mathcal{P})C$$

and setting $C' = (Ж - z)C$ one has

$$\frac{1}{\hat{Ж} - z} C' = \frac{1}{Ж - z} C' + \frac{1}{\hat{Ж} - z} \times (\mathcal{P}Ж + Ж\mathcal{P} - \mathcal{P}Ж\mathcal{P}) \frac{1}{Ж - z} C' \tag{3.7}$$

Since each of the three terms of the last equation belongs to $\mathfrak{D}(\hat{Ж})$, we have

$$\begin{aligned}&\mathcal{P}Ж(\mathbb{I}_{\tau c} - \mathcal{P}) \frac{1}{\hat{Ж} - z} C' \\ &= \mathcal{P}Ж(\mathbb{I}_{\tau c} - \mathcal{P}) \frac{1}{Ж - z} C' \\ &\quad + \mathcal{P}Ж(\mathbb{I}_{\tau c} - \mathcal{P}) \frac{1}{\hat{Ж} - z} (\mathbb{I}_{\tau c} - \mathcal{P}) Ж\mathcal{P} \frac{1}{Ж - z} C' \\ &\quad + \mathcal{P}Ж(\mathbb{I}_{\tau c} - \mathcal{P}) \frac{1}{\hat{Ж} - z}\, \mathcal{P}Ж \frac{1}{Ж - z} C'\end{aligned}$$

and since

$$\mathcal{P}Ж(\mathbb{I}_{\tau c} - \mathcal{P}) \frac{1}{Ж - z} = \mathcal{P} + z\mathcal{P} \frac{1}{Ж - z} - \mathcal{P}Ж\mathcal{P} \frac{1}{Ж - z}$$

and

$$\mathcal{P}Ж(\mathbb{1}_{\tau c} - \mathcal{P})\frac{1}{\hat{Ж} - z}\mathcal{P}Ж\frac{1}{Ж - z}C' = 0$$

because $(\mathbb{1}_{\tau c} - \mathcal{P})$ commutes with $1/(\hat{Ж} - z)$, one has at last

$$[\mathfrak{N}_1(z) - \mathcal{P}]C' = [z + \mathfrak{M}(z)]\mathcal{P}\frac{1}{Ж - z}C', \qquad \forall\, C' \in \mathcal{R}(Ж' - z).$$

By I. 4, $\overline{\mathcal{R}(Ж' - z)} = \tau c$, and by $\mathfrak{N}_1(z) \in \mathcal{B}(\tau c)$ it follows

$$\mathfrak{N}_1(z) - \mathcal{P} = \overline{[z + \mathfrak{M}(z)]\mathcal{P}[1/(Ж - z)]}.$$

Noting that $\overline{\mathcal{C}\cdot\mathcal{E}} \subset \overline{\mathcal{C}}\,\mathcal{E}$ if $\mathcal{C}$ is a closable and $\mathcal{E}$ a bounded operator, it follows the statement.

4. If $\operatorname{Im} z \neq 0$ one has

$$\mathcal{R}(\overline{\mathfrak{M}}(z) + z) = \tau c. \tag{3.8}$$

In fact, using (3.6) one easily gets

$$\mathcal{P} = -(z + \overline{\mathfrak{M}}(z))\mathcal{P}\frac{1}{Ж - z}\mathcal{P}.$$

Hence

$$(\mathbb{1}_{\tau c} - \mathcal{P}) = (z + \overline{\mathfrak{M}}(z)\,(\mathbb{1}_{\tau c} - \mathcal{P})\,1/z$$

and

$$\mathbb{1}_{\tau c} = \mathbb{1}_{\tau c} - \mathcal{P} + \mathcal{P}$$
$$= (z + \overline{\mathfrak{M}}(z))\mathcal{P}\frac{1}{Ж - z}\mathcal{P} + (z + \overline{\mathfrak{M}}(z))(\mathbb{1}_{\tau c} - \mathcal{P})\frac{1}{z},$$

then it follows (3.8).

5. (a) $\operatorname{Im} z > 0, \operatorname{Im}\lambda > 0 \Rightarrow \exists\ (\lambda + \overline{\mathfrak{M}}(z))^{-1}$
(b) $\operatorname{Im} z < 0, \operatorname{Im}\lambda < 0 \Rightarrow \exists\ (\lambda + \overline{\mathfrak{M}}(z))^{-1}$ (3.9)

*Proof:*
(a) Let us consider the operator $i\mathfrak{M}(z)$. $i\mathfrak{M}(z)$ is a densely defined operator in $\tau c$, hence, by Lemma A. 4 (see Appendix), it can be considered as a densely defined operator in $\sigma c$.

$i\mathfrak{M}(z)$ is a dissipative operator in $\sigma c$; in fact, by I. 1, (2.11), and (3.1) we have $\forall\, C \in \mathfrak{D}(\mathfrak{M}(z)) = \mathfrak{D}(Ж\mathcal{P})$

$$\operatorname{Re}\langle C, i\mathfrak{M}(z)C\rangle = -\operatorname{Im} z\left\|\frac{1}{\hat{Ж} - z}(\mathbb{1}_{\tau c} - \mathcal{P})Ж\mathcal{P}C\right\|_2^2 \leqslant 0.$$

$i\mathfrak{M}(z)$ as operator defined in $\sigma c$ is closable, and its closure $i\overline{\mathfrak{M}}^*(z)$ is a dissipative operator (Ref. 15, Lemma 3.3, 3.4).[16]

Then it exists $(\mu + i\overline{\mathfrak{M}}^*(z))^{-1}\ \forall\mu,\ \operatorname{Re}\mu \leq 0$ (Ref. 15, Lemma 3.1), and $(\mu + i\overline{\mathfrak{M}}^*(z))^{-1} \supset (\mu + i\overline{\mathfrak{M}}(z))^{-1}$ (Lemma A5 in Appendix)—where $\supset$ is the ordinary relation of containment between operators defined in $\sigma c$.

Thus $\operatorname{nul}[\mu + i\overline{\mathfrak{M}}(z)] = 0$

From this equation it follows that $(\mu + i\overline{\mathfrak{M}}(z))^{-1}$ exists and that generally it is an unbounded operator defined in $\tau c$.

We could easily proove that $(\mu + i\overline{\mathfrak{M}}(z))^{-1}$ as operator defined in $\sigma c$ is on the contrary bounded.

(b) The proof is exactly the same as the former one, where in this case $-i\mathfrak{M}(z)$ is a dissipative operator in $\sigma c$.

6. $\forall z$, $\operatorname{Im} z \neq 0$ one has

$$\frac{1}{z + \overline{\mathfrak{M}}(z)} \in \mathcal{B}(\tau c) \tag{3.10}$$

The proof follows immediately from (3.8), (3.9) and from the fact that $\{1/[z + \overline{\mathfrak{M}}(z)]\} \in \mathcal{C}(\tau c)$.

7. $\forall z$, $\operatorname{Im} z \neq 0$ one has

$$\frac{1}{z + \overline{\mathfrak{M}}(z)}(\mathfrak{N}_1(z) - \mathcal{P}) = \mathcal{P}\frac{1}{Ж - z}. \tag{3.11}$$

The equation follows immediately from (3.6) and (3.8). As a consequence of (3.11) we have

$$\frac{1}{z + \overline{\mathfrak{M}}(z)}\mathcal{P} = -\mathcal{P}\frac{1}{Ж - z}\mathcal{P} \tag{3.11'}$$

$$\frac{1}{z + \overline{\mathfrak{M}}(z)}\mathfrak{N}_1(z) = \mathcal{P}\frac{1}{Ж - z}(\mathbb{1}_{\tau c} - \mathcal{P}) \tag{3.11''}$$

8. $\mathfrak{D}(\overline{\mathfrak{M}}(z))$ is independent of $z$ and

$\mathfrak{D}(\overline{\mathfrak{M}}(z)) =: \{C \in \tau c \mid \mathcal{P}C \in \mathcal{P}\mathfrak{D}(Ж)\}$.

*Proof:* Let $\Sigma =: \{C \in \tau c \mid \mathcal{P}C \in \mathcal{P}\mathfrak{D}(Ж)\}$.

By (3.11') and $(\mathfrak{M}(z)\mathcal{P} - \mathcal{P}\mathfrak{M}(z)) \subset 0$ it follows that

$$\mathcal{P}\frac{1}{z + \overline{\mathfrak{M}}(z)} = -\mathcal{P}\frac{1}{Ж - z}\mathcal{P}.$$

Hence

$$C \in \mathcal{R}\left(\frac{1}{z + \overline{\mathfrak{M}}(z)}\right)$$
$$\Rightarrow \mathcal{P}C \in \mathcal{R}\left(-\mathcal{P}\frac{1}{Ж - z}\mathcal{P}\right) \subset \mathcal{R}\left(\mathcal{P}\frac{1}{Ж - z}\right) = \mathcal{P}\mathfrak{D}(Ж).$$

Thus

$$C \in \mathfrak{D}(\overline{\mathfrak{M}}(z)) \Rightarrow C \in \Sigma.$$

On the other hand, by (3.11) it follows that $C \in \Sigma \Rightarrow \mathcal{P}C \in \mathfrak{D}(\overline{\mathfrak{M}}(z))$. Besides, $\forall A \in \tau c\ (\mathbb{1}_{\tau c} - \mathcal{P})A \in \mathfrak{D}(\overline{\mathfrak{M}}(z))$ and $\overline{\mathfrak{M}}(z)\,(\mathbb{1}_{\tau c} - \mathcal{P})A = 0$. Then $C \in \Sigma \Rightarrow C \in \mathfrak{D}(\overline{\mathfrak{M}}(z))$ and $\overline{\mathfrak{M}}(z)C = \overline{\mathfrak{M}}(z)\mathcal{P}C$.

Therefore, $\mathfrak{D}(\overline{\mathfrak{M}}(z)) = \Sigma$.

9. $\mathfrak{N}_2(z)$ is a closable operator, and $\mathfrak{D}(\overline{\mathfrak{N}_2(z)})$ is independent of $z$.

In fact, $[\mathfrak{N}_1^+(\bar{z})]_{\tau c} \supset \mathfrak{N}_2(z)$ and $[\mathfrak{N}_1^+(\bar{z})]_{\tau c}$ is closed by Lemma A2 in Appendix.

The second statement is an immediate consequence of the equation

$$\mathfrak{N}_2(z) = \left[\mathbb{1}_{\tau c} + (z - z')\,(\mathbb{1}_{\tau c} - \mathcal{P})\frac{1}{\hat{Ж} - z}\right]\mathfrak{N}_2(z').$$

10. $\forall z$, $\operatorname{Im} z \neq 0$ one has

$$(\mathbb{1}_{\tau c} - \mathcal{P})\frac{1}{Ж - z} = (\mathbb{1}_{\tau c} - \mathcal{P})\frac{1}{\hat{Ж} - z}(\mathbb{1}_{\tau c} - \mathcal{P})$$
$$- \overline{\mathfrak{N}_2(z)}\frac{1}{z + \overline{\mathfrak{M}}(z)}[\mathfrak{N}_1(z)\,(\mathbb{1}_{\tau c} - \mathcal{P}) - \mathcal{P}]. \tag{3.12}$$

*Proof:* By (3.7) one gets $\forall\, C^1 \in \mathcal{R}(Ж' - z)$

$$(\mathbb{1}_{\tau c} - \mathcal{P})\frac{1}{Ж - z}C^1 = (\mathbb{1}_{\tau c} - \mathcal{P})\frac{1}{\hat{Ж} - z}C^1$$
$$- [(\mathbb{1}_{\tau c} - \mathcal{P})\frac{1}{\hat{Ж} - z}(\mathbb{1}_{\tau c} - \mathcal{P})Ж\mathcal{P}]\frac{1}{Ж - z}C^1$$

and taking into account (3.4) and (3.11), we have

$$(\mathbb{1}_{\tau c} - \mathcal{P}) \frac{1}{Ж - z} C^1 = (\mathbb{1}_{\tau c} - \mathcal{P}) \frac{1}{\hat{Ж} - z} C^1 - \mathfrak{N}_2(z) \frac{1}{z + \overline{\mathfrak{M}}(z)} \times [\mathfrak{N}_1(z)(\mathbb{1}_{\tau c} - \mathcal{P}) - \mathcal{P}] C^1.$$

Since $(\mathbb{1}_{\tau c} - \mathcal{P}) \left[\frac{1}{Ж - z} - \frac{1}{\hat{Ж} - z}\right] \in \mathcal{B}(\tau c)$, we have also

$$\overline{\mathfrak{N}_2(z) \frac{1}{z + \overline{\mathfrak{M}}(z)} [\mathfrak{N}_1(z) - \mathcal{P}]} \in \mathcal{B}(\tau c).$$

Hence

$$\overline{\mathfrak{N}_2(z)} \frac{1}{z + \overline{\mathfrak{M}}(z)} [\mathfrak{N}_1(z) - \mathcal{P}] \in \mathcal{B}(\tau c) \tag{3.13}$$

and (3.12) holds.

11. Taking into account the analyticity of the resolvent operator, in the two half-planes $\operatorname{Im} z > 0$ and $\operatorname{Im} z < 0$ we have:

(a) $\mathfrak{N}_1(z)$ is a holomorphic operator-valued function. This is an immediate consequence of the equation

$$\mathfrak{N}_1(z) = \mathfrak{N}_1(z^1) + (z - z^1)\mathfrak{N}_1(z^1) \frac{1}{\hat{Ж} - z} (\mathbb{1}_{\tau c} - \mathcal{P}).$$

(b) $\forall A \in \mathfrak{D}(Ж\mathcal{P}) = \mathfrak{D}(\mathfrak{N}_2(z)) = \mathfrak{D}(\mathfrak{M}(z))$, $\mathfrak{N}_2(z)A$ and $\mathfrak{M}(z)A$ are holomorphic vector-valued functions. The holomorphism of $\mathfrak{M}(z)A$ follows immediately from the equation

$$\mathfrak{M}(z) = \mathfrak{M}(z^1) + (z - z^1)\, \mathfrak{N}_1(z)\, \mathfrak{N}_2(z).$$

(c) The following operator-valued functions are holomorphic:

$$\frac{1}{z + \overline{\mathfrak{M}}(z)} \mathcal{P}, \quad \frac{1}{z + \overline{\mathfrak{M}}(z)} \mathfrak{N}_1(z),$$

$$\overline{\mathfrak{N}_2(z)} \frac{1}{z + \overline{\mathfrak{M}}(z)} \mathfrak{N}_1(z), \quad \overline{\mathfrak{N}_2(z)} \frac{1}{z + \overline{\mathfrak{M}}(z)} \mathcal{P}.$$

By (3.10) and (3.13) it follows that all these operators belongs to $\mathcal{B}(\tau c)$.

The statement then follows from (3.11'), (3.11''), and (3.12).

Now we put

$$\phi(t) =: \mathcal{P}\, W(t) = \mathcal{P}\mathfrak{U}(t) W(0)$$

and

$$\Gamma(t) =: (\mathbb{1}_{\tau c} - \mathcal{P})\, W(t) = (\mathbb{1}_{\tau c} - \mathcal{P})\, \mathfrak{U}(t) W(0),$$

where $W(0)$ is the statistical operator at time $t = 0$.

By (2.7) we have the following equations for $\phi(t)$ and $\Gamma(t)$:

$$\phi(t) = \lim_{\xi \to \infty} \frac{1}{2\pi i} \frac{1}{\xi} \int_0^\xi dw \int_{-w+i\epsilon}^{w+i\epsilon} dz \exp(-izt)\, \mathcal{P} \frac{1}{Ж - z} W(0), \tag{3.14}$$

$$\Gamma(t) = \lim_{\xi \to \infty} \frac{1}{2\pi i} \frac{1}{\xi} \int_0^\xi dw \int_{-w+i\epsilon}^{w+i\epsilon} dz \times \exp(-izt)\, (\mathbb{1}_{\tau c} - \mathcal{P}) \frac{1}{Ж - z} W(0). \tag{3.15}$$

Taking into account Eq. (3.11) and (3.12) we have

$$\phi(t) = \lim_{\xi \to \infty} \frac{1}{2\pi i} \frac{1}{\xi} \int_0^\xi dw \int_{-w+i\epsilon}^{w+i\epsilon} dz \times \exp(-izt) \frac{1}{z + \overline{\mathfrak{M}}(z)} [\mathcal{P} - \mathfrak{N}_1(z)] W(0), \tag{3.16}$$

$$\Gamma(t) = \lim_{\xi \to \infty} \frac{1}{2\pi i} \frac{1}{\xi} \int_0^\xi dw \int_{-w+i\epsilon}^{w+i\epsilon} dz \times \exp(-izt) \Big( (\mathbb{1}_{\tau c} - \mathcal{P}) \frac{1}{Ж - z} (\mathbb{1}_{\tau c} - \mathcal{P}) - \overline{\mathfrak{N}_2(z)} . \times \frac{1}{z + \overline{\mathfrak{M}}(z)} [\mathfrak{N}_1(z)(\mathbb{1}_{\tau c} - \mathcal{P}) - \mathcal{P}] \Big) W(0). \tag{3.17}$$

With the simplifying condition

$$\phi(0) = \mathcal{P} W(0) = W(0), \tag{3.18}$$

one gets the following equations:

$$\phi(t) = \frac{1}{2\pi i} \lim_{\xi \to \infty} \frac{1}{\xi} \int_0^\xi dw \int_{-w+i\epsilon}^{w+i\epsilon} dz \times \exp(-izt) \frac{1}{z + \overline{\mathfrak{M}}(z)} \phi(0), \tag{3.19}$$

$$\Gamma(t) = \frac{1}{2\pi i} \lim_{\xi \to \infty} \frac{1}{\xi} \int_0^\xi dw \int_{-w+i\epsilon}^{w+i\epsilon} dz \times \exp(-izt) \left( -\overline{\mathfrak{N}_2(z)} \frac{1}{z + \overline{\mathfrak{M}}(z)} \right) \phi(0). \tag{3.20}$$

We remark that, by (2.8), if $W(0) \in \mathfrak{D}(Ж)$ one may replace $\lim_{\xi \to \infty} (1/\xi)(1/2\pi i) \int_0^\xi dw \cdots$ with $(1/2\pi i) \lim_{w \to \infty} \cdots$ in all Eqs. (3.14)-(3.20).

Equation (3.19) is a closed time evolution equation for $\phi(t)$.

It is easy to check that $\phi(t)$ obeys the so called *generalized master equation*.

From a more general point of view Eqs. (3.16), (3.17) supply a starting point to develope the theory of independent subdynamics in the formalism of the pair of spaces $(\tau c, \mathbb{B})$.

## ACKNOWLEDGMENTS

We are very much indebted to Professor L. Lanz for many enlightening discussions on the subject of this paper. We wish to thank Professor R. Cirelli for his stimulating interest and Dr. L. A. Lugiato for helpful suggestions.

## APPENDIX: REMARKS ON THE TRACE-CLASS OPERATORS

Let $\mathfrak{H}$ be a Hilbert space. $\tau c$ is the linear space of all linear bounded operators defined on $\mathfrak{H}$ such that $\operatorname{tr}((A^+A)^{1/2}) < \infty$.

Putting $\forall A \in \tau c \|A\|_1 =: \operatorname{tr}((A^+A)^{1/2})$, $\tau c$ becomes a Banach space (Ref. 17, 3, Theorem).

Let $\sigma c$ be the Hilbert space of all Hilbert-Schmidt operators defined on $\mathfrak{H}$.

$\forall A \in \tau c$ we have $A \in \sigma c$ and $\|A\| \leq \|A\|_2 \leq \|A\|_1$ if $\|A\|_2$ is the norm in $\sigma c$ (Ref. 17, 3., Theorem 4).

Let $\mathbb{C}$ be the linear space of all completely continuous operators on $\mathfrak{H}$.

$\mathbb{C}$ is a closed linear manifold of $\mathbb{B} = \mathfrak{B}(\mathfrak{H})$; so it is a Banach space if the bound of an operator stands for its norm.

The following statements are true (Ref. 17, Chap. 4):

(1) If $\mathfrak{H}$ is infinite dimensional, then $\mathbb{C}$ is not the conjugate space of any Banach space.

(2) The dual space of $\mathbb{C}$ is $\tau c$.
Every bounded functional on $\mathbb{C}$ may be represented by

$$\mathcal{L}(B) = \mathrm{tr}\,(A^+B), \qquad B \in \mathbb{C},$$

where $A$ is a suitable "trace-class" operator.

(3) The dual space of $\tau c$ is $\mathbb{B}$.
Every bounded functional on $\tau c$ may be represented by

$$\mathcal{L}(W) = \mathrm{tr}\,(B^+W), \qquad W \in \tau c,$$

where $B$ is a siutable bounded operator of $\mathbb{B}$.

(4) The dual space of $\mathbb{B}$ is $\tau c \oplus \mathbb{C}^\perp$, that is every bounded linear functional $\mathfrak{F}$ on $\mathbb{B}$ may be represented in one and only one way in the form $\mathfrak{F} = \mathfrak{F}_1 + \mathfrak{F}_2$ where $\mathfrak{F}_1 \in \tau c$ and $\mathfrak{F}_2 \in \mathbb{C}^\perp$. Moreover, $\|\mathfrak{F}\| = \|\mathfrak{F}_1\| + \|\mathfrak{F}_2\|$.

Some other remarks on the $\tau c$ space:

(a) if $\mathfrak{H}$ is a separable Hilbert space, $\tau c$ is a separable Banach space.

Indeed, let $\varphi \otimes \overline{\psi}\,(\varphi, \psi \in \mathfrak{H})$ be the operator defined as

$$(\varphi \otimes \overline{\psi})f =: (f, \psi)\varphi, \quad \forall f \in \mathfrak{H}.$$

$\forall \varphi, \psi \in \mathfrak{H}$ we have $\varphi \otimes \overline{\psi} \in \tau c$.

If $\{u_i\}$ is a countable set of vectors which is dense in $\mathfrak{H}$, one can easily see that the set $\{u_i \otimes \bar{u}_k\}$ is dense in the set of all $\varphi \otimes \overline{\psi}$ in the sense of the topology of $\tau c$. The statement follows immediately from the fact that the linear manifold spanned by the set of all $\varphi \otimes \overline{\psi}$ is dense in $\tau c$.

(b) $\tau c$ is the linear manifold in $\mathbb{B}$ spanned by the statistical operators.

In fact, every "trace-class" operator is a linear combination of two self-adjoint operators which belong to $\tau c$, everyone of which in turn is the difference of two positive selfadjoint "trace-class" operators.

Finally, for every self-adjoint, positive, trace-class operator $D$ we can obviously write $D = \mathrm{tr}\,D\,(D/\mathrm{tr}\,D)$ and $D/\mathrm{tr}\,D$ is a statistical operator.

We now prove some useful lemmas:

*Lemma* A1:

(a) $S \lhd \mathbb{B} \Rightarrow S \cap \tau c \lhd \tau c$,

(b) $S \lhd \mathbb{B} \Rightarrow S \cap \sigma c \lhd \sigma c$,

(c) $S \lhd \sigma c \Rightarrow S \cap \tau c \lhd \tau c$

We will prove (a) only, the other ones can be proved in the same way:

$$\{A_n\} \subset S \cap \tau c, A \in \tau c \|A_n - A\|_1 \to 0$$
$$\Rightarrow A \in \mathbb{B}, \|A_n - A\| \leq \|A_n - A\|_1 \to 0$$
$$\Rightarrow A \in S \Rightarrow A \in S \cap \tau c.$$

Let $\mathfrak{F}$ a linear operator defined in $\mathbb{B}$. We will denote with $\mathfrak{F}_{\tau c}$ the restriction of $\mathfrak{F}$ to the linear manifold $\mathfrak{D}(\mathfrak{F}_{\tau c}) =: \{A | A \in \mathfrak{D}(\mathfrak{F}) \cap \tau c, \mathfrak{F}A \in \tau c\}$. We note that $\mathfrak{D}(\mathfrak{F}_{\tau c})$ may become $\{0\}$ and that $\mathfrak{F} \in \mathfrak{B}(\mathbb{B})$ does *not* imply $\mathfrak{F}_{\tau c} \in \mathfrak{B}(\tau c)$.

*Lemma* A2

$$\mathfrak{F} \in \mathfrak{C}(\mathbb{B}) \Rightarrow \mathfrak{F}_{\tau c} \in \mathfrak{C}(\tau c).$$

*Proof:* Let $\{A_n\} \subset \mathfrak{D}(\mathfrak{F}_{\tau c})$ and $A, B \in \tau c$.
Assume besides that $\|A_n - A\|_1 \to 0$ and $\|\mathfrak{F}_{\tau c}A_n - B\|_1 \to 0$. Then $\|A_n - A\| \leq \|A_n - A\|_1 \to 0$ and $\|\mathfrak{F}_{\tau c}A_n - B\| \leq \|\mathfrak{F}_{\tau c}A_n - B\|_1 \to 0$; therefore, $A \in \mathfrak{D}(\mathfrak{F}) \cap \tau c$ and $\mathfrak{F}A = B$ thus $A \in \mathfrak{D}(\mathfrak{F}_{\tau c})$ and $\mathfrak{F}_{\tau c}A = B$.

*Lemma* A3: Assume $\mathfrak{F}$ is a closable operator in $\mathbb{B}$, $\mathfrak{F}_{\tau c}$ is a closable operator as well.

*Proof:*

$$\mathfrak{F} \subset \bar{\mathfrak{F}} \in \mathfrak{C}(\mathbb{B}) \Rightarrow \mathfrak{F}_{\tau c} \subset (\bar{\mathfrak{F}})_{\tau c} \in \mathfrak{C}(\tau c),$$

thus $\mathfrak{F}_{\tau c}$ can be extended to a closed operator, then it is closable.

*Lemma* A4: Let $R$ be a linear manifold dense in $\tau c$, then $R$ is also a linear manifold dense in $\sigma c$.

*Proof:* Let us denote the closure of $R$ in $\sigma c$ with $\bar{R}^*$

$\forall A \in \tau c$, $\exists \{A_n\} \subset R$ such that $\|A_n - A\|_1 \to 0$.
But $\|A_n - A\|_2 \leq \|A_n - A\|_1 \to 0$ thus $A \in \bar{R}^*$ and then $\bar{R}^* \supset \tau c$ and $\bar{R}^* \supset \overline{\tau c}^* = \sigma c$ that is $\bar{R}^* = \sigma c$.

*Corollary:* Let $\mathfrak{F}$ be an operator densely defined in $\tau c$, then $\mathfrak{F}$ is also densely defined in $\sigma c$.

*Lemma* A5: Let $\mathfrak{F}$ be a closable operator defined in $\tau c$ and $\bar{\mathfrak{F}}$ be its closure. If $\mathfrak{F}$ considered as an operator defined in $\sigma c$ is closable, then it holds that $\bar{\mathfrak{F}}^* \supset \bar{\mathfrak{F}}$, where $\bar{\mathfrak{F}}^*$ is the closure of $\mathfrak{F}$ considered as operator defined in $\sigma c$ and the relation $\supset$ is the usual one between operators defined in $\sigma c$.

*Proof:* Let us suppose that $\binom{A}{B} \in \mathfrak{G}(\bar{\mathfrak{F}})$.
Then

$$\exists A_n \in \mathfrak{D}(\mathfrak{F})$$

with

$$\left\| \binom{A_n}{\mathfrak{F}A_n} - \binom{A}{B} \right\|_1 \to 0,$$

but

$$\left\| \binom{A_n}{\mathfrak{F}A_n} - \binom{A}{B} \right\|_2 \leq \left\| \binom{A_n}{\mathfrak{F}A_n} - \binom{A}{B} \right\|_1 \to 0;$$

thus

$$\binom{A}{B} \in \mathfrak{G}(\bar{\mathfrak{F}}^*),$$

and then

$$\mathfrak{G}(\bar{\mathfrak{F}}^*) \supset \mathfrak{G}(\bar{\mathfrak{F}}).$$

*P. Cotta-Ramusino gratefully acknowledges a fellowship of the Consiglio Nazionale delle Ricerche Comitato Nazionale per le Scienze Matematiche

[1]I. Prigogine, C. George, F. Henin, Physica **45**, 418 (1969).
[2]L. Lanz, L. A. Lugiato, G. Ramella, Physica **54**, 94 (1971).
[3]C. George, I. Prigogine, L. Rosenfeld, Det. Kong. Dansk Vidensk. Sels. **38**, 12 (1972).
[4]G. Ludwig, *Makroskopische Systeme und Quantenmechanik,* Notes in Mathematical Physics, Universität Marburg NMP 5, April 1972.
[5]L. Lanz, L. A. Lugiato, G. Ramella, A. Sabbadini, "The embedding of unstable non-relativistic particles into Galilean Quantum Field Theories," Int. J. Theor. Phys. (to be published).
[6]L. Lanz and L. A. Lugiato, Physica **44**, 532 (1969).
[7]A. P. Grecos, Physica **51**, 50 (1971).
[8]A. P. Grecos and I. Prigogine, Proc. Natl. Acad. Sci. USA 69, 6, 1629 (1972).
[9]F. Mayné and I. Prigogine, Physica **63**, 1 (1973).
[10]J. W. Middleton and W. C. Scieve, Physica **63**, 139 (1973).
[11]G. Ludwig, *Deutung des Begriffs "physikalische Theorie" und axiomatische Grundlegung der Hilbertraumstruktur der Quantenmechanik durch Hauptsätze* des Messens (Springer-Verlag, Berlin, 1970), Vol. 4.
[12]I. M. Gel'fand and N. Y. Vilenkin, Les distributions, applications de l'analyse harmonique (Dunod, Paris, 1967), Vol. 4.
[13]E. Hille and R. S. Phillips, *Functional analysis and semigroups* (American Mathematical Society, Providence, R. I., 1957).
[14]We make use of the following statement: let $T$ be a closable and invertible operator defined in a Banach space $X$. $(\bar{T})^{-1}$ exists iff $u_n \in D(T)$, $u_n \to v$, $Tu_n \to 0 \Rightarrow v = 0$. In the proof of the statement $\delta$, we have used also the fact that $\overline{\mathfrak{K} - z} = \bar{\mathfrak{K}} - z$.
[15]G. Lumer and R. S. Phillips, Pacific J. Math. **11**, 679 (1961).
[16]One must pay attention not to mistake $\overline{\mathfrak{M}}^*(z)$, that is the closure of $\mathfrak{M}(z)$ as operator defined in $\sigma c$, for $\overline{\mathfrak{M}}(z)$ that is the closure of $\mathfrak{M}(z)$ as operator defined in $\tau c$.
[17]R. Schatten, *Norm ideals of completely continuous operators* (Springer Verlag, Berlin, 1960).

# On some nonlinear evolution equations in quantum field theory

A. Tesei*

*Equipe de Recherche Associée 215, Université de Paris VI, Paris, France*
(Received 18 June 1973; revised manuscript received 22 August 1973)

The field equations of a fermion field with a scalar self-interaction are studied in a three- and four-dimensional space–time. The Faedo–Galerkin method is used to prove the existence of a sequence of approximated solutions which converges in a suitable topology. In a different topology, the existence and uniqueness of a local solution is proved using the contraction principle.

## INTRODUCTION

In the past few years, very important results have been obtained in constructive Hamiltonian field theory: Following mainly the path proposed in the middle sixties by Wightman,[1] nontrivial quantum fields have been proved to exist, satisfying the Wightman axioms.[2]

Another approach, which consists in studying the field partial differential equations, is possible; indeed, there is no *a priori* reason to prefer the Hamiltonian approach to the one through field equations. However, a very good practical reason for doing so becomes apparent if we think of the well-known Wightman's theorem,[3] according to which in a local quantum field theory with a unique vacuum and positive nontrivial spectrum all bounded operator fields (i.e., functions) are trivial—or, informally stated, "fields must be distributions." As a consequence, in the approach through field equations we are concerned with the problem of defining, and of dealing with, nonlinear functions of distributions, appearing as nonlinear terms in the field equations. This problem was studied by Segal,[4] who obtained remarkable results.

A different way of overcoming this difficulty could consist in developing the following program:

(i) writing down and giving existence and uniqueness theorems for the Cauchy problem relative to field differential equations, where the field is taken to be an operator-valued *function*;

(ii) letting the Cauchy datum converge to a distribution, and proving that the solution correspondingly converges to a distribution, to be taken as the field of physical interest.

This program has been recently put forward by Dell'Antonio.[5] Independently, Salusti and the present author were able to prove a global existence and uniqueness theorem for the field equations of the Thirring and Federbush models.[6]

In the following (as it was the case in Ref. 6) we shall be concerned only with item (i) of the above program, or, in other words, we shall always be dealing with operator-valued *functions;* in particular, the solutions of the field equations we will study are operator-valued functions. We shall speak sometimes of "fields," but it should be clear from the outset that, according to the above remarks, this term is not to be taken in the usual sense in the present context.

Some comments are in order. To our knowledge, no rigorous results are known on the link between the Hamiltonian and the field differential equations approach. We remark that, in item (i) of the above program, the solution of the field equation is not requested to have physical properties (causality, for instance); instead, these properties should be recovered *after* proving the existence of the limiting distribution[7] [item (ii)]. Similarly, the solution is not *a priori* requested to have definite covariance properties; instead, one asks under which conditions the nonlinear semigroup obtained solving the field equations would give rise to the time evolution implemented by a (approximate) Hamiltonian.[7] In a sense, one might expect to obtain via the above program a more general class of solutions than the ones given by the Hamiltonian formalism; these (the only ones of physical interest) could be singled out imposing the physical requirements.

To summarize, it seems to be interesting to study field differential equations (which at the present stage enter in quantum field theory only as identities satisfied *a posteriori* by the fields constructed via the Hamiltonian formalism). The present paper is concerned with item (i) in a particular nontrivial model, with the purpose of testing the feasibility of the above program.

## 1. STATEMENT OF THE PROBLEM AND RESULTS

In the present note we are concerned with the following Cauchy problem in three- and four-dimensional space–time:

$$\begin{aligned} &i\gamma^\mu \partial_\mu \psi + m\psi + g(\bar{\psi}\psi)\psi = \varphi, \\ &\psi(o, \mathbf{x}) = \psi_0(\mathbf{x}). \end{aligned} \tag{1.1}$$

The above problem describes, when $\varphi = 0$, a fermion field with a scalar self-interaction, and seems to be the most straightforward generalization of Thirring's model to a higher number of space–time dimensions. The quantities $\psi(t, \mathbf{x})$, $\varphi(t, \mathbf{x})$ are functions which take values in $B(\mathcal{H})$, the algebra of linear bounded operators on the Hilbert space $\mathcal{H}$; $m, g$ are real constants. As in Ref. 6, no assumption is made about commutation or anticommutation properties, or causality. By adopting a particular representation of the Dirac matrices $\gamma^\mu$, the problem (1.1) can be rewritten in explicit form as in Eqs. (4.2) (four-dimensional case) and (4.12) (three-dimensional case).

In the present case, the earlier adopted semigroup methods[6] turn out to be not useful; thus we develop a modified version of the Faedo–Galerkin method particularly fitted to our problem, or we make use of the Banach fixed point principle. In this way we are able to prove the existence of a (nonunique) sequence of global approximated solutions, converging in a sense to be specified below, or, in a different Banach space, the existence of a unique local solution.

In Sec. 2 we give a short outline of the Faedo–Galerkin method, as well as the Banach fixed point principle. In Sec. 3 we introduce some function spaces and state their principal properties. In Sec. 4 the Faedo–Galerkin method is applied to prove the existence of the above referred sequence of approximate solutions. In Sec. 5 we use Banach's fixed point principle and some semigroup techniques to prove the existence of a local unique solution. Section 6 is devoted to concluding remarks.

## 2. MATHEMATICAL TOOLS

For convenience of the reader we give a short account of the main mathematical techniques which will be used in what follows. We first review briefly the Faedo–Galerkin method, stressing particularly its relevance for the theory of abstract evolution equations. In this respect see Refs. 8, 9.

Let $X$ denote a separable normed space with the basis $\{\chi_j\}$; denoting by $X_n \subset X$ a subspace spanned by the vectors $\chi_1, \chi_2, \ldots, \chi_n$, we can state the Faedo–Galerkin method as follows. Suppose we are interested in the equation $F(x) = 0$, where $F: X \to X'$ is a mapping from $X$ into the dual space $X'$: An approximate solution is found in the form

$$x_n = \sum_{1}^{n}{}_j a_{nj}\chi_j,$$

the coefficients $a_{nj}$ satisfying the system of equations

$$\left\langle F\left(\sum_{1}^{n}{}_j a_{nj}\chi_j\right), \chi_i\right\rangle = 0, \qquad i = 1, \ldots, n. \tag{2.1}$$

In general, $\langle y, x\rangle$ is the value of the linear functional $y \in X'$ on the vector $x \in X$.

The system (2.1) is equivalent to the equation

$$P_n^* F(P_n x) = 0, \tag{2.2}$$

where $P_n$ is a projector from $X$ to $X_n$ and $P_n^*$ is the conjugate operator of $P_n$. The solutions of Eq. (2.2) are called Faedo–Galerkin approximations. In the particular case $F: X \to X$, the Faedo–Galerkin approximations are solutions of the equation $F(x) = 0$ restricted to the subspace $X_n$, i.e.,

$$P_n F(P_n x) = 0. \tag{2.3}$$

This is the case in which we will be interested in the following.

The foregoing frame turns out to be relevant to state existence theorems for partial differential equations. Indeed, if Faedo–Galerkin approximations for such equations exist, we have a sequence of "approximate" solutions: We can ask whether this sequence converges to a vector of $X$, which is an actual solution of the given equation. This point can be investigated in several ways, mainly by compactness or monotonicity methods. Typically, compactness proofs are given in three steps:

(i) existence of a sequence $\{x_n\}$ of Faedo–Galerkin approximations;

(ii) *a priori* estimate, uniform with respect to $n$, showing that the Faedo–Galerkin approximations belong to a bounded set of $X$;

(iii) using compactness theorems to extract a converging subsequence $\{x_{n_K}\}$, whose limiting point is proved to be a solution of the given equation.

We remark that compactness proofs say nothing about the uniqueness of the solution. In Sec. 4, our treatment of the problem will follow the foregoing three steps of the compactness methods; as for (i), however, there are some nontrivial differences which will be discussed in Sec. 6.

An independent approach is given by the well-known contraction principle. Let $X$ be a complete metric space, $\Omega$ a closed set in $X$, $P$ an operation transforming $\Omega$ in itself. Let, moreover, $P$ be a contraction mapping, i.e.,

$$\rho(P(x), P(x')) \le \alpha\rho(x, x'), \qquad x, x' \in \Omega, \tag{2.4}$$

where $\rho(\cdot,\cdot)$ denotes the metric in $X$, and $\alpha < 1$ is a real constant independent of $x$ and $x'$. Then the following theorem holds.

*Theorem 2.1:* If $P$ is a contraction mapping, a unique solution $x^*$ of the equation $x = P(x)$ exists in $\Omega$.

In other words, if $P$ is a contraction, a unique fixed point $x^*$ of $P$ exists. The contraction principle is often referred to also as Banach fixed point principle; other fixed point principles exist.[10]

## 3. FUNCTION SPACES

Let $B(\mathcal{H})$ denote the algebra of all bounded operator on the Hilbert space $\mathcal{H}$. We consider the linear space $C_0^\infty(\mathbb{R}^n; B(\mathcal{H}))$, i.e., the linear space of all infinitely differentiable functions with compact support in $\mathbb{R}^n$ which take values in $B(\mathcal{H})$. For any $f \in C_0^\infty(\mathbb{R}^n; B(\mathcal{H}))$, a real-valued function $|f|_*$ is defined by:

$$|f|_* = \left\| \int_{\mathbb{R}_n} dx f^+(x) f(x)\right\|^{1/2}, \tag{3.1}$$

$f^+(x)$ denoting the adjoint of the operator $f(x)$ and $\|\cdot\|$ the norm in $B(\mathcal{H})$.

We want to prove that $|f|_*$ is a norm in $C_0^\infty(\mathbb{R}; B(\mathcal{H}))$; for this purpose the following lemma is useful.

*Lemma 3.1:* Let $X$ be a complex linear space. Let us consider an application $[\cdot,\cdot]: X \times X \to B(\mathcal{H})$ which satisfies the following conditions:

(1) $[x,y] = [y,x]^+$;

(2) $[x,x] \geq 0$;

(3) $[x,x] = 0$ if and only if $x = 0$;

(4) $[\lambda x + \mu y, z] = \lambda[x,z] + \mu[y,z], \quad \lambda, \mu \in \mathbb{C}$.

Then the following inequality holds:

$$\tfrac{1}{2}\|[x,y] + [y,x]\| \le \|[x,x]\|^{1/2}\|[y,y]\|^{1/2}. \tag{3.2}$$

*Proof:* For any $\lambda, \mu \in \mathbb{R}$, $\alpha \in \mathcal{H}$, we have $\lambda^2([x,x]\alpha,\alpha) + \lambda\mu(\{[x,y] + [y,x]\}\alpha,\alpha) + \mu^2([y,y]\alpha,\alpha) \geq 0$, $(\cdot,\cdot)$ denoting the scalar product in $\mathcal{H}$. It follows that

$$\tfrac{1}{2}|(\{[x,y] + [y,x]\}\alpha,\alpha)| \le ([x,x]\alpha,\alpha)^{1/2}([y,y]\alpha,\alpha)^{1/2}$$

whence the result, $\frac{1}{2}\{[x,y] + [y,x]\}$ being a bounded self-adjoint operator by hypothesis. ∎

*Proposition 3.1:* $C_0^\infty(\mathbb{R}^n; B(\mathcal{H}))$ is a normed linear space with the norm $|f|_*$.

*Proof:* To prove the triangular inequality, it suffices to remark that the application of $C_0^\infty(\mathbb{R}^n; B(\mathcal{H})) \times C_0^\infty(\mathbb{R}^n; B(\mathcal{H}))$ in $B(\mathcal{H})$ defined by

$$[f,g] = \int_{\mathbb{R}_n} dx f^+(x) g(x) \tag{3.3}$$

satisfies the hypothesis of Lemma 3.1. The other properties of the norm are trivially verified. ■

*Definition:* The completion of $C_0^\infty(\mathbb{R}^n; B(\mathcal{H}))$ in the norm (3.1) is a Banach space which will be denoted as $L_*^2(\mathbb{R}^n; B(\mathcal{H}))$.

Obviously, $L_*^2(\mathbb{R}^n; B(\mathcal{H}))$ contains the Hilbert space $L^2(\mathbb{R}^n) \otimes \mathbb{1}$ of complex-valued functions in $\mathbb{R}^n$. It is worth noticing that, besides $L_*^2(\mathbb{R}^n; B(\mathcal{H}))$, another "$L^2$ space" for operator-valued functions in $\mathbb{R}^n$ can be defined as follows:

$$L^2(\mathbb{R}^n; B(\mathcal{H})) = \{f: \mathbb{R}^n \to B(\mathcal{H});\ \int_{\mathbb{R}_n} dx\, \|f(x)\|^2 < \infty\}.$$

This Banach space might seem the most natural extension of the usual definition to operator-valued functions. However, the space $L_*^2(\mathbb{R}^n; B(\mathcal{H}))$ turns out to be more convenient, mainly because the application (3.3)—which has the formal properties of an "operator-valued scalar product"—allows us to derive easily *a priori* estimates.

*Proposition 3.2:* $L^2(\mathbb{R}^n; B(\mathcal{H})) \subseteq L_*^2(\mathbb{R}^n; B(\mathcal{H}))$.

We recall the definition of the following Banach spaces:

$$L^p(\mathbb{R}^n; B(\mathcal{H})) = \{f: \mathbb{R}^n \to B(\mathcal{H});\ \int_{\mathbb{R}_n} dx\, \|f(x)\|^p < \infty\},$$

$$|f|_p^p = \int_{\mathbb{R}_i} dx\, \|f(x)\|^p, \qquad 1 \le p < \infty;$$

$$H^1(\mathbb{R}^n; B(\mathcal{H})) = \left\{f: \mathbb{R}^n \to B(\mathcal{H});\ \sum_1^n{}_k \int_{\mathbb{R}_n} dx \left\| \frac{\partial f(x)}{\partial x^k} \right\|^2 < \infty \right\},$$

$$|f|_{1,2}^2 = \sum_1^n{}_k \int_{\mathbb{R}_n} dx \left\| \frac{\partial f(x)}{\partial x^k} \right\|^2 .$$

For $p = 2$ we have the already mentioned space $L^2(\mathbb{R}^n; B(\mathcal{H}))$. In the same way as $L_*^2(\mathbb{R}^n; B(\mathcal{H}))$, we could define the Banach spaces $L_*^p(\mathbb{R}^n; B(\mathcal{H}))$ $(1 \le p < \infty)$ and $H_*^1(\mathbb{R}^n; B(\mathcal{H}))$.

As already remarked, $L_*^2(\mathbb{R}^n; B(\mathcal{H}))$ contains the Hilbert space $L^2(\mathbb{R}^n) \otimes \mathbb{1}$ of complex-valued functions in $\mathbb{R}^n$. Let us choose in this space a complete orthonormal system, e.g., the Chebyshev–Hermite functions: it is easily seen that linear combinations of these functions with operatorial coefficients are dense in $L_*^2(\mathbb{R}^n; B(\mathcal{H}))$.

Indeed, infinitely differentiable functions with compact support are dense in $L_*^2(\mathbb{R}^n; B(\mathcal{H}))$ (this follows from the same definition of $L_*^2(\mathbb{R}^n; B(\mathcal{H}))$, and polynomials with operational coefficients are dense in $C_0^\infty(\mathbb{R}^n; B(\mathcal{H}))$ in the uniform norm, thus in $L_*^2$ norm.

A remarkable property of $L_*^2(\mathbb{R}^n; B(\mathcal{H}))$ is that the Plancherel theorem holds true. For any $f \in L^1(\mathbb{R}^n; B(\mathcal{H}))$ let us define the Fourier transform $\hat{f}$ in the usual way:

$$\hat{f}(\xi) = (2\pi)^{-n/2} \int_{\mathbb{R}_n} dx\, e^{-i\langle \xi, x\rangle} f(x), \tag{3.4}$$

where

$$\langle \xi, x\rangle = \sum_1^n{}_j\, \xi_j x_j .$$

Then we have the following:

*Theorem 3.1:* With each $f \in L_*^2(\mathbb{R}^n; B(\mathcal{H}))$ a function $\hat{f} \in L_*^2(\mathbb{R}^n; B(\mathcal{H}))$ can be associated in such a way that:

(i) $\quad |f|_* = |\hat{f}|_*;$ (3.5)

(ii) if $f \in L^1(\mathbb{R}^n; B(\mathcal{H})) \cap L_*^2(\mathbb{R}^n; B(\mathcal{H}))$, $\hat{f}$ is the Fourier transform (3.4) of $f$.

*Proof:* The proof will follow in a standard way as soon as we prove Eq. (3.5) for the dense set given by linear combinations with operational coefficients of Chebyshev–Hermite functions, i.e.,

$$f(x) = \Sigma_i T_i u_i(x), \qquad T_i \in B(\mathcal{H}).$$

For this purpose it suffices to notice that:

$$\begin{aligned} \int_{\mathbb{R}_n} dx\, f^*(x) f(x) &= \sum_1^n{}_{i,j}\, T_i^* T_j \int_{\mathbb{R}_n} dx\, u_i(x) u_j(x) \\ &= \sum_1^n{}_{i,j}\, T_i^* T_j \int_{\mathbb{R}_n} d\chi\, \hat{u}_i(-\chi) \hat{u}_j(\chi) \\ &= \int_{\mathbb{R}_n} d\chi\, \hat{f}^*(\chi) \hat{f}(\chi), \end{aligned}$$

where use has been made of the Parseval's relation. Taking the operator norm of both sides we get the result. ■

We are now in position to introduce another Banach space which turns out to be useful. Let us consider the following linear space:

$$\hat{S}(\mathbb{R}^n; B(\mathcal{H})) = \{f: \mathbb{R}^n \to B(\mathcal{H}); \hat{f} \in L^1(\mathbb{R}^n; B(\mathcal{H}))\}; \tag{3.6}$$

$\hat{S}(\mathbb{R}^n; B(\mathcal{H}))$ is a normed linear space with respect to the norm:

$$|f|_{1,\Lambda} = |\hat{f}|_1 . \tag{3.7}$$

*Definition:* The completion of $\hat{S}(\mathbb{R}^n; B(\mathcal{H}))$ in the norm (3.7) is a Banach space, which will be denoted as $\hat{L}^1(\mathbb{R}^n; B(\mathcal{H}))$.

*Proposition 3.3:* $\hat{L}^1(\mathbb{R}^n; B(\mathcal{H}))$ is an algebra with respect to the usual product.

*Proof:* $L^1(\mathbb{R}^n; B(\mathcal{H}))$ is a (noncommutative) algebra with respect to the convolution, so we have

$$|fg|_{1,\Lambda} = |\widehat{fg}|_1 = |\hat{f} * \hat{g}|_1 \le |\hat{f}|_1 |\hat{g}|_1 = |f|_{1,\Lambda} |g|_{1,\Lambda}. \ ■$$

In the following we will be dealing with spinors in a four-dimensional (three-dimensional) space–time, which can be viewed as functions from $[o, T] \times \mathbb{R}^3$ ($[o, T] \times \mathbb{R}^2$) in $\oplus_{i=1}^4 B(\mathcal{H})$ ($\oplus_{i=1}^2 B(\mathcal{H})$). Then it is useful to introduce in a general way the Banach spaces $L_*^2(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H}))$, with norm defined

$$|\psi|_* = \left\| \sum_1^m{}_k \int_{\mathbb{R}^n} dx\, \psi_k^*(x) \psi_k(x) \right\|^{1/2},$$

and $\hat{L}^1(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H}))$, whose norm is

$$|\psi|_{1,\Lambda} = \sum_1^m{}_k |\hat{\psi}_k|_1 .$$

Similar definitions hold for the spaces $L^p(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H}))$ and $H^1(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H}))$. We will also be dealing with the Banach spaces $C([o, T]; X)$, $L^p([o, T]; X)$ $(1 \le p \le \infty)$, where $X$ is a Banach space and the norms are defined in the usual way.

## 4. COMPACTNESS METHOD

We first study the problem (1.1) in a four-dimensional space–time. The three-dimensional case, which can be treated in a similar way, will be considered afterwards.

After multiplication by $-i\gamma^0$, the Cauchy problem (1.1) can be rewritten as

$$\begin{cases}\dfrac{\partial\psi}{\partial t}+\sum_1^3{}_k\,\alpha^k\,\dfrac{\partial\psi}{\partial x^k}+im\gamma^0\psi-ig(\overline{\psi}\psi)\gamma^0\psi=-i\gamma^0\varphi,\\ \psi(o,\mathbf{x})=\psi_0(\mathbf{x}).\end{cases}\tag{4.1}$$

Here $\psi, \varphi$ are four component spinors, i.e., functions from $[o, T]\times\mathbb{R}^3$ in $\oplus_{i=1}^4 B(\mathcal{H})$, and $\alpha^k=\alpha^{k+}$ denote the three Dirac space matrices. By adopting a particular representation of the $\gamma^\mu$ matrices,[11] the problem can be written in the following explicit form:

$$\begin{aligned}&\frac{\partial\psi_1}{\partial t}+\frac{\partial\psi_4}{\partial x}-i\frac{\partial\psi_4}{\partial y}+\frac{\partial\psi_3}{\partial z}\\&\qquad+im\psi_1-ig(\psi_1^+\psi_1+\psi_2^+\psi_2-\psi_3^+\psi_3-\psi_4^+\psi_4)\psi_1=-i\varphi_1,\\&\frac{\partial\psi_2}{\partial t}+\frac{\partial\psi_3}{\partial x}+i\frac{\partial\psi_3}{\partial y}-\frac{\partial\psi_4}{\partial z}\\&\qquad+im\psi_2-ig(\psi_1^+\psi_1+\psi_2^+\psi_2-\psi_3^+\psi_3-\psi_4^+\psi_4)\psi_2=-i\varphi_2,\\&\frac{\partial\psi_3}{\partial t}+\frac{\partial\psi_2}{\partial x}-i\frac{\partial\psi_2}{\partial y}+\frac{\partial\psi_1}{\partial z}\\&\qquad-im\psi_3+ig(\psi_1^+\psi_1+\psi_2^+\psi_2-\psi_3^+\psi_3-\psi_4^+\psi_4)\psi_3=i\varphi_3,\\&\frac{\partial\psi_4}{\partial t}+\frac{\partial\psi_1}{\partial x}+i\frac{\partial\psi_1}{\partial y}-\frac{\partial\psi_2}{\partial z}\\&\qquad-im\psi_4+ig(\psi_1^+\psi_1+\psi_2^+\psi_2-\psi_3^+\psi_3-\psi_4^+\psi_4)\psi_4=i\varphi_4,\\&\begin{pmatrix}\psi_1\\\psi_2\\\psi_3\\\psi_4\end{pmatrix}(o,x)=\begin{pmatrix}\psi_{01}\\\psi_{02}\\\psi_{03}\\\psi_{04}\end{pmatrix}(\mathbf{x}).\end{aligned}\tag{4.2}$$

This is the Cauchy problem for a system of nonlinear partial differential equations in normal form. We remark that, in deriving the foregoing form of the problem, anticommutation relations for the fields have never been introduced. According to the general discussion of Sec. 2, we need three preliminary steps to prove the main theorem of this section. We assume that the Cauchy data $\psi_{0i}$ $(i=1,\ldots,4)$ belong to $L_*^2(\mathbb{R}^3;\oplus_{i=1}^4 B(\mathcal{H}))$.

(i) As we remarked already, the Hilbert space $L^2(\mathbb{R}^3)$ of complex-valued functions is contained in $L_*^2(\mathbb{R}^3; B(\mathcal{H}))$.

Let us take a basis $\{\chi_j\}$ in $L^2(\mathbb{R}^3)$, whose elements belong to $H^1(\mathbb{R}^3)\cap L^4(\mathbb{R}^3)$ (e.g., we can take the Chebyshev–Hermite functions), and look for approximated solutions of the problem (4.2) of the following form:

$$\begin{aligned}&\psi_i^{(m)}(t,\mathbf{x})=\sum_1^m{}_l\,T_{lm}^{(i)}(t)\chi_l(\mathbf{x}),\\&T_{lm}^{(i)}:[o,T]\to B(\mathcal{H}),\qquad i=1,\ldots,4.\end{aligned}\tag{4.3}$$

In other words, we take linear combinations of complex-valued functions with time dependent operator-valued coefficients $T_{lm}^{(i)}(\cdot)$; these coefficients are to be determined from the conditions:

$$\begin{aligned}&\left[\chi_j,\frac{\partial\psi_i^{(m)}}{\partial t}\right]+\sum_1^3{}_k\left[\chi_j,\left(\alpha^K\frac{\partial\psi^{(m)}}{\partial x^k}\right)_i\right]+im[\chi_j,(\gamma^0\psi^{(m)})_i]\\&\quad-ig[\chi_j,(\psi_1^{(m)+}\psi_1^{(m)}+\psi_2^{(m)+}\psi_2^{(m)}-\psi_3^{(m)+}\psi_3^{(m)}\\&\quad-\psi_4^{(m)+}\psi_4^{(m)})(\gamma^0\psi^{(m)})_i]=-i[\chi_j,(\gamma^0\varphi)_i],\\&\quad i=1,\ldots,4,\quad j=1,\ldots,m,\end{aligned}\tag{4.4}$$

where $[\cdot,\cdot]$ is the application (3.3) (for $n=3$). Then we have, for any $m$, a system of $4m$ ordinary first-order differential equations in normal form for the $4m$ unknowns $T_{lm}^{(i)}(\cdot)$, to be studied with the Cauchy datum:

$$\begin{aligned}&\psi_i^{(m)}(o,\mathbf{x})=\psi_{0i}^{(m)}(\mathbf{x}),\\&\psi_{0i}^{(m)}(\mathbf{x})=\sum_1^m{}_l\,V_{lm}^{(i)}X_l(\mathbf{x}),\\&\left|\sum_1^m{}_l\,V_{lm}^{(i)}\chi_l-\psi_{0i}\right|_*\to0\quad\text{when } m\to\infty.\end{aligned}\tag{4.5}$$

It is easily seen that the nonlinear terms in Eqs. (4.4) are locally Lipschitz continuous in the operatorial norm, thus in an interval $[o,t_m)$ a unique solution of the problem (4.4), (4.5) exists.

(ii) In the following we will assume that $\varphi$ belongs to $L^1([o,T]; L_*^2(\mathbb{R}^3;\oplus_{i=1}^4 B(\mathcal{H}))$. Let us consider in the system (4.4) the equation of indices $(i,j)$: Multiplying by $T_{jm}^{(i)+}(t)$ from left and summing over both indices, we obtain in compact form

$$\left[\psi^{(m)},\frac{\partial\psi^{(m)}}{\partial t}\right]+\sum_1^3{}_k\left[\psi^{(m)},\alpha^k\frac{\partial\psi^{(m)}}{\partial x^k}\right]+im[\overline{\psi}^{(m)},\psi^{(m)}]\\-ig[\overline{\psi}^{(m)},(\overline{\psi}^{(m)}\psi^{(m)}\psi^{(m)}]=-i[\overline{\psi}^{(m)},\varphi],$$

where we have posed $\overline{\psi}=\psi^+\gamma^0$ and

$$[\varphi,\psi]=\sum_1^4{}_l[\varphi_l,\psi_l].$$

Taking the adjoint equation and summing, we obtain simply:

$$\frac{\partial}{\partial t}[\psi^{(m)},\psi^{(m)}]=-i\{[\overline{\psi}^{(m)},\varphi]-[\varphi,\overline{\psi}^{(m)}]\}.\tag{4.6}$$

In the particular case $\varphi=0$, (4.6) expresses the conservation of the total charge of the field. Taking the operatorial norm and using (3.2), we get

$$\frac{\partial}{\partial t}\|[\psi^{(m)},\psi^{(m)}]\|\le\left\|\frac{\partial}{\partial t}[\psi^{(m)},\psi^{(m)}]\right\|\le2\|[\psi^{(m)},\psi^{(m)}]\|^{1/2}\\\times\|[\varphi,\varphi]\|^{1/2},$$

or

$$\frac{\partial}{\partial t}|\psi^{(m)}(t)|_*\le|\varphi(t)|_*.$$

It follows that

$$|\psi^{(m)}(t)|_*\le C+\int_0^T ds|\varphi(s)|_*.\tag{4.7}$$

The foregoing inequality shows that $t_m=T$; moreover, when letting $m$ go to infinity, $\psi^{(m)}$ belongs to a bounded set in $L^\infty([o,T]; L_*^2(\mathbb{R}^3;\oplus_{i=1}^4 B(\mathcal{H})))$.

(iii) Let $\{e_k\}$ be a complete orthonormal system in the Hilbert space $\mathcal{H}$. We consider the set of operator-valued functions in $\mathbb{R}^n$ defined as follows:

$$Q\left(\mathbb{R}^n;\bigoplus_{i=1}^m B(\mathcal{H})\right)\\=\left\{f:\mathbb{R}^n\to\bigoplus_{i=1}^m B(\mathcal{H});\sum_1^\infty{}_k([f,f]e_k,e_k)^{1/2}<\infty\right\},\tag{4.8}$$

with the obvious position:

$$[f,g]=\sum_1^m{}_i[f_i,g_i]=\sum_1^m{}_i\int_{\mathbb{R}^n}dx f_i^+(x)g_i(x).$$

*Lemma 4.1:* $Q(\mathbb{R}^n;\oplus_{i=1}^m B(\mathcal{H}))$ is a normed linear space with the norm

$$|f|_{X_2}=\sum_1^\infty{}_k([f,f]e_k,e_k)^{1/2}.\tag{4.9}$$

*Proof:* It suffices to prove the triangular inequality. For any $f, g \in Q(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H}))$ we have

$$([f+g, f+g]e_k, e_k) \le ([f,f]e_k, e_k) + ([g,g]e_k, e_k) + |(\{[f,g]+[g,f]\}e_k, e_k)| \le ([f,f]e_k, e_k) + ([g,g]e_k, e_k) + 2([f,f]e_k, e_k)^{1/2},$$

whence

$$([f+g, f+g]e_k, e_k)^{1/2} \le ([f,f]e_k, e_k)^{1/2} + ([g,g]e_k, e_k)^{1/2}.$$

The result follows immediately. ■

*Definition:* The completion of $Q(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H}))$ in the norm (4.9) is a Banach space which will be denoted as $X_2(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H}))$.

*Theorem 4.1:* $L_*^2(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H}))$ is embedded in the dual space $X_2'(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H}))$ of $X_2(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H}))$.

*Proof:* For any $f \in L_*^2(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H})), g \in X_2(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H}))$, we define the bilinear form

$$\langle f, g\rangle = \tfrac{1}{2} \sum_1^\infty {}_k (\{[f,g] + [g,f]\} e_k, e_k). \tag{4.10}$$

The result follows as soon as we prove that

$$|\langle f, g\rangle| \le |f|_* \; |g|_{X_2}. \tag{4.11}$$

In fact we have

$$|\langle f, g\rangle| \le \tfrac{1}{2}\sum_1^\infty {}_k \; |([f,g]e_k, e_k) + ([g,f]e_k, e_k)| \le \sum_1^\infty {}_k \; ([f,f]e_k, e_k)^{1/2}([g,g]e_k, e_k)^{1/2} \le \|[f,f]\|^{1/2} \sum_1^\infty {}_k \; ([g,g]e_k, e_k)^{1/2} = |f|_* \, |g|_{X_2},$$

where use was made of the self-adjointness and positivity of $[f,f]$.

Then every $f \in L_*^2(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H}))$ defines a continuous linear functional $\tilde{f}$ on $X_2(\mathbb{R}^n; \oplus_{i=1}^m B(\mathcal{H}))$ in the following way:

$$\tilde{f}(g) = \langle f, g\rangle, \quad |f|_{X_2'} \le |f|_* . \qquad ■$$

We can now state the following theorem.

*Theorem 4.3:* Let be given

$$\varphi \in L^1\Big([o, T]; L_*^2\Big(\mathbb{R}^3; \bigoplus_{i=1}^4 B(\mathcal{H})\Big)\Big),$$

$$\psi_0 \in L_*^2\Big(\mathbb{R}^3; \bigoplus_{i=1}^4 B(\mathcal{H})\Big).$$

In $L^\infty([o,T]; X_2'(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{H})))$ there exists a sequence of approximated solutions of the problem (4.2), converging in the weak-* topology.

*Proof:* According to a theorem of Alaoglu,[12] the closed unit sphere of the dual space of a Banach space is compact in the weak-* topology. According to theorem 4.2, $L^\infty([o,T]; L_*^2(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{H})))$ is embedded in $L^\infty([o,T]; X_2'(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{H})))$, i.e., in the dual space of the Banach space $L^1([o,T]; X_2(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{H})))$. On the other hand, inequality (4.9) shows that all the approximated solutions $\psi^{(m)}$ of the problem (4.2) are contained in a bounded set of $L^\infty([o,T]; L_*^2(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{H})))$, thus in a bounded set of $L^\infty([o,T]; X_2'(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{H})))$. Then, by Alaoglu's theorem, in $L^\infty([o,T]; X_2'(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{H})))$ a subsequence $\psi^{(m_k)}$ of approximated solutions exists, converging in the weak-* topology. ■

Similar results hold when studying the Cauchy problem (1.1) in a three-dimensional space–time. In this case it is possible to choose as a representation for the $\gamma^\mu$ ($\mu = 1, 2, 3$) the Pauli matrices,[11] so that the problem can be rewritten in the following way:

$$\begin{aligned}
&\frac{\partial\psi_1}{\partial t} - i\frac{\partial\psi_2}{\partial x} + \frac{\partial\psi_2}{\partial y} + im\psi_1 - ig(\psi_1^*\psi_1 - \psi_2^*\psi_2)\psi_1 = -i\varphi_1,\\
&\frac{\partial\psi_2}{\partial t} - i\frac{\partial\psi_1}{\partial x} - \frac{\partial\psi_1}{\partial y} - im\psi_2 + ig(\psi_1^*\psi_1 - \psi_2^*\psi_2)\psi_2 = i\varphi_2,\\
&\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}(o, \mathbf{x}) = \begin{pmatrix}\psi_{01}\\ \psi_{02}\end{pmatrix}(\mathbf{x}).
\end{aligned} \tag{4.12}$$

Again, no anticommutation relations have been assumed.

It is easily seen that the Faedo–Galerkin method can be applied to the problem (4.12) in the same way as for the four-dimensional case. Then we state without proof the following theorem, analogous to Theorem 4.2.

*Theorem 4.3:* Let be given

$$\varphi \in L^1\Big([o, T]; L_*^2\Big(\mathbb{R}^2; \bigoplus_{i=1}^2 B(\mathcal{H})\Big)\Big),$$

$$\psi_0 \in L_*^2\Big(\mathbb{R}^2; \bigoplus_{i=1}^2 B(\mathcal{H})\Big).$$

In $L^\infty([o,T]; X_2'(\mathbb{R}^2; \oplus_{i=1}^2 B(\mathcal{H})))$ there exists a sequence of approximated solutions of the problem (4.12), converging in the weak-* topology.

## 5. LOCAL SOLUTION

In this section we prove the existence and uniqueness of a local solution of the Cauchy problems (1.1).

For this purpose we remark that the problem [for instance written as in Eq. (4.2)] is of the form

$$\frac{du}{dt} + Lu + Tu = f, \qquad u(0) = u_0, \tag{5.1}$$

where $u, f$ are functions defined on $[o, T]$ taking values in a Banach space $X$, $L$ is a linear operator in $X$, $T$ a nonlinear operator in the same space, and $u_0$ belongs to $X$.

Let us consider the linear operator $L$ in $\hat{L}^1(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{H}))$:

$$L = \begin{pmatrix}
im & o & \frac{\partial}{\partial z} & \frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\\
o & im & \frac{\partial}{\partial x} + i\frac{\partial}{\partial y} & -\frac{\partial}{\partial z}\\
\frac{\partial}{\partial z} & \frac{\partial}{\partial x} - i\frac{\partial}{\partial y} & -im & o\\
\frac{\partial}{\partial x} + i\frac{\partial}{\partial y} & -\frac{\partial}{\partial z} & o & -im
\end{pmatrix} \tag{5.2}$$

*Lemma 5.1:* (i) $L$ is a closed operator with dense domain; (ii) the resolvent $(\lambda I - L)^{-1}, \lambda \in \mathbb{R}_+$, exists such that

$$\|(\lambda I - L)^{-1}\| \leq 3/\lambda \tag{5.3}$$

$\|\cdot\|$ denoting the operatorial norm in $\hat{L}^1(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{K}))$.

*Proof:* The proof of (i) is trivial. As for (ii) we have to study the following system of partial differential equations:

$$\begin{aligned}
&(\lambda - im)\psi_1 - \frac{\partial\psi_3}{\partial z} - \left(\frac{\partial\psi_4}{\partial x} - i\,\frac{\partial\psi_4}{\partial y}\right) = \varphi_1,\\
&(\lambda - im)\psi_2 - \left(\frac{\partial\psi_3}{\partial x} + i\,\frac{\partial\psi_3}{\partial y}\right) + \frac{\partial\psi_4}{\partial z} = \varphi_2,\\
&-\frac{\partial\psi_1}{\partial z} - \left(\frac{\partial\psi_2}{\partial x} - i\,\frac{\partial\psi_2}{\partial y}\right) + (\lambda + im)\psi_3 = \varphi_3,\\
&-\left(\frac{\partial\psi_1}{\partial x} + i\,\frac{\partial\psi_1}{\partial y}\right) + \frac{\partial\psi_2}{\partial z} + (\lambda + im)\psi_4 = \varphi_4,
\end{aligned} \tag{5.4}$$

where $\psi \in D_L$, $\varphi \in \hat{L}^1(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{K}))$ and $\lambda \in \mathbb{R}_+$. Taking the Fourier transform of both sides, we get

$$\begin{aligned}
\hat{\psi}_1 &= [\lambda^2 + \omega^2(\xi)]^{-1}\{\lambda\hat{\varphi}_1 + i\xi_3\hat{\varphi}_3 + (i\xi_1 + \xi_2)\hat{\varphi}_4\},\\
\hat{\psi}_2 &= [\lambda^2 + \omega^2(\xi)]^{-1}\{\lambda\hat{\varphi}_2 + (i\xi_1 - \xi_2)\hat{\varphi}_3 - i\xi_3\hat{\varphi}_4\},\\
\hat{\psi}_3 &= [\lambda^2 + \omega^2(\xi)]^{-1}\{i\xi_3\hat{\varphi}_1 + (i\xi_1 + \xi_2)\hat{\varphi}_2 + \lambda\hat{\varphi}_3\},\\
\hat{\psi}_4 &= [\lambda^2 + \omega^2(\xi)]^{-1}\{(i\xi_1 - \xi_2)\hat{\varphi}_1 - i\xi_3\hat{\varphi}_2 + \lambda\hat{\varphi}_4\},
\end{aligned} \tag{5.5}$$

where $\omega(\xi) = (m^2 + \xi_1^2 + \xi_2^2 + \xi_3^2)^{1/2}$. This proves the existence of a unique solution of the system (5.4); moreover, from (5.5) we easily obtain

$$|\psi|_{1,\Lambda} \leq (3/\lambda)\,|\varphi|_{1,\Lambda}\,. \tag{5.6}$$

This completes the proof. ■

*Lemma 5.2:* The nonlinear operator $T$ in $\hat{L}^1(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{K}))$,

$$T\begin{pmatrix}\psi_1\\ \psi_2\\ \psi_3\\ \psi_4\end{pmatrix} = ig\begin{pmatrix} -[\psi_1^+\psi_1 + \psi_2^+\psi_2 - \psi_3^+\psi_3 - \psi_4^+\psi_4]\psi_1\\ -[\psi_1^+\psi_1 + \psi_2^+\psi_2 - \psi_3^+\psi_3 - \psi_4^+\psi_4]\psi_2\\ [\psi_1^+\psi_1 + \psi_2^+\psi_2 - \psi_3^+\psi_3 - \psi_4^+\psi_4]\psi_3\\ [\psi_1^+\psi_1 + \psi_2^+\psi_2 - \psi_3^+\psi_3 - \psi_4^+\psi_4]\psi_4 \end{pmatrix}, \tag{5.7}$$

is locally Lipschitz continuous.

*Proof:* According to Proposition 3.3, $L^1(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{K}))$ is an algebra with respect to the usual product, so that $T$ is well defined. We have to prove that

$$\begin{aligned}
\sum_{1}^{4}{}_k\,&|[\psi_1^+\psi_1 + \psi_2^+\psi_2 - \psi_3^+\psi_3 - \psi_4^+\psi_4]\psi_k\\
&- [\varphi_1^+\varphi_1 + \varphi_2^+\varphi_2 - \varphi_3^+\varphi_3 - \varphi_4^+\varphi_4]\varphi_k|_{1,\Lambda}\\
&\leq M(r)\sum_{1}^{4}{}_k\,|\psi_k - \varphi_k|_{1,\Lambda}
\end{aligned}$$

when $|\psi|_{1,\Lambda} < r$, $|\varphi|_{1,\Lambda} < r$. The proof is trivial and will not be given. ■

Let us consider the linear operator $L'$ in $C([o, T]; \hat{L}^1(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{K})))$ defined as follows:

$$\begin{aligned}
D_{L'} = \Big\{ &u \in C\Big([o, T]; \hat{L}^1\Big(\mathbb{R}^3; \bigoplus_{i=1}^4 B(\mathcal{K})\Big)\Big);\ u(t) \in D_L,\\
&t \to Lu(t) \in C\Big([o, T]; \hat{L}^1\Big(\mathbb{R}^3; \bigoplus_{i=1}^4 B(\mathcal{K})\Big)\Big)\Big\},\\
&(L'u)(t) = Lu(t), \quad u \in D_{L'},
\end{aligned}$$

the linear operator $L$ being defined in (5.2). Let us, moreover, introduce the nonlinear operator $T'$ in $C([o, T]; \hat{L}^1(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{K})))$:

$$(T'u)(t) = Tu(t),$$

the operator $T$ being defined in (5.7).

*Remark:* The results established in Lemmas 5.1, 5.2 for the operators $L$ and $T$ still hold true for the operators $L'$ and $T'$. The existence of a unique local mild solution[13] is then proved by the following theorem.

*Theorem 5.1:* In $C([o, T]; \hat{L}^1(\mathbb{R}^3; \oplus_{i=1}^4 B(\mathcal{K})))$ there exists a unique local mild solution of the problem (5.1), the operators $L$ and $T$ being defined in (5.2) and (5.7), respectively.

*Proof:* The only significant point in the proof is to remark that, according to Lemmas 5.1, 5.2 and the above remark, the nonlinear operator $T'(\lambda I - L')^{-1}$, $\lambda \in \mathbb{R}_4$, is a local contraction for sufficiently large $\lambda$, i.e.,

$$\sup_{t\in[o,T]} |T(\lambda I - L)^{-1}u(t) - T(\lambda I - L)^{-1}v(t)|_{1,\Lambda} \leq \sup_{t\in[o,T]} |u(t) - v(t)|_{1,\Lambda}$$

for sufficiently large $\lambda$, and $|u(t)|_{1,\Lambda}, |v(t)|_{1,\Lambda} \leq K$. Then the proof is given by a simple application of the Banach fixed point principle. ■

## 6. CONCLUSIONS

It has been shown that a nonunique converging sequence of global approximated solutions of the Cauchy problem (1.1) exists in a three- and four-dimensional space–time. Independently, a unique local solution of the same problem has been proved to exist in a more restricted Banach space. In doing so, the semigroup methods we applied in the two-dimensional case turned out to be not of use, so that we have been led to apply to our problem other mathematical tools: In particular the Faedo–Galerkin method, which has the advantage of being independent of the number of space dimensions.

Some conclusions can be drawn. The application of a general fixed point principle only gives a correspondingly weak result, i.e., the local existence of the solution. In our opinion, a more direct and useful approach to quantum field equations can be obtained combining general differential equations methods with known results on the operator algebra. In the present case, giving a rigorous meaning to a priori estimates on the solution led us to introduce the space $L_*^2$. In this space it turned out to be possible to repeat the steps of the classical Faedo–Galerkin method; nontrivial differences with respect to the classical case were the use of the "operator-valued scalar product"—which is not the usual duality mapping—as well as the choice of a dense uncountable set in $L_*^2$. On the other hand, we could not prove that the limiting point of the sequence of approximated solutions is an actual solution of our problem, mainly because no Banach space is known in which $L_*^2$ has compact injection.[8]

These considerations suggest that $L_*^2$ is a very natural space to work with in dealing with differential equations for operator-valued functions; moreover, a systematic analysis of its properties (or else of the properties of all the $*$-spaces introduced in Sec. 3) should be of relevance for further developments of the present approach.


## ACKNOWLEDGMENTS

We thank Professor G. Da Prato for his continuous interest in this work, and Professor G. F. Dell'Antonio for very helpful discussions.


*On leave from Istituto per le Applicazioni del Calcolo, C. N. R., Roma, Italy.


[1]A. S. Wightman, "Introduction to some aspects of the relativistic dynamics of quantized fields," in *1964 Cargèse Lectures* (Gordon & Breach, New York, 1965).

[2]See, in particular: J. Glimm and A. Jaffe, Phys. Rev. **176**, 1945 (1968); J. Glimm and A. Jaffe, Ann. Math. **91**, 362 (1970); J. Glimm and A. Jaffe, Acta Math. **17**, 203 (1970); J. Glimm and A. Jaffe, J. Math. Phys. **13**, 1568 (1972). A review as of 1971 is in *Mathematics of Contemporary Physics,* edited by R. Streater (Academic, New York, 1972).

[3]A. S. Wightman, Ann. Inst. H. Poincaré **1**, 403 (1964)

[4]I. E. Segal, J. Funct. Anal. **4**, 404 (1969); **6**, 29 (1970)

[5]G. F. Dell'Antonio, "A model field theory: the Thirring model," in 1973 Schladming Winter School (see in particular Sec. 5).

[6]E. Salusti and A. Tesei, Nuovo Cimento A **2**, 122 (1971). For a related problem, see P. de Mottoni and A. Tesei, Nuovo Cimento A **15**, 536 (1973).

[7]For some results, although rather preliminary, see P. de Mottoni and E. Salusti, Ann. Inst. H. Poincaré A **15**, 363 (1971); "Einstein's causality as a consequence of canonical commutation relations," Ann. Inst. H. Poincaré A **18**, 241 (1973); P. de Mottoni and A. Tesei, Lett. Nuovo Cimento **5**, 341 (1972); P. de Mottoni, Ann. Inst. H. Poincaré A **16**, 253 (1972).

[8]J. L. Lions, *Quelques méthodes de résolution des problèmes aux limites non linéaires* (Dunod, Paris, 1969).

[9]M. M. Vainberg, Sov. Math. Dokl. **12**, 525 (1971).

[10]L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces* (Pergamon, New York, 1964).

[11]See, for instance N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience, New York, 1959).

[12]N. Dunford and J. T. Schwartz, *Linear Operators- Part I* (Interscience, New York, 1958).

[13]F. E. Browder, Ann. Math. **80**, 485 (1964).

# Equilibrium states for the infinite ideal Bose gas

**R. P. Moya**[*†]

*Department of Physics, Queen Mary College, University of London, London E1 4NS, England*
(Received 15 June 1973)

We construct a set of equilibrium states for the ideal Bose gas at temperature $\beta^{-1}$ and chemical potential $\mu$. Our choice for the equilibrium states is based on the Kubo–Martin–Schwinger conditions. In particular, we define $\mathcal{E}_{\beta,\mu}$ to be the set of states which satisfy the KMS conditions suitably defined. In the condensed phase a certain subset $\Delta_{\beta,\mu}$ of $\mathcal{E}_{\beta,\mu}$ correspond to states whose mean local densities are not uniformly bounded with respect to the volume. We, therefore, propose that $\mathcal{E}_{\beta,\mu}\setminus\Delta_{\beta,\mu} \equiv \tilde{\mathcal{E}}_{\beta,\mu}$ corresponds to the set of equilibrium states. Then it is shown that $\tilde{\mathcal{E}}_{\beta,\mu}$ contains, as special cases, the equilibrium states obtained via thermodynamical limit arguments by Araki–Woods and Lewis–Pulé. The extremal elements of the convex set $\mathcal{E}_{\beta,\mu}$ are obtained explicitly.

## 1. INTRODUCTION

It is often useful in statistical mechanics to study the properties of infinite systems so as to describe such phenomena as phase transitions in a mathematically sharp manner.[1,2] Of particular importance is the problem of specifying the equilibrium states of infinite systems. One possibility would be to construct the Gibbs state for a finite volume and then consider the limit in which the volume tends to infinity while the intensive variables remain finite. In this way it can be shown,[3] subject to certain assumptions, that time evolution can be realized as a strongly continuous group of automorphisms of the $C^*$ algebra $\mathfrak{A}$ of quasilocal observables and that the limit state $\phi$ satisfies the K.M.S. conditions. It has been shown,[4] however, that some systems, including the free Bose gas, do not satisfy those assumptions. In particular the automorphism property does not hold. Dubin and Sewell[4] have proposed a weaker set of assumptions which do not necessarily imply that time evolution can be realized as a strongly continuous group of automorphisms of $\mathfrak{A}$ but do enable them to recover the principal results of Haag, Hugenholtz, and Winnink.[3] It follows from their assumptions that time evolution can however be realized as a strongly continuous group of automorphisms of the von Neumann algebra $\pi_\phi(\mathfrak{A})''$ where $(\Omega_\phi, \pi_\phi(\cdot), \mathfrak{H}_\phi)$ is the Gelfand–Naimark–Segal (G.N.S.) triple associated with $\phi$. It follows also that $\tilde{\phi}$ the canonical extension of $\phi$ to $\pi_\phi(\mathfrak{A})''$ satisfies the K.M.S. conditions.

An alternative to the thermodynamical limit approach would be to have a global principle for determining the equilibrium states of the infinite system without ever having to resort to such a limit. A global approach would have certain advantages, namely that problems arising from a possible choice of boundary conditions do not occur.[5a] It has been suggested[6,7,8] that the K.M.S. conditions might be used as such a principle.

Thus motivated, we shall investigate a set of states $\mathcal{E}_{\beta,\mu}$ for the free Bose gas such that for $\phi \in \mathcal{E}_{\beta,\mu}$, $\tilde{\phi}$ the extension of $\phi$ to $\pi_\phi(\mathfrak{A})''$ satisfies the K.M.S. conditions with respect to the free time evolution suitably defined.[9] In Sec. 2, we shall introduce our mathematical notation and shall give the definition of our $C^*$ algebra of quasilocal observables for the free Bose gas. In Sec. 3, we define the set $\mathcal{E}_{\beta,\mu}$ and make use of the techniques of Rocca, Sirugue and Testard[10] to obtain $\mathcal{E}_{\beta,\mu}$ explicitly. In Sec. 4, we discuss the mean local densities associated with the elements of $\mathcal{E}_{\beta,\mu}$ and then we obtain as a specific example the states obtained by Araki and Woods,[5b] and Lewis and Pulé.[11] Section 5 contains our conclusions.

## 2. MATHEMATICAL DEFINITIONS

Let $\Gamma$ denote the set of all bounded open subsets $\Lambda$ of $R^3$, where $R$ denotes the real line. We denote by $\mathfrak{D}$ the L. Schwartz space of infinitely differentiable functions with compact supports in $R^3$, and by $\mathcal{S}$ the L. Schwartz space of infinitely differentiable functions of fast decrease on $R^3$. $\mathfrak{D}(\Lambda)$ denotes those $\mathfrak{D}$ class functions with support in $\Lambda$, $\Lambda \in \Gamma$. By $\mathfrak{z}$ $(\equiv \hat{\mathfrak{D}})$ we mean those $\mathcal{S}$ class functions whose Fourier transforms belong to $\mathfrak{D}$.

If $\mathcal{K}$ is a complex pre-Hilbert space with inner product $\langle\cdot,\cdot\rangle$, then a representation of the C.C.R. over $\mathcal{K}$ is defined to be a map

$$h \mapsto W(h)$$

of $\mathcal{K}$ into the unitary operators $\mathfrak{U}(\mathfrak{H})$ on some Hilbert space $\mathfrak{H}$, satisfying

$$W(h_1)W(h_2) = W(h_1 + h_2)\exp(i/2)\,\mathrm{Im}\langle h_1, h_2\rangle,$$
$$\forall h_1, h_2 \in \mathcal{K},$$

the Weyl relations, and for which the mapping

$$\lambda \mapsto W(\lambda h)$$

of $R$ into $\mathfrak{U}(\mathfrak{H})$ is strongly continuous. The representation is said to be cyclic if there exists a unit vector $\Omega$ in $\mathfrak{H}$ such that the linear span of $\{W(h)\Omega; h \in \mathcal{K}\}$, is dense in $\mathfrak{H}$. An important example is the Fock representation which we shall denote by $(W_F, \mathfrak{H}_F\Omega_F)$ with $\mathcal{K}$ taken to be $\mathcal{L}^2(R^3)$. We define $\mu: \mathcal{K} \to \mathbb{C}$ by the formula

$$\mu(h) = (\Omega, W(h)\Omega) \qquad \forall h \in \mathcal{K};$$

then $\mu$ is called a generating functional for a cyclic representation of the C.C.R. over $\mathcal{K}$. For further details of the representation theory of the C.C.R. that we will use we refer the reader to Refs. 12, 13, 14. The $C^*$ algebra $\mathfrak{A}$ of quasilocal observables for the free Bose gas is taken to be the $C^*$ inductive limit[14] of the local von Neumann algebras $\mathfrak{A}(\Lambda)$, where the $\mathfrak{A}(\Lambda)$'s are defined by

$$\mathfrak{A}(\Lambda) = \{W_F(h): h \in \mathfrak{D}(\Lambda)\}''.$$

## 3. THE SET $\mathcal{E}_{\beta,\mu}$

Define $\mathcal{E}_{\beta,\mu}$ to be the set of states $\omega$ of $\mathfrak{A}$ satisfying the following four conditions with $\beta, \mu \in R$, $\beta > 0$, $\mu \leq 0$, and $(\Omega_\omega, \pi_\omega(\cdot), \mathfrak{H}_\omega)$ the G.N.S. triple associated with $\omega$.

1. The mean number of particles in each bounded region $\Lambda$ is finite for the state $\omega$ and thus[15] $\omega$ is locally normal.

2. If $\mu_\omega(h) = \omega(W_F(h)) \equiv (\Omega_\omega, \pi_\omega(W_F(h))\Omega_\omega)$, $\forall h \in \mathfrak{D}$ then $\mu_\omega$ has a unique extension $\tilde{\mu}_\omega$ to a generating functional of a cyclic representation of the C.C.R. over $\mathcal{S}$. This is a purely technical assumption.[16] We denote the corresponding Weyl operators on $\mathfrak{H}_\omega$ by $W_\omega(\cdot)$.

3. Define $\alpha_t$ by $t \in R, \alpha_t$: $W_\omega(h) \to W_\omega(T_t h), \forall h \in \mathcal{S}$ where

$$\widehat{T_t h}(p) = \exp i(p^2/2 - \mu)t\hat{h}(p), \quad \forall h \in \mathcal{S}, \quad t \in R, \quad \mu \leq 0$$

and $\hat{}$ denotes Fourier transformation. $\alpha_t$ then characterizes the free time evolution of the Weyl operators $W_\omega$. Then our third condition is that $\omega$ is stationary with respect to this free time evolution, i.e.,

$$\tilde{\mu}_\omega(T_t h) = \tilde{\mu}_\omega(h), \quad \forall h \in \mathcal{S}.$$

If we define $\mathfrak{M}_\omega(\mathcal{S})$ to be the linear span of $\{W_\omega(h): h \in \mathcal{S}\}$, then by using 1 and 2 it may easily be shown that $\pi_\omega(\mathfrak{A})'' = \mathfrak{M}_\omega(\mathcal{S})''$ and that the time translational invariance of $\tilde{\mu}_\omega$ enables us to extend $\alpha_t$ uniquely to a strongly continuous group of automorphisms $\tilde{\alpha}_t$ of $\mathfrak{M}_\omega(\mathcal{S})''$ and hence also of $\pi_\omega(\mathfrak{A})''$.

4. $\tilde{\omega}$ the extension of $\omega$ to $\pi_\omega(\mathfrak{A})''$ defined by

$$\omega(\pi_\omega(\mathfrak{A})'') = (\Omega_\omega, \pi_\omega(\mathfrak{A})''\Omega_\omega)$$

satisfies the K.M.S. conditions with respect to $\tilde{\alpha}_t$ at temperature $(\beta)^{-1}$ and chemical potential $\mu$.

For simplicity we shall consider the cases $\mu \neq 0$ and $\mu = 0$ separately. In their paper, Rocca, Sirugue and Testard[10] define a class of quasifree time evolutions which correspond to automorphisms of a certain $C^*$ algebra for a Bose gas. For an arbitrary evolution in this class, they obtain explicitly a certain set of states which satisfy the K.M.S. conditions with respect to the corresponding automorphism. A simple adaptation of that part of their work which deals with generating functionals, enables us to write $\tilde{\mu}_\omega, \omega \in \mathcal{E}_{\beta,\mu}$ explicitly as follows.

*Case I* $\mu \neq 0$: In this case $\tilde{\mu}_\omega$ is determined uniquely and is given by

I. $\tilde{\mu}_\omega(h) = \exp(-\frac{1}{4}\|h\|_2^2 - \frac{1}{2}\langle\hat{h}, \hat{p}_\mu \hat{h}\rangle), \quad \forall h \in \mathcal{S},$

where $\hat{p}_\mu \hat{h}(p) = (\exp[\beta(p^2/2 - \mu)] - 1)^{-1}\hat{h}(p), \quad \forall h \in \mathcal{S}.$

*Case II* $\mu = 0$: This case is complicated by the occurrence of Bose condensation with the consequence that the generating functionals are no longer uniquely determined. The general form for $\tilde{\mu}_\omega, \omega \in \mathcal{E}_{\beta,\mu}$ is given by

II. $\tilde{\mu}_\omega(h) = \int_{\mathcal{L}} \exp[-\frac{1}{2}\overline{S}_\beta(h,h) + i\sigma(h)]dm_\omega(\sigma), \quad \forall h \in \mathfrak{z},$

where the bilinear form $\overline{S}_\beta$ on $\mathfrak{z} \times \mathfrak{z}$ is defined in Ref. 10, p. 130, and is essentially of the same form as in the exponent for Case I, except that the zero momentum part has been projected out. The space $\mathcal{L}$ is the space of all real time translationally invariant linear forms on $\mathfrak{z}; m_\omega(\cdot)$ is a positive normalized measure on $\mathcal{L}$.

It follows easily[7] from the definition of $T_t$ and the time translational invariance of $\sigma \in \mathcal{L}$, that the general form for $\sigma(h), h \in \mathfrak{z}$ is given by

$$\sigma(h) = \mathrm{Re}\left[\int \left(C_0 + \sum_{r=1}^{n} u_r(\mathbf{x})\right) h(\mathbf{x})d^3x\right],$$

where $C_0 \in \mathbb{C}$ and $u_r(\mathbf{x})$ is a harmonic polynomial of order $r, n$ is a finite integer. If $\sigma$ were, in addition, space translationally invariant, we would then have

$$\sigma(h) = \mathrm{Re}(C_0 \int h(\mathbf{x})d^3x).$$

We shall denote the subset of space translationally invariant elements of $\mathcal{L}$ by $\mathcal{L}_S$. Using Lemma 2.3 of Ref. 5 and standard continuity arguments, it can easily be shown that $\tilde{\mu}_\omega$ may be uniquely extended to a generating functional over $\mathcal{S}$.

## 4. MEAN LOCAL DENSITIES

*Case I* $\mu \neq 0$: It has been shown[5,11] that $\tilde{\mu}_\omega$ corresponds to a constant finite mean density $\bar{\rho}$ with $\bar{\rho} < \rho_c$ where $\rho_c$ is the critical density.

*Case II* $\mu = 0$: Firstly, we consider the generating functional

$$\mu(h) = \exp[-\tfrac{1}{2}\overline{S}_\beta(h,h) + i\sigma(h)], \quad \forall h \in \mathfrak{z}, \quad \sigma \in \mathcal{L}$$

corresponding to the measure $m_\omega(\cdot)$ having point support. Then a simple calculation of the mean local density[18] for the region $\Lambda$ shows that it is not uniformly bounded with respect to the volume $V(\Lambda)$ in the case where $\sigma \in \mathcal{L}\setminus\mathcal{L}_S$.

We regard this situation as being pathological from the physical point of view and so restrict our attention to those $\tilde{\mu}_\omega$ with $dm_\omega(\cdot)$ concentrated on $\mathcal{L}_S$.

With $dm_\omega(\cdot)$ concentrated on $\mathcal{L}_S$ it can easily be shown that we may write

$$\tilde{\mu}_\omega(h) = \int_{R\times S^1} \exp[-\tfrac{1}{2}\overline{S}_\beta(h,h) + ir\,\mathrm{Re}(\hat{h}(0)e^{-i\theta})] \times dm_{(r,\theta)} \forall h \in \mathcal{S},$$

where $c_0 = re^{-i\theta}$ (polar decomposition) and $m_{(r,\theta)}$ is a positive normalized measure on $R \times S^1$, where $S^1$ denotes the unit circle.

If we choose

$$dm_{(r,\theta)} = \exp(-r^2/r_0^2)d(r^2)/r_0^2 \,.\, d\theta/2\pi, \quad r_0 \in R,$$

then we arrive at the generating functional obtained by Lewis and Pulé[11] which corresponds to a constant finite mean density $\bar{\rho}$ with $\bar{\rho} \geq \rho_c$. Similarly, a suitable choice of the measure $m_{(r,\theta)}$ leads us to the generating functional obtained by Araki and Woods.[5]

## 5. CONCLUSIONS

It can easily be shown, using standard arguments,[14,15] that for $\mu \leq 0$ the generating functionals $\tilde{\mu}_\omega$ so obtained may be uniquely extended to locally normal states on the $C^*$ algebra $\mathfrak{A}$ and indeed these states satisfy the conditions 1 to 4, thus establishing the consistency of the definition of $\mathcal{E}_{\beta,\mu}$,

Denote by $\Delta_{\beta,\mu}$ the subset of $\mathcal{E}_{\beta,\mu}$, the states of which lack the uniform bound property. We propose that $\mathcal{E}_{\beta,\mu}\setminus\Delta_{\beta,\mu} \equiv \tilde{\mathcal{E}}_{\beta,\mu}$ corresponds to the set of equilibrium states for the free Bose gas. We suggest that for a general quantum statistical mechanical system a global approach based on the K.M.S. conditions should include also an axiom which ensures uniformly bounded mean local densities.


## ACKNOWLEDGMENTS

The author is greatly indebted to Dr. G. L. Sewell for constant help and encouragement. The author would also like to thank the S.R.C. for a research studentship.

* Now at the Mathematics Department, Bedford College, Regents Park, London, N.W.1

† This work forms part of the author's Ph.D. thesis, University of London, 1973.

[1]D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
[2]T. D. Lee and C. N. Yang, Phys. Rev. 87, 404 (1952).
[3]R. Haag, N. M. Hugenholtz, and M Winnink, Commun. Math. Phys. 5, 215 (1967).
[4]D. A. Dubin and G. L. Sewell, J. Math. Phys. 11, 2990 (1970).
[5a]R. L. Dobrushin. J. Funct. Anal. Appl. 2, 44 (1968).
[5b]H. Araki and E. J. Woods, J. Math. Phys. 4, 637 (1963).
[6]O. E. Lanford in *Cargese Lectures in Physics,* edited by D. Kastler (Gordon and Breach, New York, 1970), p. 113.
[7]G. G. Emch, H. J. F. Knops, and E. Verboven, J. Math. Phys. 11, 1655 (1970).
[8]G. L. Sewell, Lectures presented at the 1971 Boulder Summer Institute for Theoretical Physics.
[9]In our scheme time translations correspond to automorphisms of $\pi_\phi(\mathfrak{A})''$. For further justification of this point we refer to the last paragraph of Sec. 2 of G. L. Sewell's "States and Dynamics of Infinitely Extended Physical Systems" (preprint).
[10]F. Rocca, M. Sirugue, and D. Testard, Commun. Math. Phys. 19, 119 (1970).
[11]J. T. Lewis and J. Pulé, "The Equilibrium States of the Free Bose Gas" (preprint).
[12]I. E. Segal, Can. J. Math 13, 1 (1961).
[13]J. M. Chaiken, Commun. Math. Phys. 8, 164 (1968).
[14]G. G. Emch, Algebraic Methods in Statistical Mechanics and Quantum Field Theory (Wiley, New York, 1972).
[15]G. F. Dell'Antonio, S. Doplicher, and D. Ruelle, Commun. Math. Phys. 2, 223 (1966).
[16]For further justification for using the Weyl algebra over $\mathcal{S}$ class functions we refer to G. L Sewell, see Ref. 9.
[17]J. P. Marchand, Distributions: an outline (North-Holland, Amsterdam), pp. 37-39.
[18]O. E. Lanford and D. Robinson, Commun. Math Phys. 24, 193 (1972).

# Structural equations for Killing tensors of order two. I

**I. Hauser and R. J. Malhiot**

*Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616*
(Received 20 August 1973)

For any Killing tensor $K_{\alpha\beta}$ of order two, a system of linear homogeneous first-order differential equations of the form $(F_A)_{;\alpha} = \Sigma_B(\Gamma_{\alpha A B} F_B)$ is derived. $F_1$, $F_2$, ... are the components of the tensors $K_{\alpha\beta}$, $L_{\alpha\beta\gamma} = 2K_{\gamma[\beta;\alpha]}$, and $M_{\alpha\beta\gamma\delta} = (1/2)(L_{\alpha\beta[\gamma;\delta]} + L_{\gamma\delta[\alpha;\beta]})$. The coefficients $\Gamma_{\alpha A B}$ are linear expressions in the Riemann tensor and its covariant derivative. These equations are analogous to those satisfied by a Killing vector $K_\alpha$ and the Killing bivector $\omega_{\alpha\beta} = K_{\beta;\alpha}$, with $L_{\alpha\beta\gamma}$ and $M_{\alpha\beta\gamma\delta}$ playing roles analogous to $\omega_{\alpha\beta}$. The tensor $L_{\alpha\beta\gamma}$ has the symmetries $L_{\alpha\beta\gamma} = -L_{\beta\alpha\gamma}$ and $L_{[\alpha\beta\gamma]} = 0$, and $M_{\alpha\beta\gamma\delta}$ has the symmetries of the Riemann tensor. Several relations similar to those satisfied by covariant derivatives of Killing vectors are derived. Perspectives for further work are briefly discussed with the idea of using the equations to investigate space-times which admit Killing tensors of order two.

## 1. INTRODUCTION

Any symmetric tensor $K_{\alpha\beta}$ which satisfies the condition

$$K_{(\alpha\beta;\gamma)} = 0 \tag{1}$$

is called a *Killing tensor of order two*. $K_{\alpha\beta}$ will be called *redundant* if it is equal to some linear combination with constant coefficients of the metric tensor $g_{\alpha\beta}$ and of terms of the form $A_{(\alpha}B_{\beta)}$, where $A_\alpha$ and $B_\beta$ are Killing vectors.

We recall[1] that $K_{\alpha\beta}$ is a Killing tensor if and only if, for any geodesic motion of a test particle with a world velocity $p^\alpha$, the scalar $K_{\alpha\beta}p^\alpha p^\beta$ is a constant of the motion. The recent renewal of interest[2-8] in Killing tensors was largely inspired by Carter's discovery[9] of a quadratic constant of the motion peculiar to the Kerr metric; i.e., the Kerr metric admits a nonredundant Killing tensor.

The criterion of existence of Killing tensors may lead to other interesting space-times. This conjecture has motivated the authors to set up a new system of equations for investigating Riemannian geometries which admit Killing tensors of the second order. Our formalism, in addition to other differences from those used in the past,[1] is not restricted to an orthonormal tetrad relative to which the tensor has its Jordan canonical form. Instead, we work with an arbitrary natural, orthonormal, or null tetrad, which can be chosen at will to fit the exigencies of a given problem. This is the first of several papers on the subject.

In this paper, we introduce the formalism in terms of a natural tetrad in an arbitrary Riemannian space.[10] The objective is to derive and briefly discuss equations similar to the following well-known differential equations which hold for any Killing vector $K_\alpha$:

$$K_{\beta;\alpha} = \omega_{\alpha\beta} = -\omega_{\beta\alpha}, \tag{2}$$

$$\omega_{\alpha\beta;\gamma} = R_{\alpha\beta\gamma\delta}K^\delta. \tag{3}$$

Equations (2) and (3) may be regarded as a system of linear homogeneous first-order equations in the components $K_{\alpha\beta}$, $\omega_{\alpha\beta}$.

In Sec. 2, equations analogous to the above ones for a Killing vector are derived for a Killing tensor $K_{\alpha\beta}$, with the following two tensors playing roles analogous to that of the bivector $\omega_{\alpha\beta}$:

$$L_{\alpha\beta\gamma} = K_{\beta\gamma;\alpha} - K_{\alpha\gamma;\beta}, \tag{4}$$

$$M_{\alpha\beta\gamma\delta} = \tfrac{1}{2}(L_{\alpha\beta[\gamma;\delta]} + L_{\gamma\delta[\alpha;\beta]}). \tag{5}$$

The result of the derivation in Sec. 2 is a system of linear homogeneous first-order equations in which $K_{\beta\gamma;\alpha}$, $L_{\alpha\beta\gamma;\delta}$, and $M_{\alpha\beta\gamma\delta;\mu}$ are each equal to a linear combination of the components of $K_{\alpha\beta}$, $L_{\alpha\beta\gamma}$, and $M_{\alpha\beta\gamma\delta}$. These differential equations will be called the *structural equations for a Killing tensor of order two*.

Some interesting properties of the tensors $L_{\alpha\beta\gamma}$ and $M_{\alpha\beta\gamma\delta}$ are also derived in Sec. 2. In particular, we show that $M_{\alpha\beta\gamma\delta}$ has the same symmetries as the Riemann tensor[11] and that the covariant derivatives of $K_{\alpha\beta}$ and $L_{\alpha\beta\gamma}$ satisfy relations reminiscent of those satisfied by Killing vectors.

In Sec. 3, we give the flat space solution of our structural equations, discuss Lie derivatives of a Killing tensor with respect to a Killing vector, and briefly discuss perspectives for applications of the structural equations. In the immediate sequel to this paper, we will interpret $L_{\alpha\beta\gamma}$ and $M_{\alpha\beta\gamma\delta}$ geometrically and develop the null tetrad form of the structural equations in space-time.

We now begin our derivations by taking note of the symmetry properties of $L_{\alpha\beta\gamma}$ and $M_{\alpha\beta\gamma\delta}$, after which we turn to the main task of computing the covariant derivatives of $K_{\alpha\beta}$, $L_{\alpha\beta\gamma}$, and $M_{\alpha\beta\gamma\delta}$.

## 2. THE STRUCTURAL EQUATIONS

The symmetries of $L_{\alpha\beta\gamma}$ and $M_{\alpha\beta\gamma\delta}$ follow directly from their definitions and hold for any symmetric tensor $K_{\alpha\beta}$ whether or not it is a Killing tensor.

Equations (4) and (5) imply

$$L_{\alpha\beta\gamma} = L_{[\alpha\beta]\gamma}, \tag{6}$$

$$L_{[\alpha\beta\gamma]} = 0. \tag{7}$$

Therefore, in space-time, $L_{\alpha\beta\gamma}$ has 20 independent components. From Eq. (5),

$$M_{\alpha\beta\gamma\delta} = M_{[\alpha\beta][\gamma\delta]} = M_{\gamma\delta\alpha\beta}. \tag{8}$$

Also, if we use Eqs. (4) and (5) to express $M_{\alpha\beta\gamma\delta}$ in terms of second covariant derivatives of $K_{\alpha\beta}$,

$$M_{\alpha\beta\gamma\delta} = \tfrac{1}{2}(K_{\beta\gamma;(\alpha\delta)} + K_{\alpha\delta;(\beta\gamma)} - K_{\alpha\gamma;(\beta\delta)} - K_{\beta\delta;(\alpha\gamma)}). \tag{9}$$

From this, we obtain the identity

$$M_{\alpha\beta\gamma\delta} + M_{\gamma\alpha\beta\delta} + M_{\beta\gamma\alpha\delta} = 0. \tag{10}$$

Therefore, $M_{\alpha\beta\gamma\delta}$ has the same symmetries as the Riemann tensor.

As regards the covariant derivative of $K_{\alpha\beta}$, the definition (1) of a Killing tensor and the definition (4) of $L_{\alpha\beta\gamma}$ imply

$$K_{\beta\gamma;\alpha} = \tfrac{2}{3} L_{\alpha(\beta\gamma)}. \tag{11}$$

Conversely, Eq. (11) and the conditions $L_{\alpha\beta\gamma} = -L_{\beta\alpha\gamma}$ and $L_{[\alpha\beta\gamma]} = 0$ imply Eqs. (1) and (4). The similarity of Eq. (11) to the Killing vector equation (2) is clear.

We recall that the divergence of any Killing vector is zero. The following analogous relation for Killing tensors is easily derived from Eq. (1):

$$(K_{\alpha}{}^{\beta} + \tfrac{1}{2} K_{\mu}{}^{\mu}\delta_{\alpha}^{\beta})_{;\beta} = 0. \tag{12}$$

From Eqs. (9) and (12), we then obtain the interesting looking equations

$$M_{\alpha\gamma} = M_{\alpha\beta\gamma}{}^{\beta} = -\tfrac{3}{4}(\Delta K_{\alpha\gamma} + K_{\beta}{}^{\beta}{}_{;\alpha\gamma}),$$
$$M = M_{\alpha}{}^{\alpha} = -\tfrac{3}{2}\Delta K_{\alpha}{}^{\alpha}, \tag{13}$$

where $\Delta K_{\alpha\ldots\beta} = (K_{\alpha\ldots\beta})_{;\mu}{}^{\mu}$.

The covariant derivative of $L_{\alpha\beta\gamma}$ is derived by computing $L_{\alpha\beta(\gamma;\delta)}$ and $L_{\alpha\beta[\gamma;\delta]}$ and then summing these tensors. From Eqs. (4) and (1), after an appropriate grouping of terms and use of the Riemann tensor symmetries, we obtain

$$L_{\alpha\beta(\gamma;\delta)} = K_{\beta\gamma;[\alpha\delta]} + K_{\alpha\gamma;[\delta\beta]} + K_{\delta\alpha;[\gamma\beta]}$$
$$+ K_{\beta\delta;[\alpha\gamma]} + K_{\gamma\delta;[\alpha\beta]}$$
$$= -2R_{\alpha\beta\mu(\gamma}K_{\delta)}{}^{\mu} - 2K^{\mu}{}_{[\alpha}R_{\beta](\gamma\delta)\mu}. \tag{14}$$

Equation (4) implies

$$L_{\alpha\beta[\gamma;\delta]} - L_{\gamma\delta[\alpha;\beta]}$$
$$= K_{\beta\gamma;[\alpha\delta]} - K_{\alpha\gamma;[\beta\delta]} + K_{\alpha\delta;[\beta\gamma]} - K_{\beta\delta;[\alpha\gamma]}$$
$$= -R_{\gamma\delta\mu[\alpha}K_{\beta]} + R_{\alpha\beta\mu[\gamma}K_{\delta]}{}^{\mu}. \tag{15}$$

From Eqs. (14), (15), and (5), we have our result:

$$L_{\alpha\beta\gamma;\delta} = -3R_{\alpha\beta\mu(\gamma}K_{\delta)} - \tfrac{9}{4}K^{\mu}{}_{[\delta}R_{\alpha\beta]\gamma\mu} \tag{16}$$
$$-\tfrac{3}{4}K^{\mu}{}_{[\gamma}R_{\alpha\beta]\delta\mu} + M_{\alpha\beta\gamma\delta}.$$

Equation (16) has the corollary

$$L_{\alpha\beta}{}^{\mu}{}_{;\mu} = R_{\beta\mu}K^{\mu}{}_{\alpha} - R_{\alpha\mu}K^{\mu}{}_{\beta}, \tag{17}$$

a relation which shows that $L_{\alpha\beta\gamma}$ has a formal resemblance to an angular momentum density.

We next obtain an expression for the covariant derivative of $M_{\alpha\beta\gamma\delta}$. Our derivation starts with Eq. (14) which, with the aid of Eqs. (6), (7), (11), and the Bianchi identity, implies

$$L_{\alpha\beta(\gamma;\mu)\delta} - L_{\alpha\beta(\delta;\mu)\gamma} = R_{\alpha\beta\gamma\delta;\nu}K^{\nu}{}_{\mu}$$
$$-2R_{\mu\nu\alpha\beta;[\gamma}K_{\delta]}{}^{\nu} - R_{\mu\nu\gamma\delta;[\alpha}K_{\beta]}{}^{\nu} - K^{\nu}{}_{[\alpha}R_{\beta]\mu\gamma\delta;\nu}$$
$$-R_{\alpha\beta\mu}{}^{\nu}L_{\gamma\delta\nu} + \tfrac{1}{3}\delta^{\phi\chi}_{\alpha\beta}\delta^{\psi\omega}_{\gamma\delta}(R_{\phi\chi\psi}{}^{\nu}L_{\omega(\mu\nu)} + 2R_{\phi(\mu\psi)}{}^{\nu}L_{\omega(\nu\chi)}).$$

An alternative expression for the tensor in the above equation is obtained as follows:

$$L_{\alpha\beta(\gamma;\mu)\delta} - L_{\alpha\beta(\delta;\mu)\gamma}$$
$$= L_{\alpha\beta\mu;[\gamma\delta]} + L_{\alpha\beta\gamma;[\mu\delta]} + L_{\alpha\beta\delta;[\gamma\mu]} + L_{\alpha\beta[\gamma;\delta]\mu}$$
$$= R_{\gamma\delta\mu}{}^{\nu}L_{\alpha\beta\nu} + \tfrac{1}{2}\delta^{\phi\chi}_{\alpha\beta}\delta^{\psi\omega}_{\gamma\delta}(R_{\mu\psi\phi}{}^{\nu}L_{\chi\nu\omega} - \tfrac{1}{2}R_{\psi\omega\phi}{}^{\nu}L_{\chi\nu\mu})$$
$$+ L_{\alpha\beta[\gamma;\delta]\mu}.$$

After equating the preceding two expressions and using the Bianchi identity and Eqs. (5) to (7), we obtain our result:

$$M_{\alpha\beta\gamma\delta;\mu} = R_{\alpha\beta\gamma\delta;\nu}K^{\nu}{}_{\mu}$$
$$+ (\delta^{\phi\chi}_{\alpha\beta}\delta^{\psi\omega}_{\gamma\delta} + \delta^{\psi\omega}_{\alpha\beta}\delta^{\phi\chi}_{\gamma\delta})[\tfrac{1}{2}(R_{\phi\chi\psi\mu;\nu} - \tfrac{3}{4}R_{\phi\chi\psi\nu;\mu})K_{\omega}{}^{\nu}$$
$$-\tfrac{1}{4}R_{\phi\chi\mu}{}^{\nu}L_{\psi\omega\nu} + \tfrac{1}{3}(R_{\phi\mu\psi}{}^{\nu} + R_{\phi\psi\mu}{}^{\nu})L_{\omega\nu\chi}$$
$$+ \tfrac{1}{24}R_{\phi\chi\psi}{}^{\nu}(5L_{\mu\nu\omega} + 7L_{\omega\mu\nu})]. \tag{18}$$

Equations (11), (16), and (18) are our main results. As we noted in Sec. 1, they will be called the *structural equations for a Killing tensor of order two*. In the general case, the first integrability condition for these equations is obtained by computing $M_{\alpha\beta\gamma\delta;[\mu\nu]}$, and this will be discussed in a sequel to the present paper. Note that in the special case where $K_{\alpha\beta}$ is a parallel tensor field, $L_{\alpha\beta\gamma}$ and $M_{\alpha\beta\gamma\delta}$ vanish, and Eqs. (16) and (18) are the first and second integrability conditions for Eq. (11). Other special cases worth noting are those for which $L_{\alpha\beta\gamma}$ or $M_{\alpha\beta\gamma\delta}$ is a parallel tensor field.

## 3. DISCUSSION

To give us a little feeling for the structural equations and to test their merits in one simple case, let us consider their solution in a flat space. Choose Cartesian coordinates $x^{\alpha}$, whereupon covariant differentiation becomes ordinary differentiation, and Eqs. (11), (16), and (18) are easily integrated.

The general solution is

$$M_{\alpha\beta\gamma\delta} = A_{\alpha\beta\gamma\delta},$$
$$L_{\alpha\beta\gamma} = B_{\alpha\beta\gamma} + A_{\alpha\beta\gamma\delta}x^{\delta},$$
$$K_{\beta\gamma} = s_{\beta\gamma} + \tfrac{2}{3}B_{\alpha(\beta\gamma)}x^{\alpha} + \tfrac{1}{3}A_{\alpha\beta\gamma\delta}x^{\alpha}x^{\delta}, \tag{19}$$

where $s_{\beta\gamma}$, $B_{\alpha\beta\gamma}$, and $A_{\alpha\beta\gamma\delta}$ are arbitrary uniform fields having the same symmetries as $K_{\beta\gamma}$, $L_{\alpha\beta\gamma}$, and $M_{\alpha\beta\gamma\delta}$, respectively. It is common knowledge that the above Killing tensor is redundant.

The above flat space solution demonstrates, by the way, that $K_{\alpha\beta}$, $L_{\alpha\beta\gamma}$, and $M_{\alpha\beta\gamma\delta}$ contain the minimal number of independent components $F_A$ which can occur in any generally applicable equations of the form

$$(F_A)_{;\alpha} = \textstyle\sum_B \Gamma_{\alpha AB}F_B, \tag{20}$$

where $F_1, F_2, \ldots$ include the components of $K_{\alpha\beta}$ and where $\Gamma_{\alpha AB}$ are coefficients which are linear expressions in the Riemann tensor and its covariant derivative. Equation (20) is the form of our structural equations. In space-time, the number of independent components $F_A$ is fifty.

With so many components to handle, a thorough analysis of space-times which admit Killing tensors is not going to be completed overnight. The only objectives which can be attained in a reasonable time involve special cases.

One good subject for study would be a space-time which admits some given group of isometries in addition to a nonredundant Killing tensor. For example, we can consider any axially symmetric stationary space-time and impose the condition that it admit a nonredundant Killing tensor.

In this connection, it is useful to recall that the Poisson bracket of a linear constant of geodesic motion and a quadratic constant of geodesic motion is a quadratic constant of geodesic motion. In other words, for any Killing vector $\xi^\alpha$ and Killing tensor $K_{\alpha\beta}$, the Lie derivative

$$\mathcal{L}_\xi K_{\alpha\beta}$$

is also a Killing tensor; the corresponding $L_{\alpha\beta\gamma}$ and $M_{\alpha\beta\gamma\delta}$ are given by

$$\mathcal{L}_\xi L_{\alpha\beta\gamma} \text{ and } \mathcal{L}_\xi M_{\alpha\beta\gamma\delta},$$

because covariant differentiation and Lie differentiation with respect to a Killing vector commute. A special case which may be relatively simple to handle in calculations is one where the Lie derivatives of the Killing tensor with respect to some abelian group of isometries is zero. This is the case with the Kerr metric Killing tensor found by Carter.[9]

Another good subject for investigation is the family of algebraically special space-times which admit Killing tensors. Examples of what can be done in this area are the existence theorems obtained by Walker and Penrose[2] for conformal Killing tensors in type D vacuums. The authors are currently looking at the problem with the aid of a null tetrad version of our structural equations. The details of this null tetrad form, which is a Killing tensor parallel of the Newman-Penrose equations, will be given in the immediate sequel to this paper.


## ACKNOWLEDGMENTS

We thank Dr. Richard A. Isaacson and Dr. Frederick J. Ernst for their encouragement and many helpful discussions.



[1] See, for instance, L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, N. J., 1966), p. 128 et seq.

[2] M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).

[3] R. Geroch, J. Math. Phys. **11**, 1955 (1970).

[4] I. Hauser and R. J. Malhiot, Bull. Am. Phys. Soc. **16**, 116 (1971).

[5] L. P. Hughston, R. Penrose, P. Sommers, and M. Walker, Commun. Math. Phys. **27**, 303 (1972).

[6] P. Sommers, J. Math. Phys. **14**, 787 (1973).

[7] I. Hauser and R. J. Malhiot, J. Math. Phys. **15**, (1974) (to be published).

[8] L. Hughston and P. Sommers, Commun. Math. Phys. (to be published).

[9] B. Carter, Phys. Rev. **174**, 1559 (1968).


[10] We choose the sign of the Riemann tensor so that $K_{\alpha;[\beta\gamma]} = \frac{1}{2} R_{\beta\gamma\alpha\mu} K^\mu$; also $R_{\alpha\beta} = R^\mu{}_{\alpha\mu\beta}$.

[11] The tensor $M_{\alpha\beta\gamma\delta}$ is closely related to, but not identical with, the difference of the two Riemann tensors obtained by regarding $g_{\alpha\beta}$ and $K_{\alpha\beta}$ (in the special case when $K_{\alpha\beta}$ is nonsingular) as alternative metrics on the same differentiable manifold. The precise relation will be discussed in the sequel to this paper. The expression for this difference between the Riemann tensors is given, for example, in Ex. 18 on p. 33 of Ref. 1.

# The rest frame in stationary space-times with axial symmetry

**Richard D. Greene* and Engelbert L. Schucking†**

*Department of Physics, New York University, New York, New York 10003*

**C. V. Vishveshwara**

*Department of Physics, Boston University, Boston, Massachusetts 02215*

(Received 26 July 1973)

A rest frame in a stationary, axially symmetric space-time is defined as a synchronizable frame which is as nearly Killing as possible. This is a possible generalization of the Newtonian rest frame. A kinematical theorem giving the condition for the existence of a rest frame whose time vector is a linear combination of the Killing vectors is proved. The frame is also unique. The condition is shown to be weaker than the assumption of orthogonal transitivity. The surfaces of simultaneity of the rest frame are the surfaces of constancy of a particular Killing time coordinate, and its time vector is the component of the time Killing vector orthogonal to the angular Killing vector. Some properties of the rest frame are then discussed; it is shown that the frame is well-behaved down to the event horizon, where its time vector becomes null. Under a suitable condition on the event horizon, the time vector coincides with a Killing vector there. The gravitational redshift relation in the rest frame is derived. There is a dependence on the angular momentum of the geodesic. Furthermore, the event horizon is shown to be an infinite redshift surface for the rest frame observers. Finally, the three-vector potential of Landau and Lifshitz is interpreted and shown to be closely related to the rest frame, and a corresponding four-vector potential is invariantly defined.

## I. INTRODUCTION

Consider a space-time that is stationary,[1] axially symmetric, and flat at infinity. These properties imply that there exist the following two vector fields: a vector $\xi^\alpha$, which is timelike at infinity and has unit length there, and which generates the infinitesimal motions of a translational symmetry; and a vector $\eta^\alpha$, which generates rotations with closed orbits about a symmetry axis. Flatness at infinity further implies that $(-\eta^\alpha\eta_\alpha)^{-1/2}(\xi^\beta\eta_\beta)$ goes to zero at infinity. $\eta^\alpha$ is normalized so that a change in the corresponding group parameter by $2\pi$ exactly covers one orbit. These two vectors are called Killing vectors. The symmetries are expressed mathematically by saying that the metric $g_{\alpha\beta}$ has Lie derivative zero along $\xi^\alpha$ and $\eta^\alpha$:

$$\begin{aligned} \underset{\xi}{\pounds} g_{\alpha\beta} &= 2\xi_{(\alpha;\beta)} = 0, \\ \underset{\eta}{\pounds} g_{\alpha\beta} &= 2\eta_{(\alpha;\beta)} = 0. \end{aligned} \tag{1}$$

Furthermore, a theorem of Carter[2] guarantees that $\xi^\alpha$ and $\eta^\alpha$ commute:

$$\underset{\xi}{\pounds}\eta^\alpha = -\underset{\eta}{\pounds}\xi^\alpha = \xi^\beta\eta^\alpha{}_{,\beta} - \eta^\beta\xi^\alpha{}_{,\beta} = 0. \tag{2}$$

When the infinitesimal two-surfaces orthogonal to $\xi^\alpha$ and $\eta^\alpha$ are surface-forming, the space-time is said to be orthogonally transitive. In this case, space-time is filled by a two-parameter family of two-surfaces which are everywhere orthogonal to both Killing vectors. The condition that we have orthogonal transitivity is that the Ricci tensor be invertible in the group.[3-6] For such space-times, Bardeen[7-9] has found a vector field $\zeta^\alpha$ which is irrotational. He calls it the local nonrotating frame. $\zeta^\alpha$ is defined as

$$\begin{aligned} \zeta^\alpha &= \xi^\alpha - (\xi^\beta\eta_\beta/\eta^\gamma\eta_\gamma)\eta^\alpha, \\ \zeta^\alpha\zeta_\alpha &= \xi^\alpha\xi_\alpha - (\xi^\alpha\eta_\alpha)^2/\eta^\beta\eta_\beta. \end{aligned} \tag{3}$$

$\zeta^\alpha$ is the projection of $\xi^\alpha$ orthogonal to $\eta^\alpha$. Since $\zeta^\alpha$ is irrotational, observers who follow world lines along $\zeta^\alpha$ without rotating with respect to the neighboring observers, feel no Coriolis forces. Thus, the dragging of inertial frames is eliminated. Irrotation is equivalent to local hypersurface orthogonality, i.e., in some neighbourhood about each point, the infinitesimal three-surfaces orthogonal to $\zeta^\alpha$ are surface-forming. But $\zeta^\alpha$ is actually globally hypersurface orthogonal. The local hypersurfaces extend globally, and therefore define spaces of simultaneity for the $\zeta^\alpha$-observers (their world-time clocks are synchronized by the hypersurfaces). We will say that $\zeta^\alpha$ is "synchronizable," as opposed to just irrotational. (See the Appendix for an example of a frame which is irrotational but not synchronizable.) The $\zeta^\alpha$ frame is a possible generalization of the Newtonian nonrotating rest frame.

This paper will consider the definition of a rest frame in asymptotically flat, stationary, axially symmetric space-times without orthogonal transitivity. We will show that a synchronizable frame, when it exists, is uniquely defined by $\zeta^\alpha$, and that local hypersurface orthogonality of $\zeta^\alpha$ is sufficient for existence. The discussion will be completely kinematic and independent of any field equations. Carter[3] has discussed the physical significance of orthogonal transitivity for such space-times in general relativity; vacuum is generally orthogonally transitive;[5] circulatory matter currents which break the discrete inversion symmetry ($\xi^\alpha \to -\xi^\alpha$, $\eta^\alpha \to -\eta^\alpha$) can for instance destroy orthogonal transitivity. Therefore our results may apply to rotating, convecting astrophysical systems.

## II. THE REST FRAME

First we prove a lemma which generalizes a well-known property of Killing vectors.[10]

*Lemma:* If the vector field $w^\alpha$ is never null in some region, then $\bar{w}^\alpha = (w^\beta w_\beta)^{-1} w^\alpha$ satisfies

$$\bar{w}_{[\alpha,\beta]} = 0 \tag{4}$$

if and only if

$$w_{[\alpha}w_{\beta,\gamma]} = 0$$
and
$$(w_\alpha w_{(\beta;\gamma)} - w_\beta w_{(\alpha;\gamma)})w^\gamma = 0. \tag{5}$$

(If $w^\alpha$ is a Killing vector, $w_{(\alpha;\beta)} = 0$, then this reduces to the statement that $w^\alpha$ is hypersurface orthogonal if and only if it is closed when divided by its square.)

*Proof:* Assume Eq. (4). First expand this, and then take an antisymmetrized product with $w_\alpha$:

$$(w^\gamma w_\gamma)w_{[\alpha,\beta]} + (w^\gamma w_\gamma)_{,[\alpha}w_{\beta]} = 0,$$
$$w_{[\alpha}w_{\beta,\gamma]} = 0. \tag{6}$$

But
$$(w^\gamma w_\gamma)_{,\alpha} = 2w_{\gamma;\alpha}w^\gamma = 2w^\gamma(w_{(\alpha;\gamma)} - w_{[\alpha,\gamma]}). \tag{7}$$

Substituting, we obtain

$$3w^\gamma w_{[\alpha}w_{\beta,\gamma]} + w^\gamma(w_{(\alpha;\gamma)}w_\beta - w_{(\beta;\gamma)}w_\alpha) = 0 \tag{8}$$

which establishes Eq. (5). Conversely, if (5) holds, then (8) and (7) combine to produce (6) and hence (4).

We can now state and prove the main theorem of this paper.

*Theorem 1:* In any stationary, axially symmetric space–time that is flat at infinity, the Killing vectors $\xi^\alpha$ and $\eta^\alpha$ satisfy

$$\bar{\zeta}_{[\alpha,\beta]} = 0 \tag{9}$$

if and only if there exists a vector field $w^\alpha$ satisfying the following:

(i) $\underset{\xi}{\pounds} w^\alpha = \xi^\beta w^\alpha{}_{,\beta} - w^\beta\xi^\alpha{}_{,\beta} = 0$, $\underset{\eta}{\pounds} w^\alpha = \eta^\beta w^\alpha{}_{,\beta} - w^\beta\eta^\alpha{}_{,\beta} = 0$;

(ii) $w^\alpha = \xi^\alpha + B\eta^\alpha$, $w^\alpha \to \xi^\alpha$ at infinity;

(iii) outside of some world tube, there exist global functions $\psi$ and $\chi$ such that $w_\alpha = \psi\chi_{,\alpha}$.

Furthermore, if Eq. (9) is satisfied, then $w^\alpha = \zeta^\alpha$ is the unique vector field satisfying these three properties.

The three properties say that $w^\alpha$ must be invariant under the transformations of the group, must lie on orbits of the group, and must be globally hypersurface orthogonal.

*Proof:* That $\zeta^\alpha$ is locally hypersurface orthogonal if and only if Eq. (9) holds follows from the lemma by showing that

$$\zeta_{(\alpha;\beta)}\zeta^\beta = 0, \tag{10}$$

$$\left(\xi - \frac{\xi^\gamma\eta_\gamma}{\eta^\delta\eta_\delta}\eta\right)_{(\alpha;\beta)} = -\left(\frac{\xi^\gamma\eta_\gamma}{\eta^\delta\eta_\delta}\right)_{,(\beta}\eta_{\alpha)} \tag{11}$$

from Eq. (1). But $\zeta^\alpha$ is orthogonal to $\eta^\alpha$, and

$$\left(\frac{\xi^\gamma\eta_\gamma}{\eta^\delta\eta_\delta}\right)_{,\alpha}\left(\xi^\alpha - \frac{\xi^\rho\eta_\rho}{\eta^\sigma\eta_\sigma}\eta^\alpha\right) = \underset{\xi}{\pounds}\left(\frac{g_{\alpha\beta}\xi^\alpha\eta^\beta}{g_{\gamma\delta}\eta^\gamma\eta^\delta}\right) - \frac{\xi^\rho\eta_\rho}{\eta^\sigma\eta_\sigma}\underset{\eta}{\pounds}\left(\frac{g_{\alpha\beta}\xi^\alpha\eta^\beta}{g_{\gamma\delta}\eta^\gamma\eta^\delta}\right) = 0 \tag{12}$$

from Eqs. (1) and (2). $\zeta^\alpha$ obviously satisfies conditions (i) and (ii). If one removes a world tube containing all singularities of space–time and of the form $\bar{\zeta}_\alpha$ (there is flatness at infinity), then this becomes a closed one-form [when Eq. (9) holds] on a simply connected manifold, and is therefore exact. So

$$\bar{\zeta}_\alpha = \chi_{,\alpha}, \quad \zeta_\alpha = \left[\xi^\beta\xi_\beta - \frac{(\xi^\beta\eta_\beta)^2}{\eta^\gamma\eta_\gamma}\right]\chi_{,\alpha}. \tag{13}$$

Hence $w^\alpha = \zeta^\alpha$ satisfies all three conditions if and only if Eq. (9) is satisfied.

To complete the proof, we show that any other form which satisfies conditions (i) and (ii) cannot satisfy (iii). Suppose that $\xi^\alpha + B\eta^\alpha$, $B \neq -(\xi^\alpha\eta_\alpha)/(\eta^\beta\eta_\beta)$, satisfies (i) (i.e., $B_{,\alpha}\xi^\alpha = B_{,\alpha}\eta^\alpha = 0$) and is locally hypersurface orthogonal. The last property implies that

$$\begin{aligned} 0 &= 3(\xi + B\eta)_{[\alpha}(\xi + B\eta)_{\beta,\gamma]}\eta^\gamma \\ &= (\xi_\gamma + B\eta_\gamma)\eta^\gamma(\xi + B\eta)_{[\alpha,\beta]} \\ &\quad + [(\xi_\alpha + B\eta_\alpha)(\xi + B\eta)_{[\beta,\gamma]} \\ &\quad - (\xi_\beta + B\eta_\beta)(\xi + B\eta)_{[\alpha,\gamma]}]\eta^\gamma. \end{aligned} \tag{14}$$

On the other hand,

$$\begin{aligned} &[(\xi^\delta + B\eta^\delta)\eta_\delta]^2\{[(\xi^\gamma + B\eta^\gamma)\eta_\gamma]^{-1}(\xi + B\eta)_{[\alpha}\}_{,\beta]} \\ &\quad = [(\xi^\gamma + B\eta^\gamma)\eta_\gamma](\xi + B\eta)_{[\alpha,\beta]} \\ &\quad - (\xi + B\eta)_{[\alpha}[(\xi^\gamma + B\eta^\gamma)\eta_\gamma]_{,\beta]}. \end{aligned} \tag{15}$$

The above expression will be zero by virtue of Eq. (14) provided that

$$-2(\xi + B\eta)_{[\beta,\gamma]}\eta^\gamma = -2(\xi_{[\beta,\gamma]}\eta^\gamma + B\eta_{[\beta,\gamma]}\eta^\gamma) + (\eta^\gamma\eta_\gamma)B_{,\beta} \tag{16}$$

equals

$$[(\xi^\gamma + B\eta^\gamma)\eta_\gamma]_{,\beta} = (\xi^\gamma\eta_\gamma)_{,\beta} + B(\eta^\gamma\eta_\gamma)_{,\beta} + (\eta^\gamma\eta_\gamma)B_{,\beta}. \tag{17}$$

But by Eqs. (1) and (2),

$$\begin{aligned} &(\xi^\gamma\eta_\gamma)_{,\beta} = \xi_{\gamma;\beta}\eta^\gamma + \eta_{\gamma;\beta}\xi^\gamma = \xi_{[\gamma,\beta]}\eta^\gamma + \eta_{[\gamma,\beta]}\xi^\gamma, \\ &(\eta^\gamma\eta_\gamma)_{,\beta} = 2\eta_{[\gamma,\beta]}\eta^\gamma, \\ &0 = \xi^\gamma\eta_{\beta;\gamma} - \eta^\gamma\xi_{\beta;\gamma} = \xi_{[\gamma,\beta]}\eta^\gamma - \eta_{[\gamma,\beta]}\xi^\gamma. \end{aligned} \tag{18}$$

So Eq. (16) equals (17) and $[(\xi^\gamma + B\eta^\gamma)\eta_\gamma]^{-1}(\xi_\alpha + B\eta_\alpha)$ is a closed form. Integrating this form over an orbit of $\eta^\alpha$, we obtain

$$\oint_{\eta\text{-orbit}} [(\xi^\beta + B\eta^\beta)\eta_\beta]^{-1}(\xi_\alpha + B\eta_\alpha)dx^\alpha = \int_0^{2\pi}[(\xi^\beta + B\eta^\beta)\eta_\beta]^{-1}(\xi_\alpha + B\eta_\alpha)\eta^\alpha d\phi = 2\pi. \tag{19}$$

Therefore, $\xi_\alpha + B\eta_\alpha$ is proportional to a closed but inexact form, and so is not globally hypersurface orthogonal. This completes the proof of the theorem.

From this point on, we shall assume that Eq. (9) holds. Orthogonal transitivity is sufficient but not necessary. One can see this as follows. Since the two-surface spanned by $\xi^\alpha$ and $\eta^\alpha$ is the same as that spanned by $\zeta^\alpha$ and $\eta^\alpha$, the conditions[11] for orthogonal transitivity can be written

$$\zeta_{[\alpha}\eta_\beta\zeta_{\gamma,\delta]} = \zeta_{[\alpha}\eta_\beta\eta_{\gamma,\delta]} = 0. \tag{20}$$

The first implies[12] that everywhere off the axis

$$\zeta_{[\alpha}\zeta_{\beta,\gamma]} = \theta_{[\alpha\beta}\eta_{\gamma]}, \tag{21}$$

where $\theta_{\alpha\beta}$ is some two-form, $\theta_{\alpha\beta} = -\theta_{\beta\alpha}$. Furthermore, since in Eq. (21) we can replace $\theta_{\alpha\beta}$ by

$$\theta_{\alpha\beta} + 2(\eta^\delta\eta_\delta)^{-1}\eta^\gamma\theta_{\gamma[\alpha}\eta_{\beta]} \tag{22}$$

we can assume that $\theta_{\alpha\beta}\eta^\alpha = 0$. Therefore,

$$3\zeta_{[\alpha}\zeta_{\beta,\gamma]}\eta^\gamma = (\eta^\gamma\eta_\gamma)\theta_{\alpha\beta}. \tag{23}$$

But

$$3\zeta_{[\alpha}\zeta_{\beta,\gamma]}\eta^\gamma = (\zeta_\alpha\zeta_{[\beta,\gamma]} - \zeta_\beta\zeta_{[\alpha,\gamma]})\eta^\gamma,$$

$$\begin{aligned}\zeta_{[\alpha,\beta]}\eta^\beta &= \left(\xi_{[\alpha,\beta]} - \frac{\xi^\gamma\eta_\gamma}{\eta^\delta\eta_\delta}\eta_{[\alpha,\beta]} - \left(\frac{\xi^\gamma\eta_\gamma}{\eta^\delta\eta_\delta}\right)_{,[\beta}\eta_{\alpha]}\right)\eta^\beta \\ &= -\tfrac{1}{2}\Bigg[(\xi^\beta\eta_\beta)_{,\alpha} - \frac{\xi^\gamma\eta_\gamma}{\eta^\delta\eta_\delta}(\eta^\beta\eta_\beta)_{,\alpha} + \left(\frac{\xi^\gamma\eta_\gamma}{\eta^\delta\eta_\delta}\right)_{,\beta}\eta^\beta\eta_\alpha \\ &\quad - (\eta^\beta\eta_\beta)\left(\frac{\xi^\gamma\eta_\gamma}{\eta^\delta\eta_\delta}\right)_{,\alpha}\Bigg] \\ &= -\tfrac{1}{2}\left[(\xi^\beta\eta_\beta)_{,\alpha} - (\xi^\beta\eta_\beta)_{,\alpha} + \underset{\eta}{£}\left(\frac{\xi^\beta\eta_\beta}{\eta^\gamma\eta_\gamma}\right)\eta_\alpha\right] = 0. \end{aligned} \tag{24}$$

Therefore, $\theta_{\alpha\beta} = 0$, and Eq. (21) implies that $\zeta^\alpha$ is hypersurface orthogonal.

On the other hand, given that Eq. (9) holds, all that we know is that $\zeta_{[\alpha}\eta_\beta\zeta_{\gamma,\delta]} = 0$. Equation (18) gives information about the components of $\eta_{[\alpha,\beta]}$ along $\xi^\alpha$ and $\eta^\alpha$, but nothing about its other independent component, which is of prime importance in the evaluation of $\zeta_{[\alpha}\eta_\beta\eta_{\gamma,\delta]}$. Thus, we do not necessarily have orthogonal transitivity. Equation (9) is a weaker condition.

Next we prove that the function $\chi$, defined in Eq. (13), is a time coordinate adapted to the Killing frame. By this we mean that there exists a coordinate system $(x^0 = \chi, x^1, x^2, x^3)$ such that $\xi^\alpha = \delta^\alpha_0, \eta^\alpha = \delta^\alpha_3$. $\chi$ is, of course, unique up to an additive constant. From the definition of $\chi_{,\alpha}$ and Eqs. (1) and (2) we have

$$\underset{\xi}{£}\ \chi_{,\alpha} = \underset{\eta}{£}\ \chi_{,\alpha} = 0. \tag{25}$$

Moreover,

$$\xi^\alpha\chi_{,\alpha} = \xi^\alpha\left(\xi^\rho\xi_\rho - \frac{(\xi^\rho\eta_\rho)^2}{\eta^\sigma\eta_\sigma}\right)^{-1}\left(\xi_\alpha - \frac{\xi^\beta\eta_\beta}{\eta^\gamma\eta_\gamma}\eta_\alpha\right) = 1, \tag{26}$$

$$\eta^\alpha\chi_\alpha = 0.$$

Therefore, $\chi$ is a Killing time coordinate $T$, and we have

$$\zeta_\alpha = \left(\xi^\beta\xi_\beta - \frac{(\xi^\beta\eta_\beta)^2}{\eta^\gamma\eta_\gamma}\right)T_{,\alpha}. \tag{27}$$

$T_{,\alpha}$ is the unique Killing basis form which is a linear combination of $\xi_\alpha$ and $\eta_\alpha$. [If Eq. (9) does not hold, such a $T_{,\alpha}$ does not exist.]

$-(\xi^\alpha\eta_\alpha)/(\eta^\beta\eta_\beta)$ is the coordinate angular velocity of the $\zeta^\alpha$ frame in the Killing frame $\xi^\alpha$. The four-velocity of a $\zeta^\alpha$ observer, and the acceleration, expansion, rotation, and shear of the $\zeta^\alpha$ frame are respectively defined as

$$u^\alpha = (\zeta^\beta\zeta_\beta)^{-1/2}\zeta^\alpha = \left(\xi^\beta\xi_\beta - \frac{(\xi^\beta\eta_\beta)^2}{\eta^\gamma\eta_\gamma}\right)^{-1/2}\left(\xi^\alpha - \frac{\xi^\rho\eta_\rho}{\eta^\sigma\eta_\sigma}\eta^\alpha\right),$$

$$a^\alpha = u^\alpha{}_{;\beta}u^\beta, \quad \theta = u^\alpha{}_{;\alpha}, \quad \omega_{\alpha\beta} = u_{[\alpha,\beta]} - a_{[\alpha}u_{\beta]},$$

$$\sigma_{\alpha\beta} = u_{(\alpha;\beta)} - a_{(\alpha}u_{\beta)} - \tfrac{1}{3}\theta(g_{\alpha\beta} - u_\alpha u_\beta). \tag{28}$$

Calculating these, using Eqs. (1) and (2), one obtains

$$a_\alpha = -\left\{\ln\left[\xi^\beta\xi_\beta - \frac{(\xi^\beta\eta_\beta)^2}{\eta^\gamma\eta_\gamma}\right]^{1/2}\right\}_{,\alpha} = -[\ln(\zeta^\beta\zeta_\beta)^{1/2}]_{,\alpha},$$

$$\theta = 0, \quad \omega_{\alpha\beta} = (\zeta^\delta\zeta_\delta)^{1/2}\ \bar{\zeta}_{[\alpha,\beta]} = 0, \tag{29}$$

$$\sigma_{\alpha\beta} = (\zeta^\gamma\zeta_\gamma)^{-1/2}\left(-\frac{\xi^\rho\eta_\rho}{\eta^\sigma\eta_\sigma}\right)_{,(\alpha}\eta_{\beta)}$$

The expression for $\omega_{\alpha\beta}$ tells us what we already know, that the frame is irrotational if and only if Eq. (9) holds, Since $\theta = 0$, the time-dependent terms of the metric components in the $\zeta^\alpha$-frame result from the shearing of the $\zeta^\alpha$ congruence. [Since $\zeta^\alpha$ and $\eta^\alpha$ commute, we can retain $\eta^\alpha$ as a basis vector and the axial symmetry remains in the metric components. See, for example, Eq. (36).] The expression for the acceleration shows that since these observers are the generalization of Newtonian rest observers, the generalization of the classical gravitational potential should be

$$\ln(\zeta^\alpha\zeta_\alpha)^{1/2} = \ln\left[\xi^\alpha\xi_\alpha - \frac{(\xi^\alpha\eta_\alpha)^2}{\eta^\beta\eta_\beta}\right]^{1/2} \tag{30}$$

and not the traditional $\ln(\xi^\alpha\xi_\alpha)^{1/2}$ obtained in the Killing frame.

We now prove[13]

*Theorem 2:* The $\zeta^\alpha$ frame is well-behaved down to an event horizon, where $\zeta^\alpha$ becomes null.

*Proof:* $\eta^\alpha$, having closed orbits on which $\eta^\alpha\eta_\alpha$ is constant, must be spacelike everywhere (except on the symmetry axis where the orbits degenerate to points and $\eta^\alpha = 0$). Since $\eta^\alpha\eta_\alpha < 0$, $\zeta^\alpha$ defines a well-behaved frame down to where it becomes null (assuming, of course, that there are no naked singularities of space-time). Consider the surfaces of constant $\zeta^\alpha\zeta_\alpha$. (The acceleration shows that $(\zeta^\beta\zeta_\beta)_{,\alpha}$, the normal, can vanish in a region only if the $\zeta^\alpha$ curves are geodesics.) We calculate $(\zeta^\beta\zeta_\beta)_{,\alpha}(\zeta^\gamma\zeta_\gamma)^{;\alpha}$ using Eq. (9) and that $(\zeta^\beta\zeta_\beta)_{,\alpha}\zeta^\alpha = 0$ by Eqs. (1) and (2).

$$\begin{aligned}&\zeta_\alpha = (\zeta^\beta\zeta_\beta)\bar{\zeta}_\alpha, \quad \zeta_{[\alpha,\beta]} = (\zeta^\gamma\zeta_\gamma)_{,[\beta}\bar{\zeta}_{\alpha]}, \\ &\zeta_{[\alpha,\beta]}\zeta^{\alpha;\beta} = (2\zeta^\delta\zeta_\delta)^{-1}(\zeta^\beta\zeta_\beta)_{,\alpha}(\zeta^\gamma\zeta_\gamma)^{;\alpha}, \\ &(\zeta^\beta\zeta_\beta)_{,\alpha}(\zeta^\gamma\zeta_\gamma)^{;\alpha} = 2(\zeta^\delta\zeta_\delta)\zeta_{[\rho,\sigma]}\zeta^{\rho;\sigma}.\end{aligned} \tag{31}$$

Therefore, when $\zeta^\alpha$ becomes null, so does the surface of constant $\zeta^\alpha\zeta_\alpha$. This is a one-way surface which does not extend to infinity where $\zeta^\alpha\zeta_\alpha = 1$. Hence, it is an event horizon. The $\zeta^\alpha$ frame is well-behaved inside the ergosphere, where $\xi^\alpha$ is spacelike.

We can also generalize the theorem[3] that $\zeta^\alpha$ coincides with a Killing vector on the horizon when space-time is orthogonally transitive.

*Theorem 3:* If $\zeta^\alpha \neq 0$ and $\zeta_{[\alpha}\eta_\beta\eta_{\gamma,\delta]} = 0$ on the event horizon, then $\zeta^\alpha$ coincides with a Killing vector there.

Note that we will use the condition that $\zeta_{[\alpha}\eta_\beta\eta_{\gamma,\delta]} = 0$ only on the horizon. Orthogonal transitivity is still a stronger condition. The assumption that $\zeta^\alpha \neq 0$ anywhere on the horizon says that $\xi^\alpha$ and $\eta^\alpha$ are independent off the axis, and that $\xi^\alpha \neq 0$ at the poles. In other words, we assume that on the horizon there are no degeneracies of the group, except for the obvious one at the poles. In the Schwarzschild solution, for example, the only place where this assumption is violated is at the center of the Kruskal diagram. These events cannot be reached by any timelike or null curve which falls into the horizon from outside. One would hope that the assumption is valid for any stationary, axially symmetric black hole formed by gravitational collapse.

*Proof:* We must show that $-(\xi^\alpha\eta_\alpha)/(\eta^\beta\eta_\beta)$ is constant on the horizon, i.e., that its gradient is normal to the surface. Since the horizon is a null surface, its normal lies in the surface, along its unique null direction.

Hence, since $\zeta_\alpha \neq 0$, $(\zeta^\beta\zeta_\beta)_{,\alpha} = D\zeta_\alpha$ for some function $D$. Thus, it is sufficient to prove that $[-(\xi^\beta\eta_\beta)/(\eta^\gamma\eta_\gamma)]_{,\alpha}$ is proportional to $\zeta_\alpha$. Off the symmetry axis we have

$$\left(\frac{\xi^\beta\eta_\beta}{\eta^\gamma\eta_\gamma}\right)_{,\alpha} = \frac{2}{\eta^\gamma\eta_\gamma}\eta_{[\beta,\alpha]}\xi^\beta - \frac{2\xi^\gamma\eta_\gamma}{(\eta^\delta\eta_\delta)^2}\eta_{[\beta,\alpha]}\eta^\beta$$
$$= (2/\eta^\gamma\eta_\gamma)\eta_{[\beta,\alpha]}\zeta^\beta \tag{32}$$

by Eq. (18). Therefore, since $\zeta^\alpha$ is null,

$$\left(-\frac{\xi^\gamma\eta_\gamma}{\eta^\delta\eta_\delta}\right)_{,[\alpha}\zeta_{\beta]} = (1/\eta^\gamma\eta_\gamma)\zeta^\delta(\eta_{[\alpha,\delta]}\zeta_\beta + \eta_{[\delta,\beta]}\zeta_\alpha),$$
$$= \frac{3}{\eta^\gamma\eta_\gamma}\zeta^\delta\zeta_{[\alpha}\eta_{\delta,\beta]}. \tag{33}$$

As in Eqs. (21)–(23), our hypothesis implies that there exists a form $\Lambda_{\alpha\beta}$, $\Lambda_{\alpha\beta} = -\Lambda_{\beta\alpha}$, $\Lambda_{\alpha\beta}\eta^\beta = 0$, such that

$$\begin{aligned}\zeta_{[\alpha}\eta_{\beta,\gamma]} &= \Lambda_{[\alpha\beta}\eta_{\gamma]},\\ \Lambda_{\alpha\beta} &= \frac{3}{\eta^\delta\eta_\delta}\zeta_{[\alpha}\eta_{\beta,\gamma]}\eta^\gamma\\ &= (1/\eta^\delta\eta_\delta)(\zeta_\alpha\eta_{[\beta,\gamma]} + \zeta_\beta\eta_{[\gamma,\alpha]})\eta^\gamma\\ &= (1/\eta^\delta\eta_\delta)(\eta^\gamma\eta_\gamma)_{,[\alpha}\zeta_{\beta]}.\end{aligned} \tag{34}$$

Finally, we find that

$$\begin{aligned}\zeta_{[\alpha}\eta_{\beta,\gamma]} &= (1/\eta^\delta\eta_\delta)(\eta^\rho\eta_\rho)_{,[\alpha}\zeta_\beta\eta_{\gamma]},\\ \zeta_{[\alpha}\eta_{\beta,\gamma]}\zeta^\gamma &= 0.\end{aligned} \tag{35}$$

Therefore from Eq. (33), $[-(\xi^\beta\eta_\beta)/(\eta^\gamma\eta_\gamma)]_{,\alpha}$ and $\zeta_\alpha$ are dependent; and since $\zeta_\alpha \neq 0$, the proportionality is proved.

As an example, we transform the Kerr metric[14-16] to the rest frame. In the orthogonally transitive Killing frame[17] and in the rest frame, respectively, the metric is

$$ds^2 = (1 - 2mr\rho^{-2})dt^2 + 4mar\sin^2\theta\rho^{-2}\,dt\,d\phi - \rho^2\Delta^{-1}dr^2 - \rho^2 d\theta^2 - \sin^2\theta A\rho^{-2}d\phi^2,$$

$$\rho^2 = r^2 + a^2\cos^2\theta,\ \Delta = r^2 - 2mr + a^2,$$

$$A = (r^2 + a^2)^2 - a^2\Delta\sin^2\theta,$$

$$\begin{aligned}ds^2 = {} & \Delta\rho^2 A^{-1}dt^2 - \sin^2\theta A\rho^{-2}d\bar\phi^2 + 4mat\sin^2\theta Q\rho^{-2}A^{-1}dr\,d\bar\phi\\ & - 4ma^3tr\Delta\sin2\theta\sin^2\theta\rho^{-2}A^{-1}d\theta\,d\bar\phi\\ & + 8m^2a^4t^2r\Delta\sin2\theta\sin^2\theta Q\rho^{-2}A^{-3}dr\,d\theta\\ & - (\rho^2\Delta^{-1} + 4m^2a^2t^2\sin^2\theta Q^2\rho^{-2}A^{-3})dr^2\\ & - (\rho^2 + 4m^2a^6t^2r^2\Delta^2\sin^2 2\theta\sin^2\theta\rho^{-2}A^{-3})d\theta^2,\\ \bar\phi = {} & \phi - 2matrA^{-1},\\ Q = {} & (r^2 + a^2)(3r^2 - a^2) - a^2(r^2 - a^2)\sin^2\theta.\end{aligned} \tag{36}$$

## III. THE GRAVITATIONAL REDSHIFT

As an application of the $\zeta^\alpha$ frame, we will find the gravitational redshift seen by the rest observers. Given a geodesic with momentum vector $p^\alpha$ and affine parameter $\lambda$, $(D/D\lambda)p^\alpha = p^\alpha{}_{;\beta}p^\beta = 0$, we have

$$\frac{d}{d\lambda}(p_\alpha\zeta^\alpha) = \left[p_\alpha\left(\xi^\alpha - \frac{\xi^\gamma\eta_\gamma}{\eta^\delta\eta_\delta}\eta^\alpha\right)\right]_{;\beta}p^\beta$$
$$= -(p_\alpha\eta^\alpha)\left(\frac{\xi^\gamma\eta_\gamma}{\eta^\delta\eta_\delta}\right)_{;\beta}p^\beta \tag{37}$$

since Eq. (1) gives the well-known results

$$(p_\alpha\xi^\alpha)_{;\beta}p^\beta = \xi_{\alpha;\beta}p^\alpha p^\beta = 0,\ (p_\alpha\eta^\alpha)_{;\beta}p^\beta = 0. \tag{38}$$

$-p_\alpha\eta^\alpha$ is just the (conserved) angular momentum $L_z$ of the geodesic, and $p_\alpha u^\alpha$ is the energy. Integrating Eq. (37), we find that at positions 1 and 2 on the geodesic,

$$\left[\xi^\alpha\xi_\alpha - \frac{(\xi^\alpha\eta_\alpha)^2}{\eta^\beta\eta_\beta}\right]_2^{1/2} E_2 - \left[\xi^\alpha\xi_\alpha - \frac{(\xi^\alpha\eta_\alpha)^2}{\eta^\beta\eta_\beta}\right]_1^{1/2} E_1$$
$$= L_z\left[\left(\frac{\xi^\alpha\eta_\alpha}{\eta^\beta\eta_\beta}\right)_2 - \left(\frac{\xi^\alpha\eta_\alpha}{\eta^\beta\eta_\beta}\right)_1\right]. \tag{39}$$

The redshift depends on $L_z$.

We can show that the event horizon is an infinite-redshift surface for the $\zeta^\alpha$ observers, provided that $\zeta^\alpha$ does not vanish there. It must be proved that for an arbitrary world line, as event 1 approaches the event horizon holding $E_1$ constant, the energy $E_2$ measured at the event 2 outside the horizon approaches zero. We can write that

$$\frac{E_2}{E_1} = \frac{(p^\alpha u_\alpha)_2}{(p^\beta u_\beta)_1} = \frac{(\zeta^\alpha\zeta_\alpha)_1^{1/2}}{(\zeta^\beta\zeta_\beta)_2^{1/2}}\,\frac{(p^\gamma\zeta_\gamma)_2}{(p^\delta\zeta_\delta)_1}. \tag{40}$$

$p^\alpha$ is timelike or null; $\zeta^\alpha$ is timelike, but approaches a null vector at 1 as 1 approaches the horizon. Both are non zero and future-pointing. Therefore,

$$\begin{aligned}&(p^\alpha\zeta_\alpha)_1 > 0,\quad (p^\alpha\zeta_\alpha)_2 > 0,\quad (\zeta^\alpha\zeta_\alpha)_2 > 0,\\ &(\zeta^\alpha\zeta_\alpha)_1 \to 0^+.\end{aligned} \tag{41}$$

The possibility that $(p^\alpha\zeta_\alpha)_1 \to 0$ can be eliminated as follows: If $p^\alpha\zeta_\alpha = 0$ on the horizon, then $p^\alpha = S\zeta^\alpha$ there. $p^\alpha$ would represent a photon which instantaneously moves in the horizon. But the $\zeta^\alpha$ curves (properly parametrized) are null geodesics in the horizon. One can see this by considering the normal to the horizon, $(\zeta^\beta\zeta_\beta)_{,\alpha} = D\zeta_\alpha$.

$$D\zeta_\alpha = (\zeta^\beta\zeta_\beta)_{,\alpha} = 2\zeta^\beta\zeta_{\beta;\alpha} = 2\zeta^\beta(2\zeta_{(\beta;\alpha)} - \zeta_{\alpha;\beta})$$
$$= -2\zeta_{\alpha;\beta}\zeta^\beta,\quad \zeta^\alpha{}_{;\beta}\zeta^\beta = -\tfrac{1}{2}D\zeta^\alpha \tag{42}$$

by Eq. (10). Equation (42) says that with proper parametrization, the $\zeta^\alpha$ curve is a geodesic. Therefore, since there is only one geodesic through a given event in a given direction, the null geodesic along $p^\alpha$ would be the $\zeta^\alpha$ curve, and therefore the photon would have lain on the event horizon in the past as well. In other words, no photon which falls into the horizon from outside can have $p^\alpha\zeta_\alpha = 0$ on the horizon.

Therefore, $(p^\alpha\zeta_\alpha)_1 \nrightarrow 0$, and $E_2$ must approach zero.

## IV. THE VECTOR POTENTIAL

Landau and Lifshitz[18] have developed a three-dimensional equation of motion for free particles in the Killing frame of a stationary spacetime. They work in the infinitesimal three-surfaces orthogonal to $\xi^\alpha$. The rotation of the Killing frame gives a Coriolis term in the equation. The three-dimensional angular velocity (which is the three-dimensional form of the actual angular velocity) is written as the curl of a three-vector. In Killing coordinates $(t, x^i)$, this vector is

$$\tilde g_i^{(t)} = -g_{0i}/g_{00}. \tag{43}$$

(The tilde denotes three-dimensional quantities.) Under a change of gauge in choosing Killing coordinates,

$$t' = t + f^0(x^i),\qquad x^{j'} = f^j(x^i), \tag{44}$$

$\tilde{g}_i^{(t')} = \tilde{g}_j^{(t')} f^j{}_{,i}$ differs from $\tilde{g}_i^{(t)}$ by $f^0{}_{,i}$, which does not affect its curl. For the space-times we are considering in this paper, we can invariantly define a vector potential as a differential form $\lambda_\alpha$ which is orthogonal to $\xi^\alpha$, $\lambda_0 = \lambda_\alpha \xi^\alpha = 0$, and whose space components agree with the three-vector for an appropriate choice of time coordinate.

If a caret signifies projection orthogonal to $\xi^\alpha$, note that

$$\hat{t}_{,\alpha} = t_{,\alpha} - \frac{t_{,\beta}\xi^\beta}{\xi^\gamma \xi_\gamma}\,\xi_\alpha = t_{,\alpha} - \frac{1}{g_{00}}\,g_{0\alpha} = -\frac{g_{0i}}{g_{00}}\,\delta_i^\alpha. \tag{45}$$

This is the form which corresponds to $\tilde{g}_i^{(t)}$. Since we have a preferred time coordinate $T$, define

$$\begin{aligned}\lambda_\alpha = \hat{T}_{,\alpha} = \hat{\hat{\zeta}}_\alpha &= (\xi^\beta \eta_\beta)[(\xi^\gamma\eta_\gamma)^2 - (\xi^\gamma\xi_\gamma)(\eta^\delta\eta_\delta)]^{-1}\,\hat{\eta}_\alpha \\ &= (\xi^\delta\eta_\delta)[(\xi^\rho\eta_\rho)^2 - (\xi^\rho\xi_\rho)(\eta^\sigma\eta_\sigma)]^{-1}\left(\eta_\alpha - \frac{\xi^\beta\eta_\beta}{\xi^\gamma\xi_\gamma}\,\xi_\alpha\right).\end{aligned} \tag{46}$$

Moreover, the three-velocity $\tilde{v}_i$ of the $\zeta^\alpha$ frame as measured in the $\xi^\alpha$ frame is

$$\tilde{v}_i = -(\xi^\alpha\xi_\alpha)^{1/2}\left(\widetilde{\frac{\Lambda}{\zeta}}\right)_i = -(\xi^\alpha\xi_\alpha)^{1/2}\tilde{g}_i^{(T)} = (g_{00})^{-1/2} g_{0i}. \tag{47}$$

The three-vector potential is, essentially, just this three-velocity. $\tilde{v}_i$, or invariantly $\lambda_\alpha$, describes the dragging of inertial frames.

In our Killing frame, with $\eta^\alpha = \delta_3^\alpha$, $\tilde{g}^{(T)i}$ is

$$\tilde{g}^{(T)i} = -g^{0i} = -g^{03}\delta_3^i \tag{48}$$

since Eq. (27) implies that, for $i = 1$ or 2,

$$\begin{aligned} g^{0i} = T^{;\alpha}(x^i)_{,\alpha} &= \bar{\zeta}^\alpha (x^i)_{,\alpha} \\ &= (\zeta^\beta\zeta_\beta)^{-1}\left(\xi^\alpha - \frac{\xi^\gamma\eta_\gamma}{\eta^\delta\eta_\delta}\,\eta^\alpha\right)(x^i)_{,\alpha} \\ &= (\zeta^\beta\zeta_\beta)^{-1}\left(x^i{}_{,0} - \frac{\xi^\gamma\eta_\gamma}{\eta^\delta\eta_\delta}\,x^i{}_{,3}\right) = 0. \end{aligned} \tag{49}$$

We will use Eq. (48) in a moment. First note that in a general stationary spacetime, the condition that the hypersurface $t = f(x^i)$ be maximal is

$$\begin{aligned} &\delta \int (\,|\bar{g}\,|)^{1/2}\,dx^1 dx^2 dx^3 = 0, \\ &\bar{g} = \det \|\,\bar{g}_{ij}\,\|, \\ &\bar{g}_{ij} = g_{00} f_{,i} f_{,j} + g_{0j} f_{,i} + g_{0i} f_{,j} + g_{ij}. \end{aligned} \tag{50}$$

Using the Euler–Lagrange equation, one finds that this is satisfied by $f =$ constant if and only if

$$\begin{aligned} &\left[\left(-\frac{g}{g^{00}}\right)^{1/2} g^{0i}\right]_{,i} = 0, \quad n^\alpha{}_{;\alpha} = 0, \\ &n^\alpha = \frac{t^{;\alpha}}{(t_{,\beta}t^{;\beta})^{1/2}} = \frac{g^{0\alpha}}{(g^{00})^{1/2}}. \end{aligned} \tag{51}$$

Equations (48) and (51) imply that for the space–times that we have considered, the surfaces $T =$ constant are maximal hypersurfaces. Perhaps one can generalize the definition of $\zeta^\alpha$ to other stationary space–times by requiring that the surfaces of simultaneity of the rest observers be maximal.

## APPENDIX

Consider an inertial frame in special relativity. Use cylindrical coordinates $(t, z, \rho, \phi)$. Suppose we have a set of observers who rotate about the $z$ axis with angular velocity $\alpha/\rho^2$, where $\alpha$ is some nonzero constant. We are considering the region $\rho > |\alpha|$. The metric is

$$ds^2 = dt^2 - dz^2 - d\rho^2 - \rho^2 d\phi^2. \tag{A1}$$

The four-velocity of an observer is therefore

$$\begin{aligned} u^\lambda &= \left(1 - \frac{\alpha^2}{\rho^2}\right)^{-1/2}\left(\delta_0^\lambda + \frac{\alpha}{\rho^2}\,\delta_3^\lambda\right), \\ u_\lambda &= \left(1 - \frac{\alpha^2}{\rho^2}\right)^{-1/2}(\delta_\lambda^0 - \alpha\delta_\lambda^3) = \left(1 - \frac{\alpha^2}{\rho^2}\right)^{-1/2}(t - \alpha\phi)_{,\lambda}. \end{aligned} \tag{A2}$$

Therefore, $u_\lambda$ is proportional to a closed form, and the motion is irrotational.

However, $t - \alpha\phi$ is not a single-valued function. As a result, $(t - \alpha\phi)_{,\lambda}$ is not an exact form, as can be seen by integrating around a closed circle $C$ about the $z$ axis:

$$\oint_C (t - \alpha\phi)_{,\lambda}\,dx^\lambda = \int_0^{2\pi} \frac{\partial}{\partial\phi}(t - \alpha\phi)\,d\phi = -2\pi\alpha \neq 0. \tag{A3}$$

Hence, $u^\alpha$ is not globally hypersurface orthogonal.


## ACKNOWLEDGMENT

We wish to thank the referee for suggesting clarifying amendments to the last paragraph of the introduction.


* Submitted in partial fulfilment of the requirement for a degree of Doctor of Philosophy at New York University.

† Supported in part by the National Science Foundation.


[1] A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnétisme* (Masson, Paris, 1955), p. 109.
[2] B. Carter, *Comm. Math. Phys.* 17, 233 (1970).
[3] B. Carter, *J. Math. Phys.* 10, 70 (1969).
[4] W. Kundt and M. Trumper, *Z. Phys.* 192, 419 (1966).
[5] A. Papapetrou, *Ann. Inst. Poincaré* 4, 83 (1966).
[6] B. Schmidt, *Z. Naturforsch.* 22, 1351 (1967).
[7] J. M. Bardeen, *Ap. J.* 162, 71 (1970).
[8] J. M. Bardeen, W. H. Press, and S. A. Teukolsky, *Ap. J.* 178, 347 (1972).
[9] L. G. Fishbone, *"The Relativistic Roche Problem: Bodies in Equatorial, Circular Orbit Around Kerr Black Holes"*. Ph. D. thesis, University of Maryland (1972).
[10] P. G. Bergmann, I. Robinson, and E. Schucking, *Phys. Rev.* 126, 1227 (1962).
[11] J. A. Schouten, *Ricci Calculus*, (Lange, Maxwell, and Springer, New York, 1954), 2nd ed., pp. 78–83.
[12] Ibid., p. 23.
[13] The location of the event horizon for the static case, $\zeta^\alpha = \xi^\alpha$, was first discussed by C. V. Vishveshwara, *J. Math. Phys.* 9, 1319 (1968).
[14] R. P. Kerr, *Phys. Rev. Lett.* 11, 237 (1963).
[15] R. P. Kerr, *Quasi-Stellar Sources and Gravitational Collapse*, edited by I. Robinson, A. Schild, and E. Schucking (University of Chicago Press, Chicago, 1965).
[16] R. P. Kerr and A. Schild, *Applications of Nonlinear Partial Differential Equations in Mathematical Physics* (American Mathematical Society, Providence, R.I., 1965).
[17] R. H. Boyer and R. W. Lindquist, *J. Math. Phys.* 8, 265 (1967).
[18] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, New York, 1971), 3rd revised ed., pp. 251-2.

# A minimum principle for von Neumann's equation

**Sidney Golden**

*Department of Chemistry, Brandeis University, Waltham, Massachusetts 02154*
(Received 12 September 1973)

A minimum principle is described which has von Neumann's equation of motion of the statistical operator as its ultimate solution. When the statistical operator is initially restricted to be a projection of trace unity, the minimum principle has Schrödinger's time-dependent equation as its ultimate solution. As a result, the present minimum principle is shown to be a generalization of one first introduced by McLachlan, with a modification of the Dirac–Frenkel variational principle as a necessary consequence. In physical terms, the quantity that is minimized is shown to be the mean-squared deviation of an effective time-dependent Hamiltonian, which can be attributed to a time-dependent statistical operator initially chosen to describe a system, from the actual time-dependent Hamiltonian of the system. Examples of approximate solutions to von Neumann's equation and Schrödinger's equation that are obtained by use of the minimum principle confirm this identification.

## 1. INTRODUCTION

Soon after the formulation of Schrödinger's time-dependent equation of quantum mechanics, the first variational principle having the objective of determining good approximate solutions of that equation—the so-called Dirac-Frenkel variational principle[1,2]—made its appearance. At the present time, especailly as the result of recent investigations, several variational procedures are known[3-11] that serve to accomplish that objective for a variety of quantum mechanical problems. No comparable mathematical apparatus is available for determining good approximate solutions of von Neumann's equation of motion for the statistical operator,[12] however, and it is to provide some alleviation of this situation that the present paper is directed.

In the following section, *a minimum principle for von Neumann's equation* is formulated and comprises the main result of this paper; when the statistical operator is initially restricted only to be a projection of trace unity, *Schrödinger's time-dependent equation* is obtained. As a result, the present minimum principle is shown to be a generalization of one first introduced by McLachlan,[6] with a modification of the Dirac-Frenkel[1,2] variational principle as a consequence. The effect of imposing constraints which ensure that the minimum principle cannot yield exact solutions to these equations is considered, by examples, in the next section. Based upon the results that are obtained, the final section provides a physical interpretation of the quantity being minimized: It is *the mean-squared-deviation of an effective time-dependent Hamiltonian from the actual one of a system*, the former being one that is attributable to a time-dependent statistical operator initially chosen to describe the system. The effect of imposing initial conditions of continuity in time are also considered.

## 2. A MINIMUM PRINCIPLE FOR VON NEUMANN'S EQUATION

Any formulation of a minimum principle which seeks to provide good approximate solutinns of von Neumann's equation as the result of appropriate *variations in the statistical operator* must recognize, at the outset, the nonnegative property of that observable. For this reason, we begin by introducing an operator $\tau(t)$ such that

$$\tau(t)\tau^{\dagger}(t) \equiv \rho(t), \tag{2.1}$$

where $\rho(t)$ is the statistical operator of von Neumann at time $t$. In terms of $\tau$'s we next introduce the real, nonnegative quantity[13]

$$\Delta_{\rho}(t) \equiv \mathrm{Tr}\left[\left(\mathbf{H}(t) - i\hbar\frac{\partial}{\partial t}\right)\tau(t)\right]^{\dagger}\left[\left(\mathbf{H}(t) - i\hbar\frac{\partial}{\partial t}\right)\tau(t)\right], \tag{2.2}$$

where $\mathbf{H}(t)$ is the (Hermitian) Hamiltonian of the system of interest at time $t$. It is our immediate objective to obtain a solution of the variational equation

$$\delta[\Delta_{\rho}(t_0+)/\mathrm{Tr}\rho(t_0)] = 0 \tag{2.3}$$

subject to the constraints

$$\delta\tau(t_0) = \delta\tau^{\dagger}(t_0) = 0, \tag{2.4}$$

where[14]

$$\Delta_{\rho}(t_0+) \equiv \lim_{x\to 0+}\Delta_{\rho}(t_0+x). \tag{2.5}$$

The variations in $\tau(t)$, $t > t_0$, are presumed to be small and consistent with any restrictions that may later be prescribed for $\rho(t)$, but are otherwise arbitrary.

Before doing so, however, we note that the essence of the variational problem expressed by Eqs. (2.3)-(2.5) is to seek *the necessary and sufficient conditions that make* $[\Delta_{\rho}(t)/\mathrm{Tr}\rho(t_0)]$ *a minimum in the immediate neighborhood of an initial time* $t_0$, at which instant the values of $\tau$ and, hence, $\rho$ are prescribed. That the value obtained will be a minimum is ensured by the fact that $\Delta_{\rho}(t)$ is demonstrably a concave function of the variations in $\tau(t)$.

To proceed with the solution of Eqs. (2.3)-(2.4), we introduce the quantity

$$D_{\rho}(t_0;\zeta) \equiv \zeta\int_0^{\infty}dx e^{-\zeta x}\Delta_{\rho}(t_0+x), \quad \zeta > 0 \text{ and real.} \tag{2.6}$$

In these terms, the variable portion of Eq. (2.3) is expressible as

$$\delta\Delta_{\rho}(t_0+) = \lim_{\zeta\to+\infty}\delta D_{\rho}(t_0;\zeta). \tag{2.7}$$

By substituting Eq. (2.2) into Eq. (2.6), carrying out an integration by parts and invoking the initial condition of Eqs. (2.4), we obtain

$$\delta D_\rho(t_0;\zeta) = \zeta\int_0^\infty dx e^{-\zeta x}\Big\{\mathrm{Tr}\,\delta\tau^\dagger(t_0+x)\Big[\Big(\mathbf{H}(t_0+x) - i\hbar\frac{\partial}{\partial x}\Big) \times\Big(\mathbf{H}(t_0+x) - i\hbar\frac{\partial}{\partial x}\Big)\tau(t_0+x)\Big] + \mathrm{Tr}\,\delta\tau(t_0+x)\Big[\Big(\mathbf{H}(t_0+x) - i\hbar\frac{\partial}{\partial x}\Big) \times\Big(\mathbf{H}(t_0+x) - i\hbar\frac{\partial}{\partial x}\Big)\tau(t_0+x)\Big]^\dagger\Big\} + i\hbar\zeta^2\int_0^\infty dx e^{-\zeta x} \times\Big\{\mathrm{Tr}\,\delta\tau^\dagger(t_0+x)\Big[\Big(\mathbf{H}(t_0+x) - i\hbar\frac{\partial}{\partial x}\Big)\tau(t_0+x)\Big] - \mathrm{Tr}\,\delta\tau(t_0+x)\Big[\Big(\mathbf{H}(t_0+x) - i\hbar\frac{\partial}{\partial x}\Big)\tau(t_0+x)\Big]^\dagger\Big\}. \tag{2.8}$$

By passing to the limit, $\zeta\to+\infty$, and using Eq. (2.7), we find that Eq. (2.5) requires that[15]

$$\mathrm{Tr}\,\delta\tau^\dagger(t_0+)\Big[\Big(\mathbf{H}(t_0+) - i\hbar\frac{\partial}{\partial t_0}\Big)\tau(t_0+)\Big] = \mathrm{Tr}\,\delta\tau(t_0+)\Big[\Big(H(t_0+) - i\hbar\frac{\partial}{\partial t_0}\Big)\tau(t_0+)\Big]^\dagger, \tag{2.9}$$

which can be re-written as

$$\mathrm{Im}\Big\{\mathrm{Tr}\,\delta\tau^\dagger(t_0+)\Big[\Big(\mathbf{H}(t_0+) - i\hbar\frac{\partial}{\partial t_0}\Big)\tau(t_0+)\Big]\Big\} = 0. \tag{2.10}$$

When $\delta\tau(t_0+)$ is arbitrary, Eq. (2.10) yields von Neumann's equation, as we now show. To do this, we first express Eq. (2.10) in terms of the Hermitian and anti-Hermitian parts of $\delta\tau(t_0+)$. Then, since these parts are capable of independent variation, it follows that

$$i\hbar\frac{\partial}{\partial t_0}\tau(t_0+) = \mathbf{H}(t_0+)\tau(t_0+) \tag{2.11}$$

as well as the adjoint of this equation.[16] It then follows from Eq. (2.2) that

$$\Delta_\rho(t_0+)/\mathrm{Tr}\rho(t_0) = 0, \tag{2.12}$$

a not unexpected result. By postmultiplying Eq. (2.11) by $\tau^\dagger(t_0+)$ and adding the resulting equation to its adjoint and making use of Eq. (2.1), we obtain

$$i\hbar\frac{\partial\rho}{\partial t_0}(t_0+) = [\mathbf{H}(t_0+),\rho(t_0+)], \tag{2.13}$$

which is the equation we seek. It is a necessary condition for Eq. (2.12) to hold, but is not sufficient.[17]

In order to render the preceding results in somewhat more familiar terms, we restrict $\rho(t)$ initially to be *a projection of trace unity*. For that purpose, we express $\tau(t)$ in representative form as

$$\tau(t) \equiv |\Psi(t)\rangle\langle\varphi|, \tag{2.14}$$

with

$$\langle\varphi|\varphi\rangle = \langle\Psi(t_0)|\Psi(t_0)\rangle = 1. \tag{2.15}$$

As a consequence of Eq. (2.1),

$$\rho(t) \equiv |\Psi(t)\rangle\langle\Psi(t)| \tag{2.16}$$

and

$$\rho^2(t) = \langle\Psi(t)|\Psi(t)\rangle\rho(t). \tag{2.17}$$

By Eq. (2.15), $\rho(t_0)$ has the desired initial behavior.

In these terms, Eq. (2.2) is transcribable as

$$\Delta_\psi(t) \equiv \Big\langle\Big(\mathbf{H}(t) - i\hbar\frac{\partial}{\partial t}\Big)\Psi(t)\Big|\Big(\mathbf{H}(t) - i\hbar\frac{\partial}{\partial t}\Big)\Psi(t)\Big\rangle \tag{2.18}$$

and the transcription of Eqs. (2.3) and (2.4) is

$$\delta[\Delta_\psi(t_0+)/\langle\Psi(t_0)|\Psi(t_0)\rangle] = 0, \tag{2.19}$$

subject to the constraints

$$|\delta\Psi(t_0)\rangle = \langle\delta\Psi(t_0)| = 0; \tag{2.20}$$

Eq. (2.5) is unaltered. By taking advantage of the generality of the analysis leading from Eq. (2.6) to Eq. (2.10), we readily conclude that Eqs. (2.19) and (2.20) require that

$$\mathrm{Im}\Big[\Big\langle\delta\Psi(t_0+)\Big|\Big(\mathbf{H}(t_0+) - i\hbar\frac{\partial}{\partial t_0}\Big)\Psi(t_0+)\Big\rangle\Big] = 0. \tag{2.21}$$

Upon expressing Eq. (2.21) in terms of the real and imaginary parts of $|\delta\Psi(t_0+)\rangle$, and allowing these parts to vary arbitrarily and independently, we get

$$i\hbar\frac{\partial}{\partial t_0}|\Psi(t_0+)\rangle = \mathbf{H}(t_0+)|\Psi(t_0+)\rangle, \tag{2.22}$$

as well as the complex conjugate of this equation—the time-dependent equation of Schrödinger. As previously, $[\Delta_\psi(t_0+)/\langle\Psi(t_0)|\Psi(t_0)\rangle]$ acquires its minimum possible value-zero.

The relationship of the present minimum principle to previous work can now be made evident. Equation (2.18) is precisely the quantity first introduced by McLachlan,[6] and the minimum principle expressed by Eqs. (2.1)-(2.5) is a generalization of his. Equation (2.21) is similar to the Dirac-Frenkel[1,2] variational principle, but it is evidently a weaker one. Its generalization is Eq. (2.10) which, more importantly, follows as a consequence of the minimum principle. Furthermore, in contrast to previous ones, the present minimum principle makes explicit use of the variational constraints that prevail at an initial instant of time in order to obtain the behavior of the varied quantity in the immediate neighborhood of that instant.[18]

## 3. IMPOSITION OF ADDITIONAL CONSTRAINTS

Variational principles that are designed to yield known equations when the appropriate quantities are permitted to vary freely, apart from natural constraints, do not furnish thereby any real test of their adequacy. This behavior is necessary and is ensured in the construction of a variational principle, but it hardly suffices to judge the latter's utility in obtaining approximate solutions of the equations. For this reason, it is appropriate that we examine some examples in which additional constraints are imposed on the varied quantities, that ensure that they cannot yield exact solutions of the equations of interest. Thereby, the utility of the present minimum principle may be made evident.

(I) $\rho(t) = \rho_1(t)\rho_1(t)$. In this example, we seek to exploit the "best" solution for the statistical operator when it is constrainted to be representable as the product of two others, each one depending on a disjoint subset of the degrees of freedom of the system of interest.

We introduce

$$\tau_1(t)\tau_1{}^\dagger(t) \equiv \rho_1(t), \tag{3.1}$$

$$\tau_2(t)\tau_2{}^\dagger(t) \equiv \rho_2(t), \tag{3.2}$$

the subscripts denoting pertinent disjoint degrees of freedom.

In these terms, we require that

$$\rho(t) \equiv \rho_1(t)\rho_2(t) \equiv [\tau_1(t)\tau_2(t)][\tau_1(t)\tau_2(t)]^\dagger. \tag{3.3}$$

Furthermore, we assume the Hamiltonian of the system to be

$$\mathbf{H}(t) \equiv \mathbf{H}_1(t) + \mathbf{H}_{12}(t) + \mathbf{H}_2(t). \tag{3.4}$$

For the present case,

$$\Delta_{\rho_1\rho_2}(t) \equiv \mathrm{Tr}\left[\left(\mathbf{H}(t) - i\hbar\frac{\partial}{\partial t}\right)\tau_1(t)\tau_2(t)\right]^\dagger \times\left[\left(\mathbf{H}(t) - i\hbar\frac{\partial}{\partial t}\right)\tau_1(t)\tau_2(t)\right]. \tag{3.5}$$

The solution of the variational problem expressed by Eqs. (2.3)-(2.5), with the foregoing modifications of $\tau(t)$, is immediately obtainable from Eq. (2.10) and found to be

$$\mathrm{Im}\left\{\mathrm{Tr}\,\delta\tau_1^\dagger(t_0+)\tau_2^\dagger(t_0+)\left[\left(\mathbf{H}(t_0+) - i\hbar\frac{\partial}{\partial t_0}\right)\tau_1(t_0+)\tau_2(t_0+)\right]\right\} + \mathrm{Im}\left\{\mathrm{Tr}\,\delta\tau_2^\dagger(t_0+)\tau_1^\dagger(t_0+) \times\left[\left(\mathbf{H}(t_0+) - i\hbar\frac{\partial}{\partial t_0}\right)\tau_2(t_0+)\tau_1(t_0+)\right]\right\} = 0. \tag{3.6}$$

Since the variations $\delta\tau_1(t_0+)$ and $\delta\tau_2(t_0+)$ can be regarded as independent and arbitrary, each of the quantities on the left-hand side of Eq. (3.6) must vanish. As a further consequence, involving rendering the variations in terms of their Hermitian and anti-Hermitian parts, we obtain

$$i\hbar\frac{\partial\tau_1(t_0+)}{\partial t_0} = \left(\mathbf{H}_1(t_0+) + \frac{\mathrm{Tr}_2\mathbf{H}_{12}(t_0+)\rho_2(t_0+)}{\mathrm{Tr}_2\rho_2(t_0+)}\right)\tau_1(t_0+) + \left(\frac{\mathrm{Tr}_2\mathbf{H}_2(t_0+)\rho_2(t_0+)}{\mathrm{Tr}_2\rho_2(t_0+)} - i\hbar\frac{\mathrm{Tr}_2\tau_2^\dagger(t_0+)\partial\tau_2(t_0+)/\partial t_0}{\mathrm{Tr}_2\rho_2(t_0+)}\right)\tau_1(t_0+) \tag{3.7}$$

and

$$i\hbar\frac{\partial\tau_2(t_0+)}{\partial t_0} = \left(\mathbf{H}_2(t_0+) + \frac{\mathrm{Tr}_1\mathbf{H}_{12}(t_0+)\rho_1(t_0+)}{\mathrm{Tr}_1\rho_1(t_0+)}\right)\tau_2(t_0+) + \left(\frac{\mathrm{Tr}_1\mathbf{H}_1(t_0+)\rho_1(t_0+)}{\mathrm{Tr}_1\rho_1(t_0+)} - i\hbar\frac{\mathrm{Tr}_1\tau_1^\dagger(t_0+)\partial\tau_1(t_0+)/\partial t_0}{\mathrm{Tr}_1\rho_1(t_0+)}\right)\tau_2(t_0+), \tag{3.8}$$

as well as the adjoints of these equations. The various trace operations are to be carried out only over the indicated degrees of freedom.

It now follows from these equations and Eqs. (3.1) and (3.2) that

$$i\hbar\frac{\partial\rho_1(t_0+)}{\partial t_0} = \left[\mathbf{H}_1(t_0+) + \frac{\mathrm{Tr}_2\mathbf{H}_{12}(t_0+)\rho_2(t_0+)}{\mathrm{Tr}_2\rho_2(t_0+)}, \rho_1(t_0+)\right] \tag{3.9}$$

and

$$i\hbar\frac{\partial\rho_2(t_0+)}{\partial t_0} = \left[\mathbf{H}_2(t_0+) + \frac{\mathrm{Tr}_1\mathbf{H}_{12}(t_0+)\rho_1(t_0+)}{\mathrm{Tr}_1\rho_1(t_0+)}, \rho_2(t_0+)\right]. \tag{3.10}$$

Evidently, $\rho_1$ and $\rho_2$ each have equations of motion of the von Neumann type. Each is coupled to the other by an appropriate average of the interaction Hamiltonian between the two disjoint sets of degrees of freedom, involving the complementary statistical operator. Furthermore, it is to be noted that these equations are *identical to the exact equations of motion for the reduced statistical operators of a system* when the total statistical operator is representable as the product of two disjoint factors.[19] It is evident that Eqs. (3.9) and (3.10) maintain constancy of the traces of the $\rho$'s, in the immediate vicinity of the initial instant of time.

From either of Eqs. (3.7) or (3.8) it is verified that

$$i\hbar\left(\frac{\mathrm{Tr}_1\tau_1^\dagger(t_0+)\partial\tau_1(t_0+)/\partial t_0}{\mathrm{Tr}_1\rho_1(t_0+)} + \frac{\mathrm{Tr}_2\tau_2^\dagger(t_0+)\partial\tau_2(t_0+)/\partial t_0}{\mathrm{Tr}_2\rho_2(t_0+)}\right) = \frac{\mathrm{Tr}_{12}\mathbf{H}(t_0+)\rho_1(t_0+)\rho_2(t_0+)}{\mathrm{Tr}_{12}\rho_1(t_0+)\rho_2(t_0+)}. \tag{3.11}$$

As a consequence, Eqs. (3.7) and (3.8) lead to

$$i\hbar\frac{\partial}{\partial t_0}\left[\tau_1(t_0+)\tau_2(t_0+)\right] = \left[\mathbf{H}(t_0+) - \Delta\mathbf{H}_{12}(t_0+)\right]\left[\tau_1(t_0+)\tau_2(t_0+)\right], \tag{3.12}$$

where

$$\Delta\mathbf{H}_{12}(t) \equiv \mathbf{H}_{12}(t) - \left(\frac{\mathrm{Tr}_1\mathbf{H}_{12}(t)\rho_1(t)}{\mathrm{Tr}_1\rho_1(t)} + \frac{\mathrm{Tr}_2\mathbf{H}_{12}(t)\rho_2(t)}{\mathrm{Tr}_2\rho_2(t)} - \frac{\mathrm{Tr}_{12}\mathbf{H}_{12}(t)\rho_1(t)\rho_2(t)}{\mathrm{Tr}_{12}\rho_1(t)\rho_2(t)}\right). \tag{3.13}$$

From this and Eq. (3.5) we obtain

$$[\Delta_{\rho_1\rho_1}(t_0+)/\mathrm{Tr}_{12}\rho_1(t_0)\rho_2(t_0)] = \mathrm{Tr}_{12}[\Delta\mathbf{H}_{12}(t_0+)]^2\rho_1(t_0+)\rho_2(t_0+)/\mathrm{Tr}_{12}\rho_1(t_0)\rho_2(t_0). \tag{3.14}$$

Finally, when $\mathbf{H}$ is independent of time Eq. (3.12) can be exploited to establish that the energy of the system is conserved, as it should be.[20] In the interest of brevity, we omit the details.

(II) $|\Phi(t)\rangle = \sum_{m=1}^{M} C_m(t)|\varphi_m\rangle$. When the $|\varphi\rangle$'s form a complete orthonormal set of time-independent functions, this case reduces to the usual one involving Dirac's method of variation of constants.[21] For the present, however, we will only assume that the $|\varphi\rangle$'s are linearly independent, normalized to unity and finite in number.[22] In that case,

$$\Delta_\varphi(t) \equiv \sum_{m=1}^{M}\sum_{n=1}^{M}\{C_m^*(t)C_n(t)\langle\varphi_m|\mathbf{H}^2(t)|\varphi_n\rangle + i\hbar[\dot{C}_m^*(t)C_n(t) - C_m^*(t)\dot{C}_n(t)]\langle\varphi_m|\mathbf{H}(t)|\varphi_n\rangle + \hbar^2\dot{C}_m^*(t)\dot{C}_n(t)\langle\varphi_m|\varphi_n\rangle\}. \tag{3.15}$$

The variational problem expressed by Eqs. (2.19) and (2.20) is transcribable in the present terms, as is its solution, Eq. (2.21).

The result is that we must have

$$\mathrm{Im}\sum_{m=1}^{M}\delta C_m^*(t_0+)\left(\sum_{n=1}^{M}C_n(t_0+)\langle\varphi_m|\mathbf{H}(t_0+)|\varphi_n\rangle - i\hbar\sum_{n=1}^{M}\dot{C}_n(t_0+)\langle\varphi_m|\varphi_n\rangle\right) = 0. \tag{3.16}$$

Regarding the real and imaginary parts of the several

$\delta C_m(t_0+)$ as capable of independent and arbitrary variation, we obtain

$$i\hbar \sum_{n=1}^{M} \dot{C}_n(t_0+)\langle\varphi_m|\varphi_n\rangle = \sum_{n=1}^{M} C_n(t_0+)\langle\varphi_m|\mathbf{H}(t_0+)|\varphi_n\rangle, \qquad 1 \le m \le M. \quad (3.17)$$

It is readily determined from these equations that normalization of $|\Phi(t)\rangle$ remains constant in the immediate vicinity of the initial instant of time. When the $|\varphi\rangle$'s form a complete orthonormal set of time-independent functions, Eqs. (3.17) become the usual ones for an exact solution of Schrödinger's time-dependent equations.[23]

We now introduce a complete orthonormal basis $\{|\Psi\rangle\}$ such that

$$\langle\Psi_j|\Psi_k\rangle = \delta_{jk}, \quad \text{all } j \text{ and } k \quad (3.18)$$

and

$$\langle\Psi_k|\varphi_m\rangle = 0, \quad k > M, 1 \le m \le M. \quad (3.19)$$

The last relationship entails no loss of generality since the $|\varphi\rangle$'s are linearly independent, by hypothesis. Then, Eqs. (3.17) can be written as

$$\sum_{k=1}^{M}\langle\varphi_m|\Psi_k\rangle\left(i\hbar\frac{\partial}{\partial t_0}\langle\Psi_k|\Phi(t_0+)\rangle - \langle\Psi_k|\mathbf{H}(t_0+)|\Phi(t_0+)\rangle\right) = 0, \quad 1 \le m \le M. \quad (3.20)$$

Since these equations comprise a set of $M$ homogeneous equations that determine the coefficients of the $\langle\varphi_m|\Psi_k\rangle$'s and since the determinant of the $\langle\varphi_m|\Psi_k\rangle$'s, $1 \le m, k \le M$, must necessarily differ from zero, we are able to conclude that

$$i\hbar\frac{\partial}{\partial t_0}|\Phi(t_0+)\rangle = \left(\sum_{j=1}^{M}\sum_{k=1}^{M}|\Psi_j\rangle\langle\Psi_j|\mathbf{H}(t_0+)|\Psi_k\rangle\langle\Psi_k|\right)|\Phi(t_0+)\rangle. \quad (3.21)$$

As a result, we obtain that

$$[\Delta_\varphi(t_0+)/\langle\Phi(t_0)|\Phi(t_0)\rangle] = \sum_{j>M}|\langle\Psi_j|\mathbf{H}(t_0+)|\Phi(t_0+)\rangle|^2/\langle\Phi(t_0)|\Phi(t_0)\rangle \quad (3.22)$$

Finally, when $\mathbf{H}$ is independent of time Eq. (3.21) can be used to establish that the energy of the system is conserved, but we omit the details.

## 4. PHYSICAL INTERPRETATION OF THE MINIMUM PRINCIPLE

The formal nature of the preceding analysis can be mitigated somewhat by expressing the quantity being minimized in terms that ascribe a physical significance to it. This is especially desirable because of the criticism that has been given of the wavefunction formulation of the minimum principle, on physical grounds.[24]

In order to do so, we first suppose that we are given a $\tau_0(t)$. With no undue loss of generality, and guided by the form of Eq. (3.12), we further suppose that there exists *an effective time-dependent Hamiltonian* $H_0(t)$ (which may not necessarily be Hermitian, however) such that

$$\mathbf{H}_0(t) \equiv i\hbar\frac{\partial\tau_0(t)}{\partial t}[\tau_0(t)]^{-1}. \quad (4.1)$$

In such a case, making use of Eq. (2.1), we obtain from Eq. (2.2)

$$\Delta_{\rho_0}(t) \equiv \mathrm{Tr}\rho_0(t)[\mathbf{H}(t) - \mathbf{H}_0(t)]^\dagger[\mathbf{H}(t) - \mathbf{H}_0(t)]. \quad (4.2)$$

From this expression it becomes apparent that the minimum principle we have described seeks to minimize *the disparity between the actual Hamiltonian of the system of interest and an effective Hamiltonian* which can be attributed to a time-dependent statistical operator selected to describe the system. The quantity minimized is *the mean-squared-deviation of the effective Hamiltonian from the actual one*—a perfectly meaningful physical property.

To reinforce this physical interpretation, we render it in more familiar terms by dealing with the wavefunction version. Supposing that we are given a $|\Psi_0(t)\rangle$, and guided this time by the form of Eq. (3.21), we further suppose that there exists *an effective time-dependent Hamiltonian*[25]

$$\mathbf{H}_0(t) \equiv i\hbar\frac{\partial|\Psi_0(t)\rangle}{\partial t}(|\Psi_0(t)\rangle)^{-1}. \quad (4.3)$$

As a consequence, Eq. (2.18) yields

$$\Delta_{\psi_0}(t) \equiv \langle\Psi_0'(t)|(\mathbf{H}(t) - \mathbf{H}_0(t))^\dagger(\mathbf{H}(t) - \mathbf{H}_0(t))|\Psi_0(t)\rangle, \quad (4.4)$$

which is the wavefunction transcription of Eq. (4.2).

An alternative rendition is possible in wavefunction terms that exposes more clearly the physical significance of the quantity being minimized. When the Hamiltonian of the system is independent of time we may choose

$$|\Psi_0(t)\rangle = e^{-itE/\hbar}|\psi_0\rangle, \quad (4.5)$$

which, apart from the given form, is arbitrary. Then Eqs. (4.3) and (4.4) immediately yield

$$\Delta_{\psi_0} \equiv \langle\psi_0|(\mathbf{H}\text{-}E)^2|\psi_0\rangle. \quad (4.6)$$

When, apart from the normalization constraint, $|\psi_0\rangle$ and $E$ are capable of arbitrary variation, the minimization of $\Delta_{\psi_0}$ is one of minimizing *the variance in energy* associated with $|\Psi_0(t)\rangle$. Finding a vanishing value for it then amounts to finding an eigenvalue and eigenfunction for $\mathbf{H}$.[26]

There remains one final matter to be considered, having to do with the temporal domain of the solutions of the variational equations that obtain from the minimum principle. Having imposed appropriate initial constraints at $t_0$, we have no assurance that these constraints allow the quantities that are involved to have time derivatives that exist at that instant of time.[27] For this reason, the entire variational analysis of the preceding sections has been expressed in terms of the immediate vicinity of that instant, $t_0+$. The results which have been obtained as a consequence are unaltered if the physical situation is such as to require continuity of the various quantities and their time derivatives at $t_0$.

The requirement of continuity does, however, imply certain restrictions on the initial values that can be chosen for the quantities that are involved. For example, suppose that we require that

$$\tau_0(t_0+) = \tau_0(t_0) \quad (4.7)$$

and

$$\frac{\partial\tau_0(t_0+)}{\partial t_0} = \frac{\partial\tau_0(t_0)}{\partial t_0}, \quad (4.8)$$

where $\tau_0$ is a solution of the varitional problem expressed by Eq. (2.10), suitably augmented by appropriate constraints on its form. Then, by Eq. (4.1), we must have[28]

$$\mathrm{Im}\{\mathrm{Tr}\,\delta\tau_0^\dagger(t_0)[\mathbf{H}(t_0) - \mathbf{H}_0(t_0)]\tau_0(t_0)\} = 0. \tag{4.9}$$

Since this equation is to hold for arbitrary $\delta\tau(t_0)$ it must hold as well if we take

$$\delta\tau_0(t_0) = \alpha\tau_0(\mathrm{t}_0) \quad \text{or} \quad \delta\tau_0(t_0) = i\beta\tau_0(t_0) \tag{4.10}$$

where $\alpha$ and $\beta$ are arbitrary, small real numbers. But then it follows that we must have

$$\mathrm{Tr}\rho_0(t_0)[\mathbf{H}(t_0) - \mathbf{H}_0(t_0)] = 0. \tag{4.11}$$

In this case, therefore, continuity of $\tau_0(t)$ at $t_0$ requires that it be chosen so that the *effective Hamiltonian associated with it initially yields the same energy as the actual Hamiltonian of the system*.

The consequence of requiring continuity of $\partial\tau_0(t)/\partial t$ at $t_0$ involves the minimum value of $\Delta_{\rho_0}(t_0)$. To see this, we represent [analogous to Eqs. (2.6) and (2.7)]

$$\frac{\partial\Delta_{\rho_0}(t_0+)}{\partial t_0} \equiv \lim_{\zeta\to+\infty} \zeta\int_0^\infty dx\; e^{-\zeta x}\,\frac{\partial\Delta_{\rho_0}(t_0+x)}{\partial x}. \tag{4.12}$$

By carrying out an integration by parts, we obtain

$$\begin{aligned}\frac{\partial\Delta_{\rho_0}(t_0+)}{\partial t_0} &= \lim_{\zeta\to+\infty}\zeta^2\int_0^\infty dx\; e^{-\zeta x}[\Delta_{\rho_0}(t_0+x) - \Delta_{\rho_0}(t_0)]\\ &= \lim_{\zeta\to+\infty}\zeta[D_{\rho_0}(t_0;\zeta) - \Delta_{\rho_0}(t_0)].\end{aligned} \tag{4.13}$$

Since

$$\Delta_{\rho_0}(t_0+) \equiv \lim_{\zeta\to+\infty} D_{\rho_0}(t_0;\zeta),$$

a necessary condition that $[\partial\Delta_{\rho_0}(t_0+)/\partial t_0]$ exists, i.e., $\partial\tau_0(t)/\partial t$ is continuous at $t_0$, is that

$$\Delta_{\rho_0}(t_0+) = \Delta_{\rho_0}(t_0), \tag{4.14}$$

which was to be expected. However, since $\Delta_{\rho_0}(t_0)$ depends only on the initial values of the quantities involved and since $\Delta_{\rho_0}(t_0+)$ is its minimum value in the immediate vicinity of $t_0$, continuity of $\Delta_{\rho_0}(t)$ at $t_0$ then requires $\tau_0(t_0)$ to be chosen so that *the mean-squared-deviation of the effective Hamiltonian associated with it from the actual Hamiltonian of the system is initially a minimum*, viz.,

(4.15)

Although both Eqs. (4.11) and (4.15) represent constraints on the initial value of $\tau_0(t_0)$, it may not always be possible to satisfy them. For some situations it may not even be desirable to do so. However, since the variational fomulation we have described ensures that they are satisfied at $t_0+$, the solutions that have been obtained can be extended to later times with full assurance that any discontinuous behavior they exhibit will be related entirely to the intrinsic discontinuities in their Hamiltonians. By invoking Eqs. (4.11) and (4.15), the extension to times prior to $t_0$ can be carried out with the same assurance.

[1]P. A. M. Dirac, Proc. Cam. Philos. Soc. 26 376 (1930).
[2]J. Frenkel, *Wave Mechanics, Advanced General Theory* (Clarendon Press, Oxford, 1934), p. 435.
[3]B. A. Lippmann and J. Schwinger, Phys. Rev. 79, 469 (1950).
[4]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 314.
[5]M. Mittleman, Phys. Rev., 122, 499 (1961).
[6]A. D. McLachlan, Mol. Phys. 8, 39 (1964).
[7]A. Dalgarno in *Perturbation Theory and Its Application in Quantum Mechanics*, edited by C. H. Wilcox (Wiley, New York, 1966), p. 151
[8]J. Heinrichs, Phys. Rev. 172, 1315 (1968).
[9]P. O. Löwdin and P. K. Mukherjee, Chem. Phys. Lett. 14, 1 (1972).
[10]P. W. Langhoff, S. T. Epstein, and M. Karplus, Rev. Mod. Phys. 44, 602 (1972).
[11]H. Sambe, Phys. Rev. A 6, 2203 (1973).
[12]A possible variational principle for such solutions has been considered by McLachlan, Ref. 6, and found to yield unsatisfactory results.
[13]For definiteness: All traces are assumed to exist, all integrals are assumed to be absolutely convergent.
[14]In order to allow for a possible discontinuity in various quantities and their time derivatives at $t_0$, the designation $t_0+$ is used to refer to times that are greater than but in the immediate vicinity of $t_0$. Notationally, $[\partial f(t_0+)/\partial t_0]$ represents the time derivative of $f(t)$ in this vicinity. Matters bearing on this are discussed in the final section.
[15]Although we omit the details, it can be seen that setting Eq. (2.8) equal to zero for finite $\zeta$ will give an equation for $\boldsymbol{\tau}(t_0+x)$ that depends on $\zeta$. By dividing that equation by $\zeta$ and then passing to the limit $(\zeta\to+\infty)$, the effect of the first integral can be made negligibly small.
[16]Equation (2.11) is identical to one for the time-evolution operator even though, by hypothesis, $\boldsymbol{\tau}(t)$ is not unitary. This suggests that the minimum principle under consideration can be used to obtain approximations for the time-evolution operator. By working with imaginary times, Eq. (2.11) is similar to the Bloch equation, and suggests a possible adaptation of the present minimum principle for equilibrium problems. Involving what appears to be "a square root of the statistical operator", the present equation appears to be new.
[17]The lack of sufficiency is readily established by adding a real time-dependent multiple of $\boldsymbol{\tau}(t)$ to either side of Eq. (2.11), whereupon Eq. (2.13) is still obtained but Eq. (2.12) is not.
[18]In this regard, Löwdin and Mukharjee, Ref. 9, have explicitly used the existence of the norm of the time-dependent function that is varied. The counterpart here is the assumed existence of $Tr\boldsymbol{\rho}(t_0)$; the variational solution obtained then implies that $Tr\boldsymbol{\rho}(t)$ remains constant in the immediate vicinity of $t_0$.
[19]See, for example, S. Golden, *Quantum Statistical Foundations of Chemical Kinetics* (Clarendon, Oxford, 1969), p.86.
[20]It is in this respect that the minimum principle considered by McLachlan, Ref. 6, yielded unsatisfactory results.
[21]P. A. M. Dirac, Proc. R. Soc. Lond. A 112, 661 (1926).
[22]An example of this case with M = 2 has been considered in non-variational terms by E. F. Gurnee and J. L. Magee, J. Chem. Phys. 26, 1237 (1956).
[23]When the $|\varphi\rangle$'s are orthonormal but finite in number the equations reduced to those obtained by McLachlan, Ref. 6. See also, Löwdin and Mukharjee, Ref. 9.
[24]See, for example, Langhoff, Epstein, and Karplus, Ref. 10 p. 630.
[25]In the present case, the idea of interpreting the time derivative of a wavefunction in terms of a related Schrödinger equation has been used by Heinrichs, Ref. 8, Löwdin and Mukharjee, Ref. 9, and Sambe, Ref. 11. A related question in time-independent theory has been discussed by S. T. Epstein, Ref. 7, p. 49.
[26]This criterion, instead of the usual variational principle in quantum mechanics, has been used by H. Conroy, J. Chem. Phys. 41, 1327 (1964).
[27]No constraints of continuity of $\mathbf{H}(t)$, for instance, have been imposed at this stage of the analysis.
[28]Now, we are assuming that $\mathbf{H}(t)$ is a continuous operator function of time.

# Imbedding of Segal systems

**Platon C. Deliyannis**

*Department of Mathematics, Illinois Institute of Technology, Chicago, Illinois 60616*
(Received 27 March 1973; revised manuscript received 16 July 1973)

It is shown that any distributive Segal quantum system satisfying a certain continuity condition on the squaring operation can be represented as a space of Mackey observables based on some logic derived from the original system.

## INTRODUCTION

The axiomatic systems initiated by Segal (Refs. 1, 2) and Mackey (Ref. 3) have been the object of study of several mathematicians and physicists (Refs. 4 to 15 to mention only a few). In the hope of understanding the nature of these systems, the present writer has presented (Ref. 16) an axiomatic system containing both as special cases. In the present paper a study is made for the same purpose, but along different lines. It essentially amounts to showing that certain Segal systems $\mathfrak{A}$ are contained as subsystems of Mackey systems, in the sense that the elements of $\mathfrak{A}$ are actually $\sigma$-homomorphisms from the Borel sets of the line into some $\sigma$-logic $\mathcal{L}$ of events; it is also shown that the states of $\mathfrak{A}$ are obtained from probability measures on $\mathcal{L}$.

The argument we present requires two conditions to be imposed upon $\mathfrak{A}$. The first is that the operation of squaring an observable, $A \to A^2$, the basic building block of the Segal system, is continuous in the strong bounded topology (see Sec. IIA). The second is that the pseudoproduct $A \cdot B$, defined via squaring as $\frac{1}{4}[(A+B)^2 - (A-B)^2]$, is distributive over $+$. It is not difficult to show that both conditions hold for nontrivial cases, e.g., for the system of self-adjoint elements in an abstract $C^*$-algebra (see appendix B).

The proof consists in enlarging the system (without loosing any structure in the process) so that it will become strongly complete. In this it extends results obtained for various special cases (Refs. 17, 18, 19). As a result the enlarged system contains sufficiently many idempotents, which form a complete logic $\mathcal{L}$. Finally the bounded Mackey observables associated with this logic are shown to be in a one-to-one correspondence with the elements of the enlarged system.

The paper consists of four parts. In Sec. I we present a different axiomatic description of Segal systems in order to make the enlargement process mentioned above more natural. In Sec. II the enlargement is constructed, and in Sec. III we establish the main result. Some examples and comments make up Sec. IV.

## I. ALTERNATIVE DESCRIPTION OF SEGAL SYSTEMS

### A. The original Segal system

A single undefined term "observable" forms the background, and we assume the following:

*Axiom S1:* The set $\mathfrak{A}$ of all observables is a vector space over the reals.

*Axiom S2:* A map $A \to A^2$ is defined from $\mathfrak{A}$ to $\mathfrak{A}$ such that if we set $A \cdot B = \frac{1}{4}[(A+B)^2 - (A-B)^2]$ and define inductively $A^{k+1} = A^k \cdot A$, then:

(i) there exists an $I \in \mathfrak{A}$ such that $A \cdot I = A$ for all $A \in \mathfrak{A}$.

(ii) Letting $p(x) = \sum_{k=0}^{n} p_k x^k$ be any real polynomial and $p(A) = \sum_{k=0}^{n} p_k A^k$ ($A^0 = I$) we have $p(q(A)) = (p \circ q)(A)$ where $\circ$ denotes composition of functions.

*Axiom S3:* The space $\mathfrak{A}$ carries a norm $A \to \|A\|$ such that:

(i) $\|A^2\| = \|A\|^2$,

(ii) $\|A^2 - B^2\| \le \max[\|A\|^2, \|B\|^2]$,

(iii) $A \to A^2$ is continuous,

(iv) The space is complete.

A state is defined as a linear map $m: \mathfrak{A} \to$ reals such that $m(I) = 1$ and $m(A^2) \ge 0$ for all $A \in \mathfrak{A}$.

The proofs of all theorems that we shall state below can be found in Refs. 1, 2, and 4.

The following "spectral representation theorem" forms the basis of the whole Segal theory.

*Theorem 1:* For every $A \in \mathfrak{A}$ there exists a unique compact set $\sigma A$ in the real line such that the smallest vector subspace $\mathfrak{F}(A)$ of $\mathfrak{A}$ containing $I$ and $A$, closed under squaring and in the norm topology is isomorphic to the algebra of all continuous real functions defined on $\sigma A$. This isomorphism preserves the vector operations, squaring (hence the pseudoproduct $\cdot$), maps $A$ to the identity function on $\sigma A$, and transforms the norm to the functional supremum norm.

The proof involves the use of the Gelfand theorem on representations of commutative $C^*$-algebras; in establishing that the correct conditions hold for $\mathfrak{F}(A)$, part (iii) of S3 is essential. The existence of $\sigma A$ comes from the fact that $\mathfrak{F}(A)$ is an algebra with a single generator $A$ and part (ii) of S3, while uniqueness follows from the requirement that $A$ maps to the identity function on $\sigma A$.

The *spectrum* of $A$ is then defined as this set $\sigma A$, and for any continuous function $f$ on the line the observable $fA$ is defined to be that element of $\mathfrak{F}(A)$ which maps to $f$ restricted to $\sigma A$. For a polynomial $p$ the notation is consistent.

*Theorem 2:* For every $a \in \sigma A$ there exists a state $m$ such that $m(A) = a$.

The proof consists of an application of a Hahn–Banach type theorem. Of considerable importance in setting up the hypotheses of this theorem is a result of Sherman (Ref. 4) to the effect that any sum of squares in $\mathfrak{A}$ is necessarily a square.

Although it is not really crucial for our main theorem it is of interest to note that for Segal systems in which the pseudoproduct $\cdot$ is distributive, the analytical conditions S3 (iii) can be formulated in a completely algebraic way.

*Remark 1:* The operation $\cdot$ is distributive iff for any $A, B$, we have $(A+B)^2 + (A-B)^2 = 2A^2 + 2B^2$.

Necessity is obvious if we take into account that $(-A) \cdot B = -(A \cdot B)$. Sufficiency follows by using the condition to obtain $(A_1 + A_2)\cdot B + (A_1 - A_2)\cdot B = 2(A_1\cdot B)$(direct calculation), hence $(2A)\cdot B = 2(A\cdot B)$ (since by definition $O\cdot B = 0$ anyway), and finally replacing $A_1, A_2$ by $\frac{1}{2}(A_1 + A_2), \frac{1}{2}(A_1 - A_2)$.

*Remark 2:* If $\cdot$ is distributive and squaring is norm continuous, then for any $A, B$ and real $a$ we have $(aA)\cdot B = a(A\cdot B)$. The converse also holds.

To see this note that distributivity implies $(aA)\cdot B = a(A\cdot B)$ for any rational $a$. If squaring is continuous, then so is the map $a \to (aA)\cdot B$ and the conclusion holds for all real $a$. Now assume $(aA)\cdot B = a(A\cdot B)$ for all $a$, which is equivalent to $a(A + B)^2 + (aA - B)^2 = (aA + B)^2 + a(A - B)^2$. This we can write as $2C^2 + 2D^2 = (C + D)^2 + (C - D)^2$ if we set $\sqrt{a}\,(A + B) = \sqrt{2}C$, $aA - B = \sqrt{2}D$ provided that we also have $aA + B = C - D$, $\sqrt{a}\,(A - B) = C + D$; this last requirement means $\sqrt{a}/2 = (a + \sqrt{a})/2 = (1 - \sqrt{a})/2$, $a\sqrt{2} = (\sqrt{a} - a)/2$, $1/\sqrt{2} = (1 + \sqrt{a}/2$, which boils down to $a = (\sqrt{2} - 1)^2$. Thus given any $C, D$, we set $a = (\sqrt{2} - 1)^2, A = [\sqrt{2}/(1 + a)]\,[(1/\sqrt{a})C + D], B = (1/\sqrt{2})[\sqrt{a}\,C - D]$ and apply the above calculation to obtain $2C^2 + 2D^2 = (C + D)^2 + (C - D)^2$, i.e., distributively of $\cdot$ . Continuity of squaring is obtained as follows. We first show that $\|A\cdot B\| \leq \|A\|\ \|B\|$; by S3 (ii) we have that $\|A\cdot B\| \leq \frac{1}{4}$ max $[\|A + B\|^2, \|A - B\|^2]$, so that it is $\leqq 1$ in case both $\|A\|, \|B\|$ are $\leq 1$. But then $\|((1/\|A\|)A)\cdot((1/\|B\|)B)\| \leqq 1$ and using our hypothesis we can pull out the scalar factors to obtain the result stated above. Now we have $\|(A + B)^2 - A^2\| = \|(2A + B)\cdot B\| \leqq \|2A + B\|\,\|B\|$ which goes to zero as $\|B\|$ goes to zero.

*Remark 3:* The condition $(aA)\cdot B = a(A\cdot B)$ implies that for any state $m$ we have $[m(A\cdot B)]^2 \leqq m(A^2)m(B^2)$.

All we need to do is observe that the map $(a, b) \to m[(aA + bB)^2]$ will be bilinear positive; hence for all $a, b$ we have $a^2m(A^2) + 2ab\,m(A\cdot B) + b^2m(B^2) \geq 0$.

### B. The new system

This we develop in the spirit of Ref. 16, the results of which we shall use without reproducing the proofs. We shall have, as before, a single undefined term "observable" and one undefined operation carrying the pair $(f, A)$ to what we shall write as $fA$ or $f(A)$, where $f$ is an arbitrary continuous real function and $A, fA$ observables. We write $\mathfrak{M}$ for the set of all observables $\mathfrak{F}$ for the set of all continuous real functions.

*Axiom 1:* If the supports of $f_1, f_2, \cdots$ and those of $g_1, g_2, \cdots$ form locally finite systems, $f = \sum f_n$, $g = \sum g_n$, and $f_nA = g_nB$, then $fA = gB$, $(f_1f_2)A = (g_1g_2)(B)$.

Recall (Ref. 20) that a family of sets $S_\alpha$ forms a locally finite system if every point has a neighborhood intersecting a finite number only of the $S_\alpha$'s. In such a case the same holds for any compact set. Also, under the hypotheses of Axiom 1, the functions $f, g$ will be continuous.

*Axiom 2:* For any $f, g \in \mathfrak{F}$ and $A \in M$ we have $f(gA) = (f\circ g)A$, where $\circ$ denotes composition of functions.

*Axiom 3:* If $fA = fB$ for all bounded $f \in \mathfrak{F}$, then $A = B$. If 1 denotes the constant function with range $\{1\}$, then $1(A) = 1(B)$ for all $A, B \in \mathfrak{M}$.

*Axiom 4:* $\mathfrak{M}$ is a vector space over the reals such that for any $f_1, f_2, \ldots, f_n \in \mathfrak{F}$ and real $a_1, a_2, \ldots, a_n$ we have $(\sum a_if_i)A = \sum a_i(f_iA)$.

*Remark:* Writing $O$ for the constant function with range $\{O\}$, we see that $O(A)$ is the zero element of the vector space $\mathfrak{M}$, which we shall again write as $O$: we have $fA = (f + O)A = fA + O(A)$.

We shall write $\mathfrak{F}(A)$ for the set $\{fA \mid f \in \mathfrak{F}\}$ where $A$ is any member of $\mathfrak{M}$. By virtue of Axiom 1, $\mathfrak{F}(A)$ obtains a product defined by $(fA)(gA) = (fg)A$ consistently, and becomes an algebra with unity $I$ defined by $1(A)$. It is clear that for any family $A_i$ the intersection $\cap\mathfrak{F}(A_i)$ is a subalgebra of each $\mathfrak{F}(A_i)$ and that the inherited operations are consistent. Also, if $j(x) = x$ for any $x$, we have $jA = A$ for any $A$; we shall write $A^n$ instead of $j^n(A)$ and note that it is just the $n$th power of $A$ in the algebra $\mathfrak{F}(A)$.

Of fundamental importance is the spectrum of an observable, defined as follows:

*Definition 1:* (i) An open set $U$ on the real line is $A$-*null* iff for every bounded nonnegative $f \in \mathfrak{F}$ vanishing outside $U$ we have $fA = 0$ (ii) The *spectrum* of $A$ is the complement of the union of all open $A$-null sets. We write it as $\sigma A$.

In Ref. 16 we have established the following:

*Theorem 3:* Let $U, U_i$ be $A$-null, and $V \subseteq U$; then $V$ and $\cup\, U_i$ are $A$-null. Thus $\sigma A$ is closed.

*Theorem 4:* For any $f \in \mathfrak{F}, A \in \mathfrak{M}$ we have $\sigma(fA) = \overline{f(\sigma A)}$ (the closure of the set $\{f(x) \mid x \in \sigma A\}$); $A = 0$ iff $\sigma A \subseteq \{0\}$.

*Theorem 5:* The map $fA \to f\,|\,\sigma A$ (the restriction of $f$ to $\sigma A$) is an isomorphism of $\mathfrak{F}(A)$ onto the set of functions continuous on $\sigma A$.

*Definition 2:* The *norm* or *bound* $\|A\|$ of $A \in \mathfrak{M}$ is the number sup $\{|x| \ \ x \in \sigma A\}$. An observable $A$ is *bounded* if $\|A\| < +\infty$. Write $\mathfrak{B}$ for the set of all bounded observables.

By Theorem 4 we have that $\|fA\| = \sup\{|f(x)|\ |x \in \sigma A\}$ so that in particular $\|A^2\| = \|A\|^2$.

*Definition 3:* A *state* of the system is a map $m$: $\mathfrak{B} \to$ reals such that $m(I) = 1$, $m(A^2) \geq 0$, $m(aA) = am(A)$, and $m(A + B) = m(A) + m(B)$, whenever $A, B, A + B \in \mathfrak{B}$.

This last precaution is necessary, because $\mathfrak{B}$ need not as yet be closed under sums.

Note that for any state we have $|m(A)| \leq \|A\|$.

*Theorem 6:* For every state $m$ and every observable $A$ there exists a unique (finitely additive) regular probability measure $p_{A,m}$ on the ring generated by the open sets of the line such that $m(fA) = \int f(x)dp_{A,m}(x)$ for all bounded $fA$; this measure is supported by $\sigma A$ and is therefore countably additive in case $A$ is bounded.

For $A$ not bounded it may happen that $\int x dp_{A,m}(x)$ exists, in which case we shall still write it as $m(A)$.

*Definition 4:* The measure $p_{A,m}$ is the *probability distribution* of $A$ in the state $m$.

*Axiom 5:* For every $A \in \mathfrak{M}$ and every open set $U$ which is not $A$-null there exists a state $m$ for which $p_{A,m}(U) = 1$.

*Theorem 7:* For every $A \in \mathfrak{M}$ the norm $\|A\|$ is the supremum of all numbers $|m(A)|$ as $m$ varies over all states for which $mA$ exists.

*Proof:* First we show that whether $A$ is bounded or not, there exist states $m$ for which $mA$ is finite. Take $x_0 \in \sigma A$ (if $\sigma A = \phi$, then $A = 0$!) and any $\epsilon > 0$; the open interval $U = (x_0 - \epsilon, x_0 + \epsilon)$ is not $A$-null since it intersects the spectrum of $A$, hence there exists a state $m$ with $p_{A,m}(U) = 1$. Since $p_{A,m}$ is then concentrated in $U$ and for $x \in U$ we have $|x| < |x_0| + \epsilon$ the integral $\int x dp_{A,m}(x)$ is finite. Now suppose that this $x_0$ is nonnegative, so that for $x \in U$ we have $x > |x_0| - \epsilon$; then $|m(A)| = |\int x dp_{A,m}(x)| \geq |x_0| - \epsilon$, so that the supremum of all $|m(A)|$ will be $\geq |x_0|$. In case $x_0 < 0$ we consider $-A$ to reach the same conclusion. Thus for any $x_0 \in \sigma A$ we have the supremum of the various $|m(A)|$ exceeding $|x_0|$, and thus exceeding $\|A\|$. Since anyway it is also $\leq \|A\|$, we have equality.

It is clear that Axiom 4 has had no bearing on the discussion. The full Segal system actually emerges only by connecting the sum tightly to the other concepts.

*Axiom 6:* The sum of two bounded observables is bounded.

*Theorem 8:* The set $\mathfrak{B}$ of all bounded observables is a vector space such that (i) $\| \ \|$ is a norm and (ii) $\|A^2 - B^2\| \leq \max[\|A\|^2, \|B\|^2]$.

*Proof:* (i) all we need is $\|A + B\| \leq \|A\| + \|B\|$, since $\|A\| = 0$ implies $\sigma A \subseteq \{0\}$ or $A = 0$. But since $m(A + B)$ is now defined for all states, this is immediate from theorem 7. (ii) Similarly, $\|A^2 - B^2\| = \sup|m(A^2) - m(B^2)|$; for definiteness assume $\|A\| \geq \|B\|$ so that $m(A^2)$ and $m(B^2)$ do not exceed $\|A\|^2$. But $|m(A^2) - m(B^2)| \leq \max[m(A^2), m(B^2)] \leq \|A\|^2 = \max[\|A\|^2, \|B\|^2]$. QED

*Axiom 7:* The map $A \to A^2$ is continuous in the norm topology on the space of bounded observables.

We have stated this axiom last as it has not been essential in developing the basic theory. As already mentioned, however, it plays an important role in Segal's original formulation, its most crucial use being made in establishing the existence of the Segal spectrum. Segal also assumes the space to be complete; in case it is not, its completion will satisfy all axioms, and again Axiom 7 [= S3(iii)] is of essence in deriving this.

It is of some importance for the remainder of the paper to have the following theorem, in particular the implication Pythagorean $\Longrightarrow$ Axiom 5:

*Theorem 9:* In a system satisfying Axioms 1-4, Axiom 5 is equivalent to the conclusion of Theorem 7 i.e., that $\|A\|$ is the supremum of all $|m(A)|$. In a system satisfying axioms 1-4 and 6 the following are equivalent:

(i) For every bounded $A$ and $a \in \sigma A$ there exists a state $m$ for which $mA = a$.

(ii) For $A_i$ bounded, there exists an $A$ with $A^2 = \sum_{i=1}^n A_i^2$ (Pythagorean property).

(iii) For every bounded $A$ and open $U$ intersecting the spectrum of $A$ there exists a state $m$ with $p_{A,m}(U) = 1$.

*Proof:* For the first part we refer to theorem 8 of Ref. 16. So now assume (i) and let $B = \sum_{i=1}^n A_i^2$; since $m(B) \geqq 0$ for all states we have $\sigma B \subseteq [0, +\infty)$. Consider the function $f$ for which $f(x) = \sqrt{x}$ for $x \geqq 0$ and $f(x) = 0$ for $x < 0$; since $f^2|\sigma B = j|\sigma B$ we obtain from theorem 4 that $B = (jB)^2$, and we can set $A = fB$. Still under hypothesis (i) consider any bounded $A$, an open set $U$ intersecting $\sigma A$ and pick $a \in (\sigma A) \cap U$. Take a continuous function $f$ with $f(a) = 1$ and $0 \leqq f \leqq \chi_U$. Since $A$ is bounded, i.e., $\sigma A$ compact, we have $f(\sigma A) = \sigma(fA)$ so that $1 \in \sigma(fA)$; but then there exists a state $m$ for which $m(fA) = 1$, and thus $p_{A,m}(U) = \int \chi_U dp_{A,m} \geq \int f dp_{A,m} = m(fA) = 1$.

Now we shall show that (ii) implies (i). Consider the set $\mathfrak{F}(A)$; since axiom 6 is assumed, it forms a subspace of $\mathfrak{B}$ and our hypothesis is that $\mathfrak{C} \cap \mathfrak{F}(A)$ consists of all elements $fA$ for which $f|\sigma A \geq 0$, where $\mathfrak{C}$ is the cone of all sums of squares. Now $I$ is a radial point of $\mathfrak{C}$: taking any $B \in \mathfrak{B}, B \notin \mathfrak{C}$ we join it to $I$ by the segment $t \to (1 - t)I + tB = X(t)$ and observe that for $t < \|I - B\|$ we have $\|I - X(t)\| < 1$, or $\sigma(X(t)) \subseteq [0, 2]$, i.e., $X(t)$ a square. Take a point $a \in \sigma A$; the functional $fA \to f(a)$ is well defined on $\mathfrak{F}(A)$, and by our previous remark it is positive for the relative order structure imposed by $\mathfrak{C}$. Hence it extends to a state $m$ of the system (Ref. 17) for which we evidently have $m(A) = a$.

Finally we show that (iii) implies (i). Again take $a \in \sigma A$, and for each open $U$ containing $a$ choose a state $m_U$ with $p_{A,m_U}(U) = 1$. Now the space of states is evidently weakly closed in the weak topology of the dual of $\mathfrak{B}$ hence compact, and therefore the net $U \to m_U$ admits a subnet $m_j$ converging to some state $m$. We shall have $m(A) = a$. Take $\epsilon > 0$ and an index $j(\epsilon)$ such that $j > j(\epsilon)$ implies $|m(A) - m_j(A)| < \epsilon/2$. Note that for any $U$ we have $|m_U(A) - a|$ less than the measure of $U$, so by choosing a $U_\epsilon$ of measure $< \epsilon/2$ and a $U \subseteq U_\epsilon$ which is also beyond $j(\epsilon)$ (possible since $\{j\}$ is a subnet of $\{U\}$) we get $|m(A) - a| < \epsilon$.

### C. Relations between the two systems

We are now in a position to compare the two systems. Suppose that we are given a Segal system $\mathfrak{A}$. Then, via the spectral representation Theorem 1 we can define a map $(f, A) \to fA$ from $\mathfrak{F} \times \mathfrak{A}$ to $\mathfrak{A}$. The proof that Axiom 1 holds can be found in Ref. 16, while Axiom 2 is satisfied by virtue of the representation theorem itself. For Axiom 3, we see that by letting $f$ agree with the identity function on the union of the Segal spectra of $A$ and $B$ we get $A = B$; the rest of 3 and also 4 are straightforward.

Before proceeding further we note that the Segal spectrum and the one introduced in Definition 1 are identical because of Theorem 5. Hence the Segal norm is the same as that of definition 2. Evidently the states are also the same.

Axiom 5 holds by virtue of Theorems 2 and 9, and, of course Axioms 6 and 7 need no comment.

The converse is even easier to verify in view of Theorems 5, 7, and 8.

## II. COMPLETION OF A SEGAL SYSTEM

### A. Construction of the completion

Consider any distributive Segal system $\mathfrak{A}$, all of whose elements are bounded.

*Definition 5:* A net $(A_j)$ in $\mathfrak{A}$ *converges weakly* to $A \in \mathfrak{A}$ iff for each state $m$ of $\mathfrak{A}$ we have $m(A_j - A) \to 0$. A net $(A_j)$ *converges strongly* to $A \in \mathfrak{A}$ iff $m[(A_j - A)^2] \to 0$. We abbreviate these statements to $w$-lim $A_j = A$, s-lim $A_j = A$. In case the norms $\|A_j\|$ are bounded (for sufficiently large $j$) we shall prefix the adjective "bounded" and abbreviate to bw-lim $A_j = A$ and bs-lim $A_j = A$.

Note that since the system is distributive we have $|mA|^2 \leq m(A^2)$ so that strong convergence implies weak convergence.

*Definition 6:* A net $(A_j)$ is *bounded weakly Cauchy* (bw-Cauchy) iff for each state $m$ of we have $m(A_j - A_i) \to 0$ while the norms $\|A_j\|$ are bounded for large $j$. A net is *bounded strongly Cauchy* iff for each state $m$ of $\mathfrak{A}$ we have $m[(A_j - A_i)^2] \to 0$ and the norms $\|A_j\|$ are bounded for large $j$.

Since for distributive systems we have $A \to m(A^2)^{1/2}$ a seminorm, it follows that a (bounded) strongly convergent net is (bounded) strongly Cauchy. The converse will not be true in general. The purpose of this section is to enlarge the given system so that it will become "almost complete" (see Theorem 12). This will allow any locally bounded Borel function to operate on the observables of the new system so that axioms 2, 3, 4 and a weak but sufficient version of Axiom 1 hold.

To this end we introduce the following relation: Two bounded nets $(A_i)$, $(B_j)$ are *equivalent* iff for each state $m$ of $\mathfrak{A}$ we have $m[(A_i - B_j)^2] \to 0$. We write this as $(A_i) \sim (B_j)$.

The following two lemmas are easily verified:

*Lemma 1:* The relation $\sim$ is an equivalence.

*Lemma 2:* If $(A_i) \sim (B_j)$ and $(C_k) \sim (D_l)$ then $(A_i + C_k) \sim (B_j + D_l)$ and $(aA_i) \sim (aB_j)$.

Note that $(A_i + C_l)$ for example is indexed by the Cartesian product of the two index sets equipped with the product order (see Appendix A)

We shall denote by $\hat{\mathfrak{A}}$ the set of all equivalence classes of bs-Cauchy nets in $\mathfrak{A}$. This makes sense only if: $(A_i) \sim (B_j)$ and $(A_i)$ bs-Cauchy implies $(B_j)$ also bs-Cauchy. To verify this, we calculate $(B_j - B_l)^2 - (A_i - A_k)^2 = (B_j^2 - A_i^2) + (B_l^2 - A_k^2) + 2A_i \cdot A_k - 2B_j \cdot B_l$; since $|m(B_j^2 - A_i^2)|^2 \leq m[(B_j - A_i)^2]\, m[(B_j + A_i)^2] \leq Km[(B_j - A_i)^2] \to 0$ (recall that $\|A_i\|$, $\|B_j\|$ are bounded) and similarly $m(B_l^2 - A_k^2) \to 0$ we calculate $A_i \cdot A_k - B_j \cdot B_l = A_i \cdot (A_k - B_l) + (A_i - B_j) \cdot B_l$; but again as above we obtain $m[A_i \cdot (A_k - B_l)] \to 0$ and $m[(A_i - B_j) \cdot B_l] \to 0$. Thus $m[(B_j - B_l)^2 - (A_i - A_j)^2]$ tends to zero and as $(A_i)$ is Cauchy we obtain $m[(B_j - B_l)^2] \to 0$.

It is now clear how to impose the structure of a vector space on $\hat{\mathfrak{A}}$ : for $X, Y \in \hat{\mathfrak{A}}$, pick representative nets $(A_i)$, $(B_j)$; then $X + Y$ is the class of $(A_i + B_j)$ and $aX$ the class of $(aA_i)$.

Now we need to define the element $fX$, for $X \in \hat{\mathfrak{A}}$ and $f$ a continuous function. We have to show that $(A_i) \sim (B_j)$ implies $(fA_i) \sim (fB_j)$ and define $fX$ as the class of $(fA_i)$. Also, of course, we must have $(fA_i)$ bs-Cauchy to begin with. To achieve this we shall make use of the following axiom, which we shall assume from now on.

*Axiom 7*:* Suppose that bs-lim $A_i = 0$ and $(B_j)$ is bounded; then bs-lim $[(A_i + B_j)^2 - B_j^2] = 0$.

*Lemma 3:* Axiom 7* is equivalent to: if $(A_i)$ is bs-Cauchy then so is $(A_i^2)$.

*Lemma 4:* Suppose that $(A_i)$ and $(B_j)$ are equivalent bs-Cauchy nets. If $f$ is continuous, then $(fA_i)$ and $(fB_j)$ are also equivalent bs-Cauchy nets.

*Proof:* First note that if we show $(fA_i) \sim (fB_j)$ it follows that $(fA_i)$ is bs-Cauchy, by replacing $(B_j)$ with $(A_i)$ and using the Cauchy condition instead of $(A_i) \sim (B_j)$. Next we show that it suffices to prove that $(A_i^2) \sim (B_j^2)$. Assume this and observe that now we also have $(A_i^2)$, $(B_j^2)$ bs-Cauchy; assuming $(A_i^k) \sim (B_j^k)$ for $k = 1, 2, \ldots, n$ and using Lemmas 2, 3 and the relation $A^{n+1} = \frac{1}{2}[(A^n + A)^2 - A^{2n} - A^2]$, we obtain at once that $(A_i^{n+1}) \sim (B_j^{n+1})$. Thus for any polynomial $p$ we have $(p(A_i)) \sim (p(B_j))$. Consider now a continuous function $f$, a polynomial $p$ and let $\delta$ be the supremum of the set $\{|f(X) - p(X)| : |X| \leq M\}$, where $M$ is chosen to exceed the bounds of all $\|A_i\|$, $\|B_j\|$. Since the spectra of $A_i$ and $B_j$ are then in the interval $[-M, M]$ we have $\|fA_i - pA_i\|$ and $\|fB_j - pB_j\| < \delta$, $\|(f-p)^2 A_i\|$ and $\|(f-p)^2 B_j\| < \delta^2$; also, if $a$ is the bound of $f$ on $[-M, M]$, we have $\|fA_i\|$, $\|fB_j\| < a$ and $\|pA_i\|$, $\|pB_j\| < a + \delta$. Writing $fA_i - fB_j$ as $fA_i - pA_i + pA_i - pB_j + pB_j - fB_j$ we obtain $m[(fA_i - fB_j)^2] \leq m[(fA_i - pA_i)^2] + m[(pA_i - pB_j)^2] + 2[m(fA_i - pA_i)][m(pA_i - pB_j)] + m[(pB_j - fB_j)^2] + 2[m(fA_i - pB_j)][m(pB_j - fB_j)]$. Taking the previous inequalities into consideration as well as the relation $|m(C \cdot D)|^2 \leq m(C^2)m(D^2)$, we obtain $m[(fA_i - fB_j)^2] \leq 4\delta^2 + m[(pA_i - pB_j)^2] + 42\delta\{2a + \sqrt{m[(pA_i - pB_j)^2]}\}$. Given the function $f$ and $\epsilon > 0$, we choose a $\delta > 0$ so that $\delta < \epsilon/3$ and $4\delta^2 + 2\delta\,[2a + (\epsilon/3^{1/2})] < 2\epsilon/3$. Then we choose a polynomial $p$ within $\delta$ from $f$ on the interval $[-M, M]$. Finally, taking any state $m$ of $\mathfrak{A}$, we choose $i, j$ so large as to have $m[(pA_i - pB_j)^2] < \epsilon/3$; for the same $i, j$ we shall then have $m[(fA_i - fB_j)^2] < \epsilon$. So we finally turn to the relation $(A_i^2) \sim (B_j^2)$. We shall establish this using Lemma 3 and the process of "mixing" the two nets $(A_i)$, $(B_j)$. Let $I, J$ be the index set of these two nets and define $K$ as $\{(i, i_0, j_0) \mid i \geq i_0\} \cup \{(j, i_0, j_0) \mid j \geq j_0\}$; also define $(i, i_1, j_1) \geq (j, i_2, j_2)$ to mean $i_1 \geq i_2$ and $j_1 \geq j_2$ no matter what the first elements of the triples are. Clearly $K$ is a directed set. Now define $C_k$ to be $A_i$ if $k = (i, i_0, j_0)$ or $B_j$ if $k = (j, i_0, j_0)$ to obtain the "mixed" net, which is of course bounded. Consider two indices $k, k' \in K$; since $k = (i, i_0, j_0)$ or $(j, i_0, j_0)$, and $k' = (i', i_0', j_0')$ or $(j', i_0', j_0')$ the difference $C_k - C_{k'}$ will have one of the four forms $A_i - A_{i'}$, $B_j - A_{i'}$, $A_i - B_{j'}$, $B_j - B_{j'}$. Take any $n = (\ , i_1, j_1)$ and note that since $i \geq i_0$, $i' \geq i_0'$, $j \geq j_0$, $j' \geq j_0'$, the condition $k, k' \geq n$ implies $i, i' \geq i_1$ and $j, j' \geq j_1$. Since $(A_i)$ and $(B_j)$ are equivalent bs-Cauchy we see at once that $(C_k)$ is also bs-Cauchy, and by Lemma 3 we obtain $(C_k^2)$ bs-Cauchy. Now, given $\epsilon > 0$, take $n = (\ , i, j)$ so that $k, k' \geq n$ implies $m[(C_k^2 - C_{k'}^2)^2] < \epsilon$; if $i' \geq i, j' \geq j$ set $k = (i', i, j)$ and $k' = (j', i, j)$ so that $k, k' \geq n$ while $C_k = A_{i'}$ and $C_{k'} = B_{j'}$. Thus we have $m[(A_{i'}^2 - B_{j'}^2)^2] < \epsilon$, i.e., $(A_i^2) \sim (B_j^2)$.

As described previously we can now define $fX$ as the class of any net $(fA_i)$ such that $(A_i)$ determines $X$. We shall show that $\hat{\mathfrak{A}}$ thus becomes a Segal system.

*Theorem 10:* The map $(f, X) \to fX$ satisfies axioms 1 through 7, and $\hat{\mathfrak{A}}$ becomes a distributive Segal system. Furthermore the natural imbedding of $\mathfrak{A}$ into $\hat{\mathfrak{A}}$ commutes with the action of each $f$, so that $\hat{\mathfrak{A}}$ is actually an enlargement of $\mathfrak{A}$.

*Proof:* We start with axiom 2: given $X \in \hat{\mathfrak{A}}$ and continuous $f, g$ take any net $(A_i)$ which determines $X$; then $(gA_i)$ determines $gX$, so that $f(gX)$ is the class of $(f(gA_i))$ which the same as that of $((f \circ g)A_i)$, i.e., $(f \circ g)X$. For Axiom 3, suppose that $fX = fY$ for all $f$, and pick $(A_i)$, $(B_j)$ determining $X, Y$. There exists a constant $M$ exceeding the bounds of all $A_i, B_j$ and we choose for $f$ a function equal to the identity on $[-M, M]$. Then $fA_i = A_i$, $fB_j = B_j$ and since $fX$ is the class of $(fA_i)$ and $fY$ the class of $fB_j$ we have $X = fX$, $Y = fY$, or $X = Y$. The second part of 3 is obvious. For Axiom 4 we only need invoke Lemma 2. Now for Axiom 1: Suppose that $f = \sum_{n=1}^{\infty} f_n$ with the supports of the $f_n$ forming a locally finite system, $X \in \hat{\mathfrak{A}}$ and $(A_i)$ a net in the class $X$. If $M$ exceeds the bounds of all the $A_i$, we choose an $N$ such that for $n > N$ each $f_n$ vanishes on $[-M, M]$. Thus on $[-M, M]$ we have $f = \sum_{n=1}^{K} f_n$ provided $K > N$ and also $fA_i = \sum_{n=1}^{K} f_n A_i, f_n A_i = 0$ for $n > N$. Therefore $fX = \sum_{n=1}^{K} f_n X$ for any $K > N$. If we also have $g = \sum_{n=1}^{\infty} g_n$ with the supports of the $g_n$ forming a locally finite system, and $f_n X = g_n Y$ for all $n$, we obtain a similar relation $gY = \sum_{n=1}^{K} g_n Y$ for sufficiently large $K$, and thus $fX = gY$. Finally, letting $h(X) = X^2$ we see that $(f_1 f_2)X = \frac{1}{2}[(f_1 + f_2)^2 - f_1^2 - f_2^2] = \frac{1}{2}[h\circ(f_1 + f_2)X - h\circ f_1 X - h\circ f_2 X] = \frac{1}{2}[h\circ(f_1 + f_2)X - h(f_1X) - h(f_2X)]$ and since $f_k X = g_k Y$ we can retrace our steps to end up with $(g_1 g_2)Y$. To verify axiom 5, we need axiom 6 first. Let $X$ be the class of $(A_i)$ and $M$ a bound for $\|A_i\|$; since $fA_i = 0$ for $f$ vanishing outside of $[-M, M]$ we have $fX = 0$ for those $f$, hence the spectrum of $X$ is bounded. Now take any $X, Y \in \hat{\mathfrak{A}}$ be the classes of $(A_i)$, $(B_j)$ so that $X^2 + Y^2$ is the class of $(A_i^2 + B_j^2)$. Choose a nonnegative $C_{ij}$ so that $C_{ij}^2 = A_i^2 + B_j^2$; the net $(C_{ij}^2)$ being bs-Cauchy, we apply to it a continuous function which agrees with $X \to \sqrt{X}$ for $X \geq 0$ to obtain $(C_{ij})$ itself bs-Cauchy according to lemma 4.

If $Z$ is the class of $(C_{ij})$ we shall have $Z^2 = X^2 + Y^2$, so that by Theorem 9 Axiom 5 is valid.

Distributively of $\hat{\mathfrak{A}}$ follows trivially from that of $\mathfrak{A}$, as well as the relation $(aX)\cdot Y = a(X\cdot Y)$.

The last part of the theorem follows from the fact that if $A = \text{bs-lim } A_i$, then $fA = \text{bs-lim } fA_i$ as we can see immediately using Lemma 4 and the "mixing" process.

*Remark:* Given any $X \in \hat{\mathfrak{A}}$ we can find a net $(A_i)$ whose class is again $X$, for which $\|A_i\| \leq \|X\| = M$. Because take any net $(B_i)$ for $X$ and consider the function $f$ equal to the identity on $[-M, M]$, equal to $M$ on $[M, +\infty)$ and to $-M$ on $(-\infty, -M]$. Then $fX = X$, and setting $A_i = fB_i$ we have the class of $A_i$ equal to $X$ while $\|A_i\| \leq M$

### B. Properties of $\hat{\mathfrak{A}}$

In this section we shall establish a completeness theorem for $\hat{\mathfrak{A}}$; first we need to show how the states of $\mathfrak{A}$ extend to states of $\hat{\mathfrak{A}}$.

Let $m$ be a state of $\mathfrak{A}$ and $X \in \hat{\mathfrak{A}}$; for any net $(A_i)$ which determines $X$ we have at once that the numerical net $(mA_i)$ is Cauchy, since $|mA|^2 \leq m(A^2)$, and the same inequality shows that $(A_i) \sim (B_j)$ implies $\lim m(A_i) = \lim m(B_j)$. We write $mX$ for this number and observe that by Lemmas 2, 3, 4, the map $X \to mX$ will be linear, $m(fX) = \lim m(fA_i)$ and in particular $m(X^2) \geq 0$, i.e., it is a state of $\hat{\mathfrak{A}}$.

*Lemma 5:* If $m(X^2) = 0$ for all states $m$ of $\mathfrak{A}$, then $X = 0$.

*Proof:* Take any bs-Cauchy net $(A_i)$ which determines $X$, and note that $\lim m(A_i^2) = 0$; this means that the net $(A_i)$ and the net $(0)$ are equivalent, i.e., $X = 0$.

*Theorem 11:* Let $(X_i)$ be a bounded net in $\hat{\mathfrak{A}}$ such that for every state $m$ of $\mathfrak{A}$ we have $m[(X_i - X_j)^2] \to 0$. Then there exists a unique $X \in \hat{\mathfrak{A}}$ such that $m[(X - X_i)^2] \to 0$.

*Proof:* For each $i$ we choose a net $(A_{ij})_{j\in J_i}$ in $\mathfrak{A}$ which determines $X_i$ with $\|A_{ij}\| \leq \|X_i\|$. By the process in Appendix A, we convert the double net $(A_{ij})$ to a single net; thus we have the index set $K$ and for each $k \in K$ a pair of indices $i_k \in I$, $j_k \in J_{i_k}$; write $B_k$ for $A_{i_k j_k}$. We shall show that $(B_k)$ is bs-Cauchy and that its class $X$ satisfies the requirements of the theorem. We have

$$m[(B_k - B_l)^2] = m(A_{i_k j_k}{}^2) + m(A_{i_l j_l}{}^2) - 2m(A_{i_k j_k}\cdot A_{i_l j_l}).$$

From the hypothesis that $m[(X_i - X_{i'})^2] \to 0$ we have at once that $m(X_i^2 - X_{i'}^2) \to 0$ since $|m(X_i^2 - X_{i'}^2)|^2 \leq 4M^2 m[(X_i - X_{i'})^2]$ where $M$ is a bound for $\|X_i\|$. This means that $\lim_i \lim_j m(A_{ij}^2)$ exists, hence that $\lim_k m(A_{i_k j_k}{}^2)$ and $\lim_l m(A_{i_l j_l}{}^2)$ exist and are equal. Now consider the iterated limit $\lim_{i,i'} \lim_{j,j'} m(A_{ij}\cdot A_{i'j'})$; by looking at the squares we see that $\lim_{j,j'} m(A_{ij}\cdot A_{i'j'}) = m(X_i \cdot X_{i'})$. Since $2X_i\cdot X_{i'} = X_i^2 + X_{i'}^2 - (X_i - X_{i'})^2$ we have that $\lim_{i,i'} m(X_i \cdot X_{i'}) = \lim_i m(X_i^2)$. This implies that the converted single limit of $m(A_{i_k j_k}\cdot A_{i_l j_l})$ will also be $\lim_i m(X_i^2) = \lim_k m(A_{i_k j_k}{}^2)$. Putting all these together we have $m[(B_k - B_l)^2] \to 0$. We now calculate $\lim_i m[(X_i - X)^2]$. Since $(X_i - X)^2$ is the class of $(A_{ij} - B_k)^2$ ($i$ fixed) we have $m[(X_i - X)^2] = \lim_{jk} m[(A_{ij} - B_k)^2]$ which equals the iterated limit $\lim_j \lim_k m[(A_{ij} - B_k)^2]$. Thus we consider the triple iterated limit $\lim_i \lim_j \lim_k$ of the same quantity. We convert the first two to obtain the expression $\lim_l \lim_k m[(A_{i_l j_l} - B_k)^2] = \lim_l \lim_k m[(A_{i_l j_l} - A_{i_k j_k})^2]$ which exists and is zero. By the second part of the theorem in Appendix A, we conclude that the triple iterated limit mentioned above also exists and is zero. Finally we show uniqueness of $X$. If we have a $Y \in \hat{\mathfrak{A}}$ such that $m[(X_i - Y)^2] \to 0$, then, using the fact that $T \to \sqrt{m(T^2)}$ is a seminorm, we obtain $m[(X - Y)^2] = 0$ for all states $m$ of $\mathfrak{A}$, hence by Lemma 5 that $X = Y$.

It will be notationally convenient to denote the element $X$ for which $m[(X_i - X)^2] \to 0$ as $\lim_i X_i$. Note that the usual rules hold and that by a polynomial approximation we also have $\lim_i fX_i = fX$, $\lim_i f(X_i - X) = 0$.

### C. Action of the Borel functions

Our next goal is to define the action of a locally bounded Borel function on the arbitrary element $X$ of $\hat{\mathfrak{A}}$. We shall require the following lemma, which ought to be well known, but does not seem to appear in the literature.

*Lemma 6:* Let $f$ be a bounded Borel function. There exists a net of continuous functions $f_i$, uniformly bounded by the bounds of $f$, which converges to $f$ pointwise and such that for any countably additive finite Borel measure $\mu$ on the line we have $\int (f_i - f)^2 d\mu \to 0$.

*Proof:* Recall the classification of Borel functions into classes (Ref. 20): $f$ is of class 0, iff it is continuous: if $\alpha > 0$ is an ordinal, then $f$ is of class $\alpha$ iff if is a pointwise limit of a sequence of functions of class $< \alpha$, but is not itself of class $< \alpha$. The ordinals needed to exhaust all Borel functions range up to the first uncountable. Since the result holds for $f$ of class 0, we assume it holds for any $f$ of class $< \alpha$ and that $f$ is of class $\alpha$. Take a sequence of function $f$ of class $< \alpha$ converging to $f$ pointwise; replacing, if need be, $f_n$ by mid $[\inf f, f_n, \sup f]$ we have that all $f_n$ have the same bounds as $f$. By the dominated theorem we then have $\int (f_n - f)^2 d\mu \to 0$. Now for each $n$ there is a net

$(f_{nj})_{j \in J_n}$ of continuous functions satisfying the requirements of the theorem. They are evidently bounded by the bounds of $f$, and $\lim_n \lim_j f_{nj} = f$ pointwise. Now $\int (f_{nj} - f)^2 = \int (f_{nj} - f_n)^2 + \int (f_n - f)^2 - 2 \int (f_n - f)(f_{nj} - f_n)$; the last term is bounded by $4M \int |f_n - f| d\mu$ which tends to zero since $f_n$ converges boundedly to $f$. Thus $\lim_n \lim_j \int (f_{nj} - f)^2 d\mu = 0$. All we have to do now is to convert the double net $(f_{nj})$ to a single net by the process of Appendix A.

Consider a locally bounded Borel function $f$ and an element $X$ of $\hat{\mathfrak{A}}$. As we shall see, only the values of $f$ on the spectrum of $X$ will play a role in defining $fX$, and so we shall take $f$ to be zero outside $[-\|X\|, \|X\|]$. Take a net of continuous $f_i$ provided for by Lemma 6 and a state $m$ of $\mathfrak{A}$. According to Theorem 6, we have a probability measure $p_{X,m}$ at our disposal such that $m(gX) = \int g dp_{X,m}$. Thus we have at once that $\int (f_i - f_j)^2 dp_{X,m} \to 0$, i.e., $m[(f_i X - f_j X)^2] \to 0$. But the norms $\|f_i X\|$ are bounded by the bound of $|f|$, so that $\lim_i f_i X$ exists. If we consider any other net of continuous $g_k$ converging to $f$ we have $\int (f_i - g_k)^2 dp_{X,m} \to 0$ which shows that $m[(\lim_i f_i X - \lim_k g_k X)^2] = 0$ for all states $m$ of $\mathfrak{A}$; therefore the element of $\hat{\mathfrak{A}}$ obtained is independent of the choice of the net converging to $f$ and is completely and uniquely determined by $f$ and $X$. We shall write this as $fX$.

*Theorem 12:* The map $(f, X) \to fX$ defined above has the following properties:

(i) $(f \circ g)X = f(gX)$.

(ii) $(f + g)X = fX + gX$, and if $f_i X = g_i Y$, then $(f_1 f_2)X = (g_1 g_2)Y$.

(iii) If $(f_n)$ is an increasing sequence of functions with pointwise limit $f$, then $fX = \lim_n f_n X$.

*Proof:* Actually only (i) is not obvious from the definition. So we consider nets $f_i \to f$ and $g_j \to g$ of continuous functions, satisfying the requirements of lemma 6; we may take $f_i$ zero outside the interval $[-\|gX\|, \|gX\|]$. Since $\lim_j g_j X = gX$, we have for each $i$ that $\lim_j f_i(g_j X) = f_i(gX)$, so that $\lim_i \lim_j f_i(g_j X) = f(gX)$. We also have $\lim_i \lim_j f_i \circ g_j = f \circ g$, so that if we show that $\lim_i \lim_j \int (f_i \circ g_j - f \circ g)^2 d\mu = 0$ we can convert the iterated limit to a single limit to obtain $\lim_k (f_{i_k} \circ g_{j_k})X = (f \circ g)X$ and thus $f(gX) = f \circ g)X$. We have $(f_i \circ g_j - f \circ g)^2 = (f_i - f)^2 \circ g + 2(f_i \circ g_j - f \circ g)[(f_i - f) \circ g] + [f_i \circ g_j - f_i \circ g]^2$; take any finite measure $\mu$ and let $\nu$ be the measure $E \to \mu(g^{-1}E)$ so that $\int h d\nu = \int (h \circ g) d\mu$; the first term then, integrated by $\mu$ will have $\lim_i \lim_j$ equal to 0. The second term when integrated by $\mu$ will be bounded by a constant (since all functions are uniformly bounded) times $\int |f_i - f| d\nu$ which goes to zero too, since its square is bounded by $\int (f_i - f)^2 d\nu$. Now consider the last term and take a polynomial $p$ within $\delta$ from $f_i$ on an interval containing the ranges of all $g_j, g$; this last term will then be $\leqq 2\delta^2 + 2\delta$ (constant bound) $+ (p \circ g_j - p \circ g)^2$; making $\delta$ small, choosing $p$ accordingly and noting that $p \circ g_j - p \circ g$ is bounded by a constant times $g_j - g$ we see that for every $i$ we have $\lim_j \int (f_i \circ g_j - f_i \circ g)^2 d\mu = 0$, which completes the proof.

## III. IMBEDDING IN A MACKEY SYSTEM

### A. Idempotents in a distributive system

We shall first study in this section the properties of idempotent elements in a distributive system satisfying Axioms 1–6 and then apply this knowledge to the completion of a Segal system.

First note that the natural order of $\mathfrak{A}$ (given by the cone of all squares) is inherited by the set $\mathfrak{L}$ of all idempotents, and that $O$ and $I$ are the smallest and largest elements of $\mathfrak{L}$. Also $P \in \mathfrak{L}$ implies $I - P \in \mathfrak{L}$ while the map $P \to P' = I - P$ is an involution for which $P \leq Q$ implies $Q' \leq P$.

*Theorem 13:* The sum of two idempotents is an idempotent iff it is $\leq I$, iff their product is 0. Also, for $P, Q$ idempotents we have $Q - P$ an idempotent iff $P \leqq Q$; thus in particular $P \leq Q'$ iff $P \cdot Q = 0$.

*Proof:* Let $P^2 = P, Q^2 = Q, R = P + Q, T = P - Q$. By distributivity we have $2(P \cdot Q) = (P + Q)^2 - P - Q$ and $(P - Q)^2 = P + Q - 2(P \cdot Q)$ so that $T^2 = R \cdot (2I - R)$ or $(I - R)^2 = I - T^2$. Now suppose that $R \leq I$, so that $(I - T^2)^{1/2} = I - R$; then we obtain $P = \frac{1}{2}[R + T] = \frac{1}{2}[I + T - (I - T^2)^{1/2}]$ and $Q = \frac{1}{2}[R - T] = \frac{1}{2}[I - T - (I - T^2)^{1/2}]$. Idempotency of $P$ now yields $T \cdot (I - T^2)^{1/2} = 0$ which means that the spectrum of $T$ is contained in the set $\{0, -1, 1\}$; since $R = I - (I - T^2)^{1/2}$ the spectrum of $R$ is contained in $\{0, 1\}$ which means that $R$ is an idempotent.

The converse is clear, since for $R$ idempotent we have $I - R$ also, hence $I - R \geq 0$, or $R \leq I$. Equally obvious is the condition $P \cdot Q = 0$.

Now for the second part note that $Q - P$ positive is implied by idempotency; so suppose $Q \geq P$ so that the sum of the idempotents $I - Q, P$ is $\leq I$, hence $I - Q + P$ is an idempotent and $Q - P = I - (I - Q + P)$ is also.

*Theorem 14:* For any finite set of idempotents $P_i$ which are pairwise disjoint (either in the sense $P_i \leq P_j'$ or $P_i \cdot P_j = 0$ for $i \neq j$) their sum is an idempotent and is also their supremum.

*Proof:* Let $P = \sum_{i=1}^n P_i$, so that $P_i \leq P$ for all $i$; distributivity of course implies that $P$ is an idempotent, so let $R \geq P_i$ be an idempotent. Then we have that $I - R$ has product zero with each $P_i$, hence the sum $I - R + P$ will be an idempotent; this means $I - R + P \leq I$ or $P \leq R$.

So we have established that the set of all idempotents forms in a natural way an orthomodular logic.

### B. The imbedding theorem

We shall now study the logic of idempotents of the completion of a distributive Segal system $\mathfrak{A}$. This completion $\hat{\mathfrak{A}}$ will contain a large supply of idempotents due to the action of characteristic functions.

*Theorem 15:* In the set $\mathfrak{L}$ of idempotents of $\hat{\mathfrak{A}}$, any family $(P_j)$ of pairwise disjoint idempotents admits a supremum $P$; furthermore, for any state $m$ of $\mathfrak{A}$ we have $m(P) = \sum_j m(P_j)$.

*Proof:* Consider the set $K$ of all finite subsets of he given index set, partially order it by inclusion, and construct the net $k \to Q_k$, where for $k = \{j_1, j_2, \ldots, j_n\}$ we set $Q_k = \sum_r P_{j_r}$. Evidently $k_1 \leq k_2$ implies $m(Q_{k_1}) \leq m(Q_{k_2}) \leq 1$, and hence $\lim_k m(Q_k)$ exists; call it $a$. Now $(Q_{k_1} - Q_{k_2})^2 = Q_{k_1} + Q_{k_2} - 2Q_{k_1} \cdot Q_{k_2}$ and we see that $Q_{k_1} \cdot Q_{k_2} = Q_{k_1 \cap k_2}$ by the distributive law; hence, for $k_1, k_2 \geqq k$ we also have $k_1 \cap k_2 \geqq k$ and thus $m[(Q_{k_1} - Q_{k_2})^2]$ converges to $a + a - 2a = 0$. Since $\|Q_k\| \leq 1$, we have by Theorem 11 an element $P \in \hat{\mathfrak{A}}$ with $\lim_k Q_k = P$. As all $Q_k$ are idempotents, so is $P$. To show that $P$ is the supremum of the $P_j$, suppose that for some idempotent $R$ we have $P_j \le R$ (all $j$); then we also have $Q_k \le R$ for all $k$, and $R - Q_k$ is an idempotent; hence $R - P = \lim_k(R - Q_k)$ will be an idempotent, which implies $R \ge P$. On the other hand, $P - P_j = \lim_k (Q_k - P_j)$; since for $k \ge j$ we have $Q_k \ge P_j$ we obtain $P - P_j$ an idempotent and thus $P \ge P_j$. For the last part note that $m(P) = \lim_k m(Q_k)$, $m(Q_k) = \sum_r m(P_{j_r})$, or $m(P) = \sum_j m(P_j)$.

Recall that, given a $\sigma$-logic $\mathcal{L}$ we define an observable based on $\mathcal{L}$ as a $\sigma$-homorphism $u$ from the Borel sets of the line to elements of $\mathcal{L}$; a Borel function $f$ acts on such an observable by composition $fu = u \circ f^{-1}$.

*Theorem 16:* For $X \in \hat{\mathfrak{A}}$ define $u_X$ as $E \to \chi_E(X)$; then $u_X$ is a bounded observable based on $\mathcal{L}$, the map $X \to u_X$ is 1 : 1 onto the set of all bounded observables and commutes with the action of locally bounded Borel functions; finally for each state $m$ of $\mathfrak{A}$ we have $m(fX) = \int f dq_{X,m}$, where $q_{X,m} = m \circ u_X$ and is thus the probability distribution of $X$ in the state $m$.

*Proof:* By Theorem 12 and the previous theorem we see that $u_X$ is indeed a $\sigma$-homomorphism. Now partition the interval $J = [-\|X\|, \|X\|]$ into sufficiently small subintervals $J_k$ so that on $J$ we have $|j - \sum_k x_k \chi_{J_k}| < \epsilon$ for any $x_k \in J_k$ (where $j$ is the identity function). Then we obtain that $\|X - \sum_k x_k u_X(J_k)\| \le \epsilon$, which of course implies that if $u_X = u_Y$, then $\|X - Y\| \le \epsilon$, i.e., $X = Y$. So our map is 1 : 1. Now to show that this map is onto, consider any bounded observable $u$ based on $\mathcal{L}$; suppose that $u(E) = 0$ for $E$ disjoint from the interval $(-a, a)$. Take as above step functions $f_n$ converging uniformly to the identity function $j$ on $[-a, a]$, $f_n = \sum_i a_{ni} \chi_{J_{ni}}$, and set $X_n = \sum_i a_{ni} u(J_{ni}) \in \hat{\mathfrak{A}}$; note that $\|X_n\| \le a$. Since $\|X_n - X_k\| \le \sup\{|f_n(x) - f_k(x)| : x \in [-a, a]\} \to 0$, we have for each state $m$ of $\mathfrak{A}$ that $m[(X_n - X_k)^2] \to 0$; let $X = \lim_n X_n$. We shall show that $u = u_X$ and thus establish that our map is onto. Since both sides are $\sigma$-homomorphisms, it suffices to show $u(J) = u_X(J)$ for all open intervals $J \subseteq [-a, a]$. Consider continuous $g_k$ vanishing outside $J$, nonnegative and converging increasingly to $\chi_J$. For any finite measure $\mu$ we have $\int (g_k - \chi_J)^2 d\mu \to 0$, hence $u_X(J) = \lim_k g_k(X)$. On the other hand, we have $g_k(X) = \lim_n g_k(X_n)$ so that $u(J) - u_X(J) = \lim_k \lim_n [u(J) - g_k(X_n)]$. Now $X_n = \sum_i a_{ni} u(J_{ni})$ and by a polynomial approximation we have $g_k(X_n) = \sum_i g(a_{ni}) u(J_{ni} \cap J)$ since $g_k$ vanishes outside $J = \cup_i (J_{ni} \cap J)$. So we end up with the relation $[u(J) - u_X(J)]^2 = \lim_k \lim_n \sum_i [1 - g_k(a_{ni})]^2 u(J_{ni} \cap J)$. For any state $m$ of $\mathfrak{A}$ we therefore have $m[(u(J) - u_X(J))^2] = \lim_k \lim_n \sum_i [1 - g_n(a_{ni})]^2 \mu(J_{ni} \cap J)$, where $\mu = m \circ u$. The limit over $n$ will produce $\int_J (1 - g_k)^2 d\mu$ and since the limit of this over $k$ is 0 (as above) we obtain from Lemma 5 that $u(J) = u_X(J)$.

To correlate the action of $f$ we calculate: $u_{fX}(E) = \chi_E(fX) = (\chi_E \circ f) X = \chi_{f^{-1}E} X = u_X(f^{-1}E) = (u_X \circ f^{-1})E$. The relation $m(fX) = \int f dq_{X,m}$ follows at once by an argument similar to that at the very beginning of the proof.

## IV. EXAMPLES AND CONCLUDING REMARKS

### A. Examples

We present the following two examples of "complete" Segal systems. In the first all hypotheses are satisfied but the system is far from being close to a $C^*$-algebra, as it is reflexive as a Banach space. The second is not distributive, but is still imbeddable in a Mackey system; thus distributively is not necessary. Perhaps the most interesting fact about these systems is that they produce isomorphic logics, so that in general a Segal system is not determined by its idempotents.

*First example:* Let $\mathcal{H}$ be a Hilbert space with inner product $\langle | \rangle$ and $\mathfrak{A}$ the set $\mathcal{R} \times \mathcal{H}$, where $\mathcal{R}$ is the reals with the product vector space structure. We define $(t, x)^2$ to be $(t^2 + \|x\|^2, 2tx)$ and $\|(t, x)\|$ to be $|t| + \|x\|$. It follows that $(t, x)\cdot(s, y) = (ts + \langle x | y \rangle, ty + sx)$ so that it is distributive with identity $I = (1, 0)$. The states are in a 1 : 1 correspondence with the vectors $y$ of the unit ball in $\mathcal{H}$, the state determined by $y$ being $(t, x) \to t + \langle x | y \rangle$. Thus strong convergence is equivalent to convergence in the norm, which implies that axiom 7* is satisfied while $\mathfrak{A}$ is $\mathfrak{A}$ itself. The idempotents in $\hat{\mathfrak{A}}$ are easily seen to have the form $0$, $I$ or $(\frac{1}{2}, x)$ where $\|x\| = \frac{1}{2}$; also, the only order relations that hold are of the form $0 \le P \le I$ for any idempotent $P$. Finally, every element $(t, x) = (t - \|x\|)(1, 0) + 2\|x\|(\frac{1}{2}, x/2\|x\|)$ for $x \ne 0$ or $t(1, 0)$ so that every element is a linear function of some idempotent.

*Second example:* Although the construction carries over to any finite dimension we consider a three-dimensional case. Let $\mathfrak{A} = \mathcal{R}^3$ and define $(x, y, z)^2$ to be $(2xz, 2yz, x^2 + y^2 + z^2 + 2|xy|)$, $\|(x, y, z)\|$ to be $|x| + |y| + |z|$. The product is not distributive because of the presence of the absolute value in the third component of the square; the identity is $(0, 0, 1)$. Because of the finite dimension of $\mathfrak{A}$ Axiom 7* holds here, although, of course, it is largely irrelevant. The idempotents of the system have the form $0$, $I$, or $(\frac{1}{2}, y, z)$ with $|y| + |z| = \frac{1}{2}$. Again the only order relations are the trivial ones, and every element is a linear function of some idempotent.

In either case the logic $\mathcal{L}$ is the union of four-element Boolean algebras with common 0 and $I$; thus if the number of nontrivial idempotents in the two systems is the same, the logics will be isomorphic; this can be guaranteed by taking the dimension of the Hilbert space in the first example low enough. Finally note that there are states of the logic which are not Segal states of the given system.

### B. Remarks

There are several questions that arise in the present context. Here are some

1. Can the same conclusion be obtained with weaker hypotheses? What exactly is the role of the distributive law?

2. What further properties does the derived logic have? Is it full? Is it a lattice?

3. Is it possible to define a sum for the bounded observables based on a logic constructed by this process?

4. What kind of Segal systems can be recaptured from their idempotents?

## ACKNOWLEDGMENT

The writer is indebted to the referee who, by pointing out an erroneous statement in the original version, motivated the writer to obtain the desired results under significantly weaker hypotheses.

## APPENDIX A: CONVERSION OF ITERATED LIMITS

We collect here for the reader's convenience a few facts about iterated limits, which we use repeatedly in our arguments. Most of what follows can be found in Ref. 21.

A set $I$ is directed by a relation $\leqq$, if $\leq$ is transitive and for $i', i'' \in I$ there exists an $i \in I$ with $i', i'' \leqq i$. A net is a map from a directed set $I$ to some set $S$. Let $I$ be directed by $\leq$ and, for each $i \in I$, let $J_i$ be directed by some relation which we shall again write as $\leqq$ (no confusion seems likely). Write $J$ for the cartesian product $X_{i \in I} J_i$, and for $f, g \in J$ define $f \leq g$ to mean $f(i) \leq g(i)$ for each $i \in I$ ("product order" on $J$); finally, let $K = I \times J$ and impose the produce order on $K$, so that it will become directed. For $k \in K$, we have $k = (i, f)$; write $i_k$ for the first member of the pair and $j_k$ for the element $f(i_k)$ of $J_{i_k}$.

Now let $S$ be a regular Hausdorff topological space, and $s_{ij} \in S$ for each $i \in I$ and $j \in J_i$.

*Theorem:* Suppose that $\lim_j s_{ij} = s_i$ exists for each $i \in I$; them $\lim_i s_i$ exists iff $\lim_k s_{i_k, j_k}$ exists, in which case they are equal.

*Proof:* The implication from $\lim_i \lim_j s_{ij}$ to $\lim_k s_{i_k, j_k}$ is established in Ref. 21. We shall show the converse. Let $s$ be the limit of $(s_{i_k, j_k})$, take a neighborhood $U$ of $s$ and by regularity choose a closed neighborhood $V$ of $s$ contained in $U$. There is a $k_0 \in K$ such that if $k \geq k_0$, then $s_{i_k, j_k} \in V$. Write $k = (i, f)$, $k_0 = (i_0, f_0)$ to obtain $i \geq i_0$ and $f(i') \geq f_0(i')$ for all $i' \in I$. Consider an $i_1 \geq i_0$; we shall show that $s_{i_1} \in U$. Given $j \geq f_0(i_1)$ $(j \in J_{i_1})$ we construct an $f \in J$ by $f(i_1) = j$ and $f(i) = f_0(i)$ for all other $i$; thus $f \geq f_0$, which means that $(i_1, f) \geq (i_0, f_0)$ so that $s_{i_1, j_1} \in V$. In other words, $j \geq f_0(i_1)$ implies $s_{i_1, j} \in V$ and since $V$ is closed we must have its limit in $V$, i.e., $s_{i_1} \in V \subseteq U$. So given $U$ we have an index $i_0$ for which $i \geq i_0$ implies $s_i \in U$.

## APPENDIX B: CASE OF $C^* \rightarrow$ ALGEBRAS

Let $\mathfrak{A}$ be an abstract $C^*$-algebra. By the representation theorem of Gelfand it is isomorphic and isometric to an algebra of operators on a Hilbert space so that every state of $\mathfrak{A}$ has the form $A \rightarrow \langle Au | u \rangle$ for some unit vector $u$ (Ref. 22). Now let $(A_i)$ be a bounded net in $\mathfrak{A}$ and suppose that it is bs-Cauchy, i.e., $\langle (A_\iota - A_j)^2 u | u \rangle \rightarrow 0$ for all vectors $u$. This means that the net $(A_i u)$ is norm-Cauchy in the Hilbert space; let $Au$ be its limit. Evidently $A$ is a linear transformation and $\|Au\| \leq \lim\sup \|A_i u\| \leq M\|u\|$ since the net $(A_\iota)$ is bounded. By a theorem of Kaplansky (Ref. 23) we then have $A_i^2 \rightarrow A^2$ in the strong operator topology, so that in particular $\|(A_i^2 - A_j^2)u\| \rightarrow 0$, or $m[(A_i^2 - A_j^2)^2] \rightarrow 0$. Thus Axiom 7* holds for the Segal system of self-adjoint elements of a $C^*$-algebra. The above argument also shows that in this case the completion of $\mathfrak{A}$ can be identified with a system of operators.


[1] I. E. Segal, Ann. Math. **48**, 939 (1947).
[2] I. E. Segal, *A Mathematical Approach to Elementary Particles and their Fields* (University of Chicago Lecture Notes, Chicago, Ill., 1955).
[3] G. W. Mackey, *The Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963).
[4] S. Sherman, Proc. Am. Math. Soc. **2**, 31 (1951).
[5] D. Lowdenslager, Proc. Am Math. Soc. **8**, 88 (1957).
[6] S. Sherman, Ann. Math. **64**, 593 (1956).
[7] N. Zierler, Pac. J. Math. **11**, 1151 (1961).
[8] V. S. Varadarajan, Commun. Pure Appl. Math. **15**, 189 (1962).
[9] C. Piron, Helv. Phys. Acta **37**, 439 (1964).
[10] S. P. Gudder, Trans, Am. Math. Soc. **119**, 428 (1965).
[11] S. P. Gudder, Pac. J. Math. **19**, 81 (1966).
[12] A. Ramsey, J. Math. Mech. **15**, 227 (1966).
[13] M. D. MacLaren, Pac. J. Math. **14**, 597 (1964).
[14] D. J. Foulis, Proc. Am. Math. Soc. **11**, 648 (1960).
[15] C. J. T. Poole, Commun. Math. Phys. **9**, 118 (1968).
[16] P. C. Deliyannis, J. Math. Phys. **10**, 2114 (1969).
[17] R. J. Plymen, Helv. Phys. Acta **41**, 69 (1968).
[18] R. J. Plymen, Commun. Math. Phys. **8**, 132 (1968).
[19] E. T. Kehlet, Math. Scand. **25**, 57 (1969).
[20] C. Goffman, *Real Functions* (Prindle, Weber, Schmidt, Boston, 1953).
[21] J. L. Kelley, *General Topology* (Van Nostrand, New York, 1955).
[22] J. Dixmier, *Les C*-algebres* (Gauthier-Villars, Paris, 1964).
[23] I. Kaplansky, Pac. J. Math. **1**, 227 (1951).

# Erratum: Application of infinite order perturbation theory in linear systems. II [J. Math. Phys. 15, 947 (1974)]

Shi-yu Wu

*Department of Physics, University of Louisville, Louisville, Kentucky 40208*
(Received 13 September 1974)

Equation (6) should read as follows:

$$P(Z)=\sum_i c_i \frac{1}{|(A_i-Z)|^4}\, P\,\frac{1}{A_i-Z}\ .$$

# Erratum: On analytic nonlocal potentials. I. A forward dispersion relation [J. Math. Phys. 14, 1141 (1973)]

Te Hai Yao

*Mathematics Department, Bedford College, Regent's Park, London NW 1, England*
(Received 30 August 1974)

Formula (3.10) should read

$$\lim_{R\to\infty}\int_{C_3} dx'\,p(i\kappa; x, x',\ \cos\nu,\ \cos\theta')=0. \qquad (3.10)$$