Canonical formalism for degenerate Lagrangians

S. Shanmugadhasan

Division of Physics, National Research Council of Canada, Ottawa, Canada K1A 0R6
(Received 15 September 1972)

A Lagrangian is degenerate when the Hessian matrix whose elements consist of all the second-order derivatives of the Lagrangian with respect to the generalized velocities has (for simplicity) a constant singular rank everywhere in the space of the arguments of the Lagrangian. This singularity entails a definite number of first-order Lagrange equations, which act as subsidiary conditions on the coordinates and velocities. Consistency of these subsidiary conditions with the Langrange system requires them to be an invariant system with respect to the second-order Lagrange equations. An invariant system is analogous to a system of first integrals except that absolute constants appear where arbitrary constants characterize first integrals. The usual definitions of momenta and of Hamiltonian make the Hamiltonian a function of the functionally dependent canonical variables only. Introduction of the momentum variables into the subsidiary conditions on the coordinates and velocities yields under certain circumstances additional subsidiary conditions on the canonical variables only. All the subsidiary conditions on the canonical variables are determined before setting up the multiplier rule for the canonical equations of motion. The multiplier rule is exploited to deduce the invariant system among the subsidiary conditions, the explicit modifications of the canonical equations by the other susidiary conditions, and Dirac's formula for the corresponding modified Poisson brackets. The modifications are caused by the reduction in the number of independent canonically conjugate pairs. A canonical transformation adapted to the subsidiary conditions, which is found with the help of Lie's theory of function groups, transforms the canonical system from the multiplier rule into a canonical system in terms of physical variables. The invariant system is then used to reduce the order of the resulting canonical system by following Levi-Civita. This reduced canonical system is suitable for integration or quantization.

1. INTRODUCTION

A Lagrangian is called degenerate if the Hessian matrix whose elements consist of all the second-order partial derivatives of the Lagrangian with respect to the generalized velocities has (for simplicity) a constant singular rank everywhere in the space of the arguments of the Lagrangian. For such a Lagrangian the generalized momenta satisfy a number of functional-dependence relations; this number depends on the rank of the Hessian matrix. Dynamical systems with degenerate Lagrangians are of practical importance in physics.

The first problem in the general study of systems with degenerate Lagrangians is to reduce the Lagrange equations of motion to canonical form and then to write the canonical equations in terms of Poisson brackets. The theory created by Dirac,[1] Anderson and Bergmann,[2] Haag,[3] and many others is widely accepted as the solution of the problem. Kundt[4] surveys this theory in his own development. It must be emphasized, however, that this theory falls short of its goals. First, the current theory does not correctly handle the subsidiary conditions on the canonical variables that arise from the first-order Lagrange equations of motion. Since all the subsidiary conditions must be known before setting up the multiplier rule, the multiplier rule cannot be used to get these conditions. Second, the theory does not explicitly deduce the effect of the subsidiary conditions on the canonical equations of motion and on the Poisson brackets. Explicit deduction would facilitate the interpretation of the modifications caused by the subsidiary conditions. Third, the simplification of the integration problem by making use of the nature of those subsidiary conditions which remain after exploiting the multiplier rule has not been done in the general case, only for special forms of the conditions. The present paper seeks the necessary improvements to the current theory by amending an earlier attempt.[5]

Subsidiary conditions occur in two distinct types. The familiar type of subsidiary condition can be satisfied only by modifying the equations of motion. This modification amounts to applying the implicit-function theorem to the subsidiary conditions, that is, solving for some of the variables in terms of the independent variables. The other type of subsidiary condition is compatible with the equations of motion without the subsidiary condition, so that the equations of motion are not changed by the subsidiary condition. The variables can be treated as independent in spite of the subsidiary condition. But one must adjoin these subsidiary conditions to the unmodified equations of motion. Only those solutions of the unmodified equations which satisfy these subsidiary conditions are admissible. Such subsidiary conditions form what is called an invariant system[6] with respect to the differential equations of motion. The notions of invariant relation and of invariant system go back to C. J. G. Jacobi, P. Painlevé (who called them particularized first integrals), and Levi-Civita,[7] who developed some aspects of the theory of invariant systems. These important notions are treated in detail in the next section.

Subsidiary conditions appear at two stages in the canonical formulation of the equations of motion from a degenerate Lagrangian. The degeneracy of the Lagrangian gives rise to first-order Lagrange equations of motion, which act as subsidiary conditions on the coordinates and velocities. For consistency of the Lagrange equations these subsidiary conditions must form an invariant system with respect to the second-order Lagrange equations. Because of the assumption about the rank of the Hessian matrix, these subsidiary conditions cannot lower the rank of the singular Hessian matrix. So there are no other subsidiary conditions on the coordinates and velocities. On introducing the momentum variables the singular rank of the Hessian matrix yields a definite number of subsidiary conditions on the coordinates and momenta. If the momentum variables are introduced into the subsidiary conditions on the coordinates and velocities, the resulting equations in general contain some of the velocity variables. But under the right conditions they can yield subsidiary conditions on the canonical variables only. So one must see whether any additional subsidiary conditions on the canonical variables arise in this way. The complete set of subsidiary conditions on the canonical variables can be deduced in these two ways.

The multiplier rule is designed to take account of given subsidiary conditions without having to solve them explicitly for the independent variables. The actual determination of the multiplier functions in the multiplier rule gives the required information on the type of each subsidiary condition, the modifications of the canonical equations caused by the subsidiary conditions, and the consequent modification of the Poisson brackets. Thus the modification in the Poisson brackets arises from the reduction in the number of independent canonical variables. Djokić-Ristanović and Mušicki[8] have pointed out an error in the earlier deduction[5] of the formula for the modified Poisson bracket. The correct deduction gives the same formula as that postulated by Dirac.[1] Bergmann and Goldberg,[9] Mukunda and Sudarshan,[10] and Hermann[11] have studied this generalized Poisson bracket from other points of view.

The multiplier rule yields a canonical differential system with an invariant system adjoined to it. Finding the admissible solutions of these equations is not easy. Levi-Civita,[7] however, has indicated how, by passing to new canonical variables, the invariant system can be used to reduce the order of the canonical system and thereby simplify the integration problem. The method can be adapted to the problem in hand. The required canonical transformation, which is also necessary for the treatment of the modified Poisson bracket, is obtained with the help of Lie's theory[12,13] of function groups. Dirac,[1] Anderson[14], and B. S. DeWitt (in unpublished work) did this type of reduction only when the invariant system had a specially simple form. In fact, Levi-Civita's method reduces any general invariant system to this simple form by the above-mentioned canonical transformation. The new canonical variables in the reduced canonical system directly describe the allowed motions of the dynamical system and are therefore physical variables of the dynamical system. The space of these variables obviously has symplectic structure (see Kundt[4] and Künzle[15]).

2. DEFINITION OF INVARIANT SYSTEM

The summation convention is used in this paper: sum over all values of a suffix whenever it occurs twice in a term. Denote a collection of variables like

$$x_i, \quad i = 1, \ldots, n,$$

by x. A function is said to be of class $C^{(m)}$ in a domain if all its partial derivatives of order m exist and are continuous at each point of the domain.

Consider the normal system of differential equations for the x_i:

$$\frac{dx_i}{dt} = f_i(t,x), \quad i = 1, \ldots, n, \tag{1}$$

where t is the independent variable and the functions f_i are continuous and have continuous partial derivatives with respect to x on an open set in the space of the variables t, x, so that the theorem of existence and uniqueness of the solution of the differential system holds. A function $u(t,x)$ of class $C^{(1)}$ in the domain under consideration and not identically equal to a constant is called an integral of the differential system (1) if it becomes a constant with respect to t on substituting for the x_i any solution of the system whose integral curve lies wholly in the domain. The value of this constant depends on the choice of the solution of the system, being in general different for different solutions; it is a function of the initial conditions. The equality

$$u(t,x) = C, \tag{2}$$

where u is an integral of the system and C is an arbitrary constant, is called a first integral of the system. All solutions of the system satisfy the first integral. A necessary and sufficient condition for (2) to be a first integral of the system (1) is the identical vanishing of the total differential of u, or equivalently the total derivative of u with respect to t, by virtue of the system (1). That is,

$$\frac{du(t,x)}{dt} = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x_i}\frac{dx_i}{dt} = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial x_i} f_i(t,x) \equiv 0. \tag{3}$$

Now consider a relation of the form

$$v(t,x) = 0, \tag{4}$$

where the function v (not identically zero) is of class $C^{(1)}$. The relation (4) is assumed to determine a hypersurface in the $(n+1)$-dimensional (t,x)-space. Suppose that those solutions of the system (1) whose initial values, i.e., the values for a particular value of t, satisfy the relation (4) satisfy it for all values of t. That is, the function $v(t,x)$ takes the constant value zero along those integral curves of the system (1) which lie on the hypersurface defined by (4). Such a relation is called an invariant relation with respect to the differential system (1). If the C in (2) is given a fixed value C_0, then

$$u(t,x) = C_0$$

is not a first integral. But in the form

$$u(t,x) - C_0 = 0$$

it is just an invariant relation of the form (4). This is the origin of the name particularized first integral for an invariant relation. A general invariant relation does not necessarily arise from a first integral in this manner. On the other hand, a first integral can be considered as a set of ∞^1 invariant relations corresponding to the ∞^1 values of the constant C.

An invariant relation differs from a first integral by containing an absolute constant instead of an arbitrary constant. It defines a property belonging only to a part of the solutions of the system. The ∞^{n-1} solutions satisfying the invariant relation generate a hypersurface in the (t,x)-domain, while all the ∞^n solutions satisfying a first integral fill the whole (t,x)-domain. So a necessary and sufficient condition for (4) to be an invariant relation is obtained by demanding that the total differential of v, or equivalently the total derivative of v with respect to t, vanish identically only on the hypersurface determined by the invariant relation, i.e., the total derivative of v with respect to t vanishes by virtue of (1) and by virtue of (4). Thus,

$$\frac{dv(t,x)}{dt} = \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x_i} f_i(t,x) \equiv \lambda(t,x)v(t,x),$$

where $\lambda(t,x)$ is some function of class $C^{(0)}$ at least. If the invariant relation is obtained from a first integral by assigning a particular value to the arbitrary constant, then the total derivative of v with respect to t is identically zero, i.e., the function λ is identically zero.

Let

$$v_r(t,x) = 0, \quad r = 1, \ldots, m, \tag{5}$$

where the functions $v_r(t,x)$ are of class $C^{(1)}$ in the domain considered, be a system of m independent relations. These relations define a variety in the $(n+1)$-dimensional (t,x)-space. Assume for simplicity that this variety is an $(n+1-m)$-dimensional manifold. Suppose that

$$v_r(t,x(t)) = 0, \quad r = 1, \ldots, m,$$

for every t along those solutions of the system (1) whose initial values satisfy the relations (5), and therefore those solutions that lie in the manifold defined by (5). Then the system of relations (5) is called an invariant system of (1). A necessary and sufficient condition for (5) to be an invariant system of (1) is that the total derivative of the system v_r with respect to t vanishes everywhere on the manifold defined by the relations (5), i.e., vanishes in virtue of the relations (5) and not identically in (t,x). But any function vanishing on the manifold defined by (5) can be written in the form

$$\lambda_r(t,x)v_r(t,x),$$

where the λ_r are suitable functions. Thus the necessary and sufficient condition is that

$$\frac{dv_r(t,x)}{dt} = \frac{\partial v_r}{\partial t} + \frac{\partial v_r}{\partial x_i} f_i(t,x) \equiv \lambda_{rs}(t,x)v_s(t,x),$$
$$i = 1, \ldots, n \quad \text{and} \quad r,s = 1, \ldots, m.$$

Here $\lambda_{rs}(t,x)$ are suitable functions of class $C^{(0)}$ at least.

The definition of an invariant system extends directly to a canonical differential system, since this system is also a normal system. For a system of second-order differential equations, the invariant system would be a system of first-order differential equations with no arbitrary constants in them.

From the definition of an invariant system it is clear that an invariant system is compatible with the differential system. The imposition of relations on x equivalent to an invariant system does not affect the differential equations for x, which remain the same as if the x are free variables. But the invariant system restricts the allowed solutions to those lying on the manifold defined by the invariant system.

3. DEDUCTION OF THE INVARIANT SYSTEM AMONG THE LAGRANGE EQUATIONS

Let the Lagrangian of the dynamical system be $L(t,q,\dot{q})$. Here t is the time; dots denote differentiation with respect to the time t; q represents the generalized coordinates

$$q_i(t), \quad i = 1, \ldots, n;$$

and $\dot{q}$ represents the corresponding generalized velocities

$$\dot{q}_i(t), \quad i = 1, \ldots, n.$$

For the considerations of this paper it suffices to assume that the Lagrangian L is at least of class $C^{(3)}$.

According to definition the degeneracy of the Lagrangian L means that the Hessian matrix

$$\left[\frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j}\right], \quad i,j = 1, \ldots, n, \tag{6}$$

everywhere has a constant rank $n - r_1$, where the integer $r_1 > 0$.

At the outset no conditions are imposed on the variables $q, \dot{q}$. So the Lagrange equations of motion are

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0, \quad i = 1, \ldots, n, \tag{7}$$

or equivalently

$$\frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j} \ddot{q}_j = \frac{\partial L}{\partial q_i} - \frac{\partial^2 L}{\partial \dot{q}_i \partial t} - \frac{\partial^2 L}{\partial \dot{q}_i \partial q_j} \dot{q}_j, \quad i,j = 1, \ldots, n. \tag{8}$$

The q_i can be numbered so that the first $n - r_1$ rows of the matrix (6) are its independent rows. Then, by forming linear combinations of the rows with suitable functions of $t, q, \dot{q}$ as coefficients, the last r_1 rows of (6) can be converted into rows of zeros. In this way the left-hand sides of the last r_1 equations of (8) can be made into zeros while the right-hand sides give functions of $t, q, \dot{q}$. Thus the last r_1 Lagrange equations become first-order differential equations, say

$$B_a(t,q,\dot{q}) = 0, \quad a = 1, \ldots, r_1. \tag{9}$$

Assume for simplicity that all these equations involve at least some of the velocity variables.

The Lagrange equations (7) or (8) were obtained on the assumption that the $q(t), \dot{q}(t)$ were free. But r_1 of these equations reduce to the first-order equations (9), which form subsidiary conditions on the $q, \dot{q}$. For the Lagrange equations to be consistent with the conditions (9), the subsidiary conditions must be of the type that does not modify the equations of motion. So the conditions (9) must form an invariant system with respect to the Lagrange equations. Hence

$$\dot{B}_a(t,q,\dot{q}), \quad a = 1, \ldots, r_1$$

must vanish in virtue of the $n - r_1$ second-order differential equations in (8) and of the first-order equations (9), therefore by virtue of all Eqs. (8). Otherwise the Lagrange system (7) would be inconsistent. Assume that this consistency condition is satisfied. Normally the functions B_a must be at least of class $C^{(1)}$; but the consistency condition makes them functions of class $C^{(2)}$ at least.

The invariant system (9) restricts the dynamically allowed solutions to those satisfying this system. Since the invariant system is compatible with the Lagrangian, the partial derivatives of L with respect to its arguments can be computed as if the variables are free, but the domain of the derivatives is now restricted by the invariant system. According to assumption, even on this restricted domain the rank of the Hessian matrix (6) has the constant value $n - r_1$. So no other first-order Lagrange equations exist. Hence the complete system of subsidiary conditions on the coordinates and velocities consists of the invariant system (9). To satisfy the consistency condition, there must be at least as many second-order Lagrange equations as there are first-order equations. Thus the rank of the Hessian matrix must be at least $\frac{1}{2}n$.

4. THE MULTIPLIER RULE FOR THE CANONICAL EQUATIONS

The canonical formalism makes the transition from the variables $t, q, \dot{q}$, and the Lagrangian dependent on them to the canonical variables t, q, p, and the Hamiltonian dependent on them. Here p represents the collection of the generalized momentum components p_i defined by

$$p_i = \frac{\partial L}{\partial \dot{q}_i}, \qquad i = 1, \ldots, n. \tag{10}$$

If the Lagrangian is nondegenerate, all the right-hand sides are functionally independent in the $\dot{q}$, and so the implicit-function theorem[16] gives all the $\dot{q}$ uniquely in terms of t, q, and p. Then any function of the t, q, and $\dot{q}$ can be expressed as a function of t, q, and p.

If the Lagrangian is degenerate, then the momentum components are functionally dependent,[16,17] and the implicit-function theorem applies only to some of Eqs. (10). Since the Hessian matrix (6) has the rank $n - r_1$, the first $n - r_1$ Eqs. (10) can be uniquely solved (at least locally) for the $n - r_1$ components $\dot{q}_i$ in terms of t, the first $n - r_1$ components p_i, and the last r_1 components $\dot{q}_i$. Substitute these expressions for the first $n - r_1$ components $\dot{q}_i$ into the last r_1 equations of (10). Then, because of the rank of the Hessian matrix (6), the resulting equations yield r_1 relations completely independent of the $\dot{q}$. Thus the functional dependence of the variables p can be stated by expressing the last r_1 components of the p as functions of the remaining canonical variables. But the functional-dependence relations can be given in the more symmetrical form

$$\phi_b(t, q, p) = 0, \qquad b = 1, \ldots, r_1, \tag{11}$$

where the functions ϕ_b are of class $C^{(2)}$, since the right-hand sides of (10) are of class $C^{(2)}$.

Now introduce the expressions obtained above for the first $n - r_1$ components $\dot{q}_i$ into Eqs. (9). Then one gets r_1 equations involving the t, q, p, and the last r_1 components $\dot{q}_i$. Denote these equations by

$$\overline{B}_a(t, q, p, \dot{q}) = 0, \qquad a = 1, \ldots, r_1. \tag{12}$$

If these r_1 equations are functionally independent in the r_1 components $\dot{q}_i$, then they cannot yield any conditions on the canonical variables only. Suppose, on the other hand, that the functional matrix of the $\overline{B}_a$ with respect to the r_1 components $\dot{q}_i$ has the constant rank $r_1 - r_2$, where $0 < r_2 \leq r_1$, everywhere in the space of the arguments of the $\overline{B}_a$ in (12). Then the functional dependence of the $\overline{B}_a$ and Eqs. (12) together yield r_2 subsidiary conditions on the canonical variables only. Denote these conditions by

$$\phi_{b'}(t, q, p) = 0, \qquad b' = r_1 + 1, \ldots, r_3, \tag{13}$$

where one writes

$$r_3 = r_1 + r_2.$$

Thus extra subsidiary conditions on the canonical variables, caused by the equations of motion, can arise only from the first-order Lagrange equations, and that too only under the right circumstances.

In certain examples known in physics, the degenerate Lagrangian does not contain some of the components of $\dot{q}$. Then the corresponding components of p vanish, and the consequent first-order Lagrange equations do not contain the same components of $\dot{q}$. So these first-order Lagrange equations immediately yield subsidiary conditions on the canonical variables only.

The transition to canonical momenta from the velocities introduces two types of subsidiary conditions: those like (11) arise from the singularity of the Hessian matrix; and those like (13) arise only under the right circumstances from the first-order Lagrange equations, which themselves follow from the singularity of the Hessian matrix. How the conditions originate is of no importance in the deduction of the canonical equations of motion. But it is important to have the complete set of conditions (11) and (13) before deducing the canonical equations. The complete set can be written together as

$$\phi_b(t, q, p) = 0, \quad b = 1, \ldots, r_3, \tag{14}$$

where the functions ϕ_b are of class $C^{(2)}$. Assume for simplicity that the conditions (14) determine a submanifold in the (t, q, p)-space. This assumption is necessary for the validity of the multiplier rule used below.

Define the Hamiltonian H as usual by

$$H = p_i \dot{q}_i - L(t, q, \dot{q}), \qquad i = 1, \ldots, n. \tag{15}$$

Then

$$dH = -\frac{\partial L}{\partial t}\, dt - \frac{\partial L}{\partial q_i}\, dq_i + \dot{q}_i\, dp_i + \Big(p_i - \frac{\partial L}{\partial \dot{q}_i}\Big) d\dot{q}_i. \tag{16}$$

Now use the definition (10) of p in (16). Then (16) becomes

$$dH = -\frac{\partial L}{\partial t}\, dt - \frac{\partial L}{\partial q_i}\, dq_i + \dot{q}_i\, dp_i. \tag{17}$$

Hence it follows that H can be expressed as a function of only t, q, and p as a result of introducing the momenta through (10). This result is true even when L is degenerate, because the degeneracy property of L does not enter into the deduction of (17).

Introduce the expressions for the first $n - r_1$ components $\dot{q}_i$ into the right-hand side of (15). The last r_1 components $\dot{q}_i$ yet remain in H, but according to (16) these particular $\dot{q}_i$ occur with the coefficients $p_i - \partial L / \partial \dot{q}_i$. These coefficients are the left-hand sides of the functional-dependence relations solved with respect to the p_i. Thus the $\dot{q}$ variables completely drop out of the expression for H when the functional-dependence relations are taken into account. Denote the expression obtained for the Hamiltonian by $H_0(t, q, p)$. The arguments t, q, p in H_0 satisfy the functional-dependence relations, which are now taken in the symmetrical form (11). The multiplier rule enables one to treat all the arguments on the same footing.

Put $H = H_0$ in (17). Now

$$dH_0 = \frac{\partial H_0}{\partial t}\, dt + \frac{\partial H_0}{\partial q_i}\, dq_i + \frac{\partial H_0}{\partial p_i}\, dp_i,$$

because dH_0 can be written in this form even if the variables t, q, p are dependent. So Eq. (17) gives

$$-\Big(\frac{\partial H_0}{\partial t} + \frac{\partial L}{\partial t}\Big) dt - \Big(\frac{\partial H_0}{\partial q_i} + \frac{\partial L}{\partial q_i}\Big) dq_i + \Big(\dot{q}_i - \frac{\partial H_0}{\partial p_i}\Big) dp_i = 0. \tag{18}$$

Here the t, q, p satisfy the subsidiary conditions (11). Hence the differentials dt, dq, dp satisfy the equations

$$d\phi_b = \frac{\partial \phi_b}{\partial t}\, dt + \frac{\partial \phi_b}{\partial q_i}\, dq_i + \frac{\partial \phi_b}{\partial p_i} dp_i = 0, \tag{19}$$
$$b = 1, \ldots, r_1 \quad \text{and} \quad i = 1, \ldots, n,$$

at points satisfying the conditions (11). Any linear form in dt, dq, dp which vanishes at these points must have the form $\lambda_b d\phi_b$ with suitable functions $\lambda_b(t,q,p)$. Thus the left-hand side of (18) has this form. So there exist suitable multiplier functions $\lambda_b(t,q,p)$ at least of class $C^{(1)}$ such that

$$-\left(\frac{\partial H_0}{\partial t} + \frac{\partial L}{\partial t}\right) dt - \left(\frac{\partial H_0}{\partial q_i} + \frac{\partial L}{\partial q_i}\right) dq_i + \left(\dot{q}_i - \frac{\partial H_0}{\partial p_i}\right) dp_i$$
$$\equiv \lambda_b \left(\frac{\partial \phi_b}{\partial t}\, dt + \frac{\partial \phi_b}{\partial q_i}\, dq_i + \frac{\partial \phi}{\partial p_i}\, dp_i\right), \tag{20}$$
$$b = 1, \ldots, r_1 \quad \text{and} \quad i = 1, \ldots, n.$$

The uniqueness or nonuniqueness of the multiplier function $\lambda_b(t,q,p)$ associated with each ϕ_b depends on how $d\phi_b = 0$, or equivalently $\dot{\phi}_b = 0$, is satisfied. From (20) one gets

$$\dot{q}_i = \frac{\partial H_0}{\partial p_i} + \lambda_b \frac{\partial \phi_b}{\partial p_i}, \tag{21}$$

$$-\frac{\partial L}{\partial q_i} = \frac{\partial H_0}{\partial q_i} + \lambda_b \frac{\partial \phi_b}{\partial q_i}, \tag{22}$$

$$-\frac{\partial L}{\partial t} = \frac{\partial H_0}{\partial t} + \lambda_b \frac{\partial \phi_b}{\partial t}, \tag{23}$$
$$b = 1, \ldots, r_1 \quad \text{and} \quad i = 1, \ldots, n.$$

These equations, expressing the left-hand sides in terms of t, q, and p, follow from the definitions (10) and (15) and from the rank of the Hessian matrix (6).

In deducing (21)–(23) the arguments of L were not subject to any subsidiary conditions. But in the dynamical problem under consideration some of the Lagrange equations impose such subsidiary conditions (9). If these conditions yield additional subsidiary conditions on the canonical variables, then the t, q, p satisfy the complete set (14) of subsidiary conditions. In this case, the dt, dq, dp satisfy the equations

$$d\phi_b = \frac{\partial \phi_b}{\partial t} + \frac{\partial \phi_b}{\partial q_i}\, dq_i + \frac{\partial \phi_b}{\partial p_i}\, dp_i = 0, \tag{24}$$
$$b = 1, \ldots, r_3 \quad \text{and} \quad i = 1, \ldots, n.$$

Then instead of (21)–(23) one gets

$$\dot{q}_i = \frac{\partial H_0}{\partial p_i} + \lambda_b \frac{\partial \phi_b}{\partial p_i}, \tag{25}$$

$$-\frac{\partial L}{\partial q_i} = \frac{\partial H_0}{\partial q_i} + \lambda_b \frac{\partial \phi_b}{\partial q_i}, \tag{26}$$

$$-\frac{\partial L}{\partial t} = \frac{\partial H_0}{\partial t} + \lambda_b \frac{\partial \phi_b}{\partial t}, \tag{27}$$
$$b = 1, \ldots, r_3 \quad \text{and} \quad i = 1, \ldots, n,$$

for the dynamical problem.

The Lagrange equations (7) can be written as

$$\dot{p}_i = \frac{\partial L}{\partial q_i}.$$

Combining this with the expressions (25) and (26) obtained for $\dot{q}_i$ and $\partial L/\partial q_i$ in the dynamical problem gives the canonical equations of motion in the multiplier-rule form

$$\dot{q}_i = \frac{\partial H_0}{\partial p_i} + \lambda_b \frac{\partial \phi_b}{\partial p_i}, \tag{28}$$

$$\dot{p}_i = -\frac{\partial H_0}{\partial q_i} - \lambda_b \frac{\partial \phi_b}{\partial q_i}, \tag{29}$$
$$b = 1, \ldots, r_3 \quad \text{and} \quad i = 1, \ldots, n,$$

to which the relations (14) must be added. The explicit canonical equations follow only when the multiplier functions λ_b are actually determined from these equations, as is done in the next section.

The general canonical equation of motion can be written in terms of Poisson brackets. The Poisson bracket $[\xi, \eta]$ of two functions $\xi(t,q,p)$ and $\eta(t,q,p)$ of class $C^{(2)}$ is defined by

$$[\xi, \eta] = \frac{\partial \xi}{\partial q_i}\frac{\partial \eta}{\partial p_i} - \frac{\partial \xi}{\partial p_i}\frac{\partial \eta}{\partial q_i}, \quad i = 1, \ldots, n.$$

Even when t, q, p are not independent, the first differential of a function $g(t,q,p)$ of class $C^{(2)}$ has the form

$$dg = \frac{\partial g}{\partial t}\, dt + \frac{\partial g}{\partial q_i}\, dq_i + \frac{\partial g}{\partial p_i}\, dp_i,$$

and so

$$\dot{g} = \frac{\partial g}{\partial t} + \frac{\partial g}{\partial q_i}\dot{q}_i + \frac{\partial g}{\partial p_i}\dot{p}_i, \quad i = 1, \ldots, n.$$

Hence the canonical equations (28), (29) yield the general canonical equation of motion for $g(t,q,p)$ in the form

$$\dot{g} = \frac{\partial g}{\partial t} + [g, H_0] + \lambda_b[g, \phi_b] = \frac{\partial g}{\partial t} + [g, H], \tag{30}$$

$$H \overset{\text{def}}{=} H_0 + \lambda_b \phi_b,$$

$$\phi_b(t,q,p) = 0, \quad b = 1, \ldots, r_3. \tag{14}$$

Thus the multiplier rule for the subsidiary conditions replaces the Hamiltonian H_0 by H in the canonical equations.

5. EXPLICIT DETERMINATION OF THE MULTIPLIERS

The conditions (14) in the canonical equations require that the canonical equations of motion satisfy the equations

$$\dot{\phi}_b = 0, \quad b = 1, \ldots, r_3,$$

on the submanifold determined by (14). From (30) these conditions give

$$\dot{\phi}_b = \frac{\partial \phi_b}{\partial t} + [\phi_b, H_0] + \lambda_c[\phi_b, \phi_c] = 0, \quad b, c = 1, \ldots, r_3. \tag{31}$$

Thus the multiplier functions λ_c satisfy the nonhomogeneous linear system of equations

$$[\phi_b,\phi_c]\lambda_c = -\left(\frac{\partial\phi_b}{\partial t} + [\phi_b,H_0]\right), \tag{32}$$

$$\phi_b = 0, \quad b,c = 1,\ldots,r_3.$$

The coefficient matrix of the linear system (32) is the matrix $[[\phi_b,\phi_c]]$. The augmented matrix of the system (32) is obtained by adjoining the column consisting of the right-hand sides of (32) to the coefficient matrix. The system (32) has a solution if and only if the ranks of the coefficient matrix and the augmented matrix are the same. This consistency condition must be satisfied for the multiplier rule to yield consistent canonical equations, and therefore for the consistency of the original Lagrange equations. Assume that this consistency condition is satisfied. Instead of immediately solving the system (32) for the λ, however, it is simpler to replace the subsidiary conditions (14) by an equivalent system whose multiplier functions are easier to determine. The system (32) suggests the appropriate change.

Let the rank of the skew-symmetric matrix $[[\phi_b,\phi_c]]$ by virtue of Eqs. (14) be r_4, where r_4 must always be even and where $r_4 \le r_3$. If r_3 is odd, then the coefficient matrix is necessarily singular. Number the ϕ so that the first r_4 rows of the coefficient matrix and of the augmented matrix are the independent rows of these matrices. Then, by forming linear combinations of all the rows of these matrices with functions of t,q,p as coefficients, all the elements of the last $r_3 - r_4$ rows of these matrices can be converted into zeros. Hence there exist distinct functions $\mu_{\alpha b}(t,q,p)$ at least of class $C^{(1)}$ and independent of the ϕ such that

$$\mu_{\alpha b}[\phi_b,\phi_c] = [\mu_{\alpha b}\phi_b,\phi_c] = 0, \tag{33}$$

$$\mu_{\alpha b}\frac{\partial\phi_b}{\partial t} + \mu_{\alpha b}[\phi_b,H_0] = \frac{\partial(\mu_{\alpha b}\phi_b)}{\partial t} + [\mu_{\alpha b}\phi_b,H_0] = 0, \tag{34}$$

$$\alpha = r_4+1,\ldots,r_3, \quad b,c = 1,\ldots,r_3,$$

by virtue of (14). Equations (33) and (34) suggest the suitable equivalent system to replace the subsidiary conditions (14). Keep the first r_4 equations of (14) unchanged,

$$\phi_\beta \overset{\text{def}}{=} \chi_\beta = 0, \quad \beta = 1,\ldots,r_4, \tag{35}$$

but replace the last $r_3 - r_4$ equations of (14) by

$$\chi_\alpha \overset{\text{def}}{=} \mu_{\alpha b}\phi_b = 0, \tag{36}$$

$$\alpha = r_4+1,\ldots,r_3, \quad b = 1,\ldots,r_3.$$

The complete equivalent system (35) and (36) is written as

$$\chi_b = 0, \quad b = 1,\ldots,r_3, \tag{37}$$

where the functions χ_b are independent and of class $C^{(2)}$.

For the equivalent system (37) the general canonical equation (30) for $g(t,q,p)$ becomes

$$\dot g = \frac{\partial g}{\partial t} + [g,H_0] + \nu_c[g,\chi_c], \tag{38}$$

where the ν_c are new multiplier functions. The conditions (31) are now

$$\dot\chi_b = \frac{\partial\chi_b}{\partial t} + [\chi_b,H_0] + \nu_c[\chi_b,\chi_c] = 0, \tag{39}$$

$$b,c = 1,\ldots,r_3.$$

According to Eqs. (33)–(37) the χ_α have two characteristic properties:

$$[\chi_\alpha,\chi_b] = 0, \tag{40}$$

$$\frac{\partial\chi_\alpha}{\partial t} + [\chi_\alpha,H_0] = 0, \tag{41}$$

$$\alpha = r_4+1,\ldots,r_3, \quad b = 1,\ldots,r_3,$$

in virtue of the relations (37). Because of these properties the first r_4 equations of (39) become

$$\dot\chi_\beta = \frac{\partial\chi_\beta}{\partial t} + [\chi_\beta,H_0] + \nu_\gamma[\chi_\beta,\chi_\gamma] = 0, \tag{42}$$

$$\beta,\gamma = 1,\ldots,r_4,$$

and the last $r_3 - r_4$ equations of (39) for the $\dot\chi_\alpha$ are automatically satisfied for arbitrary ν_α.

The system (42) is equivalent to

$$[[\chi_\beta,\chi_\gamma]][\nu_\gamma] = -\left[\frac{\partial\chi_\beta}{\partial t} + [\chi_\beta,H_0]\right] \tag{43}$$

in matrix notation. The unique solution for the ν_γ follows at once by multiplying both sides of (43) on the left by the inverse of the nonsingular skew-symmetric matrix $[[\chi_\beta,\chi_\gamma]]$. Let the nonzero determinant of this matrix be Δ and let $C_{\beta\gamma}\Delta$ denote the cofactor of the element $[\chi_\beta,\chi_\gamma]$ in Δ. Here $C_{\beta\gamma}$ is antisymmetric in β,γ. Then the inverse of the matrix $[[\chi_\beta,\chi_\gamma]]$ is the transpose of the matrix $[C_{\beta\gamma}]$, i.e., the matrix $[-C_{\beta\gamma}]$. Therefore the solution of the system (43), or equivalently (42), is given by

$$\nu_\beta = C_{\beta\gamma}\left\{\frac{\partial\chi_\gamma}{\partial t} + [\chi_\gamma,H_0]\right\}, \quad \beta,\gamma = 1,\ldots,r_4. \tag{44}$$

Since the multipliers ν_α are arbitrary functions of t, q, and p, they can be taken to be zeros. With these values for ν_α and the expressions (44) for ν_β the explicit general canonical equation of motion for any function $g(t,q,p)$ in the presence of the subsidiary conditions (37) is

$$\dot g = \frac{\partial g}{\partial t} + [g,\chi_\beta]C_{\beta\gamma}\frac{\partial\chi_\beta}{\partial t} + [g,H_0] + [g,\chi_\beta]C_{\beta\gamma}[\chi_\gamma,H_0], \tag{45}$$

$$\beta,\gamma = 1,\ldots,r_4.$$

The above deduction shows that the two properties (40), (41) are essential for conditions (36) to form an invariant system with respect to the canonical equation (45), or equivalently to the canonical system (28), (29). It is not surprising that the conditions (36) do not modify the canonical equations. The dynamical system admits only those solutions of (45) which satisfy the invariant system (36) as possible motions.

In the terminology of Lie, a set of functions is said to be in involution if the Poisson brackets of any two functions of the set vanish identically. Analogously, a system of equations is said to be in involution if the Poisson brackets of any two left-hand sides of the equations vanish by virtue of the system of equations. Properties

(40) and (41) then assert that the invariant system (36) must be in involution with all the subsidiary conditions (37).

6. DEDUCTION AND INTERPRETATION OF THE MODIFIED POISSON BRACKETS

In deducing the canonical equation (45) the subsidiary conditions (35) can be satisfied only by choosing the multipliers ν_β so as to make $\dot{\chi}_\beta \equiv 0$. This implies that the subsidiary conditions (35) can be satisfied only by converting Eqs. (35) into the identities $\chi_\beta \equiv 0$. The only way these equations become identities is by implicitly solving for r_4 of the variables in terms of the others. Thus the modification in the canonical equations caused by the subsidiary conditions amounts to a reduction in the number of independent canonical variables. The original Poisson brackets were defined as if all the canonical variables were independent. So the modified canonical equations require modified Poisson brackets defined with respect to the reduced number of independent canonical variables. The formula for this modified Poisson bracket follows from the canonical equation (45).

It is shown in this section that the implicit solution of the conditions (35) can take place only with respect to $\frac{1}{2}r_4$ canonically conjugate pairs. This is done by means of a canonical transformation adapted to the subsidiary conditions. This canonical transformation, which is again needed in the next section, is found with the help of Lie's theory[12,13] of function groups.

r independent functions of q and p, such that the Poisson bracket of any pair of these functions is a function of the r independent functions only or is a constant, constitute an r-dimensional function group. If the functions contain t explicitly, as in the present work, it is treated as a parameter. All functions of the r independent functions belong to the function group. Thus a function group is a set of functions with the property that the Poisson bracket of any two functions of the set belongs to the set. Any set of r independent functions of the group completely specifies the group and forms a basis of the function group. Every function group can be transformed to a basis in which the Poisson brackets of the basis functions take only the values 0 and 1. Then the function group is said to be in canonical form, which is the simplest form of a function group. Those functions of the function group which are in involution with all functions of the group are called singular functions of the function group. The singular functions are in involution with one another and form a function subgroup called the null group of the given group. This null group can be at most n-dimensional. Every function group can be embedded in a $2n$-dimensional function group. In particular, any canonical function group can be embedded in a $2n$-dimensional canonical function group. It is this embedding process that is used here to find the required canonical transformation. But first it is necessary to replace the subsidiary conditions by an equivalent system whose left-hand-side functions form a canonical function group.

The skew-symmetric matrix

$$[[\chi_\beta, \chi_\gamma]], \quad \beta, \gamma = 1, \ldots, r_4 \tag{46}$$

formed with the χ of (35) is nonsingular. Therefore, by a known property of skew-symmetric matrices, there exists a nonsingular matrix

$$\Lambda \overset{\text{def}}{=} [\Lambda_{\beta\gamma}(t, q, p)]$$

such that multiplication of the matrix (46) on the left by Λ and on the right by the transpose $\tilde{\Lambda}$ of Λ gives the skew-symmetric matrix

$$J = \begin{bmatrix} 0 & E \\ -E & 0 \end{bmatrix}, \tag{47}$$

where 0 is the zero $\frac{1}{2}r_4 \times \frac{1}{2}r_4$ matrix and E is the unit $\frac{1}{2}r_4 \times \frac{1}{2}r_4$ matrix. That is, there exists Λ such that

$$[\Lambda_{\beta\gamma}][[\chi_\gamma, \chi_\delta]][\tilde{\Lambda}_{\delta\epsilon}] = [J_{\beta\epsilon}], \quad \beta, \gamma, \delta, \epsilon = 1, \ldots, r_4. \tag{48}$$

This suggests replacing the conditions (35) by the equivalent system

$$\theta_\beta \overset{\text{def}}{=} \Lambda_{\beta\gamma}\chi_\gamma = 0, \quad \beta, \gamma = 1, \ldots, r_4. \tag{49}$$

Then (48) becomes

$$[[\theta_\beta, \theta_\gamma]] = [J_{\beta\gamma}], \quad \beta, \gamma = 1, \ldots, r_4, \tag{50}$$

in virtue of Eqs. (36) and (49). Equations (36) are in involution among themselves and with the conditions (49). The conditions (36) and (49) equivalent to (14) are such that the left-hand sides of this equivalent system satisfy the conditions for a canonical function group but only by virtue of Eqs. (36) and (49).

It is known[7,13] that a system of equations in involution can be replaced by an equivalent system of equations such that the Poisson brackets of the left-hand sides vanish identically and not in virtue of the equations themselves. For the present work, however, one needs an analogous result applicable to a general system of equations like (36) and (49). Schouten and v.d. Kulk (Ref. 13, Theorem VII. 24) provide such a general theorem. This theorem asserts that one can always replace any given system of equations in t, q, p by an equivalent system of equations such that the functions on the left-hand sides of these equations form a canonical function group identically and not by virtue of the equations.

The considerations of the preceding two paragraphs show that, when only a part of the subsidiary conditions in t, q, and p, say only Eqs. (11), is known, it is in general not possible to get the other equations of the whole system, namely Eqs. (13), by forming the Poisson brackets of the left-hand sides of the known equations. This certainly holds when the system of equations contains equations like (49) or (35). Even if the whole system is in involution, the known part may belong to a subgroup and the Poisson-bracket operation may yield the subgroup but not the whole group. If the known part is part of a canonical subgroup (identically), then the Poisson-bracket operation yields no new equations.

Introduce a new notation for the functions on the left-hand sides of the system which is equivalent to (36) and (49) and which is such that these functions form a canonical function group (identically). Henceforth write

$$n_1 = n - r_3 + \tfrac{1}{2}r_4, \quad n_2 = n - \tfrac{1}{2}r_4.$$

Let

$$P_e, \quad e = n_1 + 1, \ldots, n_2,$$

denote the singular functions of this canonical function group. Let Q_f denote the functions corresponding to the first half of the θ_β in (49) and let P_f denote the functions corresponding to the last half of the θ_β in (49), where $f = n_2 + 1, \ldots, n$. Then the new equivalent system has the form

$$P_e = 0, \quad Q_f = 0, \quad P_f = 0, \tag{51}$$
$$e = n_1 + 1, \ldots, n_2, \quad f = n_2 + 1, \ldots, n.$$

The Poisson brackets

$$\left.\begin{array}{l}[P_e, P_{e'}] = 0, \quad [P_e, Q_f] = 0, \quad [P_e, P_f] = 0, \\ \quad [Q_f, Q_{f'}] = 0, \quad [Q_f, P_{f'}] = \delta_{ff'}, \quad [P_f, P_{f'}] = 0, \\ \quad e, e' = n_1 + 1, \ldots, n_2, \quad f, f' = n_2 + 1, \ldots, n,\end{array}\right\} \tag{52}$$

where $\delta_{ff'}$ is the usual Kronecker symbol, hold identically and not by virtue of Eqs. (51).

The p_e, q_f, p_f for the same values of e, f as in (51) form a canonical function subgroup of the $2n$-dimensional canonical group formed by the q, p. The P_e, Q_f, P_f form a canonical function subgroup of the $2n$-dimensional canonical function group formed by the Q, P into which this subgroup can be embedded. The transformation from t, q, p to t, Q, P provides the required canonical transformation, since it satisfies the necessary and sufficient conditions[6,13] for such a transformation. (The definition of canonical transformations and a necessary and sufficient condition for a transformation to be canonical are given in the next section.)

Now, by performing an elementary canonical transformation if necessary, i.e., by renumbering the canonical pairs or by exchanging the coordinate and momentum in a canonical pair, the P_e must be invertible functions of the p_e, and the Q_f, P_f must be invertible functions of the q_f, p_f. Hence the equations

$$Q_f = 0, \quad P_f = 0, \quad f = n_2 + 1, \ldots, n, \tag{53}$$

must be soluble for the q_f, p_f. Thus the functional matrix

$$\begin{bmatrix} \dfrac{\partial Q_f}{\partial q_{f'}} & \dfrac{\partial Q_f}{\partial p_{f'}} \\ \dfrac{\partial P_f}{\partial q_{f'}} & \dfrac{\partial P_f}{\partial p_{f'}} \end{bmatrix}, \quad f, f' = n_2 + 1, \ldots, n, \tag{54}$$

must be nonsingular. But Eqs. (53) are equivalent to Eqs. (35). Therefore, Eqs. (35) are soluble for the q_f, p_f. So the matrix

$$\begin{bmatrix} \dfrac{\partial \chi_\beta}{\partial q_f} & \dfrac{\partial \chi_\beta}{\partial p_f} \end{bmatrix}, \quad \beta = 1, \ldots, r_4, \quad f = n_2 + 1, \ldots, n, \tag{55}$$

has the same rank r_4 as (54) everywhere. Thus every nonsingular $r_4 \times r_4$ submatrix of the functional matrix

$$\begin{bmatrix} \dfrac{\partial \chi_\beta}{\partial t} & \dfrac{\partial \chi_\beta}{\partial q_i} & \dfrac{\partial \chi_\beta}{\partial p_i} \end{bmatrix}, \quad \beta = 1, \ldots, r_4, \quad i = 1, \ldots, n \tag{56}$$

has the particular form (55) in which the partial derivatives are with respect to $\frac{1}{2} r_4$ canonically conjugate pairs.

Since the matrix (55) is nonsingular everywhere, the implicit-function theorem guarantees that Eqs. (35) or (49) can be solved for the last $\frac{1}{2} r_4$ canonically conjugate pairs in the form

$$\begin{aligned} &q_f = g_f(t, q_1, \ldots, q_{n_2}, p_1, \ldots, p_{n_2}), \\ &\quad p_f = h_f(t, q_1, \ldots, q_{n_2}, p_1, \ldots, p_{n_2}), \\ &\quad f = n_2 + 1, \ldots, n. \end{aligned} \tag{57}$$

The multiplier rule for conditions (35) follows from the relation

$$-\left(\frac{\partial L}{\partial t} + \frac{\partial H_0}{\partial t}\right) dt - \left(\dot{p}_i + \frac{\partial H_0}{\partial q_i}\right) dq_i + \left(\dot{q}_i - \frac{\partial H_0}{\partial p_i}\right) dp_i$$
$$\equiv \nu_\beta \left(\frac{\partial \chi_\beta}{\partial t} + \frac{\partial \chi_\beta}{\partial q_i} dq_i + \frac{\partial \chi_\beta}{\partial p_i} dp_i\right),$$
$$\beta = 1, \ldots, r_4, \quad i = 1, \ldots, n,$$

analogous to (20). This relation implies that the rows of the $(r_4 + 1) \times (2n + 1)$ matrix

$$\begin{bmatrix} -\left(\dfrac{\partial L}{\partial t} + \dfrac{\partial H_0}{\partial t}\right) & -\left(\dot{p}_i + \dfrac{\partial H_0}{\partial q_i}\right) & \left(\dot{q}_i - \dfrac{\partial H_0}{\partial p_i}\right) \\ \dfrac{\partial \chi_\beta}{\partial t} & \dfrac{\partial \chi_\beta}{\partial q_i} & \dfrac{\partial \chi_\beta}{\partial p_i} \end{bmatrix}, \tag{58}$$
$$\beta = 1, \ldots, r_4, \quad i = 1, \ldots, n,$$

are not independent. Since the rank of the matrix (56) is known to be r_4, the matrix (58) also has the rank r_4. Therefore the multiplier rule also says that only r_4 columns of the matrix (58), namely the columns containing the r_4 columns of a nonsingular submatrix of (56), are independent and that the other columns are linear combinations of the independent columns. But, if one considers the submatrix (56) of (57), the relations between the columns are just the relations $d\chi_\beta = 0$, which are equivalent to $\dot{\chi}_\beta = 0$. So the multiplier rule amounts to solving r_4 of the differentials dt, dq, dp in the equations $d\chi_\beta = 0$ in terms of the other differentials. This in turn amounts to implicitly solving Eqs. (35) for the r_4 dependent variables among the t, q, p in terms of the independent variables among the t, q, p. This is equivalent to applying the implicit-function theorem to the equations (35). As seen already, this implicit solution can take place only with respect to $\frac{1}{2} r_4$ canonically conjugate pairs in the form (57). From the first row of the matrix it also follows that only n_2 canonically conjugate pairs of the canonical equations (28), (29) are independent.

The multiplier rule with the multipliers given by (44) ensures that only the n_2 canonically conjugate pairs on the right-hand sides of (56) are the independent canonical variables of the dynamical system. Any function of the original t, q, p then becomes a composite function of t and of the n_2 independent canonical pairs. So the Poisson bracket of any two functions of t, q, p is now equal to the Poisson bracket of the corresponding two composite functions of t and of the n_2 independent canonical pairs. Long ago J. Bertrand considered this type of generalized Poisson bracket. This modified Poisson bracket evidently satisfies all the properties of ordinary Poisson brackets, including the Jacobi–Donkin identity. The modified Poisson bracket refers to the n_2 degrees of freedom of the dynamical system. Since Eqs. (35) or (49) become identically zero in determining the independent canonical pairs, the new Poisson bracket of the left-hand side of (36) or (49) with any other function must vanish identically.

The expression for the modified Poisson bracket can be deduced from (45). The multiplier rule implicitly converts a function $f(t, q, p)$ into a composite function $\bar{f}$ of t and of the n_2 independent canonically conjugate pairs. The general canonical equation for $\bar{f}$ would be

$$\frac{d\bar{f}}{dt} = \frac{\partial \bar{f}}{\partial t} + [\bar{f}, \bar{H}_0] = \frac{\partial \bar{f}}{\partial t} + [f, H_0]^*. \tag{59}$$

Here $[f, H_0]^*$ denotes the modified Poisson bracket $[\bar{f}, \bar{H}_0]$ of the composite function $\bar{f}$ and of the composite function $\bar{H}_0$ corresponding to H_0. Note that t occurs in $\bar{f}$ through the original explicit t in the f and the explicit t in the canonical pairs (57). Comparing (59) with (45) gives

$$\frac{\partial \bar{f}}{\partial t} = \frac{\partial f}{\partial t} + [g, \chi_\beta] C_{\beta\gamma} \frac{\partial \chi_\gamma}{\partial t}, \tag{60}$$

$$[f, H_0]^* = [\bar{f}, \bar{H}_0] = [f, H_0] + [f, \chi_\beta] C_{\beta\gamma} [\chi_\gamma, H_0].$$

Thus the modified Poisson bracket of any two functions $\xi(t,q,p), \eta(t,q,p)$, with the t,q,p such that $\chi_\beta(t,q,p) \equiv 0$, is

$$[\xi, \eta]^* = [\xi, \eta] + [\xi, \chi_\beta] C_{\beta\gamma} [\chi_\gamma, \eta], \tag{61}$$

a result given by Dirac.[1] This modified Poisson bracket is considered again in the next section.

7. SIMPLIFICATION OF THE INTEGRATION PROBLEM BY PASSING TO THE PHYSICAL VARIABLES

The multiplier rule has given a canonical system together with an invariant system. But finding the solutions of this canonical system that satisfy the invariant system is in general difficult. For progress in integration or quantization it is essential to seek a simplification of the canonical system itself by making use of the invariant system.

In the theory of differential equations it is known that the order of a normal system of differential equations can be reduced if some first integrals of the system are known. The reduction in the order is equal to the number of first integrals known. One way of carrying out this reduction is to pass to a new system of variables in which the left-hand sides of the known integrals replace an equal number of the old variables while the other variables remain the same as before. Levi-Civita[7] has extended this result to the case where an invariant system replaces the system of first integrals. Levi-Civita[7] has also indicated how to carry out the reduction when one has a canonical differential system (instead of just a normal system) together with an invariant system. In this case the change of variables must be effected through a suitable canonical transformation, so that the reduced system is also canonical. This method can be adapted to the problem under consideration.

The properties of canonical transformations[6] needed for the present work are as follows. Canonical transformations are those invertible transformations of the canonical variables to new variables which take every canonical differential system into another such system. Consider the invertible transformations

$$\begin{aligned} q_i &= q_i(t,Q,P), \quad p_i = p_i(t,Q,P), \\ Q_i &= Q_i(t,q,p), \quad P_i = P_i(t,q,p), \quad i = 1,\dots,n, \end{aligned} \tag{62}$$

and the nonsingular functional matrix

$$M = \begin{bmatrix} \dfrac{\partial Q}{\partial q} & \dfrac{\partial Q}{\partial p} \\ \dfrac{\partial P}{\partial q} & \dfrac{\partial P}{\partial p} \end{bmatrix},$$

whose transpose is $\tilde{M}$. The functions on the right-hand sides of (62) are taken to be of class $C^{(2)}$. The transformation (62) is canonical if and only if the $2n \times 1$ matrix

$$\begin{bmatrix} \dot{q} - \dfrac{\partial H}{\partial p} \\ \dot{p} + \dfrac{\partial H}{\partial q} \end{bmatrix}$$

for an arbitrary Hamiltonian $H(t,q,p)$ of class $C^{(2)}$ is transformed into a $2n \times 1$ matrix of the same form, namely

$$\begin{bmatrix} \dot{Q} - \dfrac{\partial K}{\partial P} \\ \dot{P} + \dfrac{\partial K}{\partial Q} \end{bmatrix},$$

with a suitable new Hamiltonian $K(t,Q,P)$. A necessary and sufficient condition for the transformation (62) to be canonical is that the matrix relation

$$MJ\tilde{M} = \mu J$$

hold as an identity for some nonzero constant scalar μ. The matrix J here is the same as the matrix (47). This condition just states that

$$\begin{bmatrix} [Q_i, Q_j] & [Q_i, P_j] \\ [P_i, Q_j] & [P_i, P_j] \end{bmatrix} = \mu J, \quad i,j = 1,\dots,n, \tag{63}$$

identically. When this condition is satisfied, the Hamiltonian $H(t,q,p)$ goes into the new Hamiltonian $K(t,Q,P)$ given by

$$K = \mu H + R, \tag{64}$$

where H is expressed as a function of t,Q,P. The function R satisfies the identities

$$\begin{aligned} \frac{\partial Q_i(t,q,p)}{\partial t} &= \frac{\partial R(t,Q,P)}{\partial P_i}, \\ \frac{\partial P_i(t,q,p)}{\partial t} &= -\frac{\partial R(t,Q,P)}{\partial Q_i}, \end{aligned} \tag{65}$$

where the left-hand sides should be expressed in terms of t,Q,P. Hence the function R is determined by a quadrature in Q,P for fixed t, so that an arbitrary additive function of t only remains in R, and therefore in K. But, since this arbitrary function drops out of the partial derivatives $\partial K/\partial Q$ and $\partial K/\partial P$, two Hamiltonians that differ by a function of t only are equivalent. μ and R depend only on the canonical transformation and not on H. Two functions $\xi(t,q,p), \eta(t,q,p)$ can also be expressed as functions of t,Q,P. The Poisson bracket of these two functions can be evaluated with respect to the q,p or the Q,P. Then the two Poisson brackets satisfy the identity

$$[\xi, \eta]_{q,p} = \mu[\xi, \eta]_{Q,P}, \tag{66}$$

where the suffixes indicate the variables with respect to which the Poisson bracket is evaluated. Conversely, if for any two functions ξ, η this identity holds for one and the same constant μ, then the passage from the $2n$ variables q,p to the $2n$ variables Q,P is a canonical transformation characterized by the constant μ. In the present work the constant μ is taken to be 1.

The canonical transformation appropriate for the reduction has already been mentioned in the preceding section. The original subsidiary conditions must be replaced by the equivalent system (51) such that the functions on the left-hand sides of these equations form an r_3-dimensional canonical function group. This is necessary because one has to satisfy the criterion (63) for a canonical transformation. This canonical function group is then embedded in a $2n$-dimensional canonical function group.[12,13] To do this, one first determines the r_3-r_4 functions Q_e canonically conjugate to the singular functions P_e. Thus one obtains a $(2r_3\text{-}r_4)$-dimensional canonical function group without singular functions. Next one determines the $2n_1$-dimensional canonical polar group of the $(2r_3\text{-}r_4)$-dimensional function group just obtained. This standard procedure may take some effort, but it provides the $2n$-dimensional canonical function group Q,P. This $2n$-dimensional group also has no singular functions. The transformation from t,q,p to t,Q,P is the desired canonical transformation. Here the constant scalar μ is equal to 1.

Take the canonical equations for the dynamical system to be those derived from the multiplier-rule Hamiltonian H when the subsidiary conditions have the form (51). Perform the canonical transformation

$$(t,q,p) \mapsto (t,Q,P)$$

considered above. Then the canonical equations for the Q,P are derived from the new Hamiltonian

$$K(t,Q,P) = H + R, \tag{67}$$

where H is now expressed in terms of t, Q, and P, and $R(t,Q,P)$ is determined from (65) by a quadrature. The Poisson bracket of two functions ξ,η is invariant under this canonical transformation, i.e.,

$$[\xi,\eta]_{q,p} = [\xi,\eta]_{Q,P}. \tag{68}$$

Now consider the subsidiary conditions (53). As seen in the preceding section, these conditions can be satisfied only by making them into identities. Thus the variables Q_f,P_f drop out of consideration, and the canonical system now refers only to the remaining canonical pairs

$$Q_j, P_j, \quad j = 1,\ldots,n_2. \tag{69}$$

The new Hamiltonian K_1 is obtained from $K(t,Q,P)$ by using (53) in it. The functions $\xi(t,Q,P)$, $\eta(t,Q,P)$ become functions $\bar{\xi},\bar{\eta}$ depending only on t and the canonical pairs (69). Thus the Poisson bracket of the reduced canonical system, defined with respect to the canonical pairs (69), has the truncated form $[\bar{\xi},\bar{\eta}]_{Q_j,P_j}$ obtained from the original Poisson bracket $[\xi,\eta]_{Q,P}$ by omitting the contributions from the canonical pairs Q_f,P_f. This expression for the modified Poisson bracket $[\bar{\xi},\bar{\eta}]_{Q_j,P_j}$ coincides exactly with that given by formula (61) for the subsidiary conditions (53). Evidently the modified Poisson brackets of the reduced canonical system satisfy all the usual properties of ordinary Poisson brackets.

The final step is to examine the effect of the invariant system

$$P_e = 0, \quad e = n_1 + 1,\ldots,n_2, \tag{70}$$

on the reduced canonical system

$$\dot{Q}_j = \frac{\partial K_1}{\partial P_j}, \quad \dot{P}_j = -\frac{\partial K_1}{\partial Q_j}, \quad j = 1,\ldots,n_2. \tag{71}$$

obtained above. If the P_e occur in K_1, then they can all be included in the contribution (with arbitrary multipliers) from the P_e to K_1. Since this contribution from an invariant system is taken to be zero, one is justified in assuming that the P_e do not occur in K_1. Then $\partial K_1/\partial Q_e$ also does not contain the P_e. Now the conditions for the equations to be an invariant system of the canonical equations (71) are that

$$\dot{P}_e = -\frac{\partial K_1}{\partial Q_e} = 0, \quad e = n_1 + 1,\ldots,n_2, \tag{72}$$

by virtue of Eqs. (70). But the second members of (72) do not contain the P_e. So Eqs. (72) hold identically. Hence K_1 does not contain the Q_e canonically conjugate to the P_e. This means that the canonical system has been transformed into a system in which the Q_e are ignorable coordinates and the corresponding conjugate momenta P_e are equal to absolute constants instead of arbitrary constants as in the usual case of ignorable coordinates.[7] Thus the canonical system (71) gets reduced to a new canonical system for the independent canonical pairs

$$Q_k, P_k, \quad k = 1,\ldots,n_1, \tag{73}$$

with the new Hamiltonian K_2 depending on t and the canonical pairs (73). The invariant system (70) must be adjoined to this canonical system.

The dynamical system is now described by the canonical function group consisting of the P_e of (70) and the canonical pairs (73). The P_e are the singular functions of this function group. The invariant system (70) implies that each P_e admits only the single constant value zero. The dynamical system admits all solutions Q_k,P_k of the reduced canonical system, and so the canonical pairs (73) are physical variables. Since the Q_e are ignorable coordinates in the canonical system, the modified Poisson bracket obtained earlier turns out to be a Poisson bracket defined only with respect to the canonical pairs (73). The symplectic structure of the space of t and of the canonical pairs (73) is evident.[4,15] The reduced canonical system in the physical variables is quite suitable for integration or for passing to the quantum theory.

[1]P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); Proc. R. Soc. A **246**, 326 (1958); Phys. Rev. **114**, 924 (1959); *Lectures on Quantum Mechanics* (Belfer Graduate School of Science, Yeshiva University, New York, 1964); in *Contemporary Physics: Trieste Symposium 1968* (International Atomic Energy Agency, Vienna, 1969), Vol. I, p. 539.

[2]J. L. Anderson and P. G. Bergmann, Phys. Rev. **83**, 1018 (1951).

[3]R. Haag, Z. Angew. Math. Mech. **32**, 197 (1952).

[4]W. Kundt, Ergeb. Exakten Naturwiss. **40**, 107 (1966).

[5]S. Shanmugadhasan, Proc. Camb. Philos. Soc. **59**, 743 (1963).

[6]A. Wintner, *The Analytical Foundations of Celestial Mechanics* (Princeton U. P., Princeton, New Jersey, 1941).

[7]T. Levi-Civita, Prace Mat.-Fiz. **17**, 1 (1906); T. Levi-Civita and U. Amaldi, *Lezioni di Meccanica Razionale* (Nicola Zanichelli, Bologna, 1927), Vol. II, Part 2.

[8]D. Djokić-Ristanović and Dj. Mušicki, Publ. Inst. Math. (Beograd) **10** (24), 25 (1970); Publ. Inst. Math. (Beograd) **10**, (24), 38 (1970).
[9]P. G. Bergmann and I. Goldberg, Phys. Rev. **98**, 531 (1955).
[10]N. Mukunda and E. C. G. Sudarshan, J. Math. Phys. **9**, 411 (1968).
[11]R. Hermann, Phys. Rev. **177**, 2453 (1969).
[12]S. Lie, *Theorie der Transformationsgruppen* (B. G. Teubner, Leipzig, 1890), Vol. II.
[13]J. A. Schouten and W. v. d. Kulk, *Pfaff's Problem and Its Generalizations* (Clarendon Press, Oxford, 1949).
[14]J. L. Anderson, Phys. Rev. **111**, 965 (1958).
[15]H. P. Künzle, Ann. Inst. Henri Poincaré A **11**, 393 (1969).
[16]L. H. Loomis and S. Sternberg, *Advanced Calculus* (Addison-Wesley, Reading, Massachusetts, 1968).
[17]W. F. Newns, Am. Math. Monthly **74**, 911 (1967).

Relations among elements of the density matrix. II. Exotic conservation laws

James L. Park* and William Band

Washington State University, Pullman, Washington 99163
(Received 6 June 1972)

Ordinary quantal conservation laws are associated with null operators for time rate of change and are valid for causal evolution of the system through either pure or mixed states. There is, however, a larger class of quantal constants of the motion, some of which are conserved only for pure state evolution. After an analysis of the theoretical origins of these exotic conserved quantities, several illustrations are presented with empirical interpretations based on the quorum theory methods used in Paper I.

1. ORDINARY CONSERVATION PRINCIPLES

The usual procedures for the identification of conserved quantities in quantum physics are based strongly on analogies to elegant classical schemes for obtaining constants of the motion. Thus, the formal parallelism between algebraic properties of Poisson brackets and quantal commutators is often exploited in the search for conserved quantal observables; similarly, Noether's theorem is routinely extended to quantum field theory as a means of finding expressions for conservation laws. However, because these methods rely so fundamentally upon the classical framework, they fail to generate all classes of quantally conserved measurable quantities.

To see why this is the case, it is necessary to recall the rather different relationships between data and theoretical observables that distinguish classical and quantal physics. Consider mechanics. In the classical case, observables are represented as functions of state (phase) and the numerical values of these functions are identified in principle with numerical data. A constant of the motion is then simply a phase function whose total time derivative vanishes, the consequent fixed value of the function being equal to the constant measured value of the observable represented by the function.

In quantum mechanics, on the other hand, observables are represented by Hermitian operators whose relation to data is more indirect. The testable assertions of quantum theory do not refer to "values" of observables, but, rather, to mean values of statistical collectives of data gathered from ensembles of identical experiments. Thus, to say in quantum mechanics that an observable A is "conserved" can mean, in terms of data, nothing more than that $\langle A\rangle_1$, the mean value computed from A-data referring to time t_1, is equal to $\langle A\rangle_2$, computed from A-data associated with t_2, where t_1 and t_2 are arbitrary.

To emphasize the difference between this quantal statement of conservation and that usually implied in classical theory, we shall call the classical version *point-by-point conservation* and the quantum idea *conservation-in-the-mean*. Crudely stated, a point-by-point conservation law asserts that "at every measurement the conserved observable has a definite (unique) value which is independent of time," whereas conservation-in-the-mean requires only that "the mean value of measurement-results on the observable is independent of the time lapse between preparation and measurement".

Since point-by-point conservation is in fact an unphysical concept, at best an abstract idealization from the facts of life in the physical laboratory, it could be argued cogently that conservation-in-the-mean with its realistic statistical statements should be acceptable whether one is using classical or quantum theory. Nevertheless, ordinary quantal conservation theory attempts to mimic its classical counterpart in the following well-known manner.

With every Hermitian operator A there is associated another such operator called the "time rate of change of A" (symbol dA/dt) and defined by

$$\frac{dA}{dt} = \frac{1}{i\hbar}[A,H] + \frac{\partial A}{\partial t} \tag{1}$$

where H is the Hamiltonian of the system, and $\partial A/\partial t$ denotes the time derivative of the operator A, should A be defined with intrinsic time dependence. The physical significance of dA/dt rests on a theorem which establishes that the *time derivative of the mean value of A* (a number empirically obtainable by computation from A-data) is numerically equal to the quantally calculated *mean value of the operator dA/dt*.

Because the form of (1) is reminiscent of the analogous classical Poisson bracket relation from which the necessary and sufficient condition for the point-by-point conservation of a physical quantity is immediately evident (vanishing of bracket plus intrinsic derivative), conventional quantum mechanics normally declares a conserved observable A to be one for which the operator dA/dt is null. Usually $\partial A/\partial t$ is zero and the criterion for ordinary conservation becomes simply the vanishing of the commutator $[A,H]$.

Since point-by-point conservation is meaningless in quantum physics, the latter standard formulation of quantal conservation theory is overly restrictive. In fact, the vanishing of dA/dt is a sufficient but *not* a necessary condition for A to be conserved-in-the-mean. We shall refer to the mean of an observable which satisfies this *sufficient* condition as an *ordinary* conserved quantity.

Consequently, as we shall demonstrate below, there exist operators A, for which dA/dt is not the null operator, but which nevertheless are conserved-in-the-mean. Moreover, we shall find that there exist time-independent nonlinear combinations of several quantal mean values, none of which is individually conserved in any sense. We call such extraordinary quantal constants of the motion *exotic conserved quantities*.

For later reference, one characteristic feature of *ordinary* conserved quantities should be especially noted: If dA/dt vanishes, then $\langle A\rangle$ is time-independent regardless of whether the evolving quantum state is pure or mixed. By contrast, there are exotic conserved quantities which are constant only for pure state evolution.

2. EXOTIC CONSERVED QUANTITIES

In Paper I we reviewed the concept of quorum[1] and indicated how elements of the statistical matrix could be expressed as functions of quorum means. Hence, if some algebraic combination of the matrix elements were invariant under temporal evolution, that combination could be physically interpreted using quorum theory and a conserved quantity would thereby be identified as a function of the quorum means.

A. Conservation of the statistical determinant

The causal evolution of a statistical operator ρ is effected by a unitary evolution operator U determined by the system Hamiltonian; thus

$$\rho(t_2) = U(t_2, t_1)\rho(t_1)U^\dagger(t_2, t_1). \tag{2}$$

From (2) and the theory of determinants it follows immediately that

$$\det\rho(t_2) = \det\rho(t_1), \tag{3}$$

i.e., the statistical determinant is a constant of the motion.

Naturally, this theorem will produce an interesting conserved quantity only for finite-dimensional Hilbert spaces.

B. Pure state conservation of statistical minors

In our study of definiteness inequalities in Paper I, we observed that the principal minor determinants of the statistical matrix can be of special significance. It was noted in particular, that for pure quantum states all such minor determinants of dimension exceeding unity vanish. Now from (2) it is readily shown that $\rho(t_2)$ will be pure if $\rho(t_1)$ was pure; i.e., pure states evolve into pure states, a well-known quantal theorem. Hence *for pure state evolution*, the quorum means occurring in any principal minor must vary in time in such a manner that the minor determinant remains fixed at zero. We have, therefore, a prolific source of measurable quantities conserved during pure state evolution.

The question now arises as to whether the minor determinants are also conserved in the time evolution of mixed states. Investigation shows that while it is possible in specific instances for minor determinants to be conserved for both pure and mixed states, in general only the pure state conservation law holds.

For example, consider a three-dimensional Hilbert space. Let the initial statistical matrix be

$$(\rho(t_1)) = \begin{pmatrix} w & a & b \\ a^* & x & c \\ b^* & c^* & y \end{pmatrix} \tag{4}$$

and let the evolution matrix for the time interval of interest be

$$(U) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{5}$$

After substituting (4) and (5) into (2) we obtain

$$(\rho(t_2)) = \begin{pmatrix} w & b & a \\ b^* & y & c^* \\ a^* & c & x \end{pmatrix}. \tag{6}$$

The upper left minor determinant is conserved only if

$$wx - |a|^2 = wy - |b|^2. \tag{7}$$

Certainly (7) is not generally true; it essentially demands the equality of two of the principal minor determinants of $\rho(t_1)$, a necessary condition only if $\rho(t_1)$ is pure.

We conclude that the principal minors of the statistical matrix, when interpreted in terms of quorum means, will provide a family of rather anomalous constants of the motion, always conserved in pure state evolution but not necessarily conserved otherwise. Thus, the pure state definiteness equalities exhibited in Paper I are examples of exotic conservation laws.

C. Conservation of functions of ρ

The mean value of any function of $\rho, F(\rho)$, is a conserved quantity, regardless of whether or not $dF(\rho)/dt$ is null. For example, consider $F(\rho) = \rho^2$:

$$\begin{aligned} \frac{d\rho^2}{dt} &= \frac{1}{i\hbar}[\rho^2, H] + \frac{\partial\rho^2}{\partial t} \\ &= \frac{1}{i\hbar}[\rho^2, H] + 2\rho\frac{\partial\rho}{\partial t}. \end{aligned} \tag{8}$$

According to the quantal Liouville theorem,

$$\frac{\partial\rho}{\partial t} = \frac{1}{i\hbar}[H, \rho]. \tag{9}$$

Hence,

$$\begin{aligned} i\hbar\frac{d\rho^2}{dt} &= \rho^2 H - H\rho^2 + 2\rho H\rho - 2\rho^2 H \\ &= -\rho^2 H + 2\rho H\rho - H\rho^2 = [[\rho, H], \rho] \not\equiv 0. \end{aligned} \tag{10}$$

Thus, ρ^2 is not conserved in the ordinary (classically inspired) sense because its associated time rate of change operator fails to vanish. Nevertheless, $\langle\rho^2\rangle$ is a constant of the motion due to a property of the trace

$$\begin{aligned} i\hbar\frac{d}{dt}\langle\rho^2\rangle &= i\hbar\left\langle\frac{d\rho^2}{dt}\right\rangle \\ &= i\hbar\,\mathrm{Tr}\{\rho(-\rho^2 H + 2\rho H\rho - H\rho^2)\}. \end{aligned} \tag{11}$$

Since $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$, the right side of (11) is zero, even if $d\rho^2/dt \not\equiv 0$.

Similarly, it can be shown that the mean value of any function of ρ is conserved. A famous case in point is $\ln\rho$, whose mean value is proportional to the entropy in statistical mechanics.

It is possible to relate the conservation of the statistical determinant discussed above to this idea that functions of ρ generate exotic constants of motion. If there exists an operator $D(\rho)$ such that

$$\langle D(\rho)\rangle = \mathrm{Tr}[\rho D(\rho)] = \det\rho, \tag{12}$$

then the conservation of $\det\rho$ could be regarded as a consequence of the fact that $\det\rho$ is the mean value of a function of ρ.

In general, many operators $D(\rho)$ can be found which satisfy (12). Let the eigenvalues of ρ be $\{w_k\}$ and of $D(\rho)$ be $\{d_k\}$. Since $\det\rho$ is the product of the eigenvalues of ρ, and ρ and $D(\rho)$ are both diagonal in the same matrix representation, (12) may be written as

$$w_1 w_2 \cdots w_N = \sum_{k=1}^{N} w_k d_k, \tag{13}$$

where N is the dimensionality of the Hilbert space.

One solution of (13) is given by

$$d_k = \begin{Bmatrix} w_2 \cdots w_N, & k = 1 \\ 0 \quad , & k \neq 1 \end{Bmatrix}. \tag{14}$$

Many other solutions could be obtained similarly by inspection of (13).

3. ILLUSTRATIONS

Examples of exotic conserved quantities are presented below in the same format used for definiteness inequalities in Paper I.

A. Spin-½ system

Quorum: $\sigma_x, \sigma_y, \sigma_z$.

Statistical matrix:

$$(\rho) = \tfrac{1}{2}\begin{pmatrix} 1 + \langle\sigma_z\rangle & \langle\sigma_x\rangle - i\langle\sigma_y\rangle \\ \langle\sigma_x\rangle + i\langle\sigma_y\rangle & 1 - \langle\sigma_z\rangle \end{pmatrix}. \tag{15}$$

Conserved statistical determinant:

$$\det\rho = \tfrac{1}{4}(1 - \langle\sigma_z\rangle^2 - \langle\sigma_x\rangle^2 - \langle\sigma_y\rangle^2). \tag{16}$$

From the time independence of (16) it follows that the quantity

$$\Sigma \equiv \langle\sigma_x\rangle^2 + \langle\sigma_y\rangle^2 + \langle\sigma_z\rangle^2 \tag{17}$$

is conserved. It is also possible to establish the constancy of (17) by standard manipulations, starting from the observation that

$$\frac{d(\sigma_x + \sigma_y + \sigma_z)^2}{dt} = 0. \tag{18}$$

B. Harmonic oscillator with 2-level energy cutoff

(For a short explanation of the concept of cutoff observable,[2] consult Paper I.)

Quorum: $x, p, H = (p^2/2m) + (m\omega^2/2)x^2$.

Statistical matrix: Let ρ_c denote the 2×2 nonzero submatrix of ρ. The cutoff is assumed to occur after the two lowest energy levels:

$$\rho_c = \tfrac{1}{2}\begin{pmatrix} 3 - \langle K\rangle & \langle X\rangle - i\langle P\rangle \\ \langle X\rangle + i\langle P\rangle & -1 + \langle K\rangle \end{pmatrix}, \tag{19}$$

where

$$K \equiv \frac{2}{\hbar\omega}H, X \equiv \left(\frac{2m\omega}{\hbar}\right)^{1/2} x, P \equiv \left(\frac{2}{m\hbar\omega}\right)^{1/2} p. \tag{20}$$

Conserved minor determinant: The submatrix is known to be correct only for times when the cutoff exists. However, since the cutoff is in the *energy* and the energy probability distribution for the oscillator is time-independent, we conclude that (19) is valid at all times.

According to the theory of Sec. 2B, $\det\rho_c$ is conserved for all pure state evolutions. Since $\det\rho_c$ contains the mean value of H, which is conserved in the ordinary sense, it follows that the terms in $\det\rho_c$ not containing $\langle H\rangle$ must be separately conserved.

Thus, our theory predicts that the quantity

$$\tfrac{1}{4}(\langle X\rangle^2 + \langle P\rangle^2) = \frac{m\omega}{2\hbar}\langle x\rangle^2 + \frac{1}{2m\hbar\omega}\langle p\rangle^2 = (\hbar\omega)^{-1}\mathcal{E}(\langle x\rangle, \langle p\rangle), \tag{21}$$

with

$$\mathcal{E}(\langle x\rangle, \langle p\rangle) \equiv \frac{\langle p\rangle^2}{2m} + \frac{m\omega^2}{2}\langle x\rangle^2, \tag{22}$$

will be conserved in pure state evolution of the cutoff harmonic oscillator. The final result is not new. It is well known that the harmonic oscillator meets the requirements of Ehrenfest's theorem,[3] hence the classical energy function (22) with quantal means as arguments is conserved for all types of time evolution, including the pure state, cutoff case of the present example. Note that $\mathcal{E}$ is not the same thing as $\langle H\rangle$; there is no single Hermitian operator associated with $\mathcal{E}$ yet it is a meaningful physical quantity.

C. Harmonic oscillator with 3-level energy cutoff

Quorum: $x, x^2, x^3, p, p^3, xp + px, H, H^2$.

Statistical matrix: Let ρ_c denote the 3×3 nonzero submatrix of ρ. The cutoff is assumed to occur after the three lowest energy levels.

$$\rho_c = \begin{pmatrix} -\tfrac{1}{8}[8\langle K\rangle - \langle K^2\rangle - 15] & \tfrac{1}{2}[2\langle X\rangle - \langle Z\rangle - i(2\langle P\rangle - \langle Q\rangle)] & \tfrac{1}{2}[(1/\sqrt{2})(\langle Y\rangle - \langle K\rangle) - i\langle A\rangle] \\ \tfrac{1}{2}[2\langle X\rangle - \langle Z\rangle + i(2\langle P\rangle - \langle Q\rangle)] & \tfrac{1}{4}[6\langle K\rangle - \langle K^2\rangle - 5] & (1/2\sqrt{2})[\langle Z\rangle - \langle X\rangle - i(\langle Q\rangle - \langle P\rangle)] \\ \tfrac{1}{2}[(1/\sqrt{2}(\langle Y\rangle - \langle K\rangle) + i\langle A\rangle] & (1/2\sqrt{2})[\langle Z\rangle - \langle X\rangle + i(\langle Q\rangle - \langle P\rangle)] & \tfrac{1}{8}[\langle K^2\rangle - 4\langle K\rangle + 3] \end{pmatrix}, \tag{23}$$

where

$$\begin{aligned} &X \equiv (2m\omega/\hbar)^{1/2}x, \quad Y \equiv (2m\omega/\hbar)x^2, \\ &\quad Z \equiv \tfrac{1}{3}(2m\omega/\hbar)^{3/2}x^3, \\ &P \equiv (2/m\hbar\omega)^{1/2}p, \quad A = (1/\hbar\sqrt{2})(xp + px), \\ &\quad Q \equiv \tfrac{1}{3}(2/m\hbar\omega)^{3/2}p^3, \\ &K \equiv (2/\hbar\omega)H, \quad K^2 = [4/(\hbar\omega)^2]H^2. \end{aligned} \tag{24}$$

Conserved minor determinant: Consider the upper left 2×2 minor of ρ_c. The diagonal elements, being functions of H, are conserved separately. Hence we may assert that, at least for pure state evolution, the following quantity is a constant of the motion:

$$\Lambda_1 \equiv [2\langle X\rangle - \langle Z\rangle]^2 + [2\langle P\rangle - \langle Q\rangle]^2. \tag{25}$$

Similarly, from the lower right 2×2 minor another conserved quantity may be derived:

$$\Lambda_2 \equiv [\langle X\rangle - \langle Z\rangle]^2 + [\langle P\rangle - \langle Q\rangle]^2. \tag{26}$$

Subtracting Λ_2 from Λ_1 and simplifying, we get

$$\Lambda_3 = 3(\langle X\rangle^2 + \langle P\rangle^2) - 2(\langle X\rangle\langle Z\rangle + \langle P\rangle\langle Q\rangle), \tag{27}$$

which is of course likewise conserved for pure state evolution. But,

$$\langle X\rangle^2 + \langle P\rangle^2 = 4(\hbar\omega)^{-1}\mathcal{E}(\langle x\rangle, \langle p\rangle), \tag{28}$$

where $\mathcal{E}$ is defined as in (22).

Recalling again that $\mathcal{E}(\langle x\rangle, \langle p\rangle)$ is a constant of the motion because of Ehrenfest's theorem, we conclude that the following combination of quorum means is an exotic conserved quantity at least for pure state evolution

$$\begin{aligned}\Lambda &\equiv \frac{4}{3\hbar^2}(\langle X\rangle\langle Z\rangle + \langle P\rangle\langle Q\rangle)\\ &= (m\omega)^2\langle x\rangle\langle x^3\rangle + (m\omega)^{-2}\langle p\rangle\langle p^3\rangle.\end{aligned} \tag{29}$$

Note that Λ is a nonlinear function of four quorum means and that Λ has *not* been obtained by finding an operator L such that

$$\langle L\rangle = \Lambda, \quad \frac{dL}{dt} = 0. \tag{30}$$

Additional exotic conserved quantities for this system could similarly be generated from the remaining two-dimensional minor and from the three dimensional minor determinant ($\det\rho_c$).

D. Spin-1 system

Several exotic conservation laws may be obtained by calculating the determinant and minor determinants of the statistical matrix given by (20) in Paper I.

---

*Work supported by Research Corporation.

[1] W. Band and J. L. Park, Found. Phys. **1**, 211 (1971).
[2] J. L. Park and W. Band, Found. Phys. **1**, 339 (1971).
[3] A. Messiah, *Quantum Mechanics* (Amsterdam, North-Holland, 1961), p. 217.

Canonical transformations and accidental degeneracy. I. The anisotropic oscillator

J. D. Louck*, M. Moshinsky†, and K. B. Wolf

Instituto de Física and C.I.M.A.S.S., Universidad de México, México, D.F., México
(Received 20 November 1972)

The problem of accidental degeneracy in quantum mechanical systems has fascinated physicists for many decades. The usual approach to it is through the determination of the generators of the Lie algebra responsible for the degeneracy. In these papers we want to focus from the beginning on the symmetry Lie group of canonical transformations in the classical picture. We shall then derive its representation in quantum mechanics. In the present paper we limit our discussion to the anisotropic oscillator in two dimensions, though we indicate possible extensions of the reasoning to other problems in which we have accidental degeneracy.

1. INTRODUCTION

The subject of accidental degeneracy in quantum mechanical systems, i.e., degeneracies not associated with obvious groups of symmetries, has fascinated physicists for many decades. The two classical problems in this field have been the isotropic harmonic oscillators and the particle in a Coulomb potential. The nature of the Lie algebra responsible for the accidental degeneracy in these problems has been known for a long time.[1,2] However, the Lie groups of canonical transformations generated by these Lie algebras, apart from their geometrical invariance subgroups, have been discussed only recently.[3,4]

Besides the problems mentioned, there are others that present features of accidental degeneracy in the quantum picture. The question is then raised about general procedures for obtaining the Lie groups of canonical transformations responsible for these features.

To be able to focus on these procedures we decided to analyze systematically three simple problems in two dimensional configuration space that have accidental degeneracy: (1) the anisotropic oscillator when the ratio of the frequencies is rational,[1,5,6] (2) the isotropic oscillator constrained to a sector of the plane of angle π/q with q integer, (3) the Calogero problem[7] of particles moving in one dimension and interacting through potentials that depend both in the square and the inverse square of the distance between the particles. When we are dealing with three particles and eliminate the center of mass this problem can be reformulated in a two dimensional configuration space.

In this and the following paper we analyze cases (1) and (2), reserving the Calogero[7] problem for a later publication. While we shall be discussing very special systems, we will continuously try to keep in mind the general ideas behind these problems to see what is the information they supply on the abstract question of Lie groups of canonical transformations and accidental degeneracy.

2. ACCIDENTAL DEGENERACY IN AN ANISOTROPIC OSCILLATOR WHOSE FREQUENCIES HAVE A RATIONAL RATIO

The anisotropic oscillator whose frequencies have a rational ratio has been extensively discussed in the literature.[1,5,6] In the pioneering work of Jauch and Hill[1] the generators of the Lie algebra for both the isotropic and anisotropic oscillator (in the latter case for the two dimensional problem where the ratio of the frequencies was 1: 2) were obtained in the classical picture. Demkov[5] then discussed the different subsets of the set of states of the anisotropic oscillator that have the familiar degeneracy associated with $SU(2)$, and obtained the generators of this group in the quantum picture. Cisneros and McIntosh[6] greatly extend and complement the analysis in their search for a universal symmetry group in two dimensions.

From these and other papers it would appear that further discussion of the problem is unnecessary. The present approach differs though in that it goes directly into the determination of the canonical transformation that in the classical picture maps the anisotropic oscillator on the isotropic one. As the latter has a symmetry group of linear canonical transformations[3] that are a representation of $SU(2)$, we can combine them with those that give the mapping, to obtain the symmetry group of the anisotropic oscillator. Once the classical picture is clear we can pass to the creation and annihilation operators in the quantum picture which have different forms for the different subsets of states mentioned in the previous paragraph.[5] From them we can construct the generators of the $SU(2)$ group responsible for the accidental degeneracy of the two dimensional anisotropic oscillator whose frequencies have a rational ratio.

Besides its intrinsic interest, the present approach provides part of the ground work required in the next paper where we analyze the accidental degeneracy of the oscillator in a sector of angle π/q. It may also be useful in other problems[7] that have an energy spectrum similar to that of the anisotropic oscillators.[7]

Let us consider now a particle of mass unity moving in a plane under the influence of a quadratic potential whose frequencies in the X_i, $i = 1, 2$, directions are ω_i. The Hamiltonian is then

$$H = \tfrac{1}{2} (P_1^2 + \omega_1^2 X_1^2) + \tfrac{1}{2} (P_2^2 + \omega_2^2 X_2^2). \tag{2.1}$$

We shall assume, furthermore, that

$$\omega_1/\omega_2 = k_2/k_1, \quad \text{or } k_1\omega_1 = k_2\omega_2 \equiv \omega, \tag{2.2}$$

where k_1, k_2 are relatively prime integers. Without loss of generality we may take $\omega = 1$ or, equivalently,

$$\omega_i = k_i^{-1}, \quad i = 1, 2. \tag{2.3}$$

We now introduce creation and annihilation variables under the definitions

$$\eta_i \equiv (1/\sqrt{2}\,), (k_i^{-1/2} X_i - ik_i^{1/2} P_i),$$

$$\xi_i \equiv (1/\sqrt{2}\,), (k_i^{-1/2} X_i + ik_i^{1/2} P_i), \qquad i = 1, 2. \tag{2.4}$$

In the quantum mechanical picture where $[X_i, P_j] = i\delta_{ij}$ (we take $\hbar = 1$), the variables η_i, ξ_i become operators and the Hamiltonian (2. 1) takes then the form

$$H = k_1^{-1}\eta_1\xi_1 + k_2^{-1}\eta_2\xi_2 + (2k_1)^{-1} + (2k_2)^{-1}. \tag{2. 5}$$

The eigenstates of H are given by

$$|\nu_1\nu_2\rangle = (\nu_1!\,\nu_2!)^{-1/2}\eta_1^{\nu_1}\eta_2^{\nu_2}|0\rangle, \tag{2. 6}$$

where ν_1, ν_2 are nonnegative integers and $|0\rangle$ is the ground state $\pi^{-1/2}\exp[-\frac{1}{2}(X_1^2 + X_2^2)]$.

We now proceed to divide the set of states (2. 6) into subsets characterized by a pair of indices (λ_1, λ_2) given by

$$\nu_i \equiv \lambda_i \bmod k_i, \quad \lambda_i = 0, 1, 2, \cdots k_i - 1, \quad i = 1, 2. \tag{2. 7}$$

From the range of values of the λ_i we conclude that there are k_1k_2 different subsets of states, which can be characterized by the kets

$$|n_1k_1 + \lambda_1, n_2k_2 + \lambda_2\rangle \tag{2. 8}$$

for given λ_1, λ_2 restricted as in (2. 7) and for arbitrary nonnegative n_1, n_2. Immediately we see that the states (2. 8) are eigenstates of H with eigenvalues

$$E = (n_1 + n_2) + k_1^{-1}(\lambda_1 + \tfrac{1}{2}) + k_2^{-1}(\lambda_2 + \tfrac{1}{2}), \tag{2. 9}$$

and, hence, those members of the subset (2. 7) for which $n_1 + n_2$ is the same have equal energy and give rise to accidental degeneracy. It is important to stress that the accidental degeneracy is present only for states within the subset labelled by (λ_1, λ_2) and not for those belonging to different subsets even if their $n_1 + n_2$ happen to be the same. We note also that each subset of states can be put into one-to-one correspondence with the full set of states of the isotropic oscillator. Thus, we can speak of k_1k_2 copies of the fundamental degeneracy pattern.

The question of what is the Lie algebra and the Lie group responsible for this accidental degeneracy then arises. We shall endeavor to answer it both in the classical and quantum picture in the next sections.

3. THE SYMMETRY LIE ALGEBRA AND GROUP FOR THE CLASSICAL ANISOTROPIC OSCILLATOR

We proceed first to analyze the classical system. The Hamiltonian then has the form (2. 5) where we suppress the last two constant terms and in which η_i, ξ_i are the combinations (2. 4) of classical coordinates and momenta. We shall introduce a canonical transformation which maps this Hamiltonian to another one corresponding to an isotropic harmonic oscillator where the Lie algebra of the symmetry group is well known.[3]

Before proceeding with our analysis, we note that from (2. 4) the classical Poisson bracket of two observables F, G can be written as

$$\{F, G\} = \sum_i\left(\frac{\partial F}{\alpha X_i}\frac{\partial G}{\partial P_i} - \frac{\partial F}{\partial P_i}\frac{\partial G}{\partial X_i}\right) = i\sum_i\left(\frac{\partial F}{\partial\eta_i}\frac{\partial G}{\partial\xi_i} - \frac{\partial F}{\partial\xi_i}\frac{\partial G}{\partial\eta_i}\right) \tag{3. 1}$$

which implies that $\{\eta_j, \xi_k\} = i\delta_{jk}$. Thus, if we have η_j', ξ_k' as functions η_j, ξ_k such that $\{\eta_j', \xi_k'\} = i\delta_{jk}$, we can be sure that X_j', P_j' defined as

$$X_j' = (1/\sqrt{2}), (\eta_j' + \xi_j'), \quad P_j' = (i/\sqrt{2}), (\eta_j' - \xi_j'), \quad j = 1, 2 \tag{3. 2}$$

are canonically conjugate. If ξ_i' is also the complex conjugate of η_i', the canonical transformation is real.

We now consider the following canonical transformation in the classical picture

$$\eta_i' = k_i^{-1/2}(\eta_i\xi_i)^{(1-k_i)/2}\eta_i^{k_i}, \quad \xi_i' = \xi_i^{k_i}k_i^{-1/2}(\eta_i\xi_i)^{(1-k_i)/2}. \tag{3. 3}$$

From (3. 1) we see that $\{\eta_j', \xi_k'\} = i\delta_{jk}$ and, besides, as $\xi_i = \eta_i^*$ we obtain $\xi_i' = \eta_i'^*$. Furthermore,

$$H = k_1^{-1}\eta_1\xi_1 + k_2^{-1}\eta_2\xi_2 = \eta_1'\xi_1' + \eta_2'\xi_2'. \tag{3. 4}$$

Thus (3. 3) is a real canonical transformation that reduces the Hamiltonian (2. 5) to that of an isotropic harmonic oscillator.

The symmetry group[3] for the two dimensional isotropic harmonic oscillator is the unitary group $U(2)$ whose generators are

$$\eta_i'\xi_j', \quad i, j = 1, 2 \tag{3. 5}$$

and for which the Lie algebra is determined by the Poisson brackets

$$\{\eta_i'\xi_j', \eta_k'\xi_l'\} = -i(\eta_i'\xi_l'\delta_{kj} - \eta_k'\xi_j'\delta_{il}). \tag{3. 6}$$

The Lie symmetry group of the anisotropic oscillator relates the creation and annihilation classical variables $\bar\eta_i$, $\bar\xi_i$ to η_i, ξ_i through the following steps: First, inverting the relations (3. 3) and writing all variables with a bar above we get

$$\begin{aligned}\bar\eta_i &= k_i^{1/2}\,\bar\eta_i'^{(k_i+1)/2k_i}\,\bar\xi_i'^{(k_i-1)/2k_i},\\ \bar\xi_i &= k_i^{1/2}\,\bar\xi_i'^{(k_i+1)/2k_i}\,\bar\eta_i'^{(k_i-1)/2k_i}.\end{aligned} \tag{3. 7a}$$

Then we note that $\bar\eta_i'$, $\bar\xi_j'$ are related to η_i', ξ_j' by a unitary transformation generated by (3. 5) and, thus, we can write[3]

$$\bar\eta_i' = \sum_j U_{ij}\eta_j', \quad \bar\xi_i' = \sum_j U_{ij}^*\xi_j', \tag{3. 7b}$$

where $\|U_{ij}\|$ is a 2×2 unitary matrix and $\|U_{ij}^*\|$ is its complex conjugate. Finally, η_i', ξ_j' are related to η_i, ξ_j through (3. 3), which we can also write as

$$\eta_i' = k_i^{-1/2}\eta_i^{(k_i+1)/2}\xi_i^{(1-k_i)/2}, \quad \xi_i' = k_i^{-1/2}\xi_i^{(k_i+1)/2}\eta_i^{(1-k_i)/2}. \tag{3.7c}$$

It is clear that the full transformation (3. 7) leaves the Hamiltonian (3. 4) invariant and, thus, is a realization of $U(2)$ which is the symmetry Lie group of the anisotropic oscillator. We can then make use of the transformations (2. 4) and their inverse to express the elements of this group as real canonical transformations.

Having analyzed the classical Lie algebra and symmetry group, we turn now our attention to the quantum picture.

4. THE GENERATORS AND THE UNITARY REPRESENTATION OF THE SYMMETRY GROUP IN THE QUANTUM PICTURE

In the quantum picture the creation and annihilation variables η_i, ξ_i become operators. Therefore, η_i', ξ_i' of (3. 3) must also be expressed as operators that act on the states (2. 7). As η_i, ξ_i do not commute, there are ambiguities in the translation of the classical relations

(3.3) into operator form. How can we get rid of these ambiguities? We shall use the isotropic oscillator as a guide. In there $k_1 = k_2 = 1$, implying $\lambda_1 = \lambda_2 = 0$, so that we have a single set of states of the form (2.7), i.e., $|n_1 n_2\rangle$. At the same time, when $k_1 = k_2 = 1$, $\eta_i' = \eta_i$, $\xi_i' = \eta_i$, $\xi_i' = \xi_i$. Thus, in the isotropic case we have

$$\begin{aligned} \eta_1'|n_1n_2\rangle &= (n_1+1)^{1/2}|n_1+1,n_2\rangle, \\ \eta_2'|n_1n_2\rangle &= (n_2+1)^{1/2}|n_1,n_2+1\rangle, \\ \xi_1'|n_1n_2\rangle &= n_1^{1/2}|n_1-1,n_2\rangle, \\ \xi_2'|n_1n_2\rangle &= n_2^{1/2}|n_1,n_2-1\rangle. \end{aligned} \tag{4.1}$$

For the anisotropic case we require the operators η_i', ξ_j' to have the same effect on each subset of states $|n_1k_1 + \lambda_1, n_2k_2 + \lambda_2\rangle$ characterized by fixed λ_1, λ_2. This would automatically[3] guarantee that the generators $\eta_i'\xi_j'$ of $U(2)$ connect the states (of the subset of given λ_1, λ_2) for which $n_1 + n_2$ is fixed, i.e., of the same energy, showing that this symmetry group is responsible for the accidental degeneracy in the anisotropic oscillator whose frequencies have a rational ratio.

For fixed λ_1, λ_2 we shall now define the creation and annihilation operators η_i', ξ_i' as

$$\eta_i' = k_i^{-1/2}(\eta_i\xi_i - \lambda_i)^{1/2}[(\eta_i\xi_i)(\eta_i\xi_i - 1)\cdots (\eta_i\xi_i - k_i + 1)]^{-1/2}\eta_i^{k_i}, \tag{4.2a}$$

$$\xi_i' = \xi_i^{k_i}[(\eta_i\xi_i)(\eta_i\xi_i - 1)\cdots (\eta_i\xi_i - k_i + 1)]^{-1/2}(\eta_i\xi_i - \lambda_i)^{1/2}k_i^{-1/2}. \tag{4.2b}$$

We claim that (4.2) are the right quantum analogies of (3.3) when applied to the eigenstates (2.7) of the number operators $\eta_i\xi_i$, $i = 1, 2$. First, they are well defined in this basis. Second, the classical limit $\hbar \to 0$ of (4.2) is (3.3) as can be seen by keeping $\hbar$ and ω in the notation.

We now apply η_1' to the state

$$\begin{aligned} &\eta_1'|n_1k_1+\lambda_1, n_2k_2+\lambda_2\rangle \\ &= k_1^{-1/2}(\eta_1\xi_1-\lambda_1)^{1/2}[(\eta_1\xi_1)(\eta_1\xi_1-1)\cdots (\eta_1\xi_1-k_1+1)]^{-1/2}[(n_1k_1+\lambda_1)!(n_2k_2+\lambda_2)!]^{-1/2} \\ &\qquad \times \eta_1^{(n_1+1)k_1+\lambda_1}\eta_2^{n_2k_2+\lambda_2}|0\rangle \\ &= (n_1+1)^{1/2}\{[(n_1+1)k_1+\lambda_1]![n_2k_2+\lambda_2]!\}^{-1/2} \\ &\qquad \times \eta_1^{(n_1+1)k_1+\lambda_1}\eta_2^{n_2k_2+\lambda_2}|0\rangle \\ &= (n_1+1)^{1/2}|(n_1+1)k_1+\lambda_1, n_2k_2+\lambda_2\rangle. \end{aligned} \tag{4.3}$$

In a similar fashion, we can apply η_2', ξ_1', ξ_2' to (2.7) and we get kets in which, respectively, $n_1, n_2 \to n_1, n_2 + 1$; $n_1, n_2 \to n_1 - 1, n_2$; $n_1, n_2 \to n_1, n_2 - 1$, multiplied by factors $(n_2+1)^{1/2}$, $n_1^{1/2}$, $n_2^{1/2}$. It is important to stress that for each set $(\lambda_1\lambda_2)$ we have different η_i', ξ_i' as indicated in (4.2). In particular, (4.3) does not hold if the λ's for the operators and the kets do not match.

Let us introduce, for fixed λ_1, λ_2, a shorthand notation for the ket (2.7) of the form

$$|n_1k_1 + \lambda_1, n_2k_2 + \lambda_2\rangle \equiv |jm\}, \tag{4.4a}$$

in which

$$j \equiv \tfrac{1}{2}(n_1 + n_2), \qquad m \equiv \tfrac{1}{2}(n_1 - n_2). \tag{4.4b}$$

Furthermore, we denote the generators of our $SU(2)$ symmetry group by the notation

$$\begin{aligned} T_+ &= T_1 + iT_2 = \eta_1'\xi_2', \qquad T_3 = \tfrac{1}{2}(\eta_1'\xi_1' - \eta_2'\xi_2'), \\ T_- &= T_1 - iT_2 = \eta_2'\xi_1'. \end{aligned} \tag{4.5}$$

From (4.3) and similar relations, we obtain then

$$\begin{aligned} \{j'm'|T_\pm|jm\} &= [(j \mp m)(j \pm m + 1)]^{1/2}\delta_{jj'}\delta_{mm'}, \\ \{j'm'|T_3|jm\} &= m\delta_{jj'}\delta_{mm'}. \end{aligned} \tag{4.6}$$

thus seeing that in each of the k_1k_2 subsets of states of the anisotropic oscillator with fixed λ_1, λ_2, the matrix elements of the generators of $SU(2)$ have the standard form.

We can now turn to the question of the unitary representation of the $SU(2)$ symmetry group in the quantum picture. In the usual way, we define[8]

$$R(\alpha, \beta, \gamma) = e^{i\alpha T_3}e^{i\beta T_2}e^{i\gamma T_3}, \tag{4.7}$$

and have

$$R(\alpha, \beta, \gamma)|jm\}_{\lambda_1\lambda_2} = \sum_{m'} |jm'\}_{\lambda_1\lambda_2}\mathfrak{D}^j_{m'm}(\alpha\beta\gamma), \tag{4.8}$$

where the $\mathfrak{D}$'s are the Wigner matrices.[8] We stress the fact that both $|jm\}$ and $|jm'\}$ [as well as the operator $R(\alpha, \beta, \gamma)$] in (4.8) correspond to the same subset characterized by a fixed λ_1, λ_2, by now adding a subscript $\lambda_1\lambda_2$ to the kets $|jm\}$.

We, furthermore, note that

$$\begin{aligned} &{}_{\lambda_1'\lambda_2'}\{j'm'|jm\}_{\lambda_1\lambda_2} \\ &= \langle n_1'k_1 + \lambda_1', n_2'k_2 + \lambda_2'|n_1k_1 + \lambda_1, n_2k_2 + \lambda_2\rangle \\ &= \delta_{n_1'n_1}\delta_{n_2'n_2}\delta_{\lambda_1'\lambda_1}\delta_{\lambda_2'\lambda_2}. \end{aligned} \tag{4.9}$$

Actually, the matrix element (4.9) is different from 0 only if

$$n_ik_i + \lambda_i = n_i'k_i + \lambda_i' \quad \text{or} \quad (n_i - n_i')k_i = \lambda_i' - \lambda_i, \tag{4.10}$$

but as $\lambda_i, \lambda_i' = 0, 1, \cdots k_i - 1$, the Eq. (4.10) has a solution only when $n_i = n_i'$, $\lambda_i' = \lambda_i$.

From (4.8), (4.9) we then reach the conclusion that the unitary representation of the $SU(2)$ symmetry group responsible for the accidental degeneracy, with respect to the eigenstates (2.7) of the Hamiltonian λ, is given by

$${}_{\lambda_1'\lambda_2'}\{j'm'|R(\alpha, \beta, \gamma)|jm\}_{\lambda_1\lambda_2} = \delta_{\lambda_1'\lambda_1}\delta_{\lambda_2'\lambda_2}\delta_{j'j}\mathfrak{D}^j_{m'm}(\alpha\beta\gamma). \tag{4.11}$$

We have obtained, from (4.2) and (4.5) the generators of the $SU(2)$ group and from (4.11) its unitary representation. In the next section we analyze the general conclusions that one can draw from the complete analysis of the group responsible for the accidental degeneracy of the anisotropic oscillator.

5. CONCLUSIONS

From the analysis of the problem of accidental degeneracy in an anisotropic oscillator system whose ratio of frequencies is rational, one possible general procedure emerges.

We first must solve fully the quantum mechanical problem and see what is the structure of the set of states that have the same energy. If this happens to be a structure that we normally associate with an $SU(2)$ group, i.e., we have sets of states that have degeneracy $1, 2, 3, 4, \cdots$, we can look into the possibility of finding a classical canonical transformation that maps the Hamiltonian of the problem into that of a two dimensional harmonic oscillator. If the structure of accidental degeneracy is similar to that of other well-known problems in mechanics, the classical canonical transformations we may look for is the one that maps our problem into the well-known one.

Once we have this canonical transformation we can rewrite the generators of the Lie Algebra of the well-known problem in a way that makes them the generators of the symmetry group of the problem under study. We can then obtain the Lie symmetry group by a procedure similar to that in (3.7).

But our problem does not finish with the analysis in classical mechanics. We must then express the generators of our group, and frequently the creation and annihilation operators from which they are built, in the quantum picture as is done for example in (4.2). We expect this quantum mechanical formulation to reduce to the classical one in the limit $\hbar \to 0$, but we may find, as was clearly seen in Sec. 4, that the generators of the Lie algebra may have different forms in the different subsets of states of the quantum problem.

Once an explicit form of the generators is available in the quantum picture, we could pass to the determination of the unitary representation of the group of canonical transformation along the lines also discussed in Sec. 4.

While the procedure outlined for finding the groups responsible for accidental degeneracy seems fairly general, we shall show in the next article that it does not apply to some other simple problems. We shall illustrate an alternative development when we discuss in the following paper the problem of the isotropic oscillator in a sector of angle π/q where q is integer.

We are indebted to Dr. P. A. Mello for many useful discussions.

---

*Permanent address, Los Alamos Scientific Laboratory, Los Alamos, New Mexico.

†Asesor del Instituto Nacional de Energía Nuclear y miembro de El Colegio Nacional.

[1] J. M. Jauch and E. L. Hill, Phys. Rev. **57**, 641 (1940).

[2] V. Fock, Z. Phys. **98**, 145 (1935); V. Bargmann, Z. Phys. **99**, 576 (1936).

[3] M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971); J. Math. Phys. **12**, 1780 (1971).

[4] V. Bargmann, private communication.

[5] 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).

[6] A. Cisneros and H. V. McIntosh, J. Math. Phys. **11**, 870 (1970).

[7] F. Calogero, J. Math. Phys. **10**, 2191 (1969); J. Math. Phys. **12**, 419 (1971).

[8] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, N. J., 1957), Chap. 4.

Canonical transformations and accidental degeneracy. II. The isotropic oscillator in a sector

J. D. Louck*, M. Moshinsky†, and K. B. Wolf

Instituto de Física and C.I.M.A.S.S., Universidad de México, México, D.F., México

(Received 20 November 1972)

In this paper we discuss the accidental degeneracy in the problem of a particle in two dimensional oscillator potential constrained to move in a sector of angle $\pi/q, q$ integer. The degeneracy is due to both the Hamiltonian and the boundary conditions. The symmetry Lie group of canonical transformations is suggested by the explicit form of a complete nonorthonormal set of states expressed in terms of the creation operators. This group is complex and the corresponding representation in quantum mechanics is nonunitary. We discuss briefly the appearance of complex canonical transformations in physical problems.

1. INTRODUCTION

In the preceeding paper[1] we analyzed the symmetry groups of canonical transformations responsible for the accidental degeneracy of the anisotropic oscillators whose ratio of frequencies was rational. From the discussion of these problems we arrived at some general conclusions for the determination of the groups. There are, however, other problems in which accidental degeneracy is present which seem to require a different type of approach. One of these problems is the motion of a particle in a two dimensional configuration space under the action of an harmonic oscillator potential, but restricted to a sector of the plane of angle π/q, where q is a positive integer. This sector is drawn in Fig. 1 for $q = 3$ and the heavy lines indicate the infinite potential barriers that limit it. We shall analyze this problem in the present paper both because of its intrinsic interest and the insight it provides into the general problem of accidental degeneracy.

The classical trajectory is very easy to draw. The particle under an oscillator potential moving unconstrained in the full plane will have an elliptical trajectory centered at the origin of the potential. We can draw this trajectory on a transparent plastic napkin. Then folding the napkin in such a way that it sustains an angle π/q, we immediately see the orbit of the particle as modified by the barriers at the boundary of the sector. This orbit is periodic and nonergodic,[2] i.e., it does not fill all the phase space surface of constant energy. It is

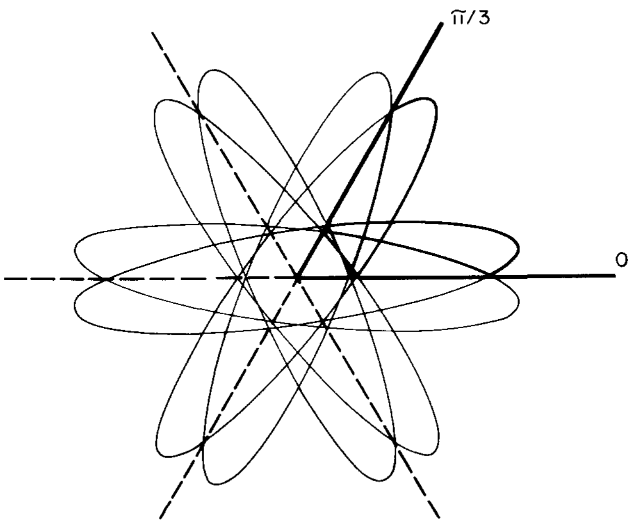

FIG. 1. Classical trajectory of a particle (bold lines) subject to an harmonic oscillator potential and restricted to a sector $\pi/3$ in the plane.

drawn in Fig. 1 for $q = 3$, where we also show the reflection of the orbit as if the barriers were mirrors.

Does the corresponding quantum mechanical problem have accidental degeneracy? In polar coordinates r, φ the Schrödinger equation (in which $\hbar$, the mass of the particle and the frequency of the oscillator are taken as 1) has solutions, subject to the condition that the wave function vanishes at $\varphi = 0, \pi/q$, of the form[3]

$$\langle r\varphi \mid \nu_1\nu_2\rangle = [2(\nu_1!)[(\nu_1 + \nu_2 q + q)!]^{-1/2} \times r^{(\nu_2+1)q} L_{\nu_1}^{(\nu_2+1)q}(r^2)\, e^{-r^2/2}\, \pi^{-1/2} \sin[q(\nu_2+1)\varphi]. \tag{1.1}$$

Notice that the state (1.1) is normalized with respect to the surface element $r dr d\varphi$ over the whole plane. We denote the state in the full Dirac notation, though later when referring to it we shall abbreviate to the ket $|\nu_1\nu_2\rangle$; the $L_{\nu_1}^{(\nu_2+1)q}(r^2)$ are associated Laguerre polynomials[4] with ν_1, ν_2 being arbitrary nonnegative integers. We write the solution in terms $\nu_2 + 1$, rather than ν_2, so that the lowest energy state of this problem corresponds to $\nu_1 = \nu_2 = 0$. The eigenvalue of the Hamiltonian[3] for the state $|\nu_1\nu_2\rangle$ is

$$E_{\nu_1\nu_2} = 2\nu_1 + q\nu_2 + q + 1. \tag{1.2}$$

We now proceed to discuss separately the cases in which q is odd and even. In the first case we divide the set of states (1.1) into $2q$ subsets characterized by λ_1, λ_2 defined by

$$\nu_1 \equiv \lambda_1 \bmod q, \quad \lambda_1 = 0, 1, 2, \cdots, q-1, \tag{1.3a}$$

$$\nu_2 \equiv \lambda_2 \bmod 2, \quad \lambda_2 = 0, 1, \tag{1.3b}$$

which implies that we may write

$$\nu_1 = qn_1 + \lambda_1, \quad \nu_2 = 2n_2 + \lambda_2, \tag{1.3c}$$

where n_1, n_2 are nonnegative integers. The energy $E_{\nu_1\nu_2}$ of (1.2) satisfies the equation

$$(E_{\nu_1\nu_2} - q - 1)/(2q) = n_1 + n_2 + (\lambda_1/q) + (\lambda_2/2). \tag{1.4}$$

For q even we can write the energy (1.2) as

$$\tfrac{1}{2}(E_{\nu_1\nu_2} - q - 1) = \nu_1 + (q/2)\nu_2. \tag{1.5}$$

We then divide the set of states (1.1) into $q/2$ subsets characterized by

$$\nu_1 \equiv \lambda_1 \bmod (q/2), \quad \lambda_1 = 0, 1, \cdots, (q/2) - 1, \tag{1.6a}$$

$$\nu_2 \equiv \lambda_2 \bmod 1, \quad \lambda_2 = 0, \tag{1.6b}$$

which implies that we may write

$$\nu_1 = (q/2)n_1 + \lambda_1, \quad \nu_2 = n_2. \tag{1.6c}$$

Thus, for q even we have

$$(E_{\nu_1\nu_2} - q - 1)/q = n_1 + n_2 + (2\lambda_1/q). \tag{1.7}$$

For both q odd and even, states corresponding to different (λ_1, λ_2) have different energies, but for a given (λ_1, λ_2) and a fixed value $n_1 + n_2 = N$ we have states that are degenerate in the energy $N + 1$ times.

In so far as the energy spectrum is concerned and the degeneracy of the states, the problem with q odd looks very similar to an anisotropic oscillator[1] whose ratio of frequencies is $(k_2/k_1) = (2/q)$, while for q even the ratio is $(k_2/k_1) = (1/[q/2])$. As the spectrum of the isotropic oscillator appears in $2q$ (q odd) or $q/2$ (q even) copies, we suspect that the group responsible for the accidental degeneracy in the present problem can be derived from $SU(2)$ by some canonical transformation.[1] Unfortunately, we cannot obtain this $SU(2)$ group by mapping the Hamiltonian of our problem on an isotropic oscillator, as the restriction on the states comes both from the Hamiltonian in the Schrödinger equation and the boundary conditions at $\varphi = 0, \pi/q$. We seem to require, then, a completely new approach and one is suggested in the next section when we express the states of the oscillator in a sector in terms of creation and annihilation operators.

2. CREATION AND ANNIHILATION OPERATORS AND THE STATES OF THE OSCILLATOR IN A SECTOR

When dealing with the two dimensional isotropic quantum oscillator it is convenient to introduce the spherical components of coordinate and momenta by the definition

$$X_\pm = (1/\sqrt{2})(X_1 \pm iX_2), \quad P_\pm = (1/\sqrt{2})(P_1 \pm iP_2), \tag{2.1}$$

where $p_i = -i\partial/\partial x_i$. From them we can in turn construct the creation operators

$$\eta_\pm = (1/\sqrt{2})(X_\pm - iP_\pm), \tag{2.2}$$

which in polar coordinates, where $x_\pm = re^{\pm i\varphi}$, take the form

$$\eta_\pm = \tfrac{1}{2} e^{\pm i\varphi}\left(r - \frac{\partial}{\partial r} \mp \frac{i}{r}\frac{\partial}{\partial\varphi}\right). \tag{2.3}$$

We note the following symmetry properties of these operators: If we have a reflection across the $X_2 = 0$ line in the plane, i.e.,

$$\varphi \to -\varphi, \quad \text{then } \eta_\pm \to \eta_\mp. \tag{2.4a}$$

If we carry out a rotation by an angle π/q, i.e.,

$$\varphi \to \varphi + (\pi/q), \quad \text{then } \eta_\pm \to e^{\pm i\pi/q}\eta_\pm, \tag{2.4b}$$

and thus, in particular, we have that when

$$\varphi \to \varphi + (\pi/q), \quad \text{then } \eta_\pm^q \to -\eta_\pm^q \tag{2.4c}$$

The Hamiltonian of the two dimensional oscillator can now be written as

$$H = \eta_+\xi_+ + \eta_-\xi_- + 1, \tag{2.5}$$

where $\xi_\pm$ is the annihilation operator

$$\xi_\pm = \eta_\pm^\dagger = \frac{1}{\sqrt{2}}(X_\mp + iP_\mp) = \tfrac{1}{2} e^{\mp i\varphi}\left(r + \frac{\partial}{\partial r} \mp \frac{i}{r}\frac{\partial}{\partial\varphi}\right). \tag{2.6}$$

In terms of $\eta_\pm, \xi_\pm$ the angular momentum takes the form

$$L = X_1P_2 - X_2P_1 = \frac{1}{i}\frac{\partial}{\partial\varphi} = \eta_+\xi_+ - \eta_-\xi_-. \tag{2.7}$$

The state (1.1) is an eigenstate of the Hamiltonian (2.5) with eigenvalue (1.2) and of the square of the angular momentum L^2 with eigenvalue $(\nu_2 + 1)^2q^2$. Thus, it can also be written in terms of creation operators as

$$\langle r\varphi \mid \nu_1\nu_2\rangle = 2^{-1/2}\{[\nu_1 + (\nu_2 + 1)q]!\,\nu_1!\}^{-1/2} \times (\eta_+\eta_-)^{\nu_1}[\eta_+^{(\nu_2+1)q} - \eta_-^{(\nu_2+1)q}]\mid 0\rangle, \tag{2.8}$$

where the symmetry properties (2.4) of $\eta_\pm$ guarantee that the wave function vanishes at $\varphi = 0, \pi/q$. The ket $|0\rangle$ is the ordinary ground state $\pi^{-1/2}\exp(-\frac{1}{2}r^2)$.

We note immediately one basic difference between the states (2.8) and those of (I.2.7) for the anisotropic oscillator. The latter can be written as

$$|n_1k_1 + \lambda_1, n_2k_2 + \lambda_2\rangle = [(n_1k_1 + \lambda_1)!(n_2k_2 + \lambda_2)!]^{-1/2} \times (\eta_1^{k_1})^{n_1}(\eta_2^{k_2})^{n_2}\eta_1^{\lambda_1}\eta_2^{\lambda_2}|0\rangle, \tag{2.9}$$

and thus almost immediately suggest the classical canonical transformation (I.3.3) [or its quantum mechanical version (I.4.2)] as, for example, the creation operator η_1' when applied to (2.9) transforms it into a state in which $n_1 \to n_1 + 1, n_2 \to n_2$.

The states (2.8) are differences of monomial products of creation operators and not just a single product of powers of basic operators as (2.9). We can though express our states in the latter form if we are willing to settle for a complete, linearly independent, but not orthonormal set of states. For this purpose let us write

$$\langle r\varphi \mid \nu_1\nu_2) = (\nu_1!\,\nu_2!)^{-1/2}(\eta_+\eta_-)^{\nu_1}(\eta_+^q + \eta_-^q)^{\nu_2}(\eta_+^q - \eta_-^q)|0\rangle, \tag{2.10}$$

where we use a round bracket for the ket $|\nu_1\nu_2)$ to distinguish the state from the one defined by $|\nu_1\nu_2\rangle$ in (2.8). As the polynomial in η_+, η_- appearing in (2.10) is homogeneous the ket $|\nu_1\nu_2)$ is an eigenstate of the Hamiltonian (2.5) with eigenvalues given by (1.2). It remains then to prove that it vanishes at $\varphi = 0, \pi/q$. We note from (2.4a) that under a change $\varphi \to -\varphi$, $|\nu_1\nu_2) \to -|\nu_1\nu_2)$ and thus $(r, 0 \mid \nu_1\nu_2) = 0$. Furthermore, from (2.4), for $\varphi \to \varphi + (\pi/q)$, $|\nu_1\nu_2) \to (-1)^{q(\nu_2+1)}|\nu_1\nu_2)$ so that

$$\langle r, \pi/q \mid \nu_1\nu_2) = (-1)^{q(\nu_2+1)}\langle r, 0 \mid \nu_1\nu_2) = 0. \tag{2.11}$$

The energy spectrum (1.2) was analyzed in the previous section and thus, again, we see that the states (2.10) are degenerate $N + 1$ times for a given (λ_1, λ_2), if $n_1 + n_2 = N$ and the relation between n_1, n_2 and ν_1, ν_2 is given by (1.3c) when q is odd, or (1.6c) when q is even.

We note that the states (2.10) corresponding to different energies are, of course, orthogonal as H of (2.5) is Hermitian. On the other hand the states of the same energy in the $N + 1$ degenerate multiplet are not orthonormal as seen from their scalar product using the commutation relations $[\xi_\pm, \eta_\pm] = 1$, $[\xi_\mp, \eta_\pm] = 0$. Thus, we still have to prove that they are linearly independent. We shall do this for q odd and a similar analysis holds for q even. From (1.3c), and as (λ_1, λ_2) is fixed, we conclude that the part of the polynomial in (2.10) that changes with each state of multiplet of energy $2qN + (2\lambda_1 + q\lambda_2 + q + 1)$ is given by

$$(\eta_+\eta_-)^{qn_1}(\eta_+^q + \eta_-^q)^{2n_2}. \tag{2.12}$$

As $n_1 + n_2 = N$, we see from (2.12) that the highest power that η_+ can take appears in the term

$$\eta_+^{q(N+n_2)}\eta_-^{qn_1}. \tag{2.13}$$

Thus, for $n_2 = N$, $n_1 = 0$ the term η_+^{2qN} is present in (2.12). For any $n_2 < N$, $n_1 = N - n_2$, this term cannot appear and thus the state (2.10) in which (for q odd) $\nu_1 = \lambda_1$, $\nu_2 = 2N + \lambda_2$, is independent from all the others corresponding to a given (λ_1, λ_2) and N. But, clearly, we can show in the same way that for $n_2 = N - 1$, $n_1 = 1$, the term $\eta_+^{q(2N-1)}\eta_-^q$ does not appear for any $n_2 < N - 1$, $n_1 = N - n_2$, and continuing in this fashion prove that all the states (2.10) are linearly independent.

The states $|\nu_1\nu_2)$ of (2.10) now have a form very similar to those of (I.2.7) in the sense that they are given by a single product of certain simple polynomial functions of the creation operators. We shall take advantage of this fact to derive, first classically and then quantum mechanically, the Lie Algebra and Lie group responsible for the accidental degeneracy of the problem of the oscillator in a sector.

3. CLASSICAL LIE ALGEBRA AND SYMMETRY GROUP FOR THE HARMONIC OSCILLATOR IN A SECTOR

In this section we shall think of $\eta_\pm, \xi_\pm$ not as operators but as classical functions of X_i, P_i as defined through (2.1), (2.2) and (2.6). From these functions we see that the Poisson bracket of any two variables F, G can now be expressed as

$$\{F, G\} = i\left(\frac{\partial F}{\partial \eta_+}\frac{\partial G}{\partial \xi_+} - \frac{\partial F}{\partial \xi_+}\frac{\partial G}{\partial \eta_+}\right) + i\left(\frac{\partial F}{\partial \eta_-}\frac{\partial G}{\partial \xi_-} - \frac{\partial F}{\partial \xi_-}\frac{\partial G}{\partial \eta_-}\right), \tag{3.1}$$

which implies $\{\eta_\pm, \xi_\pm\} = i$, $\{\eta_\mp, \xi_\pm\} = 0$.

Looking now at the states (2.10) and using as an analogy the analysis of the previous paper for the states (I.2.7), it seems appropriate to define new creation variables as

$$\eta_1 \equiv \eta_+\eta_-, \quad \eta_2 \equiv \eta_+^q + \eta_-^q. \tag{3.2}$$

Note that the creation variables defined by (3.2) are not to be confused with those given in the previous paper in terms of coordinates and momenta in the directions $i = 1, 2$. The annihilation variables ξ_1, ξ_2 corresponding to them must be canonically conjugate, i.e.,

$$\{\eta_i, \xi_j\} = i\delta_{ij}, \quad i,j = 1,2. \tag{3.3}$$

From the standpoint of commutators, this implies

$$[\xi_j, \eta_i] = \delta_{ij} \tag{3.4}$$

and we can represent ξ_j as $\partial/\partial\eta_j$. We shall use this representation to derive in a simple fashion the ξ_j. From (3.2) we have that

$$\eta_- = \eta_1/\eta_+, \; \eta_+ = \{\tfrac{1}{2}[\eta_2 + (\eta_2^2 - 4\eta_1^q)^{1/2}]\}^{1/q}, \tag{3.5}$$

and, thus,

$$\frac{\partial}{\partial \eta_j} = \frac{\partial \eta_+}{\partial \eta_j}\frac{\partial}{\partial \eta_+} + \left(\frac{1}{\eta_+}\delta_{1j} - \frac{\eta_-}{\eta_+}\frac{\partial \eta_+}{\partial \eta_j}\right)\frac{\partial}{\partial \eta_-}. \tag{3.6}$$

Interpreting now $\partial/\partial\eta_\pm$ as $\xi_\pm$ and making use of the fact that from (3.5)

$$\eta_+^q - \eta_-^q = (\eta_2^2 - 4\eta_1^q)^{1/2}, \tag{3.7}$$

we obtain

$$\xi_1 = (\eta_+^{q-1}\xi_- - \eta_-^{q-1}\xi_+)(\eta_+^q - \eta_-^q)^{-1}, \tag{3.8a}$$

$$\xi_2 = q^{-1}(\eta_+\xi_+ - \eta_-\xi_-)(\eta_+^q - \eta_-^q)^{-1}. \tag{3.8b}$$

We can now easily check that the η_i of (3.2) and ξ_i of (3.8) satisfy the Poisson bracket relation (3.3).

We note that ξ_1, ξ_2 are *not* the complex conjugates of η_1, η_2 (nor Hermitian conjugates in the quantum case) and, thus, if we make use of the customary relations between creation and annihilation variables and new coordinates and momenta, we are *not* led to real canonical transformations. As we shall show in the next section, this seems to be related with the fact that the quantum mechanical representation of the symmetry group of canonical transformations is not unitary. We furthermore show in Sec. 5 that complex canonical transformations are involved in several important problems in physics. Thus, their appearance in the symmetry group of the oscillator in a sector is not an isolated event.

From (3.2), (3.8) we can immediately check that

$$2\eta_1\xi_1 + q\eta_2\xi_2 = \eta_+\xi_+ + \eta_-\xi_-. \tag{3.9}$$

If q is odd we can divide both sides by $2q$ and the left hand side has the form (I.3.4) of the anisotropic oscillator with $k_1 = q$, $k_2 = 2$. If q is even we divide by q and again the left hand side has the form (I.3.4), but with $k_1 = (q/2)$, $k_2 = 1$.

To arrive now at the generators of the classical Lie algebra and the symmetry group for the problem of the sector we need still to transform the Hamiltonian of the anisotropic oscillator appearing in (3.9) into that of the isotropic one. As shown in the preceeding paper, we can do this if we carry out the canonical transformation

$$\eta_i' = k_i^{-1/2}(\eta_i\xi_i)^{(1-k_i)/2}\eta_i^{k_i}, \quad \xi_i' = k_i^{-1/2}\xi_i^{k_i}(\eta_i\xi_i)^{(1-k_i)/2}, \tag{3.10}$$

where k_i, $i = 1, 2$ takes the values indicated in the previous paragraph for q odd and even. Under this transformation the Hamiltonian in (3.9) becomes proportional to

$$\mathcal{H} = \eta_1'\xi_1' + \eta_2'\xi_2', \tag{3.11}$$

and thus the generators of the Lie algebra of its symmetry group[1] are given by

$$T_+ = \eta_1'\xi_2', \quad T_3 = \tfrac{1}{2}(\eta_1'\xi_1' - \eta_2'\xi_2'), T_- = \eta_2'\xi_1'. \tag{3.12}$$

We can immediately check that the Poisson brackets (3.1) of the variables $T_\pm, T_3$ and the Hamiltonian $\mathcal{H}$ are zero, while among themselves they lead to the Lie algebra of $SU(2)$.

To obtain the classical symmetry group associated with this Lie algebra of $SU(2)$ we must proceed as in Sec. 3 of the preceeding paper. We shall only outline the steps as their algebraic implementation is trivial. We relate the new creation and annihilation variables $\bar{\eta}_\pm, \bar{\xi}_\pm$ with

$\eta_\pm, \xi_\pm$ in the following way: First we invert the expressions (3.2), (3.8) to determine $\bar\eta_\pm, \bar\xi_\pm$ in terms of $\bar\eta_i, \bar\xi_i$, $i = 1, 2$. Then we invert (3.10) to obtain $\bar\eta_i, \bar\xi_i$ in terms $\bar\eta'_i, \bar\xi'_i$. The $\bar\eta'_i, \bar\xi'_i$ are related to η'_i, ξ'_i by the $U(2)$ transformation (I.3.7b). Finally, we can express η'_i, ξ'_i, $i = 1, 2$ in terms of $\eta_\pm, \xi_\pm$ through (3.10) and then (3.2), (3.8).

Having analyzed the classical Lie algebra and the symmetry group, we turn now our attention to the quantum picture.

4. THE GENERATORS AND THE REPRESENTATION OF THE SYMMETRY GROUP IN THE QUANTUM PICTURE

In the quantum picture the creation and annihilation variables return to their roles as operators, but then we must also express η_i, ξ_i, $i = 1, 2$ of (3.2), (3.8) as operators that act on the state (2.10) without ambiguities. We have no problem for the effect of η_1, η_2 of (3.2) on the state $|\nu_1\nu_2)$ of (2.10) as they are polynomial functions of η_+, η_- only and these commute. For ξ_1, ξ_2 we have both η_+, η_- and ξ_+, ξ_- in (3.8) which do not commute and, furthermore, ξ_1, ξ_2 contain the factor $\eta_+^q - \eta_-^q$ to a negative power. Yet we shall assume that ξ_1, ξ_2 as operators are given by (3.8) in the order in which η_+, η_-, ξ_+, ξ_- appear.

Due to the presence of the factor $(\eta_+^q - \eta_-^q)^{-1}$ in (3.8), the determination of the matrix elements of ξ_1, ξ_2 with respect to a complete set of orthonormal states in the sector, such as $|\nu_1\nu_2\rangle$ of (2.8), seems impossible. We note though that the states $|\nu_1\nu_2\rangle$ can be expanded in terms of the complete but not orthonormal set $|\nu_1\nu_2)$ of (2.10) with the help of transformation brackets that will be discussed below. Thus, we need only to see whether the application of ξ_1, ξ_2 to the states $|\nu_1\nu_2)$ can be carried out. As all the states $|\nu_1\nu_2)$ have a factor $(\eta_+^q - \eta_-^q)$, the $(\eta_+^q - \eta_-^q)^{-1}$ in ξ_1, ξ_2 just cancels it. Furthermore, as the commutators $[\xi_\pm, \eta_\pm] = 1$, $[\xi_\mp, \eta_\pm] = 0$, when applying the operators $\xi_\pm$ to polynomials in the creation operators $\eta_\pm$, we can replace the former by $\partial/\partial\eta_\pm$. Using these considerations, we obtain from (3.2), (3.8) and the explicit form (2.10) for the state $|\nu_1\nu_2)$ that

$$\eta_1 \,|\, \nu_1\nu_2) = (\nu_1 + 1)^{1/2} \,|\, \nu_1 + 1, \nu_2), \tag{4.1a}$$

$$\eta_2 \,|\, \nu_1\nu_2) = (\nu_2 + 1)^{1/2} \,|\, \nu_1, \nu_2 + 1), \tag{4.1b}$$

$$\xi_1 \,|\, \nu_1\nu_2) = \nu_1^{1/2} \,|\, \nu_1 - 1, \nu_2), \tag{4.1c}$$

$$\xi_2 \,|\, \nu_1\nu_2) = \nu_2^{1/2} \,|\, \nu_1, \nu_2 - 1). \tag{4.1d}$$

The behavior of the η_i, ξ_i with respect to the states $|\nu_1\nu_2)$ is then entirely similar to that of the creation and annihilation operators in the two directions $i = 1, 2$ of the anisotropic oscillator with respect to the corresponding state (I.2.6). Just as in the case of the anisotropic oscillator, we can now divide the set of states $|\nu_1\nu_2)$ of (2.10) into subsets characterized by (λ_1, λ_2). As indicated in the introduction, there will be $2q$ subsets for q odd and $(q/2)$ for q even. For each one of these subsets of states we can pass, again as in the anisotropic oscillator, from the operators η_i, ξ_i to η'_i, ξ'_i by the transformation (I.4.2), where $k_1 = q$, $k_2 = 2$ for q odd, $k_1 = q/2$, $k_2 = 1$ for q even. The quantum mechanical generators of the symmetry group of the oscillator in a sector continue to be given by (3.12), but now the η'_i, ξ'_i, $i = 1, 2$ in it, are obtained for each subset (λ_1, λ_2) of states (2.10) in terms of $\eta_\pm, \xi_\pm$ through (I.4.2) and (3.2), (3.8).

To see what is the effect of a finite group transformation of the form $R(\alpha, \beta, \gamma)$ of (I.4.2) on the states $|\nu_1\nu_2)$ of (2.10), we first rewrite them as

$$|\nu_1\nu_2) \equiv |jm\}_{\lambda_1\lambda_2}, \tag{4.2}$$

where $j = \frac{1}{2}(n_1 + n_2)$, $m = \frac{1}{2}(n_1 - n_2)$ and $n_1, n_2, \lambda_1, \lambda_2$ are related to ν_1, ν_2 by (1.3c) when q is odd and (1.6c) for q even. It is immediately clear then that, as in (I.4.7),

$$R(\alpha, \beta, \gamma) \,|\, jm\}_{\lambda_1\lambda_2} = \sum_{m'} |\, jm'\}_{\lambda_1\lambda_2} \mathfrak{D}^j_{m'm}(\alpha\beta\gamma). \tag{4.3}$$

From this result we wish now to obtain the representation of the $SU(2)$ transformation with respect to the set of orthonormal states $|\nu_1\nu_2\rangle$ of (2.8). We require first the development of the states $|jm\}_{\lambda_1\lambda_2}$ of (4.2) and (2.10) in terms of $|\nu_1\nu_2\rangle$. Using the notation (4.2), we can write

$$|jm\}_{\lambda_1\lambda_2} = \sum_{\nu_1\nu_2} |\nu_1\nu_2\rangle\langle\nu_1\nu_2 \,|\, jm\}_{\lambda_1\lambda_2}. \tag{4.4}$$

The summation extends over the finite set of states corresponding to the same energy which implies that ν_1, ν_2 correspond to the same values of λ_1, λ_2 appearing in the round ket $|\nu_1\nu_2)$ of (2.10). The transformation brackets in (4.4) can be easily obtained from the expansion of the polynomial in (2.10), and the matrix

$$\|\langle\nu_1\nu_2 \,|\, jm\}\|, \tag{4.5}$$

where we suppressed λ_1, λ_2 for a clearer notation, is invertible as the set of states (2.10) is linearly independent. Denoting by $\langle\nu_1\nu_2 \,|\, jm\}^{-1}$ the elements of the inverse matrix, we have now that

$$\begin{aligned}\langle\nu'_1\nu'_2 \,|\, R(\alpha,\beta,\gamma) \,|\, \nu_1\nu_2\rangle &= \langle\nu'_1\nu'_2 \,|\, R(\alpha,\beta,\gamma) \sum_m |\, jm\}_{\lambda_1\lambda_2}\langle\nu_1\nu_2 \,|\, jm\}^{-1} \\ &= \sum_{m', m} \langle\nu'_1\nu'_2 \,|\, jm'\} \,\mathfrak{D}^j_{m'm}(\alpha\beta\gamma)\,\langle\nu_1\nu_2 \,|\, jm\}^{-1}.\end{aligned} \tag{4.6}$$

We note that, again as in the case of the anisotropic oscillator, the matrix elements are different from zero only when $|\nu_1\nu_2\rangle, |\nu'_1\nu'_2\rangle$ belong to the same subset of states characterized by a given (λ_1, λ_2). Furthermore, the corresponding n_1, n_2 and n'_1, n'_2 related to ν_1, ν_2 and ν'_1, ν'_2 by (1.3c) or (1.6c), must satisfy $n_1 + n_2 = n'_1 + n'_2$ due to the invariance of the Hamiltonian under the transformation.

It is important to notice that the representation of the $SU(2)$ group in the quantum mechanical picture is no longer unitary due to the transformation brackets in (4.6). This seems related to the complex character of the canonical transformation as indicated in the previous section.

5. CONCLUSIONS

We can draw the following conclusions from our procedure of deriving the Lie algebra and symmetry group of a plane oscillator in a sector of angle π/q. We note first that in this problem we required the expression of the wave function in terms of creation operators acting on the lowest energy state. The states that proved useful for our purpose were the nonorthonormal ones (2.10) given as powers of certain simple polynomials in the creation operators. The form of these states then suggested the group of complex canonical transformations responsible for accidental degeneracy.

If we can decompose the states of other problems where accidental degeneracy is present in terms of powers of some basic operators acting on a ground state, we may hope that a similar procedure could give us an insight into their symmetry group. A problem with this structure is the one proposed by Calogero,[5] where particles in one dimension interact through a quadratic and inverse quadratic potentials in their relative distances. In the case when we have only three particles, and after eliminating the center of mass, we get a problem in the plane. Perelomov[6] has shown how the states of this problem can be written as products of powers of two operators acting on the ground state. The situation resembles very much that in the expression (2.10), but now the two operators are not only functions of η_+, η_- but of the coordinates r, φ as well. Thus, if we identify these two operators with η_1, η_2 as in (3.2), it is considerably more difficult to find the corresponding ξ_1, ξ_2. The problem is being studied at present and we hope to present it in a third article in this series.

The procedure followed in the present paper leads to a symmetry group which is a group of complex canonical transformations. Now normally in mechanics we are concerned with real canonical transformations and so the question arises whether the complex variety appears elsewhere than in the present problem. We wish to indicate that the simple group of complex linear canonical transformations

$$\begin{pmatrix}\bar{x}\\ \bar{p}\end{pmatrix} = \begin{pmatrix} a & ib\\ -ic & d\end{pmatrix}\begin{pmatrix}x\\ p\end{pmatrix}, \quad ad - bc = 1, \quad a,b,c,d \text{ real} \tag{5.1}$$

has a number of interesting applications.

We note first that the matrices appearing in (5.1) form a group as a product of two of the type leads to another of the same form. It is also a group of canonical transformations as $\{\bar{x}, \bar{p}\} = 1$. The representation, which is nonunitary, can be derived from the results obtained in the paper of Moshinsky and Quesne[7] for real linear canonical transformations when we replace b by ib and so, when $b \neq 0$, it takes the form

$$\langle x' \mid U \mid x''\rangle = (2\pi \mid b \mid)^{-1} \exp[-(2b)^{-1}(ax'^2 - 2x'x'' + dx''^2)]. \tag{5.2}$$

When $a = d = 0$, $b = -c = 1$ we get of the kernel of the Laplace transform, while in the corresponding real case, i.e., $\bar{x} = p, \bar{p} = -x$, the representation, which is unitary, gives the kernel of the Fourier transform.[7]

When we have

$$a = d = b = -c = 1/\sqrt{2}, \tag{5.3}$$

$\bar{x}$ is just the annihilation operator and the representation (5.2) corresponds to the states[7] for which $\bar{x}$ is diagonal, i.e., the coherent states of optics.[8]

When $a = d = 1$, $b \neq 0$ and $c = 0$, the representation is a Gaussian and so the transformation[7]

$$\mid x''\rangle = \int \mid x'\rangle\, dx' \langle x' \mid U \mid x''\rangle \tag{5.4}$$

provides a Gaussian transform of the type used in clustering theory by Brink.[9] Its inverse can then be determined purely from the fact that it corresponds to a representation of the transformation (5.1).

The expression (5.2) also appears in a very important fashion as a kernel in clustering theory as was shown by Kramer.[10] The realization that it is a representation (5.1) is very important for the factorization and products of such kernels.

Thus, the complex linear canonical transformation (5.1) and its nonunitary representation plays an important role in several branches of physics. It is, therefore, not surprising that other complex canonical transformations and their nonunitary representations appear in relation with problems such as the symmetry group of the plane oscillator in a sector of angle π/q.

*Permanent address: Los Alamos Scientific Laboratory, Los Alamos, New Mexico.
†Asesor del Instituto Nacional de Energía Nuclear y miembro de El Colegio Nacional.

1 J. D. Louck, M. Moshinsky, and K. B. Wolf, J. Math. Phys. **14**, 692 (1973).
2 P. A. Mello and M. Moshinsky, Ann. Phys. (N.Y.) **71**, 1 (1972).
3 M. Moshinsky, T. H. Seligman, and K. B. Wolf, J. Math. Phys. **13**, 901 (1972).
4 P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 784.
5 F. Calogero, J. Math. Phys. **10**, 2191 (1969); J. Math. Phys. **12**, 419 (1971).
6 A. M. Perelomov, Teor. Mat. Fiz. **6**, 364 (1971).
7 M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971); J. Math. Phys. **12**, 1780 (1971).
8 E. M. Henley and W. Thirring, *Elementary Quantum Field Theory* (McGraw-Hill, New York, 1962), Chap. 6.
9 P. M. Brink, "Generator coordinate method and geometry of alpha-clusters" in *Clustering phenomena in nuclei*, Bochum Conference, International Atomic Energy Agency, Vienna, 1969.
10 P. Kramer, private communication.

The evaluation of lattice sums. II. Number-theoretic approach

M. L. Glasser

Battelle Memorial Institute, Columbus, Ohio 43201

(Received 1 December 1972

Number theory is used to sum several series of the form $\Sigma a_{ijk} \cdots (ai^2 + bj^2 + \cdots)^{-s}$ in two and three dimensions.

I. INTRODUCTION

In the first paper of this series[1] analytic procedures were presented by whose means lattice sums of the Madelung form for crystal surfaces could be evaluated. The purpose of this paper is to supplement that discussion by showing how the results obtained there and even a much wider class of sums can be found rather simply by the use of elementary number theory. This is somewhat interesting in itself, for the theory of numbers is generally considered as being remote from the concerns of applied mathematics.

As was the case in Paper I, we simply reexpress one infinite series as a combination of others, so what is meant by "evaluating" a lattice sum should be made explicit. Here we say that a lattice sum has been evaluated exactly when it has been reexpressed in terms of familiar constants and Dirichlet L-series

$$L(s, \chi) = \sum_{n=1}^{\infty} \chi(n) n^{-s}. \tag{1}$$

The quantities $\chi(n)$, known as *Dirichlet characters*, have the properties

$$\begin{aligned} |\chi(n)| &= 0, 1, \\ \chi(mn) &= \chi(m)\chi(n). \end{aligned} \tag{2}$$

Prototypical of these series are the two functions

$$\zeta(s) = L(s, 1) \quad \text{and} \quad \beta(s) = L(s, c_1),$$

where

$$c_1(n) = \begin{cases} 0 & n \text{ even} \\ (-1)^{(n-1)/2} & n \text{ odd}, \end{cases} \tag{3}$$

of which extensive use was made in I. Every L-series satisfies a functional equation similar to that for the Riemann zeta function and is easily calculated for all real s as illustrated for $\beta(s)$ in I. (Except for the zeta function, the L-series are entire functions of the complex variable s).

Lattice sums of the Madelung type can be taken to have the form

$$\sum_{n_1, n_2, \cdots} \{Q(n_1, n_2, \cdots)\}^{-s}, \tag{4}$$

where Q is an integral quadratic form, and are called Epstein zeta functions, although little is known about the latter in general. In this paper, as in I, we are concerned with the binary case

$$Q(m,n) = am^2 + bmn + cn^2 \tag{5}$$

and primarily with the case $b = 0$. These series were apparently first considered by Dirichlet[2] about a century and a half ago, but there has been little interest in them per se among number theorists and only the case (I-2) appears to be well known.

In Sec. II we review some facts about number theory and quadratic forms; several examples are presented in Sec. III.

II. SOME ELEMENTARY NUMBER THEORY

All the facts we require are essentially contained in the first ninety pages of Dickson's *Introduction to the Theory of Numbers*[3] and we adopt his notation.

The discriminant of the form (5) is the quantity

$$d = b^2 - 4ac. \tag{6}$$

When $d < 0$, Q has the same sign for all values $(m,n) \neq (0,0)$ and is called *definite*. We shall assume that this sign is positive. If one makes a *unimodular* substitution

$$m = \alpha x + \beta y, \quad n = \gamma x + \delta y \tag{7}$$

where $\alpha, \beta, \gamma, \delta$ are integers and $\alpha\delta - \beta\gamma = 1$, the new form $Q'(x,y)$ is called *equivalent* to Q; the discriminant is preserved and it can be shown that for each d there is only a finite number of equivalence classes. This number is called the *class number* $C(d)$. If $Q(x,y) = n$, the form Q is said to represent n and the number $F_Q(n)$ of such representations is finite when $d < 0$. Thus, when $Q(m,n)$ is positive definite, for any function F,

$$S = \sum_{(m,n)\neq(0,0)} F[Q(m,n)] = \sum_{n=1}^{\infty} F_Q(n) F(n), \tag{8}$$

where the first sum is over all pairs of integers, both positive and negative. Furthermore, since when Q is equivalent to Q', we clearly have $F_Q(n) = F_{Q'}(n)$, so

$$\sum_{(m,n)\neq(0,0)} f[Q(m,n)] = \sum_{(m,n)\neq(0,0)} f[Q'(m,n)]. \tag{9}$$

For example, if we make the substitution $m = p + 2q$, $n = q$ and take $s = 2$ in (I. 2), we find the double sum

$$\sum_{p,q=1} \frac{p^4 + 26p^2q^2 + 25q^4}{(p^4 - 4p^2q^2 + 25q^4)^2} = \frac{\pi^2}{3}\left(G - \frac{13}{50}\right) \tag{10}$$

where G is Catalan's constant.

If two integers a, b leave the same remainder when divided by n, they are called congruent modulo n: $a \equiv b \pmod n$. If m is congruent to a square, modulo n, then m is called a quadratic residue of n. When n is a prime, one defines the Legendre symbol

$$(m \mid n) = \begin{cases} 0 & n \text{ divides } m \\ +1 & m \text{ is a quadratic residue} \pmod n. \\ -1 & \text{otherwise} \end{cases} \tag{11}$$

Legendre's symbol is multiplicative (D32),

$$(mn \mid p) = (m \mid p)(n \mid p) \tag{12}$$

Let $n = p_1^{e_1} \cdots p_l^{e_l}$, then the Kronecker symbol is defined as

$$(m \mid n) = (m \mid p_1)^{e_l} \cdots (m \mid p_l)^{e_l} \tag{13}$$

(D77), where those on the right are Legendre symbols.

The class number of a positive definite form of discriminant $d = -p$ is given by

$$C(d) = (1/P)(A - B),$$

where A is the sum of quadratic residues (mod p) and B is the sum of the nonquadratic residues. The calculation of this quantity for general d occupies a large portion of number theory. An extensive list of representations for classes of forms belonging to discriminants $0 > d > -400$ is given on D85.

The number w of nontrivial unimodular substitutions which leave a form invariant, and which depends only on the discriminant, is called the automorphism number and has the value

$$\begin{aligned} w &= 4 \quad \text{for the equivalence class } \{a(x^2 + y^2)\}, \\ w &= 6 \quad \text{for the equivalence class } \{a(x^2 + xy + y^2)\}, \\ w &= 2 \quad \text{for all other forms.} \end{aligned} \tag{14}$$

Dirichlet has shown that the number of times an integer k is represented by a complete set of representatives for the forms of discriminant d is

$$R(k) = w(d) \sum_{m|k} (d \mid m) \tag{15}$$

when k has no divisors in common with d and the sum is over all divisors of k. When k has common factors with $d, R(k)$ must be found by stealth as will be illustrated in Sec. III. There are two cases to consider. If the sets of integers represented by the various equivalence classes of forms for a given discriminant are disjoint, which is described by saying that there is a single class in each genus, there exists a character which takes a distinct value on each such set of integers. Thus a projection function can be formed which will vanish for all integers except those represented by a given equivalence class and which when multiplied by (15) will give the representation number for any desired form. (When the class number is unity, this is trivial.) When there is more than one class in each genus, no such characters can be found and the representation problem has not yet been solved. We shall therefore consider only the former case.

III. EXAMPLES

As a first example, we shall rederive the formulas in the table of Paper I by purely number theoretic means. Consider the forms $Q_1 = m^2 + n^2$ and $Q_2 = m^2 + 4n^2$. Since $C(-4) = C(-16) = 1$, each of these forms can be dealt with separately.

When k is odd, it can have no factor in common with either discriminant, so from (15)

$$\begin{aligned} R_1(k) &= 4 \sum_{m|k} (-4 \mid m), \\ R_2(k) &= 2 \sum_{m|k} (-16 \mid m). \end{aligned} \tag{16}$$

Since m must be odd, in which case $(2 \mid m) = \pm 1$, $(-1 \mid m) = (-1)^{(m-1)/2}$, we have

$$\begin{aligned} (-4 \mid m) &= (1 \mid m)(2 \mid m)^2 = (-1)^{(m-1)/2}, \\ (-16 \mid m) &= (-1 \mid m)(2 \mid m)^4 = (-1)^{(m-1)/2}, \end{aligned} \tag{17}$$

so

$$R_2(k) = \tfrac{1}{2} R_1(k).$$

For Q_1, if k is even, say $k = 2q$, then $2q = x^2 + y^2$ implies that $x + y = 2a$, $x - y = 2b$, say, so $q = a^2 + b^2$, $R_1(k) = R_1(q)$, and thus we can eliminate all factors of 2 from k. Hence for all k

$$R_1(k) = 4 \sum_{\substack{m|k \\ \text{odd}}} (-1)^{(m-1)/2}. \tag{18}$$

In the case of Q_2, if $k = 2q$ where q is odd, then $2q = x^2 + 4y^2$ implies that $x = 2a$ so $q = 2(a^2 + y^2)$ which is impossible, so $R_2(2q) = 0$. Proceeding similarly, we see that

$$R_2(k) = \begin{cases} \tfrac{1}{2} R_1(k) & k \text{ odd} \\ 0 & k = 2q,\ q \text{ odd.} \\ R_1(k) & k = 4q,\ q \text{ even or odd.} \end{cases} \tag{19}$$

We now have for any function F

$$\begin{aligned} \sum_{(m,n)\neq(0,0)} F[Q_1(m,n)] &= 4 \sum_{n=1}^{\infty} \sum_{\substack{m|n \\ m \text{ odd}}} (-1)^{(m-1)/2} F(n) \\ &= 4 \sum_{n=1}^{\infty} \sum_{\substack{m \text{ odd} \\ mk=n}} \sum_{k} (-1)^{(m-1)/2} F(mk) \\ &= 4 \sum_{m \text{ odd}} \sum_{k=1}^{\infty} (-1)^{(m-1)/2} F(mk). \end{aligned} \tag{20}$$

In the Madelung case where $F(n) = n^{-s}$, the sums factor and we find

$$\sum_{(m,n)\neq(0,0)} (m^2 + n^2)^{-s} = 4\zeta(s)\beta(s) \tag{21}$$

or

$$S_1 = \sum_{m,n=1} (m^2 + n^2)^{-s} = \zeta(s)\beta(s) - \zeta(2s). \tag{22}$$

Similarly, we find

$$\sum_{(m,n)\neq(0,0)} F[Q_2(m,n)] = 2 \sum_{m \text{ odd}} \sum_{k \text{ odd}} (-1)^{(m-1)/2} \left[F(mk) + 2 \sum_{l=1}^{\infty} F(2^l mk)\right]. \tag{23}$$

Hence, again in the Madelung case,

$$\sum_{(m,n)\neq(0,0)} (m^2 + 4n^2)^{-s} = 2(1 - 2^{-s} + 2^{1-2s})\zeta(s)\beta(s) \tag{24}$$

or

$$S_5 = \sum_{m,n=1} (m^2 + 4n^2)^{-s} = \tfrac{1}{2}(1 - 2^{-s} + 2^{1-2s}) \times \zeta(s)\beta(s) - \tfrac{1}{2}(1 + 2^{-2s})\zeta(2s). \tag{25}$$

The remaining sums are obtained from the elementary relations

$$\begin{aligned} S_2 &= (1 + 2^{2-2s}) S_1 - 4S_5, \\ S_3 &= S_1 - 2S_5, \\ S_4 &= (1 + 2^{-2s}) S_1 - 4S_5. \end{aligned} \tag{26}$$

As an example of the case when the class number is greater than unity we consider the series

$$S = \sum (m^2 + 16n^2)^{-s} = \sum_{n=1}^{\infty} r_1(n) n^{-s}. \tag{27}$$

This is related in a simple way to the series on the right-hand side of (I. 26) which we were unable to evaluate by the use of theta functions. The class number for discriminant $d = -64$ is 2 and as representative forms we take

$$Q_1 = m^2 + 16n^2, \quad Q_2 = 4m^2 + 4mn + 5m^2. \tag{28}$$

By examining the character table for the Abelian group of order $\sqrt{-d} = 8$, we find that

$$c_2(n) = \begin{cases} 0 & n \text{ even} \\ 1 & n \equiv \pm 1 \bmod 8 \\ -1 & n \equiv \pm 3 \bmod 8 \end{cases} \tag{29}$$

has the property that when $k = Q_1(m,n)$ is odd, $c_2(k) = 1$ whereas when $k = Q_2(m,n)$ is odd, $c_2(k) = -1$. Hence from (15), we see that when k is odd

$$r_1(k) = [c_2(k) + 1] \sum_{m|k} (-1)^{(m-1)/2}. \tag{30a}$$

Next consider what happens when k is even:

Case I: $k = 2q, q$ odd. $2q = x^2 + 16y^2$ implies that $q = 2(a^2 + 4y^2)$ which is impossible so

$$r_1(2q) = 0. \tag{30b}$$

Case II: $k = 4q, q$ odd. $4q = x^2 + 16y^2$ implies that $q = a^2 + 4y^2$ for which the answer is given in (19). Hence

$$r_1(4q) = 2 \sum_{m|q} (-1)^{(m-1)/2}. \tag{30c}$$

Case III: $k = 8q, q$ odd. This leads to $q = 2(a^2 + y^2)$ so

$$r_1(8q) = 0. \tag{30d}$$

Case IV: $k = 2^l q, q$ odd, $l \geqslant 4$. Factoring out 2's leads to the two squares problem and we find

$$r_1(2^l q) = 4 \sum_{m|q} (-1)^{(m-1)/2}. \tag{30e}$$

From (27) and (30) we have

$$S = \gamma \zeta(s) \beta(s) + L(s, c_2) L(s, c_1 c_2), \tag{31}$$

where

$$\gamma = 2^{2-4s} + (1 + 2^{1-2s})(1 - 2^{-s}),$$

and the two new L-series we require are

$$\begin{aligned} L(s, c_2) &= 1 - \frac{1}{3^s} - \frac{1}{5^s} + \frac{1}{7^s} + \frac{1}{9^s} - \cdots, \\ L(s, c_1 c_2) &= 1 + \frac{1}{3^s} - \frac{1}{5^s} - \frac{1}{7^s} + \frac{1}{9^s} + \cdots. \end{aligned} \tag{32}$$

We note that their product can be written $A^2(s) - B^2(s)$, where

$$\begin{aligned} A(s) &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(4k+1)^s}, \\ B(s) &= \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{(4k-1)^s} \end{aligned} \tag{33}$$

are special cases of the Lerch zeta function. From (31) we find that

$$S_6 = \sum_{m,n=1}^{\infty} (m^2 + 16n^2)^{-s} = \tfrac{1}{4} S - \tfrac{1}{2}(1 + 2^{-4s}) \zeta(2s),$$

$$S_7 = \sum_{\substack{m=1 \\ k \text{ odd}}}^{\infty} (m^2 + 4k^2)^{-s} = S_5 - S_6,$$

$$\begin{aligned} S_8 &= \sum_{\substack{k,l \text{ odd} \\ 1}}^{\infty} (k^2 + 4l^2)^{-s} \\ &= (1 + 2^{-2s}) S_5 - S_6 - 2^{-2s} S_1. \end{aligned}$$

We have therefore managed to evaluate the triple sum in Eq. (I. 26) which can be expressed

$$\begin{aligned} 4 \sum_{l,m,n=1}^{\infty} & (-1)^m [(l - \tfrac{1}{2})^2 + m^2 + n^2]^{-s} \\ &= 2^s (2^s + 2^{1-s} - 1) \zeta(s) \beta(s) \\ &\quad + (2^{2s} - 3) \zeta(2s) - 2^{2s} [A^2(s) - B^2(s) + \beta(2s - 1)], \\ &\qquad (\operatorname{Re} s > 1), \end{aligned} \tag{34}$$

which is the Madelung sum for an orthorhombic lattice in which planes of charges alternate in sign. This appears to be the first time such a result has been obtained and holds out the prospect that the Madelung sum for real three dimensional ionic crystal can be expressed in this way.

ACKNOWLEDGMENT

The author expresses his thanks to Professor George Mackey for indicating this approach to these sums.

---

[1]M. L. Glasser, J. Math. Phys. **14**, 409 (1973). References to formulas in this paper are preceded by the letter I.
[2]P. G. Lejeune-Dirichlet, *Vorlesungen über Zehlentheorie*, edited and supplemented by R. Dedekind (Braunschweig, 1894), 4th ed., Chap. 5.
[3]L. E. Dickson, *Introduction to the Theory of Numbers* (Dover, New York, 1957). Specific references to this book will be made by page number preceded by D.

Propagation of frequency-modulated pulses in a randomly stratified plasma*

Ioannis M. Besieris†

Courant Institute and Department of Mathematics, University Heights, New York University, 251 Mercer Street, New York, New York 10012

Frederick D. Tappert

Bell Laboratories, Whippany, New Jersey 07981
(Received 8 August 1972)

Quantitative predictions are made of the effects of randomness and dispersion on the spreading, attenuation, and modulation coefficient of a frequency-modulated wave packet propagating in a randomly stratified isotropic plasma. In the absence of random fluctuations, the recently reported findings of Millman and Bell are recovered.

1. INTRODUCTION

Scalar wave propagation in a cold, isotropic, randomly stratified plasma is described by the one-dimensional Klein–Gordon equation

$$\epsilon^2 \frac{\partial^2}{\partial t^2} u(z,t;\alpha) + \Omega^2_{\text{op}}\left(z, -i\epsilon \frac{\partial}{\partial z}; \alpha\right) u(z,t;\alpha) = 0, \tag{1.1a}$$

$$\Omega^2_{\text{op}}\left(z, -i\epsilon \frac{\partial}{\partial z}; \alpha\right) = c^2\left(-i\epsilon \frac{\partial}{\partial z}\right)^2 + \omega_p^2(z;\alpha). \tag{1.1b}$$

Here, Ω^2_{op} is a positive, self-adjoint, stochastic operator depending on a parameter $\alpha \in \mathfrak{A}$, $\mathfrak{A}$ being a probability measure space.[1] In addition, $u(z,t;\alpha)$, the real scalar random field amplitude, is an element of an infinitely dimensional vector space $\mathfrak{H}$, and c is the speed of light *in vacuo*. The quantity $\omega_p^2(z;\alpha)$, the square of the plasma frequency, is defined by

$$\omega_p^2(z;\alpha) = (4\pi e^2/m)N(z;\alpha), \tag{1.2}$$

where e and m are respectively the charge and mass of an electron, and $N(z;\alpha)$ is the electron density which is assumed to be a real nonnegative space-dependent random function. The problem (1.1) is rendered closed by specifying Cauchy initial data for u and its time derivative.

A distinguishing feature of this problem is the presence of the positive dimensionless parameter ϵ which can be taken to be inversely proportional to the scale size of the spatial inhomogeneities. As such, for a slowly varying medium, ϵ will be a small but finite quantity.[2]

It is our purpose in this paper to examine the propagation of a frequency-modulated (chirped) pulse in a randomly stratified plasma described by the Klein–Gordon equation (1.1). Towards this goal we shall make use of a recently developed general theory of wave propagation in random media, the *stochastic wave kinetic technique*.[3,4,5]

The random coefficients will be taken to vary slowly, and the correlation length of the random inhomogeneities is assumed to be large compared to a wavelength. (These notions are made mathematically precise in Ref. 4).

Strictly speaking, (1.1) describes a two-mode problem (incident and reflected waves). However, within the framework of the short wavelength approximation considered here, the reflected waves are exponentially small.[6] The theoretical results given below, therefore, pertain to a single, forward propagating wave packet.

We shall be concerned with the effects of randomness and dispersion on the spreading, attenuation, and modulation coefficient of a frequency-modulated pulse. Although dispersive effects are reversible—the receiver is usually equipped with "built-in dispersion" in order to make optimal use of the additional signal bandwidth—their stochastic counterparts lead to an irreversible degradation of the signal. The problem of making quantitative predictions of the latter effects therefore has important practical consequences. Analogous calculations have been completed in connection with the propagation of an unmodulated wave packet traversing a random, dissipative, and dispersive transmission line modelled by a generalized telegraphist's equation.[7]

In order for the discussion to be self-contained, a brief outline of the stochastic wave kinetic method as it pertains to the problem under examination is given in the following two sections. The main results of this paper together with a discussion of their significance as well as an enumeration of the conditions which must be satisfied for the theory to be applicable are given in Sec. 4. In Sec. 5, we recover the recent results of Millman and Bell[8] in connection with the effect of ionospheric dispersion on a frequency-modulated pulse by considering the dual boundary-value problem in the absence of random inhomogeneities.

2. THE ANALYTIC SIGNAL AND THE WIGNER DISTRIBUTION FUNCTION

In the following we shall be concerned with the time evolution of a "measurable" quantity. In this sense, the field $u(z,t;\alpha)$ and its intensity $u^2(z,t;\alpha)$ have little physical meaning. We may, however, consider the total wave energy and the total wave action which are given in terms of the field $u(z,t;\alpha)$ and the operator Ω_{op} by the integrals

$$E = \tfrac{1}{2} \int (u\Omega^2_{\text{op}} u + \epsilon^2 u_t^2)\, dz, \tag{2.1}$$

$$A = \tfrac{1}{2} \int (u\Omega_{\text{op}} u + \epsilon^2 u_t \Omega^{-1}_{\text{op}} u_t)\, dz, \tag{2.2}$$

respectively. In view of the assumption that the medium is time-independent, both of these quantities are conserved. The integrands of (2.1) and (2.2) are respectively the space wave energy and wave action density functions.

In order to circumvent the difficulty of working with the complicated expressions (2.1) and (2.2) directly, we shall introduce the notion of the *complex analytic signal*. This quantity is defined by means of the relation

$$\psi(z,t;\alpha) = 2^{-1/2}(\Omega^{1/2}_{\text{op}} u + i\epsilon \Omega^{-1/2}_{\text{op}} u_t). \tag{2.3}$$

The total wave energy and wave action associated with (1.1) are given in terms of ψ and $\Omega_{\rm op}$ as follows:

$$E = \int \psi^* \Omega_{\rm op} \psi dz, \qquad (2.4)$$

$$A = \int \psi^* \psi dz. \qquad (2.5)$$

The integrands $\psi^*\Omega_{\rm op}\psi$ and $\psi^*\psi$ are respectively the space wave energy and wave action densities.

The Wigner distribution function is defined next in terms of the analytic signal:

$$f(z,k,t;\alpha) = (2\pi\epsilon)^{-1} \int dy\, e^{iky/\epsilon} \psi^*(z+\tfrac{1}{2}y,t;\alpha) \times \psi(z-\tfrac{1}{2}y,t;\alpha). \qquad (2.6)$$

The total wave energy and wave action can be written in terms of the Wigner distribution function as follows:

$$E = \int dz \int dk\, \Omega(z,k;\alpha) f(z,k,t;\alpha), \qquad (2.7)$$

$$A = \int dz \int dk\, f(z,k,t;\alpha). \qquad (2.8)$$

Here, $\Omega(z,k;\alpha)$ is the Weyl transform of the operator $\Omega_{\rm op}$. By virtue of (2.7), $\Omega(z,k;\alpha)f(z,k,t;\alpha)$ can be interpreted as the wave energy density in the phase space (z,k) at time t. Its integral over k-space is the space wave energy density. Similarly, from (2.8), $f(z,k,t;\alpha)$ can be thought of as the wave action density in phase space and its integral over k-space as the space wave action density.

3. THE STOCHASTIC WAVE KINETIC EQUATION

In the wave kinetic approximation (ϵ small but finite), the Wigner distribution function obeys the Liouville equation

$$\frac{\partial}{\partial t} f(z,k,t;\alpha) + \frac{\partial}{\partial k}\Omega(z,k;\alpha)\frac{\partial}{\partial z} f(z,k,t;\alpha) - \frac{\partial}{\partial z}\Omega(z,k;\alpha)\frac{\partial}{\partial k} f(z,k,t;\alpha) = 0, \qquad (3.1)$$

to $0(\epsilon^2)$. Of course, $\Omega(z,k;\alpha)$ must be computed to the same order of accuracy. For the Klein–Gordon equation (1.1), one has

$$\Omega(z,k;\alpha) = [c^2k^2 + \omega_p^2(z;\alpha)]^{1/2} + 0(\epsilon^2). \qquad (3.2)$$

Since $\Omega(z,k;\alpha)$ is a random function of position, (3.1) is referred to as a *stochastic wave kinetic equation*.

The electron density is next separated into mean and fluctuating parts, viz.,

$$N(z;\alpha) = N_0 + \delta N(z;\alpha), \quad \langle \delta N(z;\alpha)\rangle = 0. \qquad (3.3)$$

The angular brackets in (3.3) denote a statistical average over an ensemble. The deterministic background electron density, N_0, is assumed to be homogeneous. The fluctuating part of the electron density is specified to be a random function of position with zero mean.

For spatially homogeneous fluctuations, we define the two-point correlation function by the expression

$$\Gamma(\zeta) = N_0^{-2}\langle \delta N(z;\alpha)\,\delta N(z-\zeta;\alpha)\rangle. \qquad (3.4)$$

The variance, η^2, of the fluctuations is introduced as

$$\eta^2 = \Gamma(0) \qquad (3.5)$$

and the correlation coefficient, $\gamma(\zeta)$, is given by the ratio

$$\gamma(\zeta) = \Gamma(\zeta)/\eta^2. \qquad (3.6)$$

In terms of the correlation coefficient, a correlation length, l, is defined by means of the integral

$$l = \int_0^\infty \gamma(\zeta)d\zeta. \qquad (3.7)$$

On the basis of the above assumptions and within the framework of the long time, Markovian, first-order smoothing and diffusion approximations (cf. Refs. 4, 5), the average Wigner distribution function obeys the equation

$$\left[\frac{\partial}{\partial t} + v_g\left(1 - \frac{3}{8}\eta^2 \frac{\omega_{\rm po}^4}{\Omega_0^4}\right)\frac{\partial}{\partial z}\right]\langle f(z,k,t;\alpha)\rangle = \frac{\eta^2}{4}\frac{\omega_{\rm po}^4}{\Omega_0^4}\, l\, |v_g|\, \frac{\partial^2}{\partial z^2}\langle f(z,k,t;\alpha)\rangle. \qquad (3.8)$$

Here, $\omega_{\rm po}^2 = 4\pi N_0 e^2/m$ is the square of the background plasma frequency in terms of which the background angular frequency can be written as $\Omega_0(k) = [c^2k^2 + \omega_{\rm po}^2]^{1/2}$. The group velocity, $v_g(k)$, is defined as the first derivative of $\Omega_0(k)$ with respect to k. We shall later make use of the index of dispersion, $\beta(k)$, which is defined as the second derivative of $\Omega_0(k)$.

In (3.8), the left-hand side describes convection with modified group velocity, while the term on the right-hand side describes spatial diffusion due to the random fluctuations.

4. SPACE-TIME EVOLUTION OF A FREQUENCY-MODULATED PULSE

We consider a frequency-modulated Gaussian wave packet described initially ($t = 0$) by the analytic signal

$$\psi(z,0) = \psi_0 \exp[-(z-z_0)^2/2\sigma_0^2] \times \exp\{(ik_0/\epsilon)[(z-z_0) + (\epsilon m_0/2)(z-z_0)^2 + \phi_0]\}. \qquad (4.1)$$

The constant ϵ should be regarded as a formal expansion parameter (ensuring that the wave packet is much larger than the carrier wavelength) whose numerical value is unity. The carrier wave number is k_0/ϵ, and σ_0, the standard deviation, can be considered as a measure of the pulse width. The normalization amplitude is denoted by ψ_0, and ϕ_0 represents an arbitrary phase constant. The parameter m_0, which has the dimensions of reciprocal length, measures the frequency modulation of the pulse. We mention that solutions can be obtained for other types of pulses, but the Gaussian wave packet chosen here allows one to obtain simple, easily interpretable results. The temporal derivative of u at $t = 0$ is chosen such that the wave packet propagates toward the positive z-direction.

With the normalization amplitude equal to $(\pi\sigma_0^2)^{-1/4}$, one has, corresponding to the analytic signal (4.1), the normalized (with respect to k- and z-space) initial Wigner distribution function

$$f(z,k,0) = (1/\pi\epsilon)\exp[-(z-z_0)^2/\sigma_0^2] \times \exp\{-(\sigma_0^2/\epsilon^2)[k - k_0 - k_0\epsilon m_0(z-z_0)]^2\}. \qquad (4.2)$$

The associated space wave action density (intensity or envelope function) can be found by multiplying (4.1) by its complex conjugate. Alternatively, it can be obtained by integrating $f(z,k,0)$ over k-space, viz.,

$$\rho(z,0) \equiv |\psi(z,0)|^2 = \int f(z,k,0)dk = (\pi\sigma_0^2)^{-1/2}\exp[-(z-z_0)^2/\sigma_0^2]. \qquad (4.3)$$

Furthermore, the initial mean wave number is obtained as follows:

$$K(z,0) = (1/\rho) \int kf(z,k,0)\,dk = k_0[1 + m_0(z - z_0)]. \quad (4.4)$$

Since k occurs in Eq. (3.8) only as a parameter, one may easily solve this equation exactly using the initial value given by Eq. (4.2). Since the desired information is contained in the moments with respect to k of the solution, we shall make use of the smallness of ϵ ("narrow bandwidth" approximation) to simplify the solution before computing the moments. Thus the k-dependent coefficients of $0(\eta^2)$ are replaced by constants evaluated at the carrier wave number k_0 and the background group velocity is approximated by $v_g(k) = v_{g0} + \beta_0(k - k_0)$, where $v_{g0} = v_g(k_0) = \Omega_0'(k_0)$ and $\beta_0 = \beta(k_0) = \Omega_0''(k_0)$.[9] For the problem under consideration, we obtain $v_{g0} = c^2k_0/\Omega_0$ and $\beta_0 = c^2\omega_{po}^2/\Omega_0^3$. We shall also use the abbreviations $D_0 = v_{g0}\omega_{po}^4\eta^2 l/4\Omega_0^4$ and $\gamma = \sigma_0^2 k_0 m_0$.

It may then be readily shown that the solution for the mean Wigner distribution function may be written in the form

$$\langle f(z,k,t;\alpha)\rangle = (1/\pi\epsilon)(\sigma_0/\sigma_1)\exp[-(z - Z_0)^2/\sigma^2] \times \exp\{-(\sigma^2\sigma_0^2/\sigma_1^2)[k - k_0 - k_0\epsilon m(z - Z_0)]^2/\epsilon^2\}, \quad (4.5)$$

with the definitions

$$\sigma_1^2 = \sigma_0^2 + 4D_0t(1 + \gamma^2), \quad (4.6a)$$

$$\sigma_2 = \sigma_0[\gamma + (1 + \gamma^2)(\beta_0t/\sigma_0^2)], \quad (4.6b)$$

$$\sigma^2 = (\sigma_1^2 + \sigma_2^2)/(1 + \gamma^2), \quad (4.6c)$$

$$k_0m = \sigma_2/\sigma_0\sigma^2, \quad (4.6d)$$

$$Z_0 = z_0 + v_{g0}[1 - \tfrac{3}{8}\eta^2(\omega_{po}^4/\Omega_0^4)]t. \quad (4.6e)$$

It should be noted that $\sigma_1 \to \sigma_0$, $\sigma \to \sigma_0$, and $m \to m_0$ as $t \to 0$, so that the initially prescribed Wigner distribution function may be recovered.

The mean intensity is given by the zeroth moment,

$$\langle\rho(z,t;\alpha)\rangle = \langle|\psi(z,t;\alpha)|^2\rangle = \int dk\langle f(z,k,t;\alpha)\rangle = (\pi\sigma^2)^{-1/2}\exp[-(z - Z_0)^2/\sigma^2]. \quad (4.7)$$

It is seen that within the range of validity of the approximations used in this paper, the original Gaussian pulse shape remains Gaussian for $t > 0$ and the variance increases from σ_0^2 to σ^2 [cf. Eq. (4.6c)] due to both dispersive and stochastic spreading. Furthermore, the expression for Z_0 [cf. Eq. (4.6e)] shows that the center of the wave packet moves uniformly with a velocity which is determined by both dispersive and stochastic effects. This gives rise to a stochastic group delay as well as the usual dispersive group delay. We also note that the peak power of the pulse is proportional to σ^{-1}.

The average wave number evolves in space-time according to the formula

$$\langle K(z,t;\alpha)\rangle = (1/\langle\rho\rangle)\int k\langle f(z,k,t;\alpha)\rangle\,dk = k_0[1 + m(z - Z_0)]. \quad (4.8)$$

Comparing to Eq. (4.4), we may identify m[cf. Eq. (4.6d)] as the mean modulation coefficient.

We must emphasize that σ as given by Eq. (4.6c) is the width of the average pulse, and its operational definition required that one add together many pulses under varied (random) conditions and then measure the width. This procedure yields a different result from that obtained by measuring the width of each pulse and then averaging.

Writing σ^2 out explicitly, one observes that for short times the dominant correction to σ_0^2 is the part of σ^2 linear in time, i.e.,

$$\sigma^2 \approx \sigma_0^2 + 4D_0t + 2\gamma\beta_0t = \sigma_0^2\left[1 + \frac{\omega_{po}^2}{\Omega_0^2}\frac{v_{g0}t}{\Omega_0^2}\left(\frac{\omega_{po}^2}{\Omega_0^2}\frac{\eta^2 l}{\sigma_0} + 2m_0\sigma_0\right)\right]. \quad (4.9)$$

One may then differentiate between two possibilities: if $m_0 > 0$ ("chirp down"), σ^2 is a monotonically increasing function of time and the pulse width will expand; on the other hand, if $m_0 < 0$ ("chirp up") and $\gamma < -(2D_0/\beta_0)$, the pulse width will at first decrease (pulse compression). Since, according to Eq. (4.6c), σ^2 is manifestly positive, it follows that the pulse width will ultimately expand.[10]

Similarly, for short times, the modulation coefficient is given by

$$m \approx m_0(1 - 4D_0t/\sigma_0^2) + (\beta_0t/k_0\sigma_0^4)(1 - \sigma_0^4k_0^2m_0^2). \quad (4.10)$$

In closing this section, we wish to treat several special cases.

Case (i): Randomness, zero modulation ($m_0 = 0$, $\gamma = 0$).[11]

The average intensity is given by (4.7), with

$$\sigma^2 = \sigma_0^2 + 4D_0t + \beta_0^2t^2/\sigma_0^2. \quad (4.11)$$

The term linear in time is due entirely to the random fluctuations of the electron density. The third term, an expression quadratic in time, results solely from the dispersive properties of the medium. The random and dispersive effects enter in the same way into the algebraic attenuation factor σ^{-1} appearing in (4.7).

The modulation coefficient is found to be

$$m = \beta_0t/k_0\sigma_0^2\sigma^2, \quad (4.12)$$

where σ^2 is now given by Eq. (4.11). One notes that the modulation coefficient is proportional to the index of dispersion β_0. Furthermore, since σ^2 is a monotonically increasing quadratic function of time, $m \to 0$ as $t \to \infty$.

Case (ii): No randomness ($\eta = 0$), with modulation ($m_0 \neq 0$).

The intensity function is given by (4.7), with

$$Z_0 = z_0 + v_{g0}t \quad (4.13)$$

$$\sigma^2 = \sigma_0^2[1 + 2\gamma(\beta_0t/\sigma_0^2) + (1 + \gamma^2)(\beta_0t/\sigma_0^2)^2] = \sigma_0^2[(1 + \gamma\beta_0t/\sigma_0^2)^2 + (\beta_0t/\sigma_0^2)^2]. \quad (4.14)$$

Again for short times, the term in (4.14) linear in time, viz., $2\gamma\beta_0t$ is the dominant correction to σ_0^2. Therefore, the pulse width will first decrease before it expands if $m_0 < 0$("chirp up").

The mean wave number is given by (4.8), with Z_0 as in (4.13) and modulation coefficient

$$m = \frac{m_0 + (1 + \gamma^2)(\beta_0t/\sigma_0^4k_0)}{1 + 2\gamma(\beta_0t/\sigma_0^2) + (1 + \gamma^2)(\beta_0t/\sigma_0^2)^2}. \quad (4.15)$$

Case (iii): No randomness, zero modulation ($m_0 = 0$).

Here we obtain the well-known results of dispersive spreading. The intensity assumes the form (4.7), with Z_0 given as in (4.13) and

$$\sigma^2 = \sigma_0^2 + \beta_0^2 t^2/\sigma_0^2. \tag{4.16}$$

The mean wave number and the modulation coefficient are those of Case (i).

5. THE DUAL BOUNDARY-VALUE PROBLEM

In this section we shall consider the dual boundary-value problem and, in the absence of random inhomogeneities, shall recover the recent results found by Millman and Bell[8] in connection with the effects of ionospheric dispersion on a frequency-modulated pulse.

In order to convert results for the initial-value problem (wave kinetic equation) to their counterparts for a boundary-value problem (the usual case), we introduce the time T_0 (analogous to Z_0) defined by

$$T_0 = (z - z_0)/[v_{g0}(1 - \tfrac{3}{8}\eta^2\omega_{po}^4/\Omega_0^4)]. \tag{5.1}$$

Rewriting Eq. (4.7) for the mean intensity, we obtain

$$\langle \rho(z,t;\alpha)\rangle = (\pi\tau^2)^{-1/2}\exp[-(t - T_0)^2/\tau^2], \tag{5.2}$$

where

$$\tau = \sigma/[v_{g0}(1 - \tfrac{3}{8}\eta^2\omega_{po}^4/\Omega_0^4)]. \tag{5.3}$$

Since Eq. (5.2) shows that the pulse is centered at $t = T_0$, we may replace the variable t in the slowly varying quantity σ by T_0, thereby expressing the temporal pulse width τ in terms of the distance variable $z - z_0$.

In particular, in the absence of fluctuations [$\eta = 0$, Case (ii) of Sec. 4], it follows from Eq. (4.14) that the pulse width as a function of $z - z_0$ is given by

$$\tau^2 = \tau_0^2[(1 + \gamma\beta_0 T_0/\sigma_0^2)^2 + (\beta_0 T_0/\sigma_0^2)^2], \tag{5.4}$$

where $\tau_0 = \sigma_0/v_{g0}$. It is customary to use the variable

$$q = \beta_0 T_0/v_{g0}^2, \tag{5.5}$$

which is related to the inverse square of the "slope bandwidth" of the medium,[12] and the constant

$$\mu_0 = k_0 m_0 v_{g0}^2, \tag{5.6}$$

which is the initial rate of frequency sweep. Eq. (5.4) may then be written

$$\tau^2 = \tau_0^2(1 + \mu_0 q)^2 + q^2/\tau_0^2. \tag{5.7}$$

Similarly, the rate of frequency sweep, μ, is found [cf. Eq. (4.15)] to be

$$\mu = \frac{\mu_0(1 + \mu_0 q) + q/\tau_0^4}{(1 + \mu_0 q)^2 + q^2/\tau_0^4}. \tag{5.8}$$

These results, Eqs. (5.7) and (5.8), are precisely those obtained by Millman and Bell[8] by a different method. In the absence of initial modulation ($\mu_0 = 0$), they are in agreement with Bek's analysis.[13]

6. CONCLUSIONS AND SUMMARY

The problem of wave packet propagation in a randomly stratified plasma has been solved in the diffusion approximation by means of the stochastic wave kinetic technique. Although the basic equation to leading order is formally identical to the Liouville equation, it has been shown that when solutions of this equation are restricted to be Wigner distribution functions then the stochastic wave kinetic method is capable of describing both coherent wave packet spreading and demodulation due to dispersion of the medium and incoherent wave packet spreading and demodulation due to stochastic fluctuations of the medium.

Explicit formulas describing these effects have been obtained and their significance has been discussed.

ACKNOWLEDGMENT

The authors wish to express their gratitude to Professor J. B. Keller for his encouragement and interest in this research.

* Research supported in part by the Air Force Office of Scientific Research under Grant No. AFOSR–71–2107.

† Present address: Department of Electrical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

1 All the fractional powers of Ω_{op}^2 are defined and are themselves positive, self-adjoint operators.

2 In connection with the Klein–Gordon equation describing a plasma, ϵ could alternatively be taken to be proportional to the plasma frequency. Since the solution is considered in the limit of small but finite ϵ, it holds for a large plasma frequency. Such an approach has been followed in the past in the asymptotic theory of propagation in asymptotically conservative dispersive media, e.g., R. M. Lewis, Arch. Ration. Mech. Anal. **20**, 191 (1965); B. Granoff and R. M. Lewis, Philos. Trans. R. Soc. Lond. A **262**, 387 (1967).

3 F. D. Tappert, SIAM Rev. (Chronicle) **13**, 281 (1971).

4 F. D. Tappert and I. M. Besieris, "Stochastic Wave Kinetic Equations: General Theory," to appear.

5 F. D. Tappert and I. M. Besieris, "Stochastic Wave Kinetic Equation and its Application to Wave Packet Spreading," Proc. Intern. Symp. on Electromagnetic Wave Theory, Tbilisi, USSR, Sept. 1971.

6 N. G. van Kampen, Physica (Utr.) **35**, 70 (1967); J. J. Mahony, Q. Appl. Math. **25**, 313 (1967).

7 I. M. Besieris and F. D. Tappert, "Wave packet spreading on a random transmission line," J. Appl. Phys. (May, 1972).

8 G. H. Millman and C. D. Bell, IEEE Trans. Antennas Propag. **19**, 152 (1971).

9 M. A. Leontovich and V. A. Fock, Zh. Eksp. Teor. Fiz. **16**, 557 (1946) [J. Phys. USSR **10**, 13 (1946)].

10 The problem of "designing" the modulation so that a pulse traversing a randomly stratified plasma is enhanced at the site of the receiver is presently under investigation. Unlike the chirp radar pulse discussed in this paper, the frequency modulation will be in general nonlinear and will depend on the dispersive and stochastic properties of the medium as well as on the distance between the transmitter and the receiver.

11 This case has been previously considered by the authors (cf., Ref. 5).

12 J. R. Wait, Proc. IEEE **58**, 1293 (1970).

13 S. S. Bek, "A method of analyzing the dispersive distortions of a short pulse in the ionosphere," Proc. 22nd All-Union Sci. Session Devoted to Radio Day, Moscow, USSR (1966).

Scattering into cones. II. n-body problems*

John D. Dollard

Department of Mathematics, University of Texas, Austin, Texas
(Received 20 September 1972)

An n-body scattering process is studied in the framework of nonrelativistic time-dependent scattering theory. The probability is calculated that the n bodies emerge in cones $C_1,\dots,C_n$ with apices at the origin of coordinates. Results are given for short-range and for Coulomb potentials. The results corroborate the usual interpretation of time-dependent scattering theory.

INTRODUCTION

This paper is a sequel to the paper I (Ref. 1) and has a similar purpose: to offer a geometrical account of n-body scattering processes which is rigorously derivable from nonrelativistic quantum-mechanical time-dependent scattering theory, thus corroborating the "usual" interpretation of scattering theory and at the same time calculating rigorously the probabilities of various experimentally interesting events. Specifically, in this paper we compute the probability (or, when necessary, the time-average of the probability) that starting from a given initial state at large negative times, n nonrelativistic quantum-mechanical particles will emerge from a scattering process in n given cones with vertices at the origin of coordinates (i.e., the probability that particle 1 is in cone 1, particle 2 is in cone 2, regardless of whether the particles are free or bound together to constitute a number of composite particles as $t \to +\infty$). Some would argue that there are more interesting things than this to compute, such as the probability that various composite particles emerge in various cones. As will be seen, all these probabilities will be computed in the process of finding the probability we seek. Although we deal with both Coulomb and short-range forces, we do not give in this paper an analog of Theorem 2 of Ref. 1, which allows calculation of the probability in question for potential scattering without the introduction of anomalous factors.

n-BODY SCATTERING INTO CONES

Orientation: We use units in which $\hbar = 1$. n nonrelativistic quantum-mechanical particles are described by assigning to each real number t a wave-function, i.e., a normalized element ψ_t of $\mathcal{L}^2(\mathbb{R}^{3n})$, the Hilbert space of complex-valued square-integrable functions on $3n$-dimensional Euclidean space. We write ψ_t as a function of n three-vectors $\mathbf{x}_1,\dots,\mathbf{x}_n$. $|\psi_t(\mathbf{x}_1,\dots,\mathbf{x}_n)|^2$ is the position probability density (ppd) for the particles at time t, i.e., the probability density that particle 1 is at $\mathbf{x}_1,\cdots$ particle n is at $\mathbf{x}_n$. Similarly, if $\tilde{\psi}_t$ denotes the Fourier transform of ψ_t, then $|\tilde{\psi}_t(\mathbf{k}_1,\dots,\mathbf{k}_n)|^2$ is the momentum probability density (mpd) for the particles at time t. ψ_t satisfies the Schrödinger equation

$$\psi_t = e^{-iHt}\psi_0, \tag{1}$$

where H is a self-adjoint linear transformation of the form

$$H = H_0 + V. \tag{2}$$

In (2), H_0 is given by

$$H_0 = \sum_{j=1}^{n} \frac{-\Delta_j}{2m_j}, \tag{3}$$

where m_j is the mass of the jth particle and Δ_j is the (natural self-adjoint extension of the) Laplacean in the jth coordinate. V is a multiplicative operator given by

$$V = \sum_{j=1}^{n} V_{0j}(\mathbf{x}_j) + \sum_{1\le i<j\le n} V_{ij}(\mathbf{x}_i - \mathbf{x}_j), \tag{4}$$

where each V_{ij}, $0 \le i < j \le n$, is a real-valued function on R^3, which we assume can be written as

$$V_{ij}(\mathbf{x}) = V_{ij}^{(1)}(\mathbf{x}) + V_{ij}^{(2)}(\mathbf{x}), \tag{5}$$

where $V_{ij}^{(1)}(\mathbf{x})$ is square-integrable over $\mathbb{R}^3$ and $V_{ij}^{(2)}(\mathbf{x})$ is bounded on $\mathbb{R}^3$. Then according to a theorem of Kato[2] H is self-adjoint, with the *same* domain as H_0. The assumptions we have made on the potentials are mild enough to allow most cases of physical interest, (e.g., Coulomb potentials, Yukawa potentials, etc.) We should remark at this point that the problem we will be investigating, whether n particles will be in n cones at large positive times after a scattering experiment, is most natural if we actually expect all particles to emerge in some direction. If the static potentials V_{0j} are nonzero, it is conceivable that some particles will finally be trapped near the origin, as we shall see, and do not emerge in any cone. The problem of whether the n particles are eventually in the n cones is thus most natural when all the V_{0j}, $j = 1,\dots,n$, are zero, in which case we do expect all particles to emerge in some direction. For the sake of interest, however, we carry the V_{0j} along in the analysis.

Just as in the earlier paper[1] (hereafter referred to as I), we can analyse the motion of free particles ($H = H_0$) by writing

$$e^{-iH_0t} = C_t^0\, Q_t^0, \tag{6}$$

where C_t^0 is the "classical transformation for n particles"

$$(C_t^0 f)(\mathbf{x}_1,\dots,\mathbf{x}_n) = \frac{(m_1\cdots m_n)^{3/2}}{(it)^{3n/2}} \exp\left(i\sum_{j=1}^{n} m_j x_j^2/2t\right) \times \tilde{f}\left(\frac{m_1\mathbf{x}_1}{t},\dots,\frac{m_n\mathbf{x}_n}{t}\right) \tag{7}$$

and Q_t is just multiplication by a phase:

$$(Q_t f)(\mathbf{x}_1,\dots,\mathbf{x}_n) = \exp\left(i\sum_{j=1}^{n} m_j x_j^2/2t\right) f(\mathbf{x}_1,\dots,\mathbf{x}_n). \tag{8}$$

As in I, the factorization (6) implies that

$$\lim_{t\to\pm\infty} \|e^{-iH_0t}f - C_t^0 f\| = 0 \tag{9}$$

and that asymptotically, in integrating over any measurable set S of $\mathbb{R}^{3n}$, the ppd determined from $e^{-iH_0t}f$ "can be replaced by" the absolute square of $C_t f$. (For the precise meaning of "can be replaced by", see Lemma 3 of I and the remarks after it.) Thus, n free quantum-mechanical particles with wave-function $e^{-iH_0t}f$ asymptotically behave like n free classical particles which started from the origin of coordinates at time $t = 0$ with mpd given by $|\tilde{f}(\mathbf{k}_1, \ldots, \mathbf{k}_n)|^2$. (Again, see the discussion in I.) Let us now consider n cones $C_1, \ldots, C_n$, each in three-dimensional space, defined by the equations

$$C_j = \{\mathbf{x} | \mathbf{x} \cdot \mathbf{n}_j \geq \alpha_j |\mathbf{x}|\}, \quad j = 1, \ldots, n, \tag{10}$$

where $\mathbf{n}_j$ is a unit vector in $\mathbb{R}^3$ and $0 < \alpha_j \leq 1$. We ask for the probability $P^\pm$ that if n free particles are described by the wave-function $e^{-iH_0t}f$, then as $t \to \pm\infty$ we shall find particle 1 in cone $C_1, \cdots$ particle n in cone C_n. Using the remarks we have made above, we find

$$\begin{aligned} P^\pm &= \lim_{t\to\pm\infty} \int_{C_1\times\cdots\times C_n} |(e^{-iH_0t}f)(\mathbf{x}_1, \ldots, \mathbf{x}_n)|^2 d\mathbf{x}_1, \ldots, d\mathbf{x}_n \\ &= \lim_{t\to\pm\infty} \int_{C_1\times\cdots\times C_n} |(C_t^0 f)(\mathbf{x}_1, \ldots, \mathbf{x}_n)|^2 d\mathbf{x}_1, \ldots, d\mathbf{x}_n \\ &= \int_{(\pm C_1)\times\cdots\times(\pm C_n)} |\tilde{f}(\mathbf{k}_1, \ldots, \mathbf{k}_n)|^2 d\mathbf{k}_1, \ldots, d\mathbf{k}_n, \end{aligned} \tag{11}$$

where $C_1 \times \cdots \times C_n$ is the topological product of the n cones, i.e., the set of all points $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ such that $\mathbf{x}_j \in C_j$, $j = 1, \ldots, n$; $-C_j$ is the reflection of C_j through the origin; and the final result in (11) is obtained by the change of variables $\mathbf{k}_j = m_j\mathbf{x}_j/t$, $j = 1, \ldots, n$ (see I). Thus we have arrived at the expected result: for large positive times, one of the particles will lie in a cone if and only if its momentum lies in that cone. A similar statement holds for large negative times.

So far, all has gone in complete analogy with I. In I, however, once we have analyzed the behavior of the wave-function describing a free particle, we had all necessary information on the asymptotic behavior of a scattered particle, since such a particle could exhibit only this free behavior at large positive and negative times. In the n-body theory, however, it is well known that, in general, there are many other ways that n particles can enter or leave a scattering experiment than as n free particles. In the customary terminology, there are many other channels open to the particles. We now discuss this point briefly.

A scattering experiment is described by a wave-function of the form $e^{-iHt}\psi_0$ where H is not the free Hamiltonian but instead has the form (2). It may happen that the part of H which describes the mutual interactions of a certain subset Γ of the n particles allows bound states to be formed, so that if the particles in Γ were isolated from the others it would be possible for them to travel together as a "composite particle", and this is an additional complication in n-body scattering experiments. In such experiments we may consider initial and final states in which the n (simple) particles are grouped into a number of simple particles and a number of composite particles, each composite or simple particle moving freely, and a set of particles bound near the origin by the static potentials V_{0j}. By a "fragment" we shall mean a composite or a simple particle. An initial or final state of an n-body scattering experiment is specified by telling which fragments appear and describing the state of their motion, and specifying the condition of the particles bound near the origin,[4] as follows: partition the n particles into $m + 1$ subsets $\Gamma_1, \ldots, \Gamma_m, \Gamma_{m+1}$. For $l = 1, \ldots, m$, the subset Γ_l is to contain $r_l + 1$ particles, with $r_l \geq 0$, and the particle or particles in Γ_l will constitute a fragment. The subset Γ_{m+1} is to contain the particles bound near the origin by the static potentials, and Γ_{m+1} may be empty, and will always be empty if all the static potentials V_{0j} are zero.

We do not admit the case in which Γ_{m+1} is the entire collection of n particles, because this would describe a true bound state of the n-body system, i.e., an eigenfunction of the full Hamiltonian H, which would not figure in any scattering experiment. We shall deal with these true bound states separately. We note parenthetically that we shall sometimes speak as if the subsets Γ_l contained the indices of particles instead of the particles themselves. Let $\mathbf{Y}_l$ and $\mathbf{z}_l$, $l = 1, \ldots, m$, be the center-of-mass and "internal" coordinates of the lth fragment, i.e., $\mathbf{Y}_l$ is defined by

$$\mathbf{Y}_l = \sum_{j\in\Gamma_l} \frac{m_j\mathbf{x}_j}{M_l}, \tag{12}$$

with

$$M_l = \sum_{j\in\Gamma_l} m_j, \tag{13}$$

and $\mathbf{z}_l$ stands in general for a number of three-vector variables $\mathbf{z}_l^j$, $j = 1, \ldots, r_l$, related to the $\mathbf{x}_j$ with $j \in \Gamma_l$ by equations of the form

$$\mathbf{x}_j = \mathbf{Y}_l + \sum_{k=1}^{r_l} \lambda_{jk}^l \mathbf{z}_l^k, \quad j \in \Gamma_l. \tag{14}$$

In the case that Γ_l contains only one particle there are, of course, no "internal coordinates". If Γ_l contains more than one particle then we imagine the internal coordinates to have been chosen in such a way that the Jacobian of the transformation from the set $\{\mathbf{x}_j | j \in \Gamma_l\}$ to the set $\mathbf{Y}_l, \mathbf{z}_l$ is unity. For the subset Γ_{m+1}, we denote all of its coordinates collectively by $\mathbf{z}_{m+1}$. We can now describe an initial or final scattering state of n particles by a wave-function of the form

$$\psi_t = e^{-iH_\alpha t}f(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)\phi_{\alpha_1}(\mathbf{z}_1)\cdots\phi_{\alpha_m}(\mathbf{z}_m)\phi_{\alpha_{m+1}}(\mathbf{z}_{m+1}). \tag{15}$$

In (15), f is any square-integrable function of its variables. For $l = 1, \ldots, m$, ϕ_{α_l} is the bound state of energy E_{α_l}, describing the internal structure of the lth fragment. If the lth fragment is a simple particle, then $\phi_{\alpha_l} = 1$ and $E_{\alpha_l} = 0$ by convention. For any composite particle we assume that the norm $\|\phi_{\alpha_l}\| = 1$. (The norm $\|\phi_{\alpha_l}\|$ is the norm in the space of square-integrable functions of r_l three-vector variables.) $\phi_{\alpha_{m+1}}$ is the bound state of energy $E_{\alpha_{m+1}}$, describing the way in which the particles in Γ_{m+1} are bound near the origin by the static potentials. If Γ_{m+1} is not empty, we assume that $\|\phi_{\alpha_{m+1}}\| = 1$. If Γ_{m+1} is empty, we take $\phi_{\alpha_{m+1}} = 1$ and $E_{\alpha_{m+1}} = 0$ by convention. H_α of (15) is defined by

$$H_\alpha = \sum_{l=1}^{m} \frac{-\Delta_{\mathbf{Y}_l}}{2M_l} + \sum_{l=1}^{m+1} E_{\alpha_l}. \tag{16}$$

Clearly, $e^{-iH_\alpha t}$ propagates the lth center-of-mass coordinate, $l = 1, \ldots m$, according to the free Schrödinger equation with mass M_l, and assigns to the bound state ϕ_{α_l}, $l = 1, \ldots, m + 1$, the usual time-dependence $e^{-iE_{\alpha_l}t}$, so that (15) describes a collection of m freely moving fragments and some particles which are bound near the origin by static potentials. [We note in passing that all the terms in the sum on the right-hand side of (16) permute with one another. This fact can be used to write $e^{-iH_\alpha t}$ as a product of simpler operators in a straightforward way, and we shall do this later.] It should be clear that the definition of H_α depends only on (1) the partition of the n particles into subsets $\Gamma_1, \ldots, \Gamma_{m+1}$ (this determines the $\mathbf{Y}_l$ and M_l) and (2) the selection of the bound states $\phi_{\alpha_1}, \ldots, \phi_{\alpha_{m+1}}$ (this determines the E_{α_l}). The subscript α on H_α is supposed to be a multi-index describing this partition into subsets and selection of bound states; each such partition and selection of bound states determines a *channel* of the n-particle system, that is, a way in which the n particles can enter or leave a scattering experiment. As a technical point, we remark that if the particles of subset Γ_l have several different bound states $\phi_{\alpha_l}, \phi'_{\alpha_l}, \cdots$ with the *same* energy E_{α_l}, then we select once and for all an *orthogonal* set of such bound states and use only these when defining channels. (This is necessary to guarantee the orthogonality of the $\{R^+_\alpha\}$ and of the $\{R^-_\alpha\}$ discussed later, as well as that of certain of the D_α also discussed later.) As another point, we note that in the above account two channels with the same partition into subsets $\Gamma_1, \ldots, \Gamma_{m+1}$, different bound states, but the same value of $\sum_{l=1}^{m+1} E_{\alpha l}$, are considered distinct, while some authors (cf. Jauch[3]) would lump these into a single channel. This is largely a matter of convenience.

It is not difficult to see that there is at most a countable number of channels. An individual channel is to be denoted by the index α. We shall, however, sometimes prefer to have a slightly more detailed notation for a channel, namely we shall sometimes write

$$\alpha = (p, \eta_p), \tag{17}$$

where p is an index describing the partition of n into subsets $\Gamma_1, \ldots, \Gamma_{m+1}$ and η_p is an index describing the selection of the bound states once the partition has been made. Since there are only finitely many partitions, the set of p's is finite. For fixed p, the possible selections of bound states, and hence the number of η_p's, is at most countable. Whenever, in the sequel, a sum is taken over α or p or η_p, it is to be understood that the sum is to be taken over all possible values of the index. As an example of the use of the notation (17), the reader should convince himself that if two channels have the same value of p, then the corresponding channel Hamiltonians differ only by a constant.

With these definitions in hand, we can state the following central facts from n-body scattering theory (see Cook, Hack, Jaunch and Zinnes[5]):

Proposition I: Let D_α be the closed subspace of $\mathcal{L}^2(R^{3n})$ consisting of all functions of the form $f(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)\phi_{\alpha_1}(\mathbf{z}_1)\cdots\phi_{\alpha_{m+1}}(\mathbf{z}_{m+1})$, where f runs through all square-integrable functions of $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$, and $\mathbf{Y}_l, \mathbf{z}_l$, are defined as in Eq. (15), and specified by the multi-index α. Let P_α be the projection operator for the subspace D_α. Then if the potentials V_{ij}, $0 \le i < j \le n$, satisfy certain mild restrictions [it suffices that each of them can be written as the sum of a function in $\mathcal{L}^2(\mathbb{R}^3)$ and a function in $\mathcal{L}^p(\mathbb{R}^3)$, with $2 < p < 3$], the strong limits

$$\lim_{t\to\pm\infty} e^{iHt} e^{-iH_\alpha t} P_\alpha = \Omega^\pm_\alpha \tag{18}$$

exist. (The operators $\Omega^\pm_\alpha$ are the Møller wave-matrices for our theory.) We denote by $R^\pm_\alpha$ the range of $\Omega^\pm_\alpha$, and write $F^\pm_\alpha$ for the projection on $R^\pm_\alpha$. Then $\Omega^\pm_\alpha$ is a partial isometry with initial set D_α and final set $R^\pm_\alpha$, i.e., we have

$$\Omega^{\pm *}_\alpha \Omega^\pm_\alpha = P_\alpha, \tag{19}$$

$$\Omega^\pm_\alpha \Omega^{\pm *}_\alpha = F^\pm_\alpha. \tag{20}$$

Further, as α varies the subspaces R^+_α are pairwise orthogonal, as are the R^-_α. We thus have

$$F^\pm_\alpha F^\pm_\beta = \delta_{\alpha\beta} F^\pm_\alpha. \tag{21}$$

Further, we have the intertwining relations

$$e^{-iHt}\Omega^\pm_\alpha = \Omega^\pm_\alpha e^{-iH_\alpha t}. \tag{22}$$

The proofs of the above statements are in the literature[5] and we do not give them.

We denote by $R^\pm$ the orthogonal sum of the subspaces $R^\pm_\alpha$ where α runs through all possible channels. We assume henceforth that our theory is *asymptotically complete*, i.e., *we assume* (1) that the strong limits $\Omega^\pm_\alpha$ exist, and (2) that $R^+ = R^-$ and that $R^\pm$ is the orthogonal complement of the subspace spanned by the previously mentioned true bound states of the theory, i.e., the eigenfunctions of the full Hamiltonian H, which we discarded previously in our discussion of initial and final states for scattering experiments. For important results on asymptotic completeness, see the papers of Ikebe, of Fadeev, and of Hepp, and the book of Kato.[6]

A typical element f of R^+ has the form

$$f = \sum_\alpha \Omega^+_\alpha f_\alpha \tag{23}$$

with $f_\alpha \in D_\alpha$ and

$$\sum_\alpha \|\Omega^+_\alpha f_\alpha\|^2 < \infty. \tag{24}$$

Since Ω^+_α is isometric on D_α, (24) implies that

$$\sum_\alpha \|f_\alpha\|^2 = \sum_\alpha \|\Omega^+_\alpha f_\alpha\|^2 < \infty. \tag{25}$$

The projection of f on R^+_α is $\Omega^+_\alpha f_\alpha$; whence, using (20), we obtain

$$\Omega^+_\alpha f_\alpha = F^+_\alpha f = \Omega^+_\alpha \Omega^{+*}_\alpha f. \tag{26}$$

Applying Ω^{+*}_α on the left and using (19) yields

$$P_\alpha f_\alpha = P_\alpha \Omega^{+*}_\alpha f. \tag{27}$$

Since both f_α and $\Omega^{+*}_\alpha f$ belong to D_α, (27) can be rewritten as

$$f_\alpha = \Omega^{+*}_\alpha f. \tag{28}$$

Now the f_α need not be pairwise orthogonal because the D_α are not necessarily orthogonal. However, if we write α as (p, η_p), as in (17), then for each fixed p it is true that $D_{(p,\eta_p)}$ is orthogonal to $D_{(p,\eta'_p)}$ for $\eta_p \ne \eta'_p$, since the different bound states occurring in the definitions of (p, η_p) and (p, η'_p) are orthogonal to each other. (This is partly a result of our convention on the definition of channels in case the particles in a given Γ_l have several different bound states with the same energy. See the earlier discussion of this point just after the definition of "channel".) Thus $f_{(p,\eta_p)}$ is orthogonal to $f_{(p,\eta'_p)}$.

Because of this fact and because $\sum_\alpha \|f_\alpha\|^2$ converges, we can conclude that for each fixed p the sum

$$f^{(p)} = \sum_{\eta_p} f_{(p,\eta_p)} \tag{29}$$

converges. Since there are only finitely many p's, the sum

$$\sum_p f^{(p)} = \sum_p \sum_{\eta_p} f_{(p,\eta_p)} = \sum_\alpha f_\alpha \tag{30}$$

converges. Although we have shown that the convergence takes place only when the sum is done in a special order, the reader can easily convince himself that this special order is irrelevant. Now because $e^{-iH_\alpha t}$ is a unitary operator mapping D_α into itself, it follows in a similar way that the sum

$$U_+(t)f = \sum_\alpha e^{-iH_\alpha t} f_\alpha = \sum_\alpha e^{-iH_\alpha t}\Omega_\alpha^{+*} f \tag{31}$$

converges for any t. We are now in a position to state

Lemma 1: Let $f \in R^+$ and let $U_+(t)f$ be defined by (31). Then

$$\lim_{t\to+\infty} \|e^{-iHt}f - U_+(t)f\| = 0. \tag{32}$$

The proof of this lemma is straightforward. We first assume that only a finite number of f_α's are nonzero. Then, because we are dealing with a theory in which the limits (18) exist, we have

$$\operatorname*{s-lim}_{t\to+\infty} e^{iHt}U_+(t)f = \sum_\alpha \operatorname*{s-lim}_{t\to+\infty} e^{iHt}e^{-iH_\alpha t} f_\alpha = \sum_\alpha \Omega_\alpha^+ f_\alpha = f \tag{33}$$

and (33) implies (32) in an obvious way. In the general case, in which infinitely many f_α's may be nonzero, the argument is completed using a straightforward approximation procedure, resting on the fact that we have the desired result for finitely many nonzero f_α, on (25), and on the remarks about orthogonality of certain D_α's used to show that the sum in (30) converges.

Let us suppose than an n-body scattering process is initiated with the particles in a state described by the wave-function

$$\phi_t^\beta = e^{-iH_\beta t} f_\beta, \quad \|f_\beta\| = 1 \tag{34}$$

with $f_\beta \in D_\beta$. What is meant by this statement, of course, is that the entire scattering experiment is described as usual by a wave-function ψ_t of the form

$$\psi_t = e^{-iHt}\psi_0, \quad \|\psi_0\| = 1, \tag{35}$$

but that this wave-function is to be specified by the requirement

$$\lim_{t\to-\infty} \|e^{-iHt}\psi_0 - e^{-iH_\beta t} f_\beta\| = 0. \tag{36}$$

Because the limits in (18) are assumed to exist, it is easily seen that (36) uniquely specifies the state ψ_0. In fact,

$$\psi_0 = \Omega_\beta^- f_\beta, \tag{37}$$

so that saying an n-body scattering experiment is initiated with the particles in a state described by the wave-function $e^{-iH_\beta t}f_\beta$ is the same as saying that the wave-function ψ_t at all times is given by

$$\psi_t = e^{-iHt}\psi_0 = e^{-iHt}\Omega_\beta^- f_\beta. \tag{38}$$

Clearly ψ_0 of (37) and (38) belongs to R^- and hence to R^+, since we assume $R^- = R^+$. Hence by our previous discussion $e^{-iHt}\psi_0$ will converge as $t \to +\infty$ to $U_+(t)\psi_0$. Since the expression for $U_+(t)\psi_0$ is interesting, we write it out:

$$\begin{aligned} U_+(t)\psi_0 &= \sum_\alpha e^{-iH_\alpha t}\Omega_\alpha^{+*}\psi_0 = \sum_\alpha e^{-iH_\alpha t}\Omega_\alpha^{+*}\Omega_\beta^- f_\beta \\ &= \sum_\alpha e^{-iH_\alpha t} S_{\alpha\beta} f_\beta, \end{aligned} \tag{39}$$

where

$$S_{\alpha\beta} = \Omega_\alpha^{+*}\Omega_\beta^-. \tag{40}$$

$S_{\alpha\beta}$ is sometimes called a "partial S-matrix" from channel β to channel α. An intuitive statement of our result is that if we "send in" $e^{-iH_\beta t}f_\beta$, we will "get out" $\sum_\alpha e^{-iH_\alpha t}S_{\alpha\beta}f_\beta$. We could also imagine sending in something of the form $\sum_\beta e^{-iH_\beta t}f_\beta$, with $\sum_\beta \|f_\beta\|^2 = 1$ (this is the same as saying that ψ_0 is given by $\sum_\beta \Omega_\beta^- f_\beta$, and $\|\psi_0\| = 1$). The reader is left to imagine the results for himself, as there is little to be gained from writing them down.

We now ask the following question: suppose that we are given n cones $C_1, \ldots, C_n$ in three-dimensional space, defined by the Eqs. (10). What is the probability $P(f_\beta, C_1, \ldots, C_n)$ that if we "send in" $e^{-iH_\beta t}f_\beta$, then at large positive times particle 1 will be in cone $C_1, \ldots,$ particle n in cone C_n? We can write down an expression for this probability using the fact that "sending in $e^{-iH_\beta t}f_\beta$" means that the wave-function at all times is given by (38). Then since $|\psi_t|^2$ is the ppd for the particles at time t, we have (provided, of course, that the indicated limit exists)

$$P(f_\beta; C_1, \ldots, C_n) = \lim_{t\to\infty} P(f_\beta; C_1, \ldots, C_n, t) \tag{41a}$$

with

$$P(f_\beta; C_1, \ldots, C_n, t) = \int_{C_1\times\cdots\times C_n} |(e^{-iHt}\Omega_\beta^- f_\beta)(\mathbf{x}_1, \ldots, \mathbf{x}_n)|^2 d\mathbf{x}_1, \ldots, d\mathbf{x}_n. \tag{41b}$$

In (41b), as before, $C_1 \times \cdots \times C_n$ denotes the topological product of the cones. Because $e^{-iHt}\Omega_\beta^- f_\beta$ converges strongly to $U_+(t)\Omega_\beta^- f_\beta$ as $t \to +\infty$, we may, in investigating the existence of the limit in (41a), replace the integrand on the right-hand side of (41b) by the absolute square of $(U_+(t)\Omega_\beta^- f_\beta)(\mathbf{x}_1, \ldots, \mathbf{x}_n)$. (See Lemma 3 of I.) We shall do this presently. In order to cope with the mess which will result, we first prepare some lemmas, which require the introduction of additional notation.

Let α be a channel of our system, with its associated partition $\Gamma_1^\alpha, \ldots, \Gamma_{m+1}^\alpha$ of the particles and its bound states $\phi_{\alpha_1}, \ldots, \phi_{\alpha_{m+1}}$. [Note: the number $m+1$ of subsets Γ_l^α depends on the channel α and we could indicate this by writing m as $m(\alpha)$. In order not to clutter the notation, however, we shall write simply m, except when, in dealing with several channels, it is necessary to write $m(\alpha)$ for clarity. A similar convention will be used with various other channel-dependent pieces of notation for objects of secondary interest, e.g., $\mathbf{p}_l$ of (42). Various important symbols like Γ_l^α, $\chi_{B_{\alpha l}}$ will carry the index α throughout.] We define the center-of-mass coordinates $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ by (12), and the center-of-mass momenta $\mathbf{p}_1, \ldots, \mathbf{p}_m$ by

$$\mathbf{p}_l = \sum_{j \in \Gamma_l^\alpha} \mathbf{k}_j, \quad l = 1, \ldots, m, \tag{42}$$

where $\mathbf{k}_j$ is the momentum of particle j. [Equation (42) can be interpreted either as an equation between vectors in momentum space, i.e., after Fourier transformation, or as an equation between differential operators, replacing $\mathbf{k}_j$ by $-i\nabla_j$, etc. Naturally, the two interpretations are really the same. We shall mainly be talking in terms of the first.] We define

$$T_{0l} = -\Delta_{\mathbf{Y}_l}/2M_l. \tag{43}$$

Then, write

$$T_\alpha = \sum_{l=1}^{m} T_{0l}, \tag{44a}$$

$$E_\alpha = \sum_{l=1}^{m+1} E_{\alpha_l}. \tag{44b}$$

We have, according to (16),

$$H_\alpha = T_\alpha + E_\alpha. \tag{45}$$

We are going to be interested in the topological product $C_1 \times \cdots \times C_n$ of the n cones of (10), and we introduce some notation relating this topological product to the partition $\Gamma_1^\alpha, \ldots, \Gamma_{m+1}^\alpha$. We denote by $\chi_{C_1 \cdots n}$ the characteristic function of $C_1 \times \cdots \times C_n$:

$$\chi_{C_1 \cdots n}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \begin{cases} 1 & (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in C_1 \times \cdots \times C_n \\ 0 & \text{otherwise.} \end{cases} \tag{46}$$

We shall frequently in the sequel consider characteristic functions like $\chi_{C_1 \cdots n}$ as operators on $\mathcal{L}^2(\mathbb{R}^{3n})$ in an obvious sense—namely, the operation is multiplication of an element of $\mathcal{L}^2(\mathbb{R}^{3n})$ by $\chi_{C_1 \cdots n}$. χ_{C_j} will denote the characteristic function of the set defined by the condition $\mathbf{x}_j \in C_j$:

$$\chi_{C_j}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \begin{cases} 1, & \mathbf{x}_j \in C_j \\ 0, & \text{otherwise.} \end{cases} \tag{47}$$

We also introduce the characteristic function $\chi_{B_{\alpha l}}$ of the set $B_{\alpha l}$ in which the variables with indices in Γ_l^α are restricted to their respective cones. We have

$$\chi_{B_{\alpha l}} = \prod_{j \in \Gamma_l^\alpha} \chi_{C_j}. \tag{48}$$

Clearly,

$$\chi_{C_1 \cdots n} = \prod_{j=1}^{n} \chi_{C_j} = \prod_{l=1}^{m+1} \chi_{B_{\alpha l}}. \tag{49}$$

Finally, we introduce the intersection $I_{\alpha l}$ of the cones with indices in Γ_l^α:

$$I_{\alpha l} = \bigcap_{j \in \Gamma_l^\alpha} C_j. \tag{50}$$

$I_{\alpha l}$ is, of course, a subset of $\mathbb{R}^3$. We write $\chi_{I_{\alpha l}}$ for the characteristic function of $I_{\alpha l}$:

$$\chi_{I_{\alpha l}}(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in I_{\alpha l} \\ 0, & \text{otherwise.} \end{cases} \tag{51}$$

We define $M_{I_{\alpha l}}$ to be the operation of multiplication *in momentum space* by the characteristic function of the set in which the center-of-mass momentum $\mathbf{p}_l$ is restricted to $I_{\alpha l}$, i.e., $M_{I_{\alpha l}}$ is defined by the equation

$$(\widetilde{M_{I_{\alpha l}} f})(\mathbf{k}_1, \ldots, \mathbf{k}_n) = \chi_{I_{\alpha l}}(\mathbf{p}_l) \tilde{f}(\mathbf{k}_1, \ldots, \mathbf{k}_n), \tag{52}$$

where $\mathbf{p}_l$ is defined by (42). The reader will note that, except for the definition of E_α in Eq. (44b), none of the notation defined above depends on the bound states ϕ_{α_l} of the channel α, but only on the partition $\Gamma_1^\alpha, \ldots, \Gamma_{m+1}^\alpha$.

We are now in a position to prove

Lemma 2: Let $\chi_{C_1 \cdots n}, H_\alpha, M_{I_{\alpha l}}$, and $\chi_{B_{\alpha l}}$ be as above. Then

$$\operatorname*{s-lim}_{t \to \infty} e^{iH_\alpha t} \chi_{C_1 \cdots n} e^{-iH_\alpha t} = \left(\prod_{l=1}^{m} M_{I_{\alpha l}} \right) \chi_{B_{\alpha(m+1)}}, \tag{53}$$

where "s—lim" means that the expressions on each side of (53) are to be considered in the obvious way as operators on $\mathcal{L}^2(\mathbb{R}^{3n})$ and that convergence takes place in the strong topology.

Proof: From (45) we find

$$\begin{aligned} e^{iH_\alpha t} \chi_{C_1 \cdots n} e^{-iH_\alpha t} &= e^{iT_\alpha t} e^{iE_\alpha t} \chi_{C_1 \cdots n} e^{-iT_\alpha t} e^{-iE_\alpha t} \\ &= e^{iT_\alpha t} \chi_{C_1 \cdots n} e^{-iT_\alpha t}, \end{aligned} \tag{54}$$

the last step following since E_α is just a real constant. Now because all the operators T_{0l} of (44a) permute, we have

$$e^{-iT_\alpha t} = \prod_{l=1}^{m} e^{-iT_{0l} t}. \tag{55}$$

We recall now the second expression for $\chi_{C_1 \cdots n}$ in (49). We note that T_{0l} affects only the coordinates of particles in Γ_l^α. Further, Eq. (48) shows that $\chi_{B_{\alpha l}}$ actually depends only on the coordinates of particles in Γ_l^α. Using these facts we find

$$\begin{aligned} &e^{iH_\alpha t} \chi_{C_1 \cdots n} e^{-iH_\alpha t} \\ &= \left(\prod_{l=1}^{m} e^{iT_{0l} t} \right) \left(\prod_{l=1}^{m+1} \chi_{B_{\alpha l}} \right) \left(\prod_{l=1}^{m} e^{-iT_{0l} t} \right) \\ &= \left[\prod_{l=1}^{m} \left(e^{iT_{0l} t} \chi_{B_{\alpha l}} e^{-iT_{0l} t} \right) \right] \chi_{B_{\alpha(m+1)}}. \end{aligned} \tag{56}$$

The rest of the proof consists in showing that we have

$$\operatorname*{s-lim}_{t \to \infty} e^{iT_{0l} t} \chi_{B_{\alpha l}} e^{-iT_{0l} t} = M_{I_{\alpha l}}, \quad l = 1, \ldots, m. \tag{57}$$

This will prove (53) because each of the operators $e^{iT_{0l} t} \chi_{B_{\alpha l}} e^{-iT_{0l} t}$ is a projection and thus bounded in norm by unity for all t, so that convergence of the separate factors as in (57) will imply convergence of the product as in (53).

The operator $e^{-iT_{0l} t}$ can be factored into the product of two simpler unitary operators in analogy to the way we factored $e^{-iH_0 t}$ in (6). Namely, we have

$$e^{-iT_{0l} t} = C_{l,t} Q_{l,t}, \tag{58}$$

where

$$(Q_{l,t} f)(\mathbf{x}_1, \ldots, \mathbf{x}_n) = e^{iM_l Y_l^2/2t} f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \tag{59}$$

and

$$(C_{l,t} f)(\mathbf{x}_1, \ldots, \mathbf{x}_n) = (M_l/it)^{3/2} e^{iM_l Y_l^2/2t} (F_l f)\big|_{\mathbf{p}_l = M_l \mathbf{Y}_l/t}. \tag{60}$$

The expression $F_l f$ means the Fourier transform of f with respect to the one center-of-mass variable $\mathbf{Y}_l$ [see (12)]: when f is "sufficiently nice" the definition is

$$(F_l f)(\mathbf{p}_l, \mathbf{z}_l; \mathbf{x}_j') = [1/(2\pi)^{3/2}] \int e^{-i\mathbf{p}_l\cdot\mathbf{Y}_l} f(\mathbf{x}_1, \ldots, \mathbf{x}_n)\, d\mathbf{Y}_l \tag{61}$$

and, for general $f \in \mathcal{L}^2$, $F_l f$ is obtained by a familiar limiting process. On the left-hand side of (61) $\mathbf{x}_j'$ stands for all $\mathbf{x}_j$'s not in Γ_l, and we have written $F_l f$ as a function of $\mathbf{x}_j'$, $\mathbf{p}_l$, and the "internal coordinates" $\mathbf{z}_l$ of Γ_l. As indicated in (60), the operation $C_{l,t}$ requires that $\mathbf{p}_l$ of (61) be set equal to $M_l\mathbf{Y}_l/t$. Because of (58) we have

$$e^{iT_{0l}} \chi_{B_{\alpha l}} e^{-iT_{0l}} = Q_{l,t}^* C_{l,t}^* \chi_{B_{\alpha l}} C_{l,t} Q_{l,t}. \tag{62}$$

It should be clear from (59) that

$$\operatorname*{s-lim}_{t\to\infty} Q_{l,t} = 1 \tag{63a}$$

and

$$\operatorname*{s-lim}_{t\to\infty} Q_{l,t}^* = 1. \tag{63b}$$

We now examine the behavior of the operator $C_{l,t}^* \chi_{B_{\alpha l}} C_{l,t}$. If $f \in \mathcal{L}^2(\mathbb{R}^{3n})$, then by working out the definition of the operator $C_{l,t}^*$ in a straightforward manner we can verify that, in the notation of (61), we have

$$(F_l C_{l,t}^* \chi_{B_{\alpha l}} C_{l,t} f)(\mathbf{p}_l, \mathbf{z}_l; \mathbf{x}_j') = \chi_{B_{\alpha l}}(\mathbf{x}_1, \ldots, \mathbf{x}_n)\big|_{\mathbf{Y}_l = t\mathbf{p}_l/M_l} (F_l f)(\mathbf{p}_l, \mathbf{z}_l; \mathbf{x}_j'). \tag{64}$$

Thus $F_l C_{l,t}^* \chi_{B_{\alpha l}} C_{l,t} f$ is obtained from $F_l f$ by multiplication with a simple function. We now study this function. According to the definition of $\chi_{B_{\alpha l}}$, we have

$$\chi_{B_{\alpha l}}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \begin{cases} 1, & \mathbf{x}_j \in C_j \text{ for all } j \in \Gamma_l^\alpha, \\ 0, & \text{otherwise.} \end{cases} \tag{65}$$

Henceforth we suppress the variables not in Γ_l^α, on which $\chi_{B_{\alpha l}}$ depends trivially. Using equations (14) to express the x_j's with $j \in \Gamma_l^\alpha$ in terms of $\mathbf{Y}_l$ and "internal coordinates", and setting $\mathbf{Y}_l = t\mathbf{p}_l/M_l$, we have

$$\chi_{B_{\alpha l}}(\mathbf{x}_1, \ldots, \mathbf{x}_n)\big|_{\mathbf{Y}_l = t\mathbf{p}_l/M_l} = \begin{cases} 1, & \dfrac{t\mathbf{p}_l}{M_l} + \sum_{k=1}^{r_l} \lambda_{jk}^l \mathbf{z}_l^k \in C_j \text{ for all } j \in \Gamma_l^\alpha, \\ 0, & \text{otherwise.} \end{cases} \tag{66}$$

The reader should not find it hard to convince himself that except at the point $\mathbf{p}_l = 0$ the conditions on the variables $\mathbf{z}_l^k$ in (66) become irrelevant as $t \to \infty$, since the sums of $\mathbf{z}_l^k$'s are "swamped" by the term $t\mathbf{p}_l/M_l$ so that asymptotically we need ask only whether or not $t\mathbf{p}_l/M_l \in C_j$. Furthermore, we have $(t\mathbf{p}_l/M_l) \in C_j$ if and only if $\mathbf{p}_l \in C_j$. To put the matter more briefly and precisely, the following is easy to prove:

$$\text{pointwise} \lim_{t\to+\infty} \chi_{B_{\alpha l}}(\mathbf{x}_1, \ldots, \mathbf{x}_n)\big|_{\mathbf{Y}_l = \mathbf{p}_l t/M_l} = \chi_{I_{\alpha l}}(\mathbf{p}_l), \quad (\mathbf{p}_l \neq 0), \tag{67}$$

where $\chi_{I_{\alpha l}}(\mathbf{p}_l)$ is defined by

$$\chi_{I_{\alpha l}}(\mathbf{p}_l) = \begin{cases} 1, & \mathbf{p}_l \in C_j \text{ for all } j \in \Gamma_l^\alpha, \\ 0, & \text{otherwise,} \end{cases} \tag{68}$$

and is just the characteristic function of the intersection of the cones defined in (51). Writing $\chi_{I_{\alpha l}} F_l f$ for the function whose value at the point $\mathbf{p}_l, \mathbf{z}_l, \mathbf{x}_j'$ is $\chi_{I_{\alpha l}}(\mathbf{p}_l)(F_l f)(\mathbf{p}_l, \mathbf{z}_l, \mathbf{x}_j')$, we now have

$$\begin{aligned} \|F_l C_{l,t}^* \chi_{B_{\alpha l}} C_{l,t} f - \chi_{I_{\alpha l}} F_l f\|^2 &= \int |\chi_{B_{\alpha l}}(\mathbf{x}_1 \cdots \mathbf{x}_n)\big|_{\mathbf{Y}_l = \mathbf{p}_l t/M_l} - \chi_{I_{\alpha l}}(\mathbf{p}_l)|^2 \, |(F_l f)(\mathbf{p}_l, \mathbf{z}_l; \mathbf{x}_j')|^2 \, d\mathbf{p}_l \, d\mathbf{z}_l \, d\mathbf{x}_j' \\ &\xrightarrow[t\to+\infty]{} 0. \end{aligned} \tag{69}$$

Convergence to zero in (69) is obtained from Lebesgue's dominated convergence theorem. The integrand in (69) converges pointwise to zero almost everywhere by (67), and it is bounded by the fixed integrable function $4|F_l f|^2$. Hence the result. From (69) we see that $F_l C_{l,t}^* \chi_{B_{\alpha l}} C_{l,t} f$ converges strongly to $\chi_{I_{\alpha l}} F_l f$ as $t \to +\infty$. Applying the continuous operator $F_l^{-1} = F_l^*$ to both these elements, we obtain

$$\operatorname*{s-lim}_{t\to\infty} C_{l,t}^* \chi_{B_{\alpha l}} C_{l,t} f = F_l^* \chi_{I_{\alpha l}} F_l f, \ \forall f \in \mathcal{L}^2. \tag{70}$$

A little thought will convince the reader that

$$F_l^* \chi_{I_{\alpha l}} F_l f = M_{I_{\alpha l}} f, \tag{71}$$

where $M_{I_{\alpha l}}$ is defined by (52). Thus, finally, we have

$$\operatorname*{s-lim}_{t\to\infty} C_{l,t}^* \chi_{B_{\alpha l}} C_{l,t} = M_{I_{\alpha l}}. \tag{72}$$

Combining (72) with (62) and (63) and using the fact that $Q_{l,t}$, $Q_{l,t}^*$, and $C_{l,t}^* \chi_{B_{\alpha l}} C_{l,t}$ are all bounded in norm by unity, we finally have

$$\operatorname*{s-lim}_{t\to\infty} e^{iT_{0l}t} \chi_{B_{\alpha l}} e^{-iT_{0l}t} = M_{I_{\alpha l}} \tag{73}$$

and this completes the proof of Lemma 2.

Lemma 3: Let α and β be two channels whose corresponding partitions are distinct. Let $\chi_{C_1\cdots n}$ be as before. Then

$$\operatorname*{\omega-lim}_{t\to\infty} e^{iH_\beta t} \chi_{C_{1\cdots n}} e^{-iH_\alpha t} = 0. \tag{74}$$

Proof: This lemma holds because of the fact that

$$\operatorname*{\omega-lim}_{t\to\infty} e^{iH_\beta t} e^{-iH_\alpha t} = 0. \tag{75}$$

The proof of (75) is a simple matter—we do not give a formal proof, but only remark that since the partitions for α and β are not the same, the sums of Laplaceans occurring in H_α and H_β do not cancel each other in (75), and the operator in (75) "oscillates itself to death" as $t \to \infty$, essentially because of the Riemann–Lebesgue lemma. Once (75) is established, we prove (74) by writing

$$e^{iH_\beta t} \chi_{C_{1\cdots n}} e^{-iH_\alpha t} = e^{iH_\beta t} e^{-iH_\alpha t} (e^{iH_\alpha t} \chi_{C_{1\cdots n}} e^{-iH_\alpha t}). \tag{76}$$

By Lemma 2, the factor in parentheses on the right-hand side of (76) converges strongly to a constant operator as $t \to +\infty$, while the unitary operator $e^{iH_\beta t} e^{-iH_\alpha t}$ converges weakly to zero. From these facts it follows simply that the entire product on the right-hand side of (76) converges weakly to zero, proving the lemma.

Lemma 4: Let α and β be two channels whose corresponding partitions are identical. Let $\chi_{C_{1\ldots n}}$ be as before. Then if E_α and E_β are defined as in (44b), we have

$$\text{s-}\lim_{t\to\infty}\left[e^{iH_\beta t}\chi_{C_{1\ldots n}}e^{-iH_\alpha t}-e^{i(E_\beta-E_\alpha)t}\left(\prod_{l=1}^{m}M_{I_{\alpha l}}\right)\chi_{B_{\alpha(m+1)}}\right]=0. \quad (77)$$

Proof: This is a direct consequence of Lemma 2 and the fact that, because α and β have identical partitions, H_α and H_β differ only in the energies E_α and E_β.

We can now continue with our evaluation of the probability $P(f_\beta; C_1,\ldots,C_n)$ that if we "send in" the state $e^{-iH_\beta t}f_\beta$ then for $i=1,\ldots,n$ particle i will be in cone C_i at large positive times. Inserting, as promised, $U_+(t)\Omega_\beta^- f_\beta$ for $e^{-iHt}\Omega_\beta^- f_\beta$ in the integral occurring in (41), we have

$$\begin{aligned}\lim_{t\to+\infty}&P(f_\beta;C_1,\ldots,C_n,t)\\ &\doteq\lim_{t\to+\infty}\int_{C_1\times\cdots\times C_n}|(U_+(t)\Omega_\beta^- f_\beta)(\mathbf{x}_1,\ldots,\mathbf{x}_n)|^2 d\mathbf{x}_1\cdots d\mathbf{x}_n\\ &\doteq\lim_{t\to+\infty}\int_{C_1\times\cdots\times C_n}\Big|\sum_\alpha(e^{-iH_\alpha t}S_{\alpha\beta}f_\beta)(\mathbf{x}_1,\ldots,\mathbf{x}_n)\Big|^2\\ &\quad\times d\mathbf{x}_1\ldots d\mathbf{x}_n.\end{aligned} \quad (78)$$

[Note: in (78), it is not yet clear that the indicated limits exist. We shall see that sometimes they do and sometimes they do not. To deal with both possibilities simultaneously, we have used $\doteq$ signs instead of $=$ signs in (78). These are to be interpreted as follows: $\lim_{t\to+\infty}A(t)\doteq\lim_{t\to+\infty}B(t)$ means $\lim_{t\to+\infty}(A(t)-B(t))=0$. Thus the first $\doteq$ in (78) implies that to investigate the existence of the limit of $P(f_\beta;C_1,\ldots,C_n,t)$, it suffices to investigate the existence of the limit of the integral on the right-hand side of the $\doteq$ sign, since if either quantity has a limit they both do, and the limits are equal. This was explained after (41)—for the proof that (78) follows from (41), compare the proof of Lemma 3 of I. Similar use of $\doteq$ signs is made later.]

For simplicity we shall analyze the case in which the sum over α on the right-hand side of (78) is finite. An infinite sum can then be dealt with by a straightforward but somewhat clumsy limiting process (based on the finiteness of $\sum_\alpha\|S_{\alpha\beta}f_\beta\|^2$) which we do not give explicitly. The results for an infinite sum are formally the same as those we will get for a finite sum. If the sum over α in (78) is finite, then we can write without more ado

$$\begin{aligned}\lim_{t\to+\infty}&P(f_\beta;C_1,\ldots,C_n,t)\\ &\doteq\lim_{t\to+\infty}\sum_{\alpha,\alpha'}\int_{C_1\times\cdots\times C_n}(e^{-iH_\alpha t}S_{\alpha\beta}f_\beta)(\mathbf{x}_1,\ldots,\mathbf{x}_n)\\ &\quad\times(e^{-iH_{\alpha'}t}S_{\alpha'\beta}f_\beta)(\mathbf{x}_1,\ldots,\mathbf{x}_n)d\mathbf{x}_1\cdots d\mathbf{x}_n\\ &\doteq\lim_{t\to+\infty}\sum_{\alpha,\alpha'}\left(e^{-iH_\alpha t}S_{\alpha\beta}f_\beta,\chi_{C_{1\ldots n}}e^{-iH_{\alpha'}t}S_{\alpha'\beta}f_\beta\right)\\ &\doteq\lim_{t\to+\infty}\left[\sum_\alpha\left(S_{\alpha\beta}f_\beta,e^{iH_\alpha t}\chi_{C_{1\ldots n}}e^{-iH_\alpha t}S_{\alpha\beta}f_\beta\right)\right.\\ &\quad\left.+\sum_{\alpha\neq\alpha'}\left(S_{\alpha\beta}f_\beta,e^{iH_\alpha t}\chi_{C_{1\ldots n}}e^{-iH_{\alpha'}t}S_{\alpha'\beta}f_\beta\right)\right]\\ &\doteq\lim_{t\to+\infty}\left[\sum_\alpha\left(S_{\alpha\beta}f_\beta,\prod_{l=1}^{m(\alpha)}M_{I_{\alpha l}}\chi_{B_{\alpha(m(\alpha)+1)}}S_{\alpha\beta}f_\beta\right)\right.\\ &\quad+\sum_p\sum_{\eta_p\neq\eta'_p}\exp[i(E_{\eta_p}-E_{\eta'_p})t]\\ &\quad\left.\times\left(S_{\eta_p\beta}f_\beta,\prod_{l=1}^{m(\eta_p)}M_{I_{\eta_p l}}\chi_{B_{\eta_p(m(\eta_p)+1)}}S_{\eta'_p\beta}f_\beta\right)\right],\end{aligned} \quad (79)$$

where the last line deserves some explanation, namely: To obtain it, we have made use of Lemmas 2, 3, and 4. All terms in the sum for which $\alpha=\alpha'$ have been evaluated by Lemma 1. Terms in which the partitions for α and α' are different have been omitted by Lemma 3. Writing $\alpha=(p,\eta_p)$, $\alpha'=(p',\eta'_{p'})$, the remaining terms are ones in which $p=p'$ but $\eta_p\neq\eta'_p$. In the second sum in the last line of (79) the idea is to write (p,η_p) instead of α and (p,η'_p) instead of α'—since the entry "p" carries no information here, however, we have simply replaced α by η_p and α' by η'_p. We have evaluated the terms with $p=p'$, $\eta_p\neq\eta'_p$, by Lemma 4. We have also called attention to the previously mentioned fact that m depends on α by writing it as $m(\alpha)$ [or $m(\eta_p)$]. In the analysis to follow, where we discuss just one term, we write as before m instead of $m(\alpha)$.

The task of analysing the last expression in (79) is not as hard as it looks. The first remark to be made is that the functions $S_{\alpha\beta}f_\beta$ have a very special form, since for any α we have $S_{\alpha\beta}f_\beta\in D_\alpha$. Thus we can write

$$(S_{\alpha\beta}f_\beta)(\mathbf{x}_1,\ldots,\mathbf{x}_n)=g_{\alpha\beta}(\mathbf{Y}_1,\ldots,\mathbf{Y}_m)\phi_{\alpha_1}(\mathbf{z}_1)\cdots\phi_{\alpha_{m+1}}(\mathbf{z}_{m+1}). \quad (80)$$

The operators $M_{I_{\alpha l}}$, $l=1,\ldots,m$, affect only the function $g_{\alpha\beta}$ of the center-of-mass coordinates $\mathbf{Y}_1,\ldots,\mathbf{Y}_m$ and not the functions $\phi_{\alpha_1}(\mathbf{z}_1),\ldots,\phi_{\alpha_{m+1}}(\mathbf{z}_{m+1})$. The operator $\chi_{B_{\alpha(m+1)}}$ affects only $\phi_{\alpha_{m+1}}$. In the inner product $(S_{\alpha\beta}f_\beta,\Pi_{l=1}^m M_{I_{\alpha l}}\chi_{B_{\alpha(m+1)}}S_{\alpha\beta}f_\beta)$, we can change variables to the set $\mathbf{Y}_1,\ldots,\mathbf{Y}_m$, $\mathbf{z}_1,\ldots,\mathbf{z}_m$, $\mathbf{z}_{m+1}$. The integration over $\mathbf{z}_1,\ldots,\mathbf{z}_m$ is entirely unaffected by the product of operators in the inner product and contributes a factor of unity since $\phi_{\alpha_1},\ldots,\phi_{\alpha_m}$ are normalized. Taking account of this, we have

$$\begin{aligned}&\left(S_{\alpha\beta}f_\beta,\prod_{l=1}^m M_{I_{\alpha l}}\chi_{B_{\alpha(m+1)}}S_{\alpha\beta}f_\beta\right)\\ &\quad=\int\overline{g_{\alpha\beta}(\mathbf{Y}_1,\ldots,\mathbf{Y}_m)}\prod_{l=1}^m M_{I_{\alpha l}}g_{\alpha\beta}(\mathbf{Y}_1,\ldots,\mathbf{Y}_m)\\ &\quad\times d\mathbf{Y}_1\cdots d\mathbf{Y}_m\\ &\quad\times\int\overline{\phi_{\alpha_{m+1}}(\mathbf{z}_{m+1})}\chi_{B_{\alpha(m+1)}}(\mathbf{z}_{m+1})\phi_{\alpha_{m+1}}(\mathbf{z}_{m+1})d\mathbf{z}_{m+1},\end{aligned} \quad (81)$$

where in (81) $M_{I_{\alpha l}}$ operates on the function $g_{\alpha\beta}$ in the obvious way. [Previously $M_{I_{\alpha l}}$ was defined only on $\mathcal{L}^2(\mathbb{R}^{3n})$. When acting on $S_{\alpha\beta}f_\beta$, $M_{I_{\alpha l}}$ affects only the function $g_{\alpha\beta}$. We have written the result as $M_{I_{\alpha l}}g_{\alpha\beta}$.] We can rewrite (81) using the definition of $M_{I_{\alpha l}}$ as:

$$\begin{aligned}&\left(S_{\alpha\beta}f_\beta,\prod_{l=1}^m M_{I_{\alpha l}}\chi_{B_{\alpha(m+1)}}S_{\alpha\beta}f_\beta\right)\\ &\quad=\int_{I_{\alpha_1}\times\cdots\times I_{\alpha_m}}|\widetilde{g_{\alpha\beta}}(\mathbf{p}_1,\ldots,\mathbf{p}_m)|^2 d\mathbf{p}_1\cdots d\mathbf{p}_m\\ &\quad\times\int_{B_{\alpha(m+1)}}|\phi_{\alpha_{m+1}}(\mathbf{z}_{m+1})|^2 d\mathbf{z}_{m+1},\end{aligned} \quad (82)$$

where $B_{\alpha(m+1)}$ is the subset of $\mathbb{R}^{3S}$ in which each of the $S\mathbf{x}_i$'s in Γ_{m+1}^α (i.e., the $\mathbf{x}_i$'s we have represented by $\mathbf{z}_{m+1}$) lies in its corresponding cone, and $\widetilde{g_{\alpha\beta}}$ is the Fourier transform of $g_{\alpha\beta}$ in all of its variables.

Before analysing the "cross terms" in the second sum of the right-hand side of (79), we pause to examine (82).

The inner product on the left-hand side of (82) is the contribution of the outgoing channel α to the probability we wish to compute, and we can interpret the expression for this contribution on the right-hand side of (82) as follows: if T_α is defined by (44a), then $|(e^{-iT_\alpha t} g_{\alpha\beta}(\mathbf{Y}_1,\ldots,\mathbf{Y}_m)|^2$ is the ppd for the centers of mass $\mathbf{Y}_1,\ldots,\mathbf{Y}_m$ of the fragments in α in the outgoing state $e^{-iH_\alpha t}S_{\alpha\beta}f_\beta$ and $|\widetilde{g_{\alpha\beta}}(\mathbf{p}_1,\ldots,\mathbf{p}_m)|^2$ is the (constant!) mpd for these centers of mass in the state $e^{-iH_\alpha t}S_{\alpha\beta}f_\beta$. The first integral on the right-hand side of (82) is therefore just the probability that the center-of-mass momentum of the fragment l lies in the intersection $I_{\alpha l}$ for $l=1,\ldots,m$. It is not hard to see physically that this is the same as the probability that all the particles of all the fragments $l=1,\ldots,m$ asymptotically will be found in their appropriate cones: let $i \in \Gamma_l$ be the index of a particle in the lth fragment. Since the particles constituting a fragment travel together, particle i will asymptotically be in C_i if and only if the fragment l is asymptotically in C_i, and this in turn will be true if and only if the center-of-mass momentum $\mathbf{p}_l$ of the fragment belongs to C_i. All particles in fragment l will asymptotically belong to their corresponding cones if and only if $\mathbf{p}_l$ belongs to all of the cones C_i with $i \in \Gamma_l$, that is to $I_{\alpha l}$. This confirms the interpretation of the first integral on the right-hand side of (82).

Since $|\phi_{\alpha_{m+1}}(\mathbf{z}_{m+1})|^2$ is the (constant) ppd for the particles in Γ_{m+1} in the state $e^{-iH_\alpha t}S_{\alpha\beta}f_\beta$, and $B_{\alpha(m+1)}$ is the set defined after (82), the second integral in (82) is by definition the probability that the particles in Γ_{m+1} are in their respective cones—asymptotically or not, it makes no difference, since this latter probability is time-independent. Since the right-hand side of (82) is the product of the two integrals we have discussed, we have now explained why this right-hand side gives the probability for the particles to wind up in their respective cones if the outgoing channel is α. We must now deal with the "cross-terms" in the second sum of the right-hand side of (79). In analysing a term of this sum, we shall write m instead of $m(\eta_p)$.

To analyse a typical term of the second sum on the right-hand side of (79), we first write down expressions for the elements of $\mathcal{L}^2(R^{3n})$ occurring in the inner product in this term:

$$(S_{\eta_p\beta}f_\beta)(\mathbf{x}_1,\ldots,\mathbf{x}_n) = g_{\eta_p\beta}(\mathbf{Y}_1,\ldots,\mathbf{Y}_n) \times \phi_{\eta_p 1}(\mathbf{z}_1)\cdots\phi_{\eta_p(m+1)}(\mathbf{z}_{m+1}), \tag{83a}$$

$$(S_{\eta'_p\beta}f_\beta)(\mathbf{x}_1,\ldots,\mathbf{x}_n) = g_{\eta'_p\beta}(\mathbf{Y}_1,\ldots,\mathbf{Y}_n)\phi_{\eta'_p 1}(\mathbf{z}_1)\cdots\phi_{\eta'_p(m+1)}(\mathbf{z}_{m+1}). \tag{83b}$$

Note that there are as many bound states in (83a) as in (83b) and corresponding bound states have the *same* sets of variables. This is because the channels (p,η_p) and (p,η'_p) have the same partition. [In (83), as in (79), we are using η_p as an abbreviation for (p,η_p), etc.] The selection of bound states in (83a) and (83b) must, however, be different, because in the second sum on the right-hand side of (79) we are only interested in the case $\beta_p \neq \beta'_p$. Now in the inner product occurring in the term we are analysing, the product $\prod_{l=1}^m M_{I_{\eta_p l}} \chi_{B_{\eta_p(m+1)}}$ does not affect any of the "internal variables" $\mathbf{z}_1,\ldots,\mathbf{z}_m$. Changing variables in the inner product to $\mathbf{Y}_1,\ldots,\mathbf{Y}_m$, $\mathbf{z}_1,\ldots,\mathbf{z}_m$, $\mathbf{z}_{m+1}$, we see that the integration over $\mathbf{z}_1,\ldots,\mathbf{z}_m$ produces the factor $(\phi_{\eta_p 1},\phi_{\eta'_p 1})\cdots(\phi_{\eta_p m},\phi_{\eta'_p m})$ which is either unity or zero as the two sets of m bound states in question are identical or not. If they are not identical the term we are analysing vanishes, so we proceed on the assumption that they are identical. Since the selection of bound states specified by η_p and η'_p is different, this must mean that

$$\phi_{\eta_p(m+1)} \neq \phi_{\eta'_p(m+1)}. \tag{84}$$

The bound states $\phi_{\eta_p(m+1)}$ and $\phi_{\eta'_p(m+1)}$ are then orthogonal. Arguing as in our earlier discussion of the first sum on the right-hand side of (79), we can see that the inner product in the term we are analysing now becomes

$$\begin{aligned} D(\eta_p,\eta'_p) = &\int_{I_{\eta_p 1}\times\cdots\times I_{\eta_p m}} \overline{g_{\eta_p}(\mathbf{Y}_1,\ldots,\mathbf{Y}_m)} \\ &\times g_{\eta'_p}(\mathbf{Y}_1,\ldots,\mathbf{Y}_m)\,d\mathbf{Y}_1\cdots d\mathbf{Y}_m \\ &\times \int_{B_{\eta_p(m+1)}} \overline{\phi_{\eta_p(m+1)}(\mathbf{z}_{m+1})}\,\phi_{\eta'_p(m+1)}(\mathbf{z}_{m+1})\,d\mathbf{z}_{m+1}. \end{aligned} \tag{85}$$

$B_{\eta_p(m+1)}$ is, of course, the set in which each coordinate $\mathbf{x}_i$ with $i \in \Gamma_{m+1}$ is restricted to its appropriate cone, and although the bound states in the integrand are orthogonal, the integral over only the set $B_{\eta_p(m+1)}$ need not vanish. We now have for the second sum on the right-hand side of (79) the expression

$$S_2(t) = \sum_p \sum_{\eta_p\neq\eta'_p} \exp[i(E_{\eta_p} - E_{\eta'_p})t]\,D(\eta_p,\eta'_p), \tag{86}$$

where $D(\eta_p,\eta'_p)$ is given by (85) when only the bound states $\phi_{\eta_p(m+1)}$ and $\phi_{\eta'_p(m+1)}$ differ, and is otherwise zero. Using (82), we now write

$$\begin{aligned} &\lim_{t\to+\infty} P(f_\beta;C_1,\ldots,C_n,t) \\ &\doteq \lim_{t\to+\infty}\Big(\sum_\alpha \int_{I_{\alpha 1}\times\cdots\times I_{\alpha m(\alpha)}} |\tilde g_{\alpha\beta}(\mathbf{p}_1,\ldots,\mathbf{p}_{m(\alpha)})|^2 \\ &\quad\times d\mathbf{p}_1\cdots d\mathbf{p}_{m(\alpha)} \\ &\quad\times \int_{B_{\alpha(m(\alpha)+1)}} |\phi_{\alpha_{m(\alpha)+1}}(\mathbf{z}_{m(\alpha)+1})|^2\,dz_{m(\alpha)+1} + S_2(t)\Big), \end{aligned} \tag{87}$$

where $S_2(t)$ is given by (86).

We now make several remarks. First, if all the static potentials V_{0j} were zero, then there could be no particles trapped near the origin, and the set Γ_{m+1} would always be empty. The case in which a term of $S_2(t)$ was nonzero would never arise, so that S_2 would vanish. The integral over $\mathbf{z}_{m(\alpha)+1}$ in the sum on the right-hand side of (87) would not appear. The limit indicated in (87) would then exist, and we would have the result:

Theorem 1: Suppose the Hamiltonian H of (2) describing n particles is such that the limits $\Omega^\pm_\alpha$ of (18) exist and the requirement of asymptotic completeness is satisfied. Suppose also that the *static* potentials $V_{0j}(\mathbf{x}_j)$, $j=1,\ldots,n$, in H are all identically zero. Then, with the notation defined previously, we have

$$\begin{aligned} P(f_\beta;C_1,\ldots,C_n) &= \lim_{t\to+\infty} P(f_\beta;C_1,\ldots,C_n,t) \\ &= \lim_{t\to+\infty}\int_{C_1\times\cdots\times C_n} |(e^{-iHt}\Omega^-_\beta f_\beta)(\mathbf{x}_1,\ldots,\mathbf{x}_n)|^2\,d\mathbf{x}_1\cdots d\mathbf{x}_n \\ &= \sum_\alpha \int_{I_{\alpha l}\times\cdots\times I_{\alpha m(\alpha)}} |\tilde g_{\alpha\beta}(\mathbf{p}_1,\ldots,\mathbf{p}_{m(\alpha)})|^2\,d\mathbf{p}_1\cdots d\mathbf{p}_{m(\alpha)}. \end{aligned} \tag{88}$$

The significance of the integrals occurring on the right-hand side of (88) has already been explained, and in view of this significance we ought to consider Theorem 1 as a mathematical verification of "just what we expected physically" (for the case when all the V_{0j} vanish): The probability for the particles to appear in their respective cones is a sum. Each term of the sum represents the particles emerging in a definite channel α, and gives the probability that the center-of-mass coordinate of each composite particle is in the appropriate intersection of cones. It is noteworthy that the various channels represented in (88) do not "interfere" with each other—each contributes independently to the sum. We also remark that we could write down immediately, by selecting the appropriate term of (88), the probability that the particles emerge *in a definite channel* with each composite particle in a specified cone. This is a probability frequently sought in scattering theory, and it is a pleasant feature of our approach that all such probabilities can be read off in a straightforward way from the one equation (88). The proof of Theorem 1 seems to require a lot of writing, although it is straightforward. Theorem 1 is the statement of the result on scattering into cones in the situation when the question concerning scattering into cones is most natural, as discussed at the beginning of this paper. The case when there are no static potentials is actually also the most realistic case physically, particularly if one wants to deal with the recoil of the target one is bombarding. Nevertheless, we go on to make some remarks about the situation in which the V_{0j} do not vanish.

When the V_{0j} do not vanish, the $D(\eta_p, \eta'_p)$ in (86) need not vanish, and of course in this case $S_2(t)$ will not converge as $t \to +\infty$ unless by some stroke of luck we have $E_{\eta_p} = E_{\eta'_p}$ whenever $D(\eta_p, \eta'_p) \neq 0$. The physical reason for the nonconvergence of $S_2(t)$ is not far to seek. Rather than try to disentangle the notation involved in the n-body problem, we give the reason by a simple example: consider a single particle acted on by a potential V, and suppose there are normalized bound states $\phi_1, \phi_2, \ldots$, with energies $E_1, E_2, \cdots$. We ask for the asymptotic probability that the particle is in a cone C when the state of the particle is

$$\phi_t(\mathbf{x}) = \sum_n e^{-iE_n t} C_n \phi_n(\mathbf{x}), \quad \sum_n |C_n|^2 < \infty. \tag{89}$$

We assume for convenience that the sum in (89) is finite, so that there are no later convergence difficulties. If the sum in (89) is infinite, we can use a limiting process to obtain the same results. The required probability is

$$P = \lim_{t\to+\infty} \int_C |\sum_n e^{-iE_n t} C_n \phi_n(\mathbf{x})|^2 d\mathbf{x} \tag{90}$$

and the limit in (90) does not exist because the ppd for the particle fluctuates in a regular way, so that as time goes on the probability for the particle to be in the cone does not approach a limit. It is, however, sensible to ask for the time average of the probability that the particle will be in the cone C. This is just the Cesaro limit:

$$\overline{P} = \lim_{T\to\infty} \frac{1}{T} \int_0^T \left(\int_C |\sum_n e^{-iE_n t} C_n \phi_n(\mathbf{x})|^2 d\mathbf{x}\right) dt. \tag{91}$$

We can write [note the analogy to (87)]

$$\int_C |\sum_n e^{-iE_n t} C_n \phi_n(\mathbf{x})|^2 d\mathbf{x} = \sum_n |C_n|^2 \int_C |\phi_n(\mathbf{x})|^2 d\mathbf{x} + S_2(t) \tag{92}$$

with

$$S_2(t) = \sum_{n\neq n'} e^{i(E_n - E_{n'})t} d(E_n, E_{n'}), \tag{93}$$

where

$$d(E_n, E_{n'}) = \overline{C_n} C_{n'} \int_C \overline{\phi_n(x)} \phi_{n'}(\mathbf{x}) d\mathbf{x}. \tag{94}$$

It should be clear that the Cesaro limiting process of (91) will not change the constant first term of (92), and will remove all terms of $S_2(t)$ for which $E_n \neq E_{n'}$, while leaving unchanged terms of $S_2(t)$ with $E_n = E_{n'}$. Thus we have

$$\overline{P} = \sum_n |C_n|^2 \int_C |\phi_n(\mathbf{x})|^2 d\mathbf{x} + \sum_{\substack{n\neq n' \\ E_n = E'_n}} \overline{C_n} C_{n'} \int_C \overline{\phi_n(\mathbf{x})} \phi_{n'}(\mathbf{x}) d\mathbf{x}. \tag{95}$$

We could write $\overline{P}$ in a slightly more attractive form if desired. Namely, we could group together *all* ϕ_n's with the same energy E and call the result ψ_E:

$$\psi_E(\mathbf{x}) = \sum_{E_n = E} C_n \phi_n(\mathbf{x}), \qquad \|\psi_E\|^2 = \sum_{E_n = E} |C_n|^2. \tag{96}$$

Then we would have instead of (95) the equation

$$\overline{P} = \sum_E \int_C |\psi_E(x)|^2 d\mathbf{x}. \tag{97}$$

The cross-terms in (95) arise because wave-functions ϕ_n with the same energy contribute "together" to $\overline{P}$—the ppd determined from ψ_E is not oscillatory. The expression (97) makes it somewhat clearer that the sums in (95) will converge when n is infinite.

The behavior of the particles near the origin in our n-body problem is analogous to the behavior of the particle we have just been discussing, and is the origin of the nonconvergent part $S_2(t)$ of (87). By resorting to the same method as before, we can compute the Cesaro Limit, i.e., the time-average $\overline{P}(f_\beta; C_1, \ldots, C_n)$ as $t \to +\infty$ of the probability that the particles will be in their respective cones if we "send in" $e^{-iH_\beta t} f_\beta$. The expression for $\overline{P}(f_\beta; C_1, \ldots, C_n)$ is a direct analogue of (95). Namely, we have:

Theorem 2: Suppose the Hamiltonian H satisfies the hypotheses of Theorem 1, except for the hypothesis that the static potentials V_{0j} are all zero. Then, in the notation defined previously, we have

$$\begin{aligned}
\overline{P}(f_\beta; C_1, \ldots, C_n) &= \lim_{T\to\infty} \frac{1}{T} \int_0^T P(f_\beta; C_1, \ldots, C_n, t)\, dt \\
&= \lim_{T\to\infty} \frac{1}{T} \int_0^T \left\{\int_{C_1\times\cdots\times C_n} |(e^{-iHt}\Omega_\beta f_\beta)(\mathbf{x}_1, \ldots, \mathbf{x}_n)|^2 \times d\mathbf{x}_1 \cdots d\mathbf{x}_n\right\} dt \\
&= \sum_\alpha \int_{I_{\alpha 1}\times\cdots\times I_{\alpha m(\alpha)}} |\tilde{g}_{\alpha\beta}(\mathbf{p}_1, \ldots, \mathbf{p}_{m(\alpha)})|^2 d\mathbf{p}_1 \cdots d\mathbf{p}_{m(\alpha)} \\
&\quad \times \int_{B_{\alpha(m(\alpha)+1)}} |\phi_{\alpha(m(\alpha)+1)}(\mathbf{z}_{m(\alpha)+1})|^2 d\mathbf{z}_{m(\alpha)+1} \\
&\quad + \sum_p \sum_{\substack{\eta_p \neq \eta'_p \\ E_{\eta_p} = E_{\eta'_p}}} D(\eta_p, \eta'_p).
\end{aligned} \tag{98}$$

We could absorb the cross-terms $D(\eta_p, \eta'_p)$ into the first sum just as we did with the analogous terms in (95), but this would just be an exercise in notation changing. The explanation of these cross-terms remaining in (98) is entirely analogous to the explanation of the analogous terms in (95).

With Theorems 1 and 2, we have reached our goal of computing the probability (or when necessary the time average of the probability) for the particles to be in cones $C_1, \ldots, C_n$ after a scattering experiment. We recall the expression (41) from which we started, and restate some of our results as follows:

When all the V_{0j} are zero, we have

$$P(f_\beta; C_1, \ldots, C_n) = \lim_{t\to\infty} \int_{C_1\times\cdots\times C_n} |(e^{-iHt}\Omega_\beta^- f_\beta)(\mathbf{x}_1, \ldots, \mathbf{x}_n)|^2 \, d\mathbf{x}_1 \cdots d\mathbf{x}_n \,. \tag{99}$$

When not all the V_{0j} are zero, we have

$$\overline{P}(f_\beta; C_1, \ldots, C_n) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \left(\int_{C_1\times\cdots\times C_n} |(e^{-iHt}\Omega_\beta^- f_\beta)(\mathbf{x}_1, \ldots, \mathbf{x}_n)|^2 \times d\mathbf{x}_1 \cdots d\mathbf{x}_n . \right) dt. \tag{100}$$

We now know that the limits in (99) and (100) exist. Now $e^{-iHt}e^{iH_\beta t}f_\beta$ converges strongly to $\Omega_\beta^- f_\beta$ as $t \to +\infty$. We thus also have (use unitarity of e^{-iHt})

$$\lim_{t\to+\infty} \|e^{-2iHt}e^{iH_\beta t}f_\beta - e^{-iHt}\Omega_\beta^- f_\beta\| = 0. \tag{101}$$

Because of (101), we can replace $e^{-iHt}\Omega_\beta^- f_\beta$ by $e^{-2iHt}e^{iH_\beta t}f_\beta$ in the integrands in (99) and (100) (see Lemma 3 of I). We call attention to this fact by stating it as

Theorem 3: Suppose the Hamiltonian H of (2) describing n particles is such that the limits $\Omega_\alpha^\pm$ of (18) exist and the requirement of asymptotic completeness is satisfied. Let $P(f_\beta; C_1, \ldots, C_n)$ be the probability that if scattering is initiated in the state $e^{-iH_\beta t}f_\beta$ then, for $i = 1, \ldots, n$, particle i will be in the cone C_i at large positive times (provided it makes sense to speak of this probability). Let $\overline{P}(f_\beta; C_1, \ldots, C_n)$ denote the Cesaro limit for large positive times of the probability that for $i = 1, \ldots, n$, particle i will be in cone C_i when scattering is initiated in the state $e^{-iH_\beta t}f_\beta$. Then

(A) If all the static potentials V_{0j} vanish, then it makes sense to speak of $P(f_\beta; C_1, \ldots, C_n)$, and

$$P(f_\beta; C_1, \ldots, C_n) = \lim_{t\to\infty} \int_{C_1\times\cdots\times C_n} |(e^{-2iHt}e^{iH_\beta t}f_\beta)(\mathbf{x}_1, \ldots, \mathbf{x}_n)|^2 \times d\mathbf{x}_1 \cdots d\mathbf{x}_n . \tag{102}$$

(B) If the V_{0j} do not all vanish, it makes sense to speak of $\overline{P}(f_\beta; C_1, \ldots, C_n)$ and

$$\overline{P}(f_\beta; C_1, \ldots, C_n) = \lim_{T\to+\infty} \frac{1}{T} \int_0^T \left(\int_{C_1\times\cdots\times C_n} |(e^{-2iHt}e^{iH_\beta t}f_\beta)(\mathbf{x}_1, \ldots, \mathbf{x}_n)|^2 d\mathbf{x}_1 \cdots d\mathbf{x}_n \right) dt. \tag{103}$$

Theorem 3 is the analog of Theorem 1 of I, and Eqs. (102) and (103) are the analogs of Eq. (22) of I. Theorem 3 represents a departure from the usual time-dependent n-body scattering theory because the formulas for the probabilities in (102) and (103) no longer involve Ω_β^-. Of course, existence of Ω_β^- was one of the hypotheses used in deriving these formulas. However, since Ω_β^- does not appear in (102) and (103) it is at least conceivable that these equations can be used to compute the probability in which we are interested in a case in which the MØller wave matrices Ω_β^- do not exist. One sees easily enough that (102) and (103) give the correct probabilities in the case of constant potentials V_{ij}, for which the MØller wave matrices trivially do not exist unless all $V_{ij} = 0$. This is only of academic interest, however. It is tempting to conjecture that (102) and (103) also give correct results when Coulomb potentials are involved, in analogy to Theorem 2 of I. After a considerable struggle, the present author is unable to ascertain whether or not this is true, and is willing to call it a hard problem. Let us now proceed to a brief discussion of the case when Coulomb potentials are involved.

If some of the potentials occurring in the Hamiltonian H of (2) are Coulomb potentials, then the strong limits $\Omega_\alpha^\pm$ of (18) no longer exist, and we must modify our treatment. In discussing the necessary modifications, we shall rely heavily on the treatment of Coulomb n-body problems in Ref. 7. To indicate that there are now Coulomb potentials present, we write H_C instead of H. We proceed as follows: Just as in the case when there are no Coulomb potentials present, we make a list of all the ways in which various subsets of the n particles can be bound together. We denote the bound states by $\phi_{\alpha_l}(\mathbf{z}_l)$, as before. Concerning the ϕ_{α_l} we make the technical assumption that they are "slightly better than square-integrable"—what this means is that for each ϕ_{α_l} and for each internal coordinate $\mathbf{z}_l^k$, $k = 1, \ldots, r_l$, of the set $\mathbf{z}_l$, there exists an $\epsilon > 0$ such that

$$\int |\mathbf{z}_l^k|^\epsilon \, |\phi_{\alpha_l}(\mathbf{z}_l^1, \ldots, \mathbf{z}_l^{r_l})|^2 \, d\mathbf{z}_l^1, \ldots, d\mathbf{z}_l^{r_l} < \infty. \tag{104}$$

The domains D_α are defined as before with P_α denoting the projection on D_α. In place of the usual operators $e^{-iH_\alpha t}$ we construct "distorted" operators $U_{\alpha C}(t)$ described in Ref. 7. Then under the hypothesis (104) and the hypothesis that each potential in H_C which is not a Coulomb potential can be written as the sum of a function in $\mathcal{L}^2(\mathbb{R}^3)$ and a function in $\mathcal{L}^p(\mathbb{R}^3)$, with $2 < p < 3$, we can show that the strong limits

$$\lim_{t\to\pm\infty} e^{iH_C t}U_{\alpha C}(t)P_\alpha = \Omega_{\alpha C}^\pm \tag{105}$$

exist.[7,8] The operators $\Omega_{\alpha C}^\pm$ have all the properties (19) through (22) previously listed for MØller wave-matrices. We denote by $R_{\alpha C}^\pm$ the range of $\Omega_{\alpha C}^\pm$ and by $R_C^\pm$ the orthogonal sum of all the $R_{\alpha C}^\pm$. We shall continue our discussion under the assumption that the limits in (105) do indeed exist and that the theory is asymptotically complete, i.e., that $R_C^+ = R_C^-$ and $R_C^\pm$ is the orthogonal complement of the subspace spanned by the true bound states of the theory.

When Coulomb potentials are present, one can no longer "send in $e^{-iH_\beta t}f_\beta$", i.e., one can no longer specify a scattering state of the n-body system by requiring that at large negative times the n particles were in a state described by the wave-function $e^{-iH_\beta t}f_\beta$. Instead, we must "send in $U_{C\beta}(t)f_\beta$"—the operator $U_{C\beta}(t)$ is quite a bit like $e^{-iH_\beta t}$, but is somewhat "distorted", as necessary to obtain convergence in (105). The ppd's determined by $e^{-iH_\beta t}f_\beta$ and $U_{C\beta}(t)f_\beta$ agree asymptotically, so that to the casual observer a group of particles propagated by $U_{C\beta}(t)$ looks at large positive and negative times like a group of particles propagated by $e^{-iH_\beta t}$. In any case, we may now ask the question, Suppose that an n-body scattering experiment is initiated with the particles in a state described by the wave-function $U_{C\beta}(t)f_\beta$. What is the probability $P(f_\beta; C_1, \ldots, C_n)$ that for each $j = 1, \ldots, n$ particle j will be in cone C_j at large positive times? The probability is of course given by

$$P(f; C_1, \ldots, C_n) = \lim_{t\to+\infty} \int_{C_1\times\cdots\times C_n} |(e^{-iH_C t}\Omega^-_{C\beta} f_\beta)(\mathbf{x}_1, \ldots, \mathbf{x}_n)|^2 \times d\mathbf{x}_1, \ldots, d\mathbf{x}_n, \qquad (106)$$

provided the limit in (104) exists. Actually, the situation concerning the existence or nonexistence of this limit is exactly the same as in the short-range case, and Theorems 1 and 2 hold verbatim when Coulomb potentials are present in the Hamiltonian, provided that any Møller wave-matrices arising (in the hypotheses of the theorems, the definition of $g_{\alpha\beta}$, etc.) are interpreted as the operators in (105). The fact that Theorems 1 and 2 hold when Coulomb potentials are present is analogous to the fact that it was possible to prove (29) in I. The reason that the theorems remain true is this: At large times $e^{-iH_C t}\Omega^-_{C\beta} f_\beta$ can be replaced by $\sum_\alpha U_{C\alpha}(t)\Omega^{+*}_{C\alpha}\Omega^-_{C\beta} f_\beta$. At large positive times the ppd determined by $U_{C\alpha}(t)\Omega^{+*}_{C\alpha}\Omega^-_{C\beta} f_\beta$ is the same as that determined by $e^{-iH_\alpha t}\Omega^{+*}_{C\alpha}\Omega^-_{C\beta} f_\beta$, and "thus" the final result is the same as before. We put "thus" in quotation marks because all this must be proved. The essential point, however, is the one we have mentioned, and we omit the proofs, which are mostly a combination of the proof of (29) in I and the proofs of (88) and (98) in the present paper. Unfortunately, as stated above, the author has not been able to prove that (102) and (103) give correct results in the case that Coulomb potentials are involved.

In any case, we have derived a number of results which corroborate the usual interpretation of nonrelativistic scattering theory, giving a rigorous foundation to the geometrical picture of the scattering process.

---

* Work supported in part by the U.S. Atomic Energy Commission under Grant No. AT (40–1)–3992.

[1] J. D. Dollard, Commun. Math. Phys. **12**, 193 (1969).

[2] T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951).

[3] J. M. Jauch, Helv. Phys. Acta **31**, 127 (1958); Helv. Phys. Acta **31**, 661 (1958).

[4] Of course we can also consider linear combinations of situations in which different fragments occur, etc. We are trying to resolve the situation into simple cases.

[5] J. M. Cook, J. Math. and Phys. **36**, 82 (1957); M. N. Hack, Nuovo Cimento **9**, 731 (1958); Nuovo Cimento **13**, 231 (1959); J. M. Jauch and I. Zinnes, Nuovo Cimento **11** (10), 553 (1959).

[6] T. Ikebe, Arch. Ration. Mech. Anal. **5**, 1 (1960), L. D. Fadeev, Tr. Mat. Inst. Akad. Nauk SSSR **69**, ,(1963); English transl., Scientific Translations, Jerusalem, 1965. K. Hepp, Helv. Phys. Acta **42**, fasc. 3, 425 (1969), T. Kato, *Perturbation Theory for Linear Operators*, Die Grundlehren der math. Wissenschaften, Band 132 (Springer, New York, 1966).

[7] J. D. Dollard, J. Math. Phys. **5**, 729 (1964).

[8] D. Marchesin and M. L. O'Carroll, J. Math. Phys. **13**, 982 (1972).

Maximal foliations of extended Schwarzschild space

Bruce L. Reinhart*

Department of Mathematics, University of Maryland, College Park, Maryland 20742
(Received 30 November 1972)

Smooth families of spherically symmetric maximal surfaces which are spacelike except at $r = 2m$ are explicitly constructed in Schwarzschild space. Such surfaces should be useful in the study of initial value problems.

Recent work of York[1] has indicated the usefulness of studying maximal spacelike foliations of relativity spaces. A foliation is a family of 3-surfaces filling the space, such that locally the surfaces arise as level surfaces of a function. A foliation is spacelike if each 3-surface is spacelike; it is maximal if the trace of the second fundamental form of each surface is 0. For background material on foliations, a good source is the expository paper of Haefliger.[2] In this paper, we shall find all spherically symmetric maximal foliations in extended Schwarzschild space.[3] It is proved that there are maximal foliations defined for all $r > 0$ which are spacelike except on the boundary of the black hole, $r = 2m$. At $r = 2m$, the 1-form defining the foliation enters the light cone.

Let us adopt the notation of Kruskal.[3] Then we are seeking a 1-form

$$\omega = dt + h(r)\,dr$$

such that the surfaces defined by $\omega = 0$ are maximal and spacelike. The unit normal vector to these surfaces is

$$T = \frac{1}{\sqrt{B - B^3h^2}}\left[-\frac{\partial}{\partial t} + B^2h\,\frac{\partial}{\partial r}\right],$$

where $1 - B^2h^2 > 0$ because the surfaces are spacelike. Also, an orthonormal frame tangent to the surfaces is given by

$$X_1 = \frac{\sqrt{B}}{\sqrt{1 - B^2h^2}}\left(-h\,\frac{\partial}{\partial t} + \frac{\partial}{\partial r}\right),$$

$$X_2 = \frac{1}{r}\frac{\partial}{\partial \theta}, \qquad X_3 = \frac{1}{r\sin\theta}\,\frac{\partial}{\partial \phi}.$$

Using this basis to calculate the trace of the second fundamental form, we find that it vanishes if and only if

$$h' = -\frac{2r - m}{r(r - 2m)}\,h + \frac{(r - 2m)(2r - 3m)}{r^3}\,h^3.$$

This is most easily solved by introducing the variable $\phi = Bh$, which satisfies

$$\phi' = [(2r - 3m)/r(r - 2m)](-\phi + \phi^3).$$

This equation has the solutions

$$\phi = 0, \qquad \phi = \pm\,[1 + Ar^3(r - 2m)]^{-1/2},$$

where A is a constant. The values of r for which this is defined depend upon A. However, if $0 < A < 16/27m^4$, then ϕ is defined for all $r > 0$, attains a maximum value at $r = 3m/2$, and approaches 0 as r goes to infinity. If

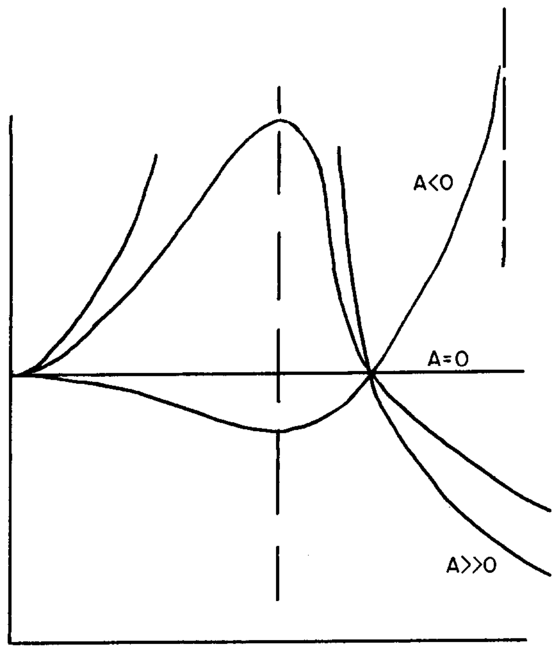

FIG. 1. ϕ as a function of r for various values of A. The location of the vertical asymptote for $A < 0$ depends upon A.

we now change to (v, u, θ, ϕ) coordinates, we find that ω is a multiple of

$$\omega_0 = (u - \phi v)\,dv + (\phi u - v)\,du.$$

As r approaches $2m$, ϕ approaches 1. Hence, the form

$$\omega_1 = [(1 + \phi^2)(u^2 + v^2) - 4uv\phi]^{-1/2}\,\omega_0$$

defines the same foliation, and approaches

$$2^{-1/2}(dv + du)$$

as r approaches $2m$ with $u > v$. It follows that the foliation can be defined on the whole space so as to be maximal and spacelike for $r \neq 2m$.

Note added in proof: Surfaces of this type have been studied independently by Estabrook *et al.*, Phys. Rev., in press. Their study focuses on computational questions, and includes a different derivation of the equations.

*This research supported by the General Research Board of the University of Maryland.

[1] J. W. York, Jr., Phys. Rev. Lett. **28**, 1082 (1972).
[2] A. Haefliger, Ann. Scuola Norm. Sup. Pisa **16**, 367 (1962).
[3] M. Kruskal, Phys. Rev. **119**, 1743 (1960).

The one-dimensional X–Y model in inhomogeneous magnetic fields*

K. M. Case and C. W. Lau

The Rockefeller University, New York, New York 10021
(Received 16 August 1972; final changes made 22 December 1972)

The one-dimensional $X-Y$ model in various kinds of inhomogeneous magnetic fields is analyzed. In each case we find the spectrum and eigenfunctions of the system, and study the equilibrium and relaxation behavior of the magnetization at any spin site.

1. INTRODUCTION

The $X-Y$ model of a one-dimensional interacting spin-$\frac{1}{2}$ system was first introduced by Lieb, Schultz, and Mattis in 1961.[1] Provided the interactions are restricted to nearest neighbors, and are under uniform external magnetic field, it is possible to obtain the spectrum and eigenfunctions of the system and even solve the Liouville equation exactly. Consequently the equilibrium and nonequilibrium properties of the system can be studied.[2]

One interesting nonequilibrium property of the model that has been studied concerns its relaxation behavior. Suppose initially the system is in thermal equilibrium in the presence of a uniform magnetic field. At time $t = 0$, the field is switched off, and we observe the evolution of certain physical observables such as magnetization. For the isotropic $X-Y$ model the total magnetization is a constant of motion and hence remains unchanged. For the anisotropic $X-Y$ model it is found that the magnetization per spin approaches a limiting value at $t \to \infty$, which, however, is not equal to its equilibrium value. To state this property concisely, one says that the system does not thermalize.[3]

Abraham et al[4] have studied the relaxation of the $X-Y$ model when the initial magnetic field is localized to only one spin site. It is found in this case that the magnetization at any spin site and the spin correlation functions approach their equilibrium values as t^{-1}.

In this paper we treat the isotropic $X-Y$ model in various kinds of spatially inhomogeneous magnetic fields. With a wide class of fields it is possible to find the spectrum and eigenfunctions of the system, and hence study its equilibrium and nonequilibrium properties. In particular, we find that in the relaxation problem, if the initial magnetic field approaches zero sufficiently rapidly at infinity, the magnetization at any spin site relaxes to its equilibrium value asymptotically as t^{-1}.

In the absence of a magnetic field and in the limit of an infinite number of spins, the spectrum of the $X-Y$ model consists of a bounded continuum. In an inhomogeneous magnetic field the nature of the spectrum in general changes. The cases of different magnetic fields considered here serve as explicit examples of how a continuous spectrum can be changed by external perturbations.

The plan of this paper is as follows. In Sec. 2 we perform the necessary transformations of variables to diagonalize the Hamiltonian, and to bring the eigenvalue equation to a form to be studied. In contrast to the usual procedure, the total number of spins is taken to be infinite right from the beginning. In Sec. 3 we consider the case of a magnetic field which is confined to a finite number of spin sites but is otherwise arbitrary. The eigenvalue equation is then cast into a form similar to that encountered in the linear theory of neutron transport. Consequently the mathematical apparatus of singular eigenfunction solution[5] can be taken over with very slight modifications. In Sec. 4, the results are generalized to the case of magnetic fields infinite in extent but approaching zero sufficiently rapidly at infinity. Then in the following three sections we consider the cases of a linearly increasing field, a quadratically increasing field, and a spatially alternating field. In each of these cases we find the spectrum and eigenfunctions explicitly, and study the equilibrium and relaxation behaviors of the magnetization. The final section briefly summarizes the results.

2. FORMULATION

The Hamiltonian which describes the isotropic $X-Y$ model with nearest neighbor interactions is given by

$$\mathcal{H}_0 = \sum_{n=-\infty}^{\infty} [S_n^x S_{n+1}^x + S_n^y S_{n+1}^y], \tag{2.1}$$

where the $\mathbf{S}_n$ are $\frac{1}{2}$ times the Pauli spin operators. This Hamiltonian commutes with $\sum_{n=-\infty}^{\infty} S_n^z$, the total spin in z direction.

Let h_n be the magnetic field (in the z direction) at the nth spin site. The complete Hamiltonian is then

$$\mathcal{H} = \mathcal{H}_0 + \sum_{n=-\infty}^{\infty} h_n S_n^z. \tag{2.2}$$

The Hamiltonian $\mathcal{H}$ can be brought to diagonal form by the following well-known transformations:

1. Define spin-raising and lowering operators b_n^+ b_n, by

$$b_n^+ = S_n^x + iS_n^y, \tag{2.3}$$

$$b_n = S_n^x - iS_n^y. \tag{2.4}$$

It follows that

$$S_n^z = b_n^+ b_n - \tfrac{1}{2}. \tag{2.5}$$

The b operators are neither Fermi nor Bose operators, but they can be expressed in terms of Fermi operators c_m, c_m^+ by the transformation

$$b_n = \exp(-\pi i \sum_{l=0}^{n-1} c_l^+ c_l) c_n, \tag{2.6}$$

$$b_n^+ = c_n^+ \exp(\pi i \sum_{l=0}^{n-1} c_l^+ c_l). \tag{2.7}$$

Under this transformation,

$$b_n^+ b_n = c_n^+ c_n, \tag{2.8}$$

and

$$b_n^+ b_{n\pm1} = c_n^+ c_{n\pm1} \tag{2.9}$$

In terms of the c operators, the Hamiltonian $\mathcal{H}$ becomes

$$\mathcal{H} = \tfrac{1}{2} \sum_{n=-\infty}^{\infty} [c_n^+ c_{n+1} + c_{n+1}^+ c_n] + \sum_{n=-\infty}^{\infty} h_n c_n^+ c_n - \tfrac{1}{2} \sum_{n=-\infty}^{\infty} h_n. \tag{2.10}$$

The constant term $-\frac{1}{2}\sum_{n=-\infty}^{\infty} h_n$ can be dropped by shifting the zero of the energy level.

2. Spatial Fourier transformation. Define

$$\psi(\theta) = \sum_{n=-\infty}^{\infty} \frac{c_n}{(2\pi)^{1/2}} e^{in\theta}, \qquad -\pi \le \theta \le \pi, \tag{2.11}$$

the inverse relation being

$$c_n = \int_{-\pi}^{\pi} \frac{d\theta}{(2\pi)^{1/2}} e^{-in\theta} \psi(\theta). \tag{2.12}$$

The $\psi(\theta)$ are Fermi operators satisfying the anticommutation relations

$$\{\psi^+(\theta), \psi(\theta')\} = \delta(\theta - \theta'),$$
$$\{\psi(\theta), \psi(\theta')\} = \{\psi^+(\theta), \psi^+(\theta')\} = 0. \tag{2.13}$$

Let $F(\theta)$ be the Fourier transform of the magnetic field,

$$F(\theta) = \sum_{n=-\infty}^{\infty} \frac{h_n}{2\pi} e^{in\theta}. \tag{2.14}$$

We have, since h_n is real,

$$F^*(\theta) = F(-\theta). \tag{2.15}$$

In this representation that Hamiltonian $\mathcal{H}$ becomes

$$\mathcal{H} = \int_{-\pi}^{\pi} d\theta \cos\theta\, \psi^+(\theta)\psi(\theta) + \int_{-\pi}^{\pi} d\theta \int_{-\pi}^{\pi} d\theta' \psi^+(\theta) F(\theta - \theta')\psi(\theta'). \tag{2.16}$$

To diagonalize $\mathcal{H}$, consider the eigenvalue problem

$$\cos\theta\, w_\lambda(\theta) + \int_{-\pi}^{\pi} d\theta' F(\theta - \theta') w_\lambda(\theta') = E_\lambda w_\lambda(\theta), \tag{2.17}$$

where $w_\lambda(\theta)$ are c-numbers, not operators. Suppose a complete orthonormal set of eigenfunctions $\{w_\lambda(\theta)\}$ can be found. Expand

$$\psi(\theta) = \int d\lambda\, a_\lambda w_\lambda(\theta),$$
$$\psi^+(\theta) = \int d\lambda\, a_\lambda^+ w_\lambda^*(\theta), \tag{2.18}$$

where the integral is understood to be integration over continuum states and summation over discrete states. Using the orthonormality of $w_\lambda(\theta)$, one can show that the coefficients a_λ are Fermi operators, and the Hamiltonian $\mathcal{H}$ can be brought to the diagonal form

$$\mathcal{H} = \int d\lambda E_\lambda a_\lambda^+ a_\lambda. \tag{2.19}$$

In the next few sections we will solve the eigenvalue problem (2.17) for different h_n and hence different F.

Before proceeding, we express the magnetization in terms of the eigenfunctions $u_\lambda(\theta)$. Let $m_n [= S_n^z]$ be the magnetization of the nth spin,

$$m_n = c_n^+ c_n - \tfrac{1}{2}. \tag{2.20}$$

In the relaxation problem the Hamiltonian is given by (2.16) for time $t \le 0$, and by (2.1) for $t > 0$. Hence if $\langle \cdots \rangle$ denotes a thermal average in the presence of the magnetic field, and $t \ge 0$, then

$$\begin{aligned}\langle m_n(t)\rangle &= \langle c_n^+(t) c_n(t)\rangle - \tfrac{1}{2} \\ &= \langle e^{i\mathcal{H}_0 t} c_n^+ c_n e^{-i\mathcal{H}_0 t}\rangle - \tfrac{1}{2} \\ &= \int_{-\pi}^{\pi} \frac{d\theta'}{(2\pi)^{1/2}} e^{in\theta'} \\ &\quad \times \int_{-\pi}^{\pi} \frac{d\theta}{(2\pi)^{1/2}} e^{-in\theta} \langle e^{i\mathcal{H}_0 t} \psi^+(\theta')\psi(\theta) e^{-i\mathcal{H}_0 t}\rangle - \tfrac{1}{2}.\end{aligned} \tag{2.21}$$

The Hamiltonian $\mathcal{H}_0$ is just the first term on the right-hand side of (2.16). Using the anticommutation relations (2.13), we have

$$e^{i\mathcal{H}_0 t} \psi^+(\theta')\psi(\theta) e^{-i\mathcal{H}_0 t} = e^{-it(\cos\theta' - \cos\theta)} \psi^+(\theta')\psi(\theta). \tag{2.22}$$

To calculate $\langle \psi^+(\theta')\psi(\theta)\rangle$, we use the eigenfunction expansion (2.18). Since a_λ are Fermi operators,

$$\langle a_{\lambda'}^+, a_\lambda\rangle = \frac{\delta(\lambda' - \lambda)}{1 + e^{\beta E_\lambda}}, \tag{2.23}$$

where $\delta(\lambda' - \lambda)$ denotes a Dirac delta-function if λ' and λ parametrize continuum states, and a Kronecker delta if both parametrize discrete states. Substituting (2.22), (2.18), and (2.23) into (2.21), we obtain

$$\begin{aligned}\langle m_n(t)\rangle &= \langle c_n^+(t) c_n(t)\rangle - \tfrac{1}{2} \\ &= \int d\lambda \frac{1}{1 + e^{\beta E_\lambda}} \int_{-\pi}^{\pi} \frac{d\theta'}{(2\pi)^{1/2}} \\ &\quad \times \int_{-\pi}^{\pi} \frac{d\theta}{(2\pi)^{1/2}} e^{in(\theta' - \theta) - i(\cos\theta' - \cos\theta)t} w_\lambda^*(\theta') w_\lambda(\theta) - \tfrac{1}{2}.\end{aligned} \tag{2.24}$$

Setting $t = 0$ gives the expression for the equilibrium magnetization of the nth spin in the presence of the external magnetic field.

Finally, one simple property of the magnetization should be noted. The Hamiltonian $\mathcal{H}_0$ in (2.1) does not distinguish between the positive and negative z axis. Write $\langle m_n(t)\rangle$ as

$$\langle m_n(t)\rangle = \frac{\mathrm{Tr}[e^{-\beta\mathcal{H}} e^{i\mathcal{H}_0 t} S_n^z e^{-i\mathcal{H}_0 t}]}{\mathrm{Tr}[e^{-\beta\mathcal{H}}]}. \tag{2.25}$$

Let $\langle m_n(t)\rangle_{(-)}$ be the magnetization obtained with all the h_n in (2.2) replaced by $-h_n$, i.e., with an overall sign change in the magnetic field. It can easily be seen from (2.25) that

$$\langle m_n(t)\rangle = -\langle m_n(t)\rangle_{(-)}. \tag{2.26}$$

This simple relation is sometimes useful in calculations related to the magnetization.

3. MAGNETIC FIELD ON FINITE NUMBER OF SITES

A. General properties of the eigenvalue equation (2.17)

Before solving the eigenvalue equation (2.17) for a given magnetic field, we note some of its general properties [which can be proved in the usual way by using (2.15)]:

(i) If an eigenfunction is square-integrable, its corresponding eigenvalue is real. The eigenfunctions that are not square-integrable will be shown to correspond to continuum modes with eigenvalues on the real line between -1 and $+1$. Hence all eigenvalues of (2.17) are real.

(ii) Eigenfunctions corresponding to different eigenvalues are orthogonal to each other.

(iii) There is a degeneracy: if $w_\lambda(\theta)$ is a solution of (2.17), then $w_\lambda^*(-\theta)$ is also a solution with the same eigenvalue. For simplicity we often consider magnetic fields which are symmetric with respect to the zeroth spin, i.e., $h_n = h_{-n}$. Then $F(\theta)$ is real and even in θ, by (2.15), and the eigenfunctions $w_\lambda(\theta)$ can be taken to be real functions. Because of the degeneracy we form linear combinations and classify the eigenfunctions as either even or odd in θ.

B. Eigenvalue equation with the magnetic field on finite number of sites

We assume $h_n = 0$ for $|n| > N$, and, for simplicity, $h_n = h_{-n}$. The function $F(\theta)$ then becomes

$$F(\theta) = \frac{h_0}{2\pi} + \sum_{n=1}^{N} \frac{h_n}{\pi} \cos n\theta = \sum_{n=0}^{N} \frac{h_n}{(1+\delta_{0,n})\pi} \cos n\theta. \tag{3.1}$$

The eigenvalue equation (2.17) becomes

$$\cos\theta\, w_\lambda(\theta) + \sum_{n=0}^{N} \frac{h_n}{(1+\delta_{0,n})\pi} \int_{-\pi}^{\pi} d\theta' \times \cos[n(\theta-\theta')] w_\lambda(\theta') = E_\lambda w_\lambda(\theta). \tag{3.2}$$

Let $u_\lambda(\theta)$ and $\bar{u}_\lambda(\theta)$ denote the even and odd eigenfunctions respectively. Expanding $\cos[n(\theta-\theta')] = \cos n\theta \cos n\theta' + \sin n\theta \sin n\theta'$, we have separate equations for the two types of eigenfunctions:

$$\cos\theta\, u_\lambda(\theta) + \sum_{n=0}^{N} \frac{h_n}{(1+\delta_{0,n})\pi} \cos n\theta \int_{-\pi}^{\pi} d\theta' \times \cos n\theta' u_\lambda(\theta') = E_\lambda u_\lambda(\theta), \tag{3.3}$$

$$\cos\theta\, \bar{u}_\lambda(\theta) + \sum_{n=0}^{N} \frac{h_n}{(1+\delta_{0,n})\pi} \sin n\theta \int_{-\pi}^{\pi} d\theta' \times \sin n\theta' \bar{u}_\lambda(\theta') = E_\lambda \bar{u}_\lambda(\theta). \tag{3.4}$$

These equations are quite similar to the eigenvalue equation that arises in linear neutron transport theory.[5] To solve these equations we therefore follow closely the method used in the latter case. We shall treat (3.3) in detail and only quote the results for (3.4), since the procedures are entirely the same in the two cases.

Let

$$s_n(E_\lambda) = \int_{-\pi}^{\pi} d\theta \cos n\theta\, u_\lambda(\theta) \tag{3.5}$$

be the Fourier coefficients of the eigenfunctions. Multiplying (3.3) throughout by $\cos l\theta$ and integrating over θ, we obtain the following recurrence relations for the coefficients:

$$s_{l+1}(E_\lambda) = [2/(1+\delta_{0,l})](E_\lambda - h_l) s_l(E_\lambda) - s_{l-1}(E_\lambda),$$
$$s_{-1} \equiv 0. \tag{3.6}$$

Clearly any $s_l(E_\lambda)$ is proportional to $s_0(E_\lambda)$, the proportionality factor being a polynomial of order l in E. The first few coefficients are given below:

$$s_1(E_\lambda) = (E_\lambda - h_0) s_0(E_\lambda),$$

$$s_2(E_\lambda) = [2(E_\lambda - h_1)(E_\lambda - h_0) - 1] s_0(E_\lambda),$$

$$s_3(E_\lambda) = [4(E_\lambda - h_2)(E_\lambda - h_1)(E_\lambda - h_0) - 2(E_\lambda - h_2) - (E_\lambda - h_0)] s_0(E_\lambda).$$

It can be seen that $u_\lambda(\theta)$ is a nontrivial solution of (3.3) only if $s_0(E_\lambda) = \int_{-\pi}^{\pi} d\theta u_\lambda(\theta)$ is nonvanishing. In this case we can set

$$s_0(E_\lambda) = 1, \tag{3.7}$$

since $u_\lambda(\theta)$ is determined only up to a multiplicative constant.

The eigenvalue equation (3.3) can be written as

$$(\cos\theta - E_\lambda) u_\lambda(\theta) = M(\cos\theta, E_\lambda), \tag{3.8}$$

where

$$M(\cos\theta, E_\lambda) = -\sum_{n=0}^{N} \frac{h_n}{(1+\delta_{0,n})\pi} \cos n\theta\, s_n(E_\lambda). \tag{3.9}$$

C. Discrete modes

Square-integrable solutions of (3.8) are discrete eigenfunctions. We use i instead of λ to parametrize these eigenfunctions and eigenvalues. Clearly either E_i lies outside the interval $[-1, +1]$ or, if it lies inside the interval, $M(E_i, E_i) = 0$. In either case,

$$u_i(\theta) = \frac{M(\cos\theta, E_i)}{\cos\theta - E_i}. \tag{3.10}$$

The eigenvalues E_i are determined by the condition (3.7):

$$1 = \int_{-\pi}^{\pi} d\theta \frac{M(\cos\theta, E_i)}{\cos\theta - E_i}. \tag{3.11}$$

Introduce the dispersion function

$$\Lambda(z) = 1 - \int_{-\pi}^{\pi} d\theta \frac{M(\cos\theta, z)}{\cos\theta - z} \tag{3.12}$$

$$= 1 + \sum_{n=0}^{N} \frac{h_n}{(1+\delta_{0,n})} s_n(z) Q_n(z), \tag{3.13}$$

where

$$Q_n(z) = \int_{-\pi}^{\pi} d\theta \frac{\cos n\theta}{\cos\theta - z}. \tag{3.14}$$

Every eigenvalue E_i is then a zero of $\Lambda(z)$. Conversely, if z_i is a zero of $\Lambda(z)$, i.e., $\Lambda(z_i) = 0$, we can use the recurrence relation (3.6) to construct a set of numbers $\{s_0(z_i), s_1(z_i), s_2(z_i), \cdots\}$, where $s_0(z_i)$ is taken to be unity. Using (3.9) and (3.10), we then construct a nontrivial eigenfunction $u_i(\theta)$ with eigenvalue z_i. Hence the discrete eigenvalues E_i are in one-to-one correspondence with the zeros of $\Lambda(z)$. This is an important observation to which we shall refer later. As an immediate consequence, the zeros of $\Lambda(z)$ must all lie on the real axis.

The function $\Lambda(z)$ is discontinuous across the cut from -1 to $+1$. Let $z \to \nu \pm i\epsilon$, where $\nu \epsilon [-1, +1]$, and denote the boundary values of the dispersion function by $\Lambda^\pm(\nu)$. It follows from (3.12) that

$$\tfrac{1}{2}[\Lambda^+(\nu) + \Lambda^-(\nu)] = 1 - 2P \int_{-1}^{1} \frac{d\mu}{\sqrt{1-\mu^2}} \frac{M(\mu, \nu)}{\mu - \nu}, \tag{3.15}$$

$$\tfrac{1}{2}[\Lambda^+(\nu) - \Lambda^-(\nu)] = -2\pi i \frac{M(\nu,\nu)}{(1-\nu^2)^{1/2}}, \quad (3.16)$$

where P denotes the Cauchy principal value.

D. Continuum modes

Suppose E_λ lies in the interval $[-1,+1]$. For convenience set

$$E_\lambda = \cos\lambda, \quad 0 \le \lambda \le \pi. \quad (3.17)$$

The general solution of (3.8) is a distribution,

$$u_\lambda(\theta) = P\frac{M(\cos\theta, \cos\lambda)}{\cos\theta - \cos\lambda} + \Gamma(\lambda)[\delta(\theta-\lambda) + \delta(\theta+\lambda)], \quad (3.18)$$

where the function $\Gamma(\lambda)$ is determined by (3.7):

$$\Gamma(\lambda) = \tfrac{1}{2}\left(1 - P\int_{-\pi}^{\pi} d\theta \frac{M(\cos\theta,\ \cos\lambda)}{\cos\theta - \cos\lambda}\right), \quad (3.19)$$

or, in terms of the dispersion function,

$$\Gamma(\lambda) = 1/4[\Lambda^+(\cos\lambda) + \Lambda^-(\cos\lambda)], \quad (3.20)$$

where (3.15) has been used. Clearly $\Gamma(\lambda)$ can always be *chosen* so that (3.20) is satisfied. Hence any E_λ on the real interval $[-1,+1]$ is an acceptable eigenvalue; the corresponding eigenfunction is given by (3.18) and (3.20).

E. Normalization integrals

The eigenfunctions $u_i(\theta)$ and $u_\lambda(\theta)$ are not properly normalized. Define the normalization integrals N_i and N_λ by

$$\int_{-\pi}^{\pi} d\theta[u_i(\theta)]^2 = N_i \quad (3.21)$$

and

$$\int_{-\pi}^{\pi} d\theta u_\lambda(\theta) u_{\lambda'}(\theta) = N_\lambda \delta(\lambda-\lambda'). \quad (3.22)$$

The properly normalized eigenfunctions are $N_i^{-1/2} u_i(\theta)$ and $N_\lambda^{-1/2} u_\lambda(\theta)$. Some care must be exercised in defining the normalization integral N_λ for the continuum eigenfunctions. We wish to use the normalization integrals to evaluate the coefficients in the expansion of a function $f(\theta)$ in terms of $u_\lambda(\theta)$,

$$f(\theta) = \int_0^{\pi} d\lambda A(\lambda) u_\lambda(\theta). \quad (3.23)$$

Multiplying by $u_{\lambda'}(\theta)$ and integrating over θ, we find that

$$\int_{-\pi}^{\pi} d\theta f(\theta) u_{\lambda'}(\theta) = \int_{-\pi}^{\pi} d\theta u_{\lambda'}(\theta) \int_0^{\pi} d\lambda A(\lambda) u_\lambda(\theta). \quad (3.24)$$

In analogy with the usual case, the left-hand side of (3.24) is defined to be $N_{\lambda'} A(\lambda')$. Thus,

$$N_\lambda = \frac{1}{A(\lambda')}\int_{-\pi}^{\pi} d\theta u_{\lambda'}(\theta) \int_0^{\pi} d\lambda A(\lambda) u_\lambda(\theta). \quad (3.25)$$

The important point to note is that the order of integration in this double integral matters, because of the singular nature of the functions $u_\lambda(\theta)$. To interchange the order of integration, one needs to use the Poincaré–Bertrand formula[5]

$$\int_0^{\pi} d\theta P\frac{1}{\cos\theta - \cos\lambda}\int_0^{\pi} d\lambda' P\frac{1}{\cos\lambda' - \cos\theta} g(\theta,\lambda')$$
$$= -\frac{\pi^2}{\sin^2\lambda} g(\lambda,\lambda) + \int_0^{\pi} d\lambda' \int_0^{\pi} d\theta P\frac{1}{\cos\theta - \cos\lambda} P\frac{1}{\cos\lambda' - \cos\theta} g(\theta,\lambda'), \quad (3.26)$$

where $g(\theta,\lambda')$ is an arbitrary function. To ensure that the quantity N_λ defined by (3.25) agrees with that defined by (3.22), the evaluation of the integral in (3.22) has to be supplemented by the following recipe regarding the product of two principal values:

$$P\frac{1}{\cos\theta-\cos\lambda} P\frac{1}{\cos\theta - \cos\lambda'} = \frac{1}{\cos\theta - \cos\lambda'}\left[P\frac{1}{\cos\theta-\cos\lambda} - P\frac{1}{\cos\theta - \cos\lambda'}\right] - \frac{\pi^2}{\sin^2\lambda}\delta(\theta-\lambda)\delta(\theta-\lambda'). \quad (3.27)$$

The recipe is a symbolic form of applying the Poincaré–Bertrand formula.

Evaluations of the normalization integrals are given in the Appendix. We state the results here.

$$N_i = -M(E_i,E_i)\frac{\partial\Lambda(z)}{\partial z}\bigg|_{z=E_i}, \quad (3.28)$$

$$N_\lambda = \tfrac{1}{2}\Lambda^+(\cos\lambda)\Lambda^-(\cos\lambda). \quad (3.29)$$

Two properties of the dispersion function $\Lambda(z)$ follow from these results:

(i) All zeros of $\Lambda(z)$ lie on the real axis *outside* the interval $[-1,+1]$.

Proof: Equation (3.12) shows that if $\Lambda(E_i) = 0$, where $E_i \in [-1,+1]$, then $M(E_i,E_i) = 0$. It follows from (3.28), then, that $N_i = \int_{-\pi}^{\pi} d\theta[u_i(\theta)]^2 = 0$. However, we have seen that every zero of $\Lambda(z)$ corresponds to a discrete eigenstate with a [nontrivial] square-integrable eigenfunction; i.e., $\int_{-\pi}^{\pi} d\theta[u_i(\theta)]^2 \ne 0$. This contradiction shows that $\Lambda(z)$ cannot have a zero within the interval $[-1,+1]$. Physically this property means that the discrete and continuum spectra do not overlap.

(ii) All zeros of $\Lambda(z)$ are simple.

Proof: If $\Lambda(z)$ has a multiple zero at E_i, then $\partial\Lambda(z)/\partial z\,|_{z=E_i} = 0$. The same argument as above shows that this leads to a contradiction.

In addition to the above two properties, we have the following:

(iii) All zeros of $\Lambda(z)$ lie on a finite segment of the real axis.

Proof: We examine the behavior of $\Lambda(z)$ [as given by (3.13)] at $|z| \to \infty$. From the recurrence relation (3.6) we have

$$s_n(z) \xrightarrow[|z|\to\infty]{} 2^{n-1} z^n. \quad (3.30)$$

From (3.14) we have

$$Q_n(z) \xrightarrow[|z|\to\infty]{} -\frac{\pi}{2^n z^{n+1}}. \quad (3.31)$$

Therefore,

$$\Lambda(z) \xrightarrow[|z|\to\infty]{} 1 - \frac{1}{2z} \sum_{n=0}^{N} \frac{h_n}{1+\delta_{0,n}}. \tag{3.32}$$

This shows that $\Lambda(\infty) = 1$. Hence the zeros of $\Lambda(z)$ must lie on a finite segment of the real axis.

F. Completeness

The eigenfunctions we have constructed form a complete set (for even functions). The proof of this important property follows closely the procedure used in the case of the neutron transport equation; instead of repeating the procedure, we refer the reader to Ref. 5 for details. Here we briefly sketch what the proof involves.

Let $f(\theta)$ be any given function,[6] even in θ. Because of evenness it is sufficient to consider only positive θ. We wish to prove that it is always possible to write $f(\theta)$ in the form

$$f(\theta) = \sum_{i=1}^{K} a_i u_i(\theta) + \int_0^{\pi} d\lambda\, a(\lambda) u_\lambda(\theta), \tag{3.33}$$

where the a are expansion coefficients to be determined, and K is the total number of discrete modes. Set

$$f'(\theta) = f(\theta) - \sum_{i=1}^{K} a_i u_i(\theta). \tag{3.34}$$

Using the explicit expression (3.18) for $u_\lambda(\theta)$, (3.33) can be written as

$$f'(\theta) = P\int_0^{\pi} d\lambda\, a(\lambda) \frac{M(\cos\theta,\ \cos\lambda)}{\cos\theta - \cos\lambda} + \tfrac{1}{4}[\Lambda^+(\cos\theta) + \Lambda^-(\cos\theta)]a(\theta). \tag{3.35}$$

If $f'(\theta)$ is regarded as a known function, (3.35) is a singular integral equation for the unknown function $a(\lambda)$. From the theory of singular integral equations, a solution exists only if the function $f'(\theta)$ satisfies K subsidiary conditions [which is connected with the existence of K simple zeros in the dispersion function $\Lambda(z)$]. These K conditions suffice to determine the K discrete coefficients a_i. Consequently the expansion (3.33) exists. Furthermore, the coefficients a_i and $a(\lambda)$ can be determined from the orthonormality of the eigenfunctions.

G. Odd eigenfunctions

The solution of (3.4) is entirely similar to (3.3), and we quote the results here.

The Fourier coefficients of the odd eigenfunctions $\bar{u}_\lambda(\theta)$,

$$\bar{s}_n(E_\lambda) = \int_{-\pi}^{\pi} d\theta\, \sin n\theta\, \bar{u}_\lambda(\theta), \tag{3.36}$$

satisfy the recurrence relation

$$\bar{s}_{n+1}(E_\lambda) = 2(E_\lambda - h_n)\bar{s}_n(E_\lambda) - \bar{s}_{n-1}(E), \qquad n \geq 1,$$
$$\bar{s}_0(E_\lambda) = 0. \tag{3.37}$$

It can be seen that every coefficient $\bar{s}_n$ is proportional to $\bar{s}_1$, the proportionality factor being a polynomial of order $n-1$ in E. As before, we set

$$\bar{s}_1(\lambda) = 1. \tag{3.38}$$

Define

$$\overline{M}(\cos\theta, E_\lambda) = -\sum_{n=1}^{N} \frac{h_n}{\pi} \sin n\theta\, \bar{s}_n(E_\lambda) \tag{3.39}$$

and

$$\overline{\Lambda}(z) = 1 - \int_{-\pi}^{\pi} d\theta\, \sin\theta \frac{\overline{M}(\cos\theta, z)}{\cos\theta - z} \tag{3.40}$$

$$= 1 + \sum_{n=1}^{N} \frac{h_n}{\pi} \bar{s}_n(z) \overline{Q}_n(z), \tag{3.41}$$

where

$$\overline{Q}_n(z) = \int_{-\pi}^{\pi} d\theta\, \sin\theta \frac{\sin n\theta}{\cos\theta - z}. \tag{3.42}$$

As before, there exist in general both discrete and continuum modes. The discrete eigenvalues are the zeros of $\overline{\Lambda}(z)$, these zeros having the same properties as in the even case. The continuum eigenvalues consist of the real interval $[-1, +1]$. The corresponding eigenfunctions are

$$\bar{u}_i(\theta) = \frac{\overline{M}(\cos\theta, \overline{E}_i)}{\cos\theta - \overline{E}_i}, \tag{3.43}$$

$$\bar{u}_\lambda(\theta) = P\, \frac{\overline{M}(\cos\theta,\ \cos\lambda)}{\cos\theta - \cos\lambda} + \frac{1}{4\sin\lambda}[\overline{\Lambda}^+(\cos\lambda) + \overline{\Lambda}^-(\cos\lambda)] \times [\delta(\theta - \lambda) - \delta(\theta + \lambda)], \tag{3.44}$$

and the normalization integrals

$$\int_{-\pi}^{\pi} d\theta [\bar{u}_i(\theta)]^2 = \overline{N}_i = -\frac{\overline{M}(\overline{E}_i, \overline{E}_i)}{(1 - \overline{E}_i^2)^{1/2}} \left.\frac{\partial \overline{\Lambda}(z)}{\partial z}\right|_{z=\overline{E}_i}, \tag{3.45}$$

$$\int_{-\pi}^{\pi} d\theta\, \bar{u}_\lambda(\theta) \bar{u}_{\lambda'}(\theta) = \overline{N}_\lambda \delta(\lambda - \lambda'), \tag{3.46}$$

$$\overline{N}_\lambda = \frac{\overline{\Lambda}^+(\cos\lambda)\overline{\Lambda}^-(\cos\lambda)}{2\sin^2\lambda}. \tag{3.47}$$

These discrete and continuum eigenfunctions together form a complete set (for odd functions).

H. An example

Only in very simple cases can one calculate the dispersion functions explicitly and hence obtain more detailed information about the distribution of discrete eigenvalues. Here we consider one such case, for which the magnetic field is

$$h_n = \begin{cases} h > 0, & |n| \leq N, \\ 0, & |n| > N. \end{cases} \tag{3.48}$$

The recurrence relations for the Fourier coefficients $s_n(z)$ and $\bar{s}_n(z)$ can be solved explicitly in this case. Let α_1 and α_2 be the roots of the quadratic equation

$$\alpha^2 - 2z\alpha + 1 = 0, \tag{3.49}$$

that is,

$$\alpha_{1,2} = z \pm \sqrt{z^2 - 1)^{1/2}} \tag{3.50}$$

For z outside the real interval $[-1, +1]$, α_1 and α_2 have different moduli, their product always being unity. We always take α_1 to be the root with modulus ≥ 1, and α_2 the root with modulus ≤ 1.

Let

$$z' = z - h. \tag{3.51}$$

The solutions of the recurrence relations (3.6) and (3.37), with the initial conditions (3.7) and (3.38), are

$$s_n(z) = \tfrac{1}{2}[\alpha_1'^n + \alpha_2'^n], \quad n \le N+1 \tag{3.52}$$

and

$$\bar{s}_n(z) = \frac{\alpha_1'^n + \alpha_2'^n}{\alpha_1' - \alpha_2'}, \quad n \le N+1, \tag{3.53}$$

where $\alpha_{1,2}'$ are defined by (3.50), with z replaced by z'. The right-hand sides of (3.52) and (3.53) are just Tchebichef polynomials of the first and second kinds, respectively, in the variable z':

$$s_n(z) = T_n(z'), \quad n \le N+1, \tag{3.54}$$

$$\bar{s}_n(z) = U_{n-1}(z'), \quad n \le N+1. \tag{3.55}$$

For $n > N+1$, the coefficients have different form, but they are not needed to evaluate the dispersion functions.

The integrals $Q_n(z)$ and $\bar{Q}_n(z)$ as defined by (3.14) and (3.42) can be evaluated to give

$$Q_n(z) = 4\pi\alpha_2^n/(\alpha_2 - \alpha_1), \quad n \ge 0, \tag{3.56}$$

$$\bar{Q}_n(z) = -2\pi\alpha_2^n, \quad n > 0. \tag{3.57}$$

Substitute (3.52), (3.53), (3.56), and (3.57) into (3.13) and (3.41) for the dispersion functions. The sums can be carried out and we find that

$$\Lambda(z) = [Q_N(z)/2\pi][\alpha_2 T_N(z') - T_{N+1}(z')], \tag{3.58}$$

$$\bar{\Lambda}(z) = [\bar{Q}_N(z)/2\pi][\alpha_2 U_{N-1}(z') - U_N(z')]. \tag{3.59}$$

Hence the zeros of $\Lambda(z)$ and $\bar{\Lambda}(z)$ are determined, respectively, by the two equations

$$\alpha_2 = T_{N+1}(z')/T_N(z'), \tag{3.60}$$

$$\alpha_2 = U_N(z')/U_{N-1}(z'). \tag{3.61}$$

One can show that

$$\frac{T_{N+1}(z')}{T_N(z')} > 1, \quad \text{for } z' > 1,$$

$$\frac{T_{N+1}(z')}{T_N(z')} < -1, \quad \text{for } z' < -1. \tag{3.62}$$

The same relations hold for the Tchebichef polynomial of the second kind. Since α_2 is always less than one in magnitude, we conclude that (3.60) and (3.61) have solutions only for

$$-1 < z' < +1$$

or

$$-1 + h < z < 1 + h. \tag{3.63}$$

Recall that the discrete eigenvalues cannot lie within the continuum $[-1, +1]$. Therefore, if $-1 + h < 1$ or $h < 2$, (3.63) can be replaced by

$$1 < z < 1 + h. \tag{3.64}$$

Equation (3.63) or (3.64) gives the bounds of the discrete eigenvalues. Their exact values can be found only by numerically solving (3.60) and (3.61).

I. Relaxation of magnetization

Returning to the case of a general magnetic field [finite in extent], the eigenfunction solutions will now be used to study the relaxation of the magnetization as formulated at the end of Sec. 2. We recall that in the relaxation problem the system is initially in thermal equilibrium in the presence of the magnetic field. At $t = 0$, the field is switched off, and we observe the evolution of the magnetization of the nth spin. The time-dependent magnetization, as given by (2.24), has four contributions, due to the even-discrete modes, even-continuum modes, odd-discrete modes, and odd-continuum modes, respectively.

Consider the even modes. Equation (2.24) gives

$$\langle C_n^+(t)C_n(t)\rangle_{\text{even}} = \sum_i \frac{B(E_i)}{N_i}\left|\int_{-\pi}^{\pi}\frac{d\theta}{(2\pi)^{1/2}}e^{in\theta - i\cos\theta\cdot t}u_i(\theta)\right|^2 + \int_0^{\pi} d\lambda\,\frac{B(\cos\lambda)}{N_\lambda}\left|\int_{-\pi}^{\pi}\frac{d\theta}{(2\pi)^{1/2}}e^{in\theta - i\cos\theta\cdot t}u_\lambda(\theta)\right|^2, \tag{3.65}$$

where

$$B(x) = 1/(1 + e^{\beta x}). \tag{3.66}$$

Clearly, as $t \to \infty$, the contributions of the discrete modes are all of order t^{-1}. The asymptotic time dependence of the continuum part, however, cannot be seen directly because of the singular nature of the eigenfunctions $u_\lambda(\theta)$. Some manipulations are needed.

Define

$$I(t, n, z) \equiv \int_{-\pi}^{\pi}\frac{d\theta}{(2\pi)^{1/2}}e^{in\theta - i\cos\theta\cdot t}\frac{M(\cos\theta, z)}{\cos\theta - z}, \tag{3.67}$$

$$F(z) \equiv -M(z,z)\Lambda(z). \tag{3.68}$$

The discrete normalization integral is given by

$$N_i = -M(E_i, E_i)\left.\frac{\partial\Lambda(z)}{\partial\Lambda}\right|_{z=E_i} = \left.\frac{\partial F(z)}{\partial z}\right|_{z=E_i} \tag{3.69}$$

where we have used $\Lambda(E_i) = 0$. The contributions of the discrete modes can be expressed as follows:

$$\sum_i \frac{B(E_i)}{N_i}\left|\int_{-\pi}^{\pi}\frac{d\theta}{(2\pi)^{1/2}}e^{in\theta - i\cos\theta\cdot t}u_i(\theta)\right|^2 = \sum_i \left.\frac{B(E_i)}{\partial F(z)/\partial z}\right|_{z=E_i} I(t,n,E_i)I(-t,n,E_i) = \frac{1}{2\pi i}\int_{C_2} dz\,\frac{B(z)}{F(z)}I(t,n,z)I(-t,n,z), \tag{3.70}$$

where C_2 is a contour around the discrete eigenvalues, as shown in Fig. 1. The contribution of the continuum modes is simply related to the contour integral in (3.70), but now with the contour C_1 surrounding the cut (Fig. 1):

$$\int_0^{\pi} d\lambda\,\frac{B(\cos\lambda)}{N_\lambda}\int_{-\pi}^{\pi}\frac{d\theta}{(2\pi)^{1/2}}\left|e^{in\theta - i\cos\theta\cdot t}u_\lambda(\theta)\right|^2 = \frac{1}{2\pi i}\int_{C_1} dz\,\frac{B(z)}{F(z)}I(t,n,z)I(-t,n,z) + \frac{1}{\pi}\int_0^{\pi} d\lambda B(\cos\lambda)\cos^2 n\lambda. \tag{3.71}$$

The validity of (3.71) can be checked by deforming the contour integral into integration above and below the cut, and using the explicit expression (3.18), (3.20) for the eigenfunctions $u_\lambda(\theta)$. Equation (3.65) then becomes

$$\langle C_n^+(t)C_n(t)\rangle_{\text{even}} = \frac{1}{2\pi i}\int_C dz\,\frac{B(z)}{F(z)}I(t,n,z)I(-t,n,z) + \frac{1}{\pi}\int_0^\pi d\lambda B(\cos\lambda)\cos^2 n\lambda, \qquad (3.72)$$

where C is the sum of the two contours C_1 and C_2.

Similarly for the odd modes, we have the result

$$\langle C_n^+(t)C_n(t)\rangle_{\text{odd}} = \frac{1}{2\pi i}\int_C dz\,\frac{B(z)}{\bar F(z)}\bar I(t,n,z)\bar I(-t,n,z) + \frac{1}{\pi}\int_0^\pi d\lambda B(\cos\lambda)\sin^2 n\lambda, \qquad (3.73)$$

where

$$\bar I(t,n,z) = \int_{-\pi}^{\pi}\frac{d\theta}{(2\pi)^{1/2}}e^{in\theta - i\cos\theta\cdot t}\,\frac{\bar M(\cos\theta,z)}{\cos\theta - z}, \qquad (3.74)$$

$$\bar F(z) = -\frac{\bar M(z,z)}{(1-z^2)^{1/2}}\bar\Lambda(z). \qquad (3.75)$$

Using the fact that

$$\frac{1}{\pi}\int_0^\pi d\lambda B(\cos\lambda)\cos^2 n\lambda + \frac{1}{\pi}\int_0^\pi d\lambda B(\cos\lambda)\sin^2 n\lambda = \frac{1}{\pi}\int_0^\pi d\lambda\, B(\cos\lambda) = 1/2, \qquad (3.76)$$

we finally have

$$\langle m_n(t)\rangle = \langle C_n^+(t)C_n(t)\rangle - \tfrac{1}{2} = \frac{1}{2\pi i}\int_C dz B(z) \times \left[\frac{I(t,n,z)I(-t,n,z)}{F(z)} + \frac{\bar I(t,n,z)\bar I(-t,n,z)}{\bar F(z)}\right] \qquad (3.77)$$

As $t\to\infty$, the integrand in (3.77) is of order t^{-1}. Hence we conclude that the magnetization approaches zero as t^{-1} for large t.

4. MAGNETIC FIELD ON INFINITE NUMBER OF SITES

The treatment and results in the last section are generalizable to the case of a magnetic field which is infinite in extent and satisfies

$$\sum_{n=0}^{\infty} h_n < \infty. \qquad (4.1)$$

Such a magnetic field poses questions of convergence of the dispersion function $\Lambda(z)$ and the M function. Let us first consider the dispersion function given by (3.13),

$$\Lambda(z) = 1 + \sum_{n=0}^{\infty}\frac{h_n}{(1+\delta_{0,n})\pi}s_n(z)Q_n(z). \qquad (4.2)$$

The coefficients $s_n(z)$ satisfy the recurrence relation (3.6), which for $n\to\infty$ becomes approximately

$$s_{n+1}(z) \cong 2zs_n(z) - s_{n-1}(z). \qquad (4.3)$$

The characteristic equation of this recurrence relation is just (3.49). Therefore,

$$s_n(z) \cong A(z)\alpha_1^n + B(z)\alpha_2^n, \qquad n\to\infty, \qquad (4.4)$$

where $A(z)$ and $B(z)$ are some functions of z, independent of n, and $\alpha_{1,2}$ are the roots [Eq. (3.50)] of the characteristic equation. We recall that $\alpha_1\alpha_2 = 1$, and that α_1 is taken to be the root with magnitude greater than or equal to one.

The quantity $Q_n(z)$ is given by (3.56),

$$Q_n(z) = 4\pi\alpha_2^n/(\alpha_2 - \alpha_1). \qquad (4.5)$$

Therefore,

$$\lim_{n\to\infty}\left|\frac{Q_{n+1}(z)s_{n+1}(z)}{Q_n(z)s_n(z)}\right| = \begin{cases}1, & \text{if } A(z)\neq 0,\\ |\alpha|^2 < 1, & \text{if } A(z) = 0.\end{cases} \qquad (4.6)$$

In either case, since (4.1) is assumed to hold, the infinite sum in $\Lambda(z)$ converges *uniformly* in the z-plane cut along the real interval $[-1,+1]$. Hence $\Lambda(z)$ is an analytic function in the cut plane. As (3.32) shows, $\Lambda(\infty) = 1$.

The function $A(z)$ in (4.4) is related to $\Lambda(z)$. To show this we use the relation

$$1 + \sum_{n=0}^{l}\frac{h_n}{(1+\delta_{0,n})\pi}s_n(z)Q_n(z) = \frac{Q_l(z)}{2\pi}[\alpha_2 s_l(z) - s_{l+1}(z)], \qquad (4.7)$$

which can be proved by using the recurrence relation (3.6) for $s_n(z)$ and (4.5) for $Q_n(z)$. Let l be large enough so that

$$\Lambda(z) \cong 1 + \sum_{n=0}^{l}\frac{h_n}{(1+\delta_{0,n})\pi}s_n(z)Q_n(z) = \frac{Q_l(z)}{2\pi}[\alpha_2 s_l(z) - s_{l+1}(z)]. \qquad (4.8)$$

From (4.4) it follows that

$$\alpha_2 s_l(z) - s_{l+1}(z) = A(z)\alpha_1^l[\alpha_2 - \alpha_1], \qquad l\to\infty \qquad (4.9)$$

or

$$\frac{Q_l(z)}{2\pi}[\alpha_2 s_l(z) - s_{l+1}(z)] = 2A(z), \qquad l\to\infty. \qquad (4.10)$$

Comparing (4.8) and (4.10) gives

$$A(z) = \tfrac{1}{2}\Lambda(z), \qquad (4.11)$$

so that

$$s_l(z) = \tfrac{1}{2}\Lambda(z)\alpha_1^l + B(z)\alpha_2^l, \qquad l\to\infty. \qquad (4.12)$$

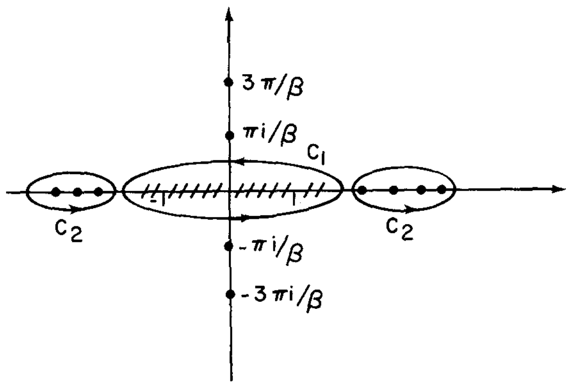

FIG. 1. Contours for evaluating integrals for relaxation of the magnetization.

This is a satisfying result, because it shows that at $z = E_i$, the discrete eigenvalues or zeros of $\Lambda(z)$, $s_l(E_i) \to 0$ as $l \to \infty$, as is required of the Fourier coefficient of an integrable function.

The M-function is defined by the generalization of (3.9):

$$M\ (\cos\theta, z) = -\sum_{n=0}^{\infty} \frac{h_n}{(1+\delta_{0,n})\pi} \cos n\theta \cdot s_n(z). \qquad (4.13)$$

This infinite sum converges for $z \in [-1, +1]$, since (4.4) and (3.50) show that

$$s_n(\cos\lambda) = A\ (\cos\lambda) e^{in\lambda} + B(\cos\lambda) e^{-in\lambda}, \qquad n \to \infty. \quad (4.14)$$

At the discrete eigenvalues $z = E_i$, the convergence is guaranteed by (4.12).

We need also to examine $M(E_i, E_i)$, which is related to the discrete normalization integral. Let $T_n(z)$ be the Tchebichef polynomial of the first kind,

$$T_n(z) = \tfrac{1}{2}[\alpha_1^n + \alpha_2^n] \xrightarrow[n\to\infty]{} \tfrac{1}{2}\alpha_1^n, \qquad z \notin [-1, +1]. \quad (4.15)$$

It follows from (4.12) and (4.15) that

$$T_n(E_i) s_n(E_i) \xrightarrow[n\to\infty]{} \tfrac{1}{2} B(E_i). \qquad (4.16)$$

Consequently the infinite sum

$$M(E_i, E_i) = -\sum_{n=0}^{\infty} \frac{h_n}{(1+\delta_{0,n})\pi} T_n(E_i) s_n(E_i) \qquad (4.17)$$

converges.

We conclude, therefore, that the eigenfunction solutions in Sec. 3 also apply to the case of a magnetic field infinite in extent and satisfying (4.1). There are a finite number of discrete modes in addition to the continuum modes. The eigenfunctions are given by the same equations (3.10), (3.18), and (3.20), where M and Λ are now defined by infinite sums. These eigenfunctions form a complete set (for even functions), and their normalization integrals are given by (3.28) and (3.29). Actually the derivation of the normalization integrals in the Appendix requires modification, since now $M(\cos\theta, z)$ does not exist for all z. However, instead of carrying out the modified derivation, we merely note that, since $\Lambda(z)$ is analytic and $M(E_i, E_i)$ converges, the final results (3.28) and (3.29) can be obtained by taking appropriate limits.

These conclusions, of course, also apply to the odd eigenfunctions.

In the relaxation problem, a difficulty is encountered because of the fact that the function $M(\cos\theta, E_\lambda)$ cannot now be extended to include complex values of its arguments. The magnetization cannot be written in the form of (3.77); instead we have

$$\langle m_n(t)\rangle = \frac{1}{2\pi i}\int_{-1}^{1} d\mu\, B(\mu)\left(\frac{I^-(t,n,\mu)I^-(-t,n,\mu)}{F^-(\mu)} + \frac{\bar{I}^-(t,n,\mu)\bar{I}^-(-t,n,\mu)}{\bar{F}^-(\mu)} - \frac{I^+(t,n,\mu)I^+(-t,n,\mu)}{F^+(\mu)} - \frac{\bar{I}^+(t,n,\mu)\bar{I}^+(-t,n,\mu)}{\bar{F}^+(\mu)}\right)$$
$$+ \sum_i B(E_i)\frac{I(t,n,E_i)I(-t,n,E_i)}{N_i} + \sum_i B(\bar{E}_i)\frac{\bar{I}(t,n,\bar{E}_i)\bar{I}(-t,n,\bar{E}_i)}{\bar{N}_i}, \qquad (4.18)$$

where, for example, $I^\pm(t, n, \mu)$ are the boundary values of the function defined in (3.67).

To study the long-time behavior of the magnetization, there seems to be no other way than a direct asymptotic evaluation of the expression (4.18). We will not present the rather lengthy calculations here, but only make one technical remark concerning the asymptotic evaluation of the integral $I^\pm(t, n, \mu)$, which is given by

$$I^\pm(t,n,\mu) = \int_{-\pi}^{\pi} \frac{d\theta}{(2\pi)^{1/2}} e^{in\theta - it\cos\theta} \frac{M(\cos\theta, \mu)}{\cos\theta - \mu \mp i\epsilon}. \qquad (4.19)$$

The factor $e^{-it\cos\theta}$ has saddle points at $\theta = 0, \pi$. For μ "close" to ± 1, we therefore have singularities "close" to the saddle points. Consequently the usual saddle-point evaluation of the integral (4.19) must be modified to take into account these singularities.

The lengthy calculations produce the expected result

$$\langle m_n(t)\rangle = O(1/t), \qquad t \to \infty. \qquad (4.20)$$

This result is expected because of the fact that in a magnetic field infinite in extent but satisfying (4.1), the spectrum of the system is essentially the same as in a magnetic field finite in extent. That is, there are a finite number of discrete eigenvalues in addition to the continuum. Consequently the relaxation behavior of the magnetization in the two cases should be the same.

5. LINEARLY INCREASING FIELD

The magnetic fields considered in the last two sections do not induce much change in the spectrum of the system: a finite number of discrete eigenvalues appear in addition to the original continuum. If the magnetic field h_n is a polynomial of order p in n, the eigenvalue equation (2.17) in general becomes a differential equation of order p, and the spectrum is radically changed. The simplest case, that of a linearly increasing field, has recently been discussed by Smith.[7]

For completeness we here briefly summarize the pertinent results.

(1) The spectrum is purely discrete and extends from $-\infty$ to $+\infty$. The eigenvalues are

$$E_l = lh, \qquad l = 0, \pm 1, \pm 2, \cdots, \qquad (5.1)$$

and the normalized eigenfunctions are

$$W_l(\theta) = (2\pi)^{-1/2} \exp[(i/h)(\sin\theta - E_l\theta)]. \qquad (5.2)$$

(2) The equilibrium magnetization under the applied field is given by

$$\langle m_n(0)\rangle = \sum_{l=-\infty}^{\infty} B(lh)[J_{l+n}(h^{-1})]^2 - \tfrac{1}{2}, \qquad (5.3)$$

where

$$B(x) = 1/(1 + e^{\beta x}) \qquad (5.4)$$

and J_n is the Bessel function of order n.

For $h > 0$, this shows that

$$\begin{aligned}\langle m_n(0)\rangle &\cong 1/2, \quad n \text{ large and positive},\\ &\cong -1/2, \quad n \text{ large and negative},\end{aligned} \qquad (5.5)$$

as would be expected physically.

At zero temperature the expression (5.3) simplifies to

$$\langle m_n(0)\rangle_{\beta\to\infty} = \tfrac{1}{2}[J_0^2(h^{-1}) + J_n^2(h^{-1})] + \sum_{l=1}^{n-1} J_l^2(h^{-1}). \qquad (5.6)$$

(3) In the relaxation problem, the time-dependent magnetization $\langle m_n(t)\rangle$ is given by

$$\langle m_n(t)\rangle = \sum_{l=-\infty}^{\infty} B(lh)\,|F_{l+n}(h^{-1},t)|^2 - \tfrac{1}{2}, \qquad (5.7)$$

where

$$F(h^{-1},t) = \int_{-\pi}^{\pi} \frac{d\theta}{\pi} e^{-it\cos\theta - (i/h)\sin\theta + il\theta}. \qquad (5.8)$$

In particular,

$$\langle m_{-n}(t)\rangle = -\langle m_n(t)\rangle, \qquad (5.9)$$

and hence the total magnetization is always zero and

$$\langle m_0(t)\rangle = 0.$$

For large t, we have

$$\langle m_n(t)\rangle \cong \frac{n}{\pi t} \sum_{l=-\infty}^{\infty} \mathcal{B}_l,\; t\to\infty, \qquad (5.10)$$

where

$$\mathcal{B}_l = B[(l-1)h] - B(lh).$$

At zero temperature this simplifies to

$$\langle m_n(t)\rangle_{\beta=\infty} \cong \frac{n}{\pi t}, \quad t\to\infty. \qquad (5.11)$$

6. QUADRATICALLY INCREASING FIELD

The case of a quadratic field leads to a well-known equation for the eigenvalue problem. With $h_n = n^2 h$, (2.14) gives

$$\begin{aligned} F(\theta) &= h \sum_{n=-\infty}^{\infty} \frac{n^2}{2\pi} e^{in\theta} \\ &= -h \frac{\partial^2}{\partial\theta} \sum_{n=-\infty}^{\infty} \frac{e^{in\theta}}{2\pi} \\ &= -h\delta''(\theta). \end{aligned} \qquad (6.1)$$

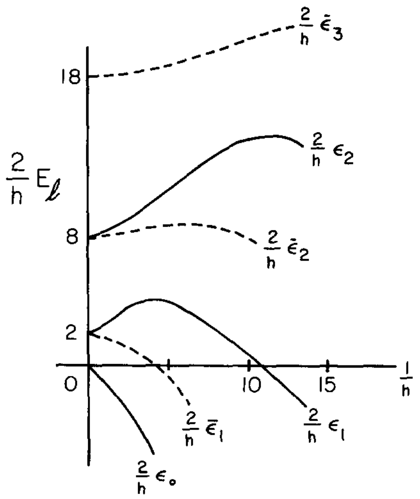

FIG. 2. Dependence of the eigenvalues on the magnetic field, taken from Ref. 6, p. 40, Fig. 8(A).

Therefore the eigenvalue equation (2.17) becomes a second-order differential equation

$$\frac{d^2 w_l(\theta)}{d\theta^2} + \left(\frac{E_l}{h} - \frac{\cos\theta}{h}\right) w_l(\theta) = 0, \qquad (6.2)$$

where again the parameter λ is replaced by l. The solutions $w_l(\theta)$ must in addition satisfy the periodic condition

$$w_l(\theta + 2\pi) = w_l(\theta). \qquad (6.3)$$

Equation (6.2) is just Mathieu's equation, which has been well investigated. Here we summarize the relevant properties of the solutions subject to the condition (6.3)[8]:

1. The spectrum is discrete and bounded from below. The eigenvalue E_l is a continuous and single-valued function of h.

2. The solutions are either even or odd in θ. Let $\left\{\begin{smallmatrix}\epsilon_l\\ \bar{\epsilon}_l\end{smallmatrix}\right\}$ be eigenvalues of the $\left\{\begin{smallmatrix}\text{even}\\ \text{odd}\end{smallmatrix}\right\}$ eigensolutions. For finite values of h, $\epsilon_l \neq \bar{\epsilon}_l$; as $h\to\infty$, both ϵ_l and $\bar{\epsilon}_l$ approach hl^2, where l is an integer, and the even and odd eigenfunctions approach multiples of $\cos l\theta$ and $\sin l\theta$, respectively. A graph of the eigenvalues as a function of h is shown in Fig. 2.

3. The eigenfunctions are in one-to-one correspondence with the trigonometric functions $\cos l\theta$ and $\sin l\theta$, and, analogous to the latter functions, form a complete set. Let $u_l(\theta)$ and $\bar{u}_l(\theta)$ denote the even and odd normalized eigenfunctions, respectively. We follow standard notations and write

$$\left.\begin{aligned} u_l(\theta) &= \pi^{-1/2}\,\mathrm{ce}_{2l}(\theta/2), \\ \bar{u}_l(\theta) &= \pi^{-1/2}\,\mathrm{se}_{2l}(\theta/2) \end{aligned}\right\}, \quad l = 0,1,2,\cdots, \qquad (6.4)$$

where $\left\{\begin{smallmatrix}\mathrm{ce}_{2l}(\theta/2)\\ \mathrm{se}_{2l}(\theta/2)\end{smallmatrix}\right\}$ is the Mathieu function which reduces to a multiple of $\left\{\begin{smallmatrix}\cos l\theta\\ \sin l\theta\end{smallmatrix}\right\}$ as $h\to\infty$. The function $\mathrm{se}_0(\theta/2)$, and hence $\bar{u}_0(\theta)$, is identically zero.

In terms of the eigenfunctions, the equilibrium magnetization becomes [(2.24) with $t=0$]

$$\langle m_n(0)\rangle = -\tfrac{1}{2} + \sum_{l=0}^{\infty}\left(\int_{-\pi}^{\pi} \frac{d\theta d\theta'}{2\pi} e^{in(\theta-\theta')} \frac{u_l(\theta)u_l(\theta')}{1+e^{\beta\bar{\epsilon}_l}} + \int_{-\pi}^{\pi} \frac{d\theta d\theta'}{2\pi} e^{in(\theta-\theta')} \frac{u_l(\theta)u_l(\theta')}{1+e^{\beta\bar{\epsilon}_l}}\right). \qquad (6.5)$$

Expanding the eigenfunctions in Fourier series (again in standard notations):

$$u_l(\theta) = \pi^{-1/2} \sum_{r=0}^{\infty} A_{2r}^{(2l)} \cos r\theta, \qquad (6.6)$$

$$\bar{u}_l(\theta) = \pi^{-1/2} \sum_{r=0}^{\infty} B_{2r}^{(2l)} \sin r\theta, \qquad (6.7)$$

leads to

$$\langle m_0(0)\rangle = -1/2 + 2\sum_{l=0}^{\infty} B(\epsilon_l)[A_0^{(2l)}]^2, \qquad (6.8)$$

$$\langle m_n(0)\rangle = -\tfrac{1}{2} + \tfrac{1}{2}\sum_{l=0}^{\infty}\left\{B(\epsilon_l)[A_{2n}^{(2l)}]^2 + B(\bar{\epsilon}_l)[B_{2n}^{(2l)}]^2\right\}, \quad n\neq 0, \qquad (6.9)$$

where $B(x)$ is defined by (5.8). The dependence on h appears through the Fourier coefficients of A's and B's.

Using the relations[6]

$$\left.\begin{array}{l} A_{2n}^{(2l)} \xrightarrow[n\to\infty]{} 0 \\ B_{2n}^{(2l)} \xrightarrow[n\to\infty]{} 0 \end{array}\right\} \quad \text{uniformly in } l, \tag{6.10}$$

one obtains the expected result:

$$\langle m_n(0)\rangle \xrightarrow[n\to\infty]{} -\tfrac{1}{2}. \tag{6.11}$$

If $h \gg 1$, all the eigenvalues are positive except ϵ_0. If in addition the temperature is taken to be absolute zero, the magnetization becomes

$$\langle m_0(0)\rangle = -\tfrac{1}{2} + 2[A_0^{(0)}]^2, \tag{6.12}$$

$$\langle m_n(0)\rangle = -\tfrac{1}{2} + \tfrac{1}{2}[A_{2n}^{(0)}]^2, \quad n \neq 0. \tag{6.13}$$

But for $h \gg 1$,[6]

$$A_0^{(0)} \cong 1/\sqrt{2},$$
$$A_{2n}^{(0)} \cong 0, \quad n \neq 0.$$

Therefore,

$$\langle m_0(0)\rangle \cong \tfrac{1}{2}, \tag{6.14}$$

$$\langle m_n(0)\rangle \cong -\tfrac{1}{2}, \quad n \neq 0. \tag{6.15}$$

All the spins are thus lined up in the negative z direction except the zeroth spin, which points in the opposite direction.

From (6.11) and the conservation of total magnetization, we expect that in the relaxation problem the magnetization at any site, $\langle m_n(t)\rangle$, will approach $-\frac{1}{2}$ as $t\to\infty$. This is easily checked. Equation (2.24) becomes in this case

$$\langle m_n(t)\rangle = -\tfrac{1}{2} + \sum_{l=0}^{\infty} \int_{-\pi}^{\pi} \frac{d\theta d\theta'}{2\pi} e^{in(\theta-\theta')-i(\cos\theta-\cos\theta')t} \times \left(\frac{u_l(\theta)u_l(\theta')}{1+e^{\beta\epsilon_l}} + \frac{\bar{u}_l(\theta)\bar{u}_l(\theta')}{1+e^{\beta\bar{\epsilon}_l}}\right). \tag{6.16}$$

Integration by parts shows that

$$\langle m_n(t)\rangle = -\tfrac{1}{2} + O(1/t), \quad t\to\infty. \tag{6.17}$$

Hence $m_n(\infty) = -\frac{1}{2}$, and all the spins are lined up in the negative z direction. We note that $\langle m_n(\infty)\rangle$ is equal to the spatial average of the initial magnetization, which has the value $-\frac{1}{2}$.

7. SPATIALLY ALTERNATING FIELD

In this final section the case of a spatially alternating magnetic field is considered. We set

$$h_n = \begin{cases} 0, & n \text{ odd}, \\ 2h > 0, & n \text{ even}, \end{cases} \tag{7.1}$$

where the factor of 2 is for later convenience. Now for an angular variable ϕ not restricted to the range $-\pi$ to π, we have the relation

$$\frac{1}{2\pi}\sum_{n=-\infty}^{\infty} e^{in\phi} = \sum_{l=-\infty}^{\infty} \delta(\phi + 2l\pi). \tag{7.2}$$

Therefore, (2.14) for F leads to

$$\begin{aligned} F(\theta-\theta') &= \frac{h}{\pi}\sum_{n\text{ even}} e^{in(\theta-\theta')} \\ &= \frac{h}{\pi}\sum_{n=-\infty}^{\infty} e^{2in(\theta-\theta')} \\ &= \frac{h}{\pi}\, 2\pi \sum_{l=-\infty}^{\infty} \delta[2(\theta-\theta') + 2l\pi] \\ &= h[\delta(\theta-\theta') + \delta(\theta-\theta'+\pi) + \delta(\theta-\theta'-\pi) \\ &\quad + \delta(\theta-\theta'+2\pi) + \delta(\theta-\theta'-2\pi)]. \end{aligned} \tag{7.3}$$

A delta function at the end-points of an integration interval is taken to be 1/2 of its usual value, i.e.,

$$\int^{\pi} d\theta f(\theta)\delta(\theta-\pi) = \tfrac{1}{2}f(\pi)$$

and

$$\int_{-\pi} d\theta f(\theta)\delta(\theta+\pi) = \tfrac{1}{2}f(-\pi). \tag{7.4}$$

As before, $\left\{\begin{smallmatrix} u_\lambda(\theta) \\ \bar{u}_\lambda(\theta)\end{smallmatrix}\right\}$ denote the $\left\{\begin{smallmatrix}\text{even}\\ \text{odd}\end{smallmatrix}\right\}$ eigenfunctions. Substitution of (7.3) into the eigenvalue equation (2.17) gives

$$\cos\theta w_\lambda(\theta) + h[w_\lambda(\theta) + \Delta(-\theta)w_\lambda(\pi+\theta) + \Delta(\theta)w_\lambda(\theta-\pi)] = E_\lambda w_\lambda(\theta). \tag{7.5}$$

where $w_\lambda(\theta)$ denotes either $u_\lambda(\theta)$ or $\bar{u}_\lambda(\theta)$, and

$$\Delta(x) = \begin{cases} 1, & x > 0, \\ 1/2, & x = 0, \\ 0, & x < 0. \end{cases}$$

We first consider the even solutions. It is convenient to consider only positive values of θ, $0 \le \theta \le \pi$. It follows from (7.4), then, that

$$\cos\theta u_\lambda(\theta) + hu_\lambda(\pi-\theta) = \epsilon_\lambda u_\lambda(\theta), \tag{7.6}$$

where we have set

$$\epsilon_\lambda = E_\lambda - h \tag{7.7}$$

and used

$$u_\lambda(-\theta) = u_\lambda(\theta). \tag{7.8}$$

Change variable $\theta \to \pi - \theta$. Equation (7.6) then takes the form

$$-\cos\theta u_\lambda(\pi-\theta) + hu_\lambda(\theta) = \epsilon_\lambda u_\lambda(\pi-\theta). \tag{7.9}$$

Eliminating $u_\lambda(\pi-\theta)$ from (7.6) and (7.9) gives

$$[\epsilon_\lambda^2 - (h^2+\cos^2\theta)]u_\lambda(\theta) = 0. \tag{7.10}$$

We set

$$\epsilon_{\lambda_\pm} = \pm(h^2+\cos^2\lambda)^{1/2}, \quad 0 \le \lambda \le \pi/2. \tag{7.11}$$

The signs $\pm$ are needed because the eigenstates are not parametrized by λ alone. Combining (7.11) and (7.10), we have

$$[\cos^2\lambda - \cos^2\theta]u_{\lambda_\pm}(\theta) = 0, \tag{7.12}$$

with the solution

$$u_{\lambda\pm}(\theta) = A_\pm(\lambda)\delta(\theta-\lambda) + B_\pm(\lambda)\delta(\theta-\pi+\lambda). \tag{7.13}$$

Substituting (7.13) back into the original Eq. (7.6) gives

$$B_\pm(\lambda) = (1/h)[\epsilon_{\lambda\pm} - \cos\lambda]A_\pm(\lambda), \tag{7.14}$$

Therefore,

$$u_{\lambda\pm}(\theta) = A_\pm(\lambda)[\delta(\theta-\lambda) + (1/h)(\epsilon_{\lambda\pm}-\cos\lambda)\delta(\theta-\pi+\lambda)]. \tag{7.15}$$

The function $A_\pm(\lambda)$ is determined by the normalization

$$\int_{-\pi}^{\pi} d\theta u_{\lambda\pm}(\theta)u_{\lambda'\pm}(\theta) = \delta(\lambda-\lambda'), \tag{7.16}$$

and is readily found to be

$$A_\pm^2(\lambda) = \{2[1 + (1/h^2)(\epsilon_{\lambda\pm}-\cos\lambda)^2]\}^{-1}. \tag{7.17}$$

Note that the spectrum, as given by (7.7) and (7.11), consists of two separate branches:

$$E_{\lambda\pm} = h \pm (h^2 + \cos^2\lambda)^{1/2}, \quad 0 \le \lambda \le \pi/2. \tag{7.18}$$

The + branch ranges continuously from $2h$ to $h + (1+h^2)^{1/2}$, while the − branch ranges from $-(1+h^2)^{1/2} + h$ to 0.

We now turn to the odd eigenfunctions. It follows from (7.5) that for positive values of θ, i.e., $0 \le \theta \le \pi$, we have

$$\cos\theta \bar{u}_\lambda(\theta) - h\bar{u}_\lambda(\pi-\theta) = \bar{\epsilon}_\lambda \bar{u}_\lambda(\theta), \tag{7.19}$$

where the energy eigenvalue is given by

$$\bar{E}_\lambda = \bar{\epsilon}_\lambda + h. \tag{7.20}$$

Equation (7.19) is the same as (7.6) for the even eigenfunctions, except for $h \to -h$. Consequently, $\bar{\epsilon}_{\lambda\pm}$ is the same as $\epsilon_{\lambda\pm}$, and

$$\bar{E}_{\lambda\pm} = -h \pm (h^2+\cos^2\lambda)^{1/2}, \quad 0 \le \lambda \le \pi/2. \tag{7.21}$$

For positive values of θ, the odd eigenfunctions $u_{\lambda\pm}(\theta)$ are given by (7.15) and (7.17) with all h replaced by $-h$. For negative values of θ we use

$$\bar{u}_{\lambda\pm}(-\theta) = -\bar{u}_{\lambda\pm}(\theta). \tag{7.22}$$

It is straightforward but somewhat lengthy to prove that $\{u_{\lambda\pm}(\theta), \bar{u}_{\lambda\pm}(\theta)\}$ form a complete set. We only quote the results here, which follow directly from using (7.15), (7.17), and the corresponding equations for the odd eigenfunctions:

$$\int_0^{\pi/2} d\lambda[u_{\lambda+}(\theta)u_{\lambda+}(\theta') + u_{\lambda-}(\theta)u_{\lambda-}(\theta') + \bar{u}_{\lambda+}(\theta)\bar{u}_{\lambda+}(\theta') + \bar{u}_{\lambda-}(\theta)\bar{u}_{\lambda-}(\theta')] = \delta(\theta-\theta'). \tag{7.23}$$

Consequently, any function $f(\theta)$ can be expanded in the eigenfunctions $\{u_{\lambda\pm}(\theta), \bar{u}_{\lambda\pm}(\theta)\}$.

The eigensolutions obtained above can now be used to study the magnetization of the system, as given by (2.24). Consider first the contribution of the even modes,

$$\begin{aligned}\langle c_n^+(t)c_n(t)\rangle_{\text{even}} &= \frac{2}{\pi}\int_0^\pi d\theta_1 \int_0^\pi d\theta_2 \cos n\theta_1 \cos n\theta_2 e^{-it(\cos\theta_1-\cos\theta_2)} \\ &\times \int_0^{\pi/2} d\lambda[B(E_{\lambda+})u_{\lambda+}(\theta_1)u_{\lambda+}(\theta_2) + B(E_{\lambda-})u_{\lambda-}(\theta_1)u_{\lambda-}(\theta_2)].\end{aligned} \tag{7.24}$$

Using the explicit form (7.15) and simplifying, we obtain

$$\begin{aligned}\langle c_n^+(t)c_n(t)\rangle_{\text{even}} &= \frac{2}{\pi}\int_0^{\pi/2} d\theta \cos^2 n\lambda \Big\{B(E_{\lambda+})A_+^2(\lambda)\Big[\frac{1}{2A_+^2(\lambda)} \\ &+ \frac{2}{h}(-1)^n(\epsilon_{\lambda+}-\cos\lambda)\cos(2t\cos\lambda)\Big] + B(E_{\lambda-})A_-^2(\lambda) \\ &\times \Big[\frac{1}{2A_-^2(\lambda)} + \frac{2}{h}(-1)^n(\epsilon_{\lambda-}-\cos\lambda)\cos(2t\cos\lambda)\Big]\Big\}.\end{aligned} \tag{7.25}$$

The contribution of the odd modes is of the same form, but with $\cos^2 n\lambda \to \sin^2 n\lambda$ in the integrand. Hence we have

$$\begin{aligned}\langle c_n^+(t)c_n(t)\rangle &= \langle c_n^+(t)c_n(t)\rangle_{\text{even}} + \langle c_n^+(t)c_n(t)\rangle_{\text{odd}} \\ &= \frac{1}{\pi}\int_0^{\pi/2} d\lambda[B(E_{\lambda+}) + B(E_{\lambda-})] \\ &+ \frac{4}{\pi h}(-1)^n \int_0^{\pi/2} d\lambda \cos(2t\cos\lambda) \\ &\times [B(E_{\lambda+})A_+^2(\lambda)(\epsilon_{\lambda+}-\cos\lambda) \\ &+ B(E_{\lambda-})A_-^2(\lambda)(\epsilon_{\lambda-}-\cos\lambda)].\end{aligned} \tag{7.26}$$

Notice that the *spatially-averaged* magnetization $\bar{m} \equiv \frac{1}{2}[\langle m_n(t)\rangle + \langle m_{n+1}(t)\rangle]$ is independent of time:

$$\bar{m} = \frac{1}{\pi}\int_0^{\pi/2} d\theta[B(E_{\lambda+}) + B(E_{\lambda-})] - \tfrac{1}{2}. \tag{7.27}$$

To obtain results of simple form, we consider the special case $h \gg 1$, so that

$$\epsilon_{\lambda\pm} \cong \pm h,$$

$$E_{\lambda+} \cong 2h, \quad E_{\lambda-} \cong 0,$$

$$A_\pm^2(\lambda) \cong \tfrac{1}{4},$$

as can be seen from (7.11), (7.19), and (7.20). Therefore,

$$[B(E_{\lambda+})A_+^2(\lambda)(\epsilon_{\lambda+}-\cos\lambda) + B(E_{\lambda-})A_-^2(\lambda)(\epsilon_{\lambda-}-\cos\lambda)] \cong -h/8. \tag{7.28}$$

It follows from (7.26) and (7.27) that

$$\begin{aligned}\langle m_n(t)\rangle &= \langle c_n^+(t)c_n(t)\rangle - \tfrac{1}{2} \\ &\cong \bar{m} - \frac{(-1)^n}{2\pi}\int_0^{\pi/2} d\lambda \cos(2t\cos\lambda).\end{aligned}$$

The integral is proportional to a Bessel function of zeroth order:

$$\int_0^{\pi/2} d\lambda \cos(2t\cos\lambda) = \frac{\pi}{2}J_0(2t). \tag{7.29}$$

Therefore,

$$\langle m_n(t)\rangle \cong \bar{m} - [(-1)^n/4]J_0(2t), \quad h \gg 1. \tag{7.30}$$

This result holds for *all* $t \ge 0$. Asymptotically,

$$\langle m_n(t)\rangle \cong \bar{m} - \frac{(-1)^n}{4}\frac{\cos(2t-\pi/4)}{(\pi t)^{1/2}}, \quad t \to \infty. \tag{7.31}$$

For general value of h, it also follows from (7.26) that asympotically

$$\langle m_n(t)\rangle = \bar{m} + (-1)^n \xi(h) \frac{\cos(2t - \pi/4)}{\sqrt{t}}, \qquad t \to \infty, \tag{7.32}$$

where the constant ξ depends on the value of h only.

8. SUMMARY

We have found the spectrum and eigenfunctions of the one-dimensional $X-Y$ model in various kinds of magnetic fields. If the magnetic field is finite in extent or, if infinite, approaches zero sufficiently rapidly so that $\sum_n h_n$ converges, then the spectrum consists of the bounded continuum (which is the spectrum in the absence of a magnetic field) plus a *finite* number of discrete eigenvalues *outside* the continuum. If the magnetic field increases linearly or quadratically with the spin sites, then the spectrum is much more radically changed and becomes completely discrete. Finally, for the case of a spatially alternating field, the spectrum consists of two separate continuum branches.

In the relaxation problem, the magnetization at any spin site relaxes to a limiting value $\langle m_n(\infty)\rangle$ with an asymptotic t^{-1} dependence in each of the above cases except that of an alternating field, where the asymptotic time dependence is $t^{-1/2}$. The fact that the initial magnetic field affects the asymptotic time dependence of the magnetization is understandable; in the extreme case of an initial uniform magnetic field, the magnetization at any spin site stays constant even after the field is switched off.

The limiting value of the magnetization, $\langle m_n(\infty)\rangle$, is nonvanishing for the case of a quadratically increasing field or an alternating field, but equal to zero for the other cases. This result is again physically expected. The total magnetization cannot change in the relaxation problem. Hence $\langle m_n(\infty)\rangle$ must be equal to the spatial average of the initial magnetization, which is precisely our result.

APPENDIX: NORMALIZATION INTEGRALS

In this appendix the two normalization integrals (3.28), (3.29) will be derived.

Consider

$$J(z,z') \equiv \int_{-\pi}^{\pi} d\theta \frac{M(\cos\theta, z)M(\cos\theta, z')}{(\cos\theta - z)(\cos\theta - z')}. \tag{A1}$$

Using (3.9) for M and the decomposition

$$\frac{1}{(\cos\theta - z)(\cos\theta - z')} = \frac{1}{z - z'}\left(\frac{1}{\cos\theta - z} - \frac{1}{\cos\theta - z'}\right),$$

we rewrite $J(z,z')$ in the form

$$J(z,z') = -\sum_{n=0}^{N} \frac{h_n}{(1+\delta_{0,n})\pi} \frac{1}{z-z'}[s_n(z')K_n(z) - s_n(z)K_n(z')], \tag{A2}$$

where

$$K_n(z) = \int_{-\pi}^{\pi} d\theta \cos n\theta \frac{M(\cos\theta, z)}{\cos\theta - z} \tag{A3}$$

$$= -\sum_{l=0}^{N} \frac{h_l s_l(z)}{(1+\delta_{0,l})\pi} \int_{-\pi}^{\pi} d\theta \frac{\cos n\theta \cos l\theta}{\cos\theta - z}. \tag{A4}$$

For definiteness, consider $n < N$. The final result is independent of this assumption.

Recalling the definition of $Q_n(z)$ [(3.14) and (3.56)], we have

$$\int_{-\pi}^{\pi} d\theta \frac{\cos n\theta \cos l\theta}{\cos\theta - z} = \begin{cases} \frac{1}{2}[Q_{n+l}(z) + Q_{n-l}(z)], & l \le n, \\ \frac{1}{2}[Q_{l+n}(z) + Q_{l-n}(z)], & l \ge n, \end{cases} \tag{A5}$$

$$= \begin{cases} Q_n(z)T_l(z), & l \le n, \\ Q_l(z)T_n(z), & l \ge n, \end{cases} \tag{A6}$$

where $T_n(z)$ is the Tchebichef polynomial of the first kind,

$$T_n(z) = 1/2[\alpha_1^n + \alpha_2^n] \tag{A7}$$

and

$$\alpha_{1,2} = z \pm \sqrt{(z^2-1)^{1/2}}. \tag{A8}$$

Therefore,

$$\begin{aligned} K_n(z) &= -\sum_{l \ge n} \frac{h_l s_l(z)Q_l(z)T_n(z)}{(1+\delta_{0,n})\pi} - \sum_{l<n} \frac{h_l s_l(z)Q_n(z)T_l(z)}{(1+\delta_{0,l})\pi} \\ &= -\sum_{l=0}^{N} \frac{h_l s_l(z)Q_l(z)T_n(z)}{(1+\delta_{0,l})\pi} + \sum_{l<n} \frac{h_l s_l(z)Q_l(z)T_n(z)}{(1+\delta_{0,l})\pi} \\ &\quad - \sum_{l<n} \frac{h_l s_l(z)Q_n(z)T_l(z)}{(1+\delta_{0,l})\pi}. \end{aligned} \tag{A9}$$

The right-hand side can be put into more compact form as follows:

$$\begin{aligned} -\sum_{l=0}^{N} \frac{h_l s_l(z)Q_l(z)T_n(z)}{(1+\delta_{0,l})\pi} &= -T_n(z)[\Lambda(z) - 1] \\ &= -T_n(z)\Lambda(z) + T_n(z), \end{aligned} \tag{A10}$$

where (3.13) has been used. Using the recurrence relation (3.6) satisfied by the coefficients $s_l(z)$, one can show that

$$T_n(z)\sum_{l=0}^{n-1} \frac{h_l s_l(z)Q_l(z)}{(1+\delta_{0,l})\pi} = T_n(z)\left(\frac{1}{2\pi}Q_n(z)s_{n-1}(z) - \frac{1}{2\pi}Q_{n-1}(z)s_n(z) - 1\right) \tag{A11}$$

and

$$-Q_n(z)\sum_{l=0}^{n-1} \frac{h_l s_l(z)T_l(z)}{(1+\delta_{0,l})\pi} = -\frac{Q_n(z)}{2\pi}[T_n(z)s_{n-1}(z) - T_{n-1}(z)s_n(z)]. \tag{A12}$$

Putting (A10), (A11), and (A12) into (A9) and simplifying, one gets

$$K_n(z) = -T_n(z)\Lambda(z) + s_n(z). \tag{A13}$$

Therefore, (A2) becomes

$$\begin{aligned} J(z,z') &= \frac{1}{z-z'}\left(\sum_{n=0}^{N} \frac{h_n}{(1+\delta_{0,n})\pi} s_n(z')T_n(z)\Lambda(z) - \sum_{n=0}^{N} \frac{h_n}{(1+\delta_{0,n})\pi} s_n(z)T_n(z')\Lambda(z')\right), \\ &= \frac{1}{z-z'}[M(z',z)\Lambda(z') - M(z,z')\Lambda(z)]. \end{aligned} \tag{A14}$$

The normalization integral for the discrete eigenfunctions is

$$N_i = \int_{-\pi}^{\pi} d\theta [u_i(\theta)]^2 = \int_{-\pi}^{\pi} d\theta \left(\frac{M(\cos\theta, E_i)}{\cos\theta - E_i}\right)^2 = J(E_i, E_i). \tag{A15}$$

In (A14), set $z' = E_i$; taking limit $z \to z'$ gives the required result:

$$N_i = J(E_i, E_i) = -M(E_i, E_i)\frac{\partial \Lambda(z)}{\partial(z)}\Bigg|_{z=E_i}, \tag{A16}$$

where $\Lambda(E_i) = 0$ has been used.

Now consider the continuum eigenfunctions. Using the explicit expressions (3.18) and (3.20), we have

$$\int_{-\pi}^{\pi} d\theta u_\lambda(\theta) u_{\lambda'}(\theta) = 2\int_{-1}^{1} \frac{d\mu}{(1-\mu^2)^{1/2}} P\frac{M(\mu,\nu)}{\mu-\nu} P\frac{M(\mu,\nu')}{\mu-\nu'} + \frac{1}{2}\frac{M(\nu',\nu)}{\nu'-\nu}[\Lambda^+(\nu') + \Lambda^-(\nu')] + \frac{1}{2}\frac{M(\nu,\nu')}{\nu-\nu'}[\Lambda^+(\nu) + \Lambda^-(\nu)] + \frac{1}{8}[\Lambda^+(\nu) + \Lambda^-(\nu)]^2 \delta(\lambda - \lambda'), \tag{A17}$$

where

$$\mu = \cos\theta, \quad \nu = \cos\lambda, \quad \nu' = \cos\lambda'.$$

In accordance with the discussion in Sec. 3.5, the Poincaré–Bertrand formula is applied by writing

$$P\frac{1}{\mu-\nu} P\frac{1}{\mu-\nu'} = \frac{1}{\nu'-\nu}\left(P\frac{1}{\mu-\nu'} - P\frac{1}{\mu-\nu}\right) + \pi^2 \delta(\mu-\nu)\delta(\mu-\nu'). \tag{A18}$$

Therefore, the first term on the right-hand side of (A17) becomes

$$2\left(\frac{1}{\nu'-\nu}\int_{-1}^{1}\frac{d\mu}{(1-\mu^2)^{1/2}} P\frac{M(\mu,\nu)M(\mu,\nu')}{\mu-\nu'} - \frac{1}{\nu'-\nu}\int_{-1}^{1}\frac{d\mu}{(1-\mu^2)^{1/2}} P\frac{M(\mu,\nu)M(\mu,\nu')}{\mu-\nu}\right) + 2\frac{\pi^2 M^2(\nu,\nu)}{(1-\nu^2)^{1/2}}\delta(\nu-\nu'). \tag{A19}$$

The expression $[\cdots]$ in (A19) can be evaluated from (A14) by taking limits $z \to \nu \pm i\epsilon$, $z' \to \nu' \pm i\epsilon$,

$$2\left(\frac{1}{\nu'-\nu}\int_{-1}^{1}\frac{d\mu}{(1-\mu^2)^{1/2}} P\frac{M(\mu,\nu)M(\mu,\nu')}{\mu-\nu'} - \frac{1}{\nu'-\nu}\int_{-1}^{1}\frac{d\mu}{(1-\mu^2)^{1/2}} P\frac{M(\mu,\nu)M(\mu,\nu')}{\mu-\nu}\right) = \frac{1}{2}\frac{M(\nu,\nu')}{\nu'-\nu}[\Lambda^+(\nu) + \Lambda^-(\nu)] - \frac{1}{2}\frac{M(\nu',\nu)}{\nu'-\nu}[\Lambda^+(\nu') + \Lambda^-(\nu')]. \tag{A20}$$

Putting (A19), (A20) into (A17) and simplifying, one arrives at the final result:

$$\int_{-\pi}^{\pi} d\theta u_\lambda(\theta) u_{\lambda'}(\theta) = \frac{\Lambda^+(\nu)\Lambda^-(\nu)}{2}\delta(\lambda-\lambda'). \tag{A21}$$

Hence (3.29) is derived.

---

*This work was supported by the Air Force Office of Scientific Research, Grant 722187.

[1]E. H. Lieb, T. D. Schultz, and C. D. Mattis, Ann. Phys. (N.Y.) **16**, 941 (1961).

[2]B. McCoy, Phys. Rev. **173**, 531 (1968); S. Katsura, Phys. Rev. **127**, 1508 (1962); Th. Niemejer, Physica (Utr.) **36**, 377 (1967); Physica (Utr.) **39**, 313 (1968).

[3]E. Barouch, B. McCoy, and Max Dresden, Phys. Rev. A **2**, 1075 (1970).

[4]D. Abraham, E. Barouch, G. Gallavotti, and A. Martin-Löff, Studies Appl. Math. **50**, 121 (1971).

[5]K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960); J. R. Mika, Nuc. Sci. and Eng. **11**, 415 (1961); K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).

[6]Strictly speaking, $f(\theta)$ has to satisfy a Hölder condition, but this is a rather weak restriction, satisfied by the functions we deal with. See Ref. 5.

[7]E. R. Smith, Physica (Utr.) **53**, 289 (1971).

[8]N. W. McLachlan, *Theory and Application of Mathieu Functions* (Oxford U. P., London 1947).

Theory of nonlocal electromagnetic elastic solids*

A. Cemal Eringen

Princeton University, Princeton, New Jersey 08540
(Received 23 October 1972)

A continuum theory of nonlocal electromagnetic elastic solids is proposed. Nonlocal and local balance laws and jump conditions are obtained. Through the use of an extension of Clausius–Duhem inequality, encompassing nonlocal effects, E–M momentum and constitutive equations are derived and restricted.

1. INTRODUCTION

It is well known that the Maxwell theory of electromagnetism (classical theory) predicts a constant wave velocity and consequently a constant refractive index in an isotropic nondissipative medium. While for some gases in moderate pressures this is nearly the case, for all other substances experiments indicate that the refractive index depends on the frequency and consequently *dispersion* is the rule rather than exception. In fact, starting even with radio waves, in invisible and ultraviolet regions there is no agreement whatsoever between the measured refractive index and the calculations based on the classical theory. Moreover, the classical theory possess no mechanism for the *absorption* phenomena that occurs in the neighborhood of certain critical frequencies. To incorporate these effects either supplementary excursions are made to molecular and atomic theories to remedy partially the classical theory, cf. Born and Wolf (Ref. 1, Sec. 2. 3. 4) or quantum mechanical approaches are used, cf. Ref. 2.

For magnetic materials the existence of magnetic domains, instabilities, and spin waves cannot be explained through the classical formalism. Extensive published work in this area, in one form or another, makes use of ideas of inner structure and domains that exist in material either in the form of multipoles or atomic structures, cf. Brown[3], Tiersten[4], Eringen and Maugin[5] for continuum theories and Kittel[6], Van Vleck[7], Bloch[8] and Heisenberg[9] for quantum mechanical approaches.

Recently,[10,11] we have shown that by means of the nonlocal theories of continua one can take into account the effects of distant atomic, molecular, and granular interactions thus accounting for the complete dispersion curve of elastic waves in lattice structures. The present paper's intent is the construction of the corresponding nonlocal theory for the elastic materials subject to electromagnetic interactions. While any theory involving electromagnetic fields should be based on relativistic considerations, it is possible to construct a rational theory of electromagnetism on nonrelativistic grounds For simplicity in the physics of the matter here we avoid relativity, however, borrowing some of the critical results valid for small material velocities, v, as compared to speed of light in vacuum, $(v^2/c^2 \ll 1)$.

In Sec. 2, starting with the global electromechanical balance laws for the entire body, we obtain the *equivalent* local laws. In Sec. 3 we formulate the second law of thermodynamics valid for the entire body. These equations, as against the classical field theories, contain *nonlocal residuals* that account for the distant atomic and molecular interactions. The integrals of the nonlocal residuals over the entire body vanish. The determination of the nonlocal residuals is an integral part of the nonlocal theory.

In Sec. 4 a discussion is presented for the electromagnetic momentum and energy and various classical proposals are examined. Section 5 is devoted to the determination of the constitutive equations of local E-M elastic solids. The constitutive equations of the nonlocal E-M elastic solids are obtained and restricted through the nonlocal entropy inequality. The invariance requirements are studied. The theory is exact and valid for fields and deformations of arbitrary magnitude. We postpone the discussion of the linear theory and its application to electromagnetic waves to a future paper.

2. BALANCE LAWS

The balance laws of continuum theory of electromechanical media have been formulated both from the nonrelativistic and relativistic points of view, cf. Truesdell and Toupin[12], Dixon and Eringen[13], Grot and Eringen[14]. The integral form of these laws are:

Conservation of Mass:

$$\frac{d}{dt}\int_{\mathcal{V}-\sigma} \rho\, dv = 0. \tag{2.1}$$

Balance of Momentum:

$$\frac{d}{dt}\int_{\mathcal{V}-\sigma} \rho(\mathbf{v}+\mathbf{g}) - \oint_{\mathcal{S}-\sigma} \mathbf{t}^k da_k - \oint_{\mathcal{V}-\sigma} \rho \mathbf{f} dv = 0. \tag{2.2}$$

Balance of Moment of Momentum:

$$\frac{d}{dt}\int_{\mathcal{V}-\sigma} \rho\mathbf{x}\times(\mathbf{v}+\mathbf{g})dv - \oint_{\mathcal{S}-\sigma} \mathbf{x}\times\mathbf{t}^k da_k - \int_{\mathcal{V}-\sigma} \rho\mathbf{x}\times\mathbf{f}\,dv = 0. \tag{2.3}$$

Conservation of Energy:

$$\frac{d}{dt}\int_{\mathcal{V}-\sigma} \rho(\epsilon+\tfrac{1}{2}v^2) - \oint_{\mathcal{S}-\sigma} (\mathbf{t}^k\cdot\mathbf{v}+q^k)da_k - \int_{\mathcal{V}-\sigma} \rho(\mathbf{f}\cdot\mathbf{v}+h)dv = 0. \tag{2.4}$$

Faraday's Law:

$$\int_{\mathcal{C}-\gamma} \boldsymbol{\mathcal{E}}\cdot d\mathbf{x} + \frac{1}{c}\frac{d}{dt}\int_{\mathcal{S}-\gamma} \mathbf{B}\cdot d\mathbf{a} = 0. \tag{2.5}$$

Ampère's Law (as modified by Maxwell):

$$\int_{\mathcal{C}-\gamma} \boldsymbol{\mathcal{H}}\cdot d\mathbf{x} - \frac{1}{c}\frac{d}{dt}\int_{\mathcal{S}-\gamma} \mathbf{D}\cdot d\mathbf{a} - \frac{1}{c}\int_{\mathcal{S}-\gamma} \boldsymbol{\mathcal{J}}\cdot d\mathbf{a} = 0. \tag{2.6}$$

Gauss' Law:

$$\oint_{\mathcal{S}-\sigma} \mathbf{D}\cdot d\mathbf{a} - \int_{\mathcal{V}-\sigma} q\, dv = 0. \tag{2.7}$$

Conservation of Magnetic Flux:

$$\oint_{\mathcal{S}-\sigma} \mathbf{B}\cdot d\mathbf{a} = 0. \tag{2.8}$$

Conservation of Charge:

$$\frac{d}{dt}\int_{\mathcal{V}-\sigma} q\,dv + \int_{\mathcal{S}-\sigma} \mathcal{J}\cdot d\mathbf{a} = 0. \tag{2.9}$$

Here:

ρ ≡ mass density,
$\mathbf{g}$ ≡ E-M momentum density,
$\mathbf{t}^k$ ≡ stress vectors,
q^k ≡ total energy flux,
$\mathbf{D}$ ≡ dielectric displacement vector,
q ≡ free charge density,
$\mathbf{v} = \dot{\mathbf{x}}$ ≡ velocity field,
$\mathbf{f}$ ≡ total body force density,
ϵ ≡ total internal energy density,
h ≡ energy supply density,
$\mathbf{B}$ ≡ magnetic flux,
c ≡ speed of light in vacuum,

and

$$\mathcal{E} \equiv \mathbf{E} + \frac{1}{c}\mathbf{v}\times\mathbf{B}, \tag{2.10a}$$

$$\mathcal{H} \equiv \mathbf{H} - \frac{1}{c}\mathbf{v}\times\mathbf{D}, \tag{2.10b}$$

$$\mathcal{J} \equiv \mathbf{J} - q\mathbf{v}, \tag{2.10c}$$

in which $\mathbf{E}$ ≡ electric field, $\mathbf{H}$ ≡ magnetic field, $\mathbf{J}$ ≡ free current.

The balance laws (2.1) to (2.4) and (2.7) to (2.9) are expressed over a material volume $\mathcal{V}$ (enclosed by a surface $\mathcal{S}$) which is being swept by a discontinuity surface σ having velocity $\mathbf{u}$. The E-M laws are expressed in Heaviside-Lorentz units. Integrals in (2.5) and (2.6) are over an open material surface, $\mathcal{S}$, (enclosed within a curve $\mathcal{C}$), which is being swept by a discontinuity curve γ having velocity $\mathbf{u}$. The regions of integrations $\mathcal{V}-\sigma$, $\mathcal{S}-\sigma$, and $\mathcal{C}-\gamma$ are over all points of $\mathcal{V}$, $\mathcal{S}$ and $\mathcal{C}$ except those that are contained on σ and γ.

The *motion* carries any material point $\mathbf{X}$ in the undeformed body B to a spatial place $\mathbf{x}$ in the deformed body $\mathcal{B}$ at time t. Thus, the motion is expressed by

$$\mathbf{x} = \mathbf{x}(\mathbf{X}, t).$$

We refer both $\mathbf{x}$ and $\mathbf{X}$ to a rectangular coordinate system so that in component notation

$$x^k = x^k(X^K, t), \quad k, K = 1, 2, 3, \tag{2.11}$$

which is assumed to have unique inverse

$$X^K = X^K(x^k, t) \tag{2.12}$$

in B, at all times, except possibly some singular surfaces, lines and points. Thus, we assume

$$\det x^k{}_{,K} > 0, \quad \text{in } \mathcal{V}-\sigma. \tag{2.13}$$

All vector and tensor fields in B (and its image $\mathcal{B}$ with volume $\mathcal{V}$ and surface $\mathcal{S}$) are referred to material coordinates X^K (and spatial coordinates x^k). For future convenience, we also introduce polarization vector $\mathbf{P}$ and magnetization vector $\mathbf{M}$ by

$$\mathbf{D} = \mathbf{E} + \mathbf{P}, \quad \mathbf{B} = \mathbf{H} + \mathbf{M} \tag{2.14}$$

The integral balance laws (2.1) to (2.9) have either of the following two forms:

$$\frac{d}{dt}\int_{\mathcal{V}-\sigma}\phi\,dv - \int_{\mathcal{S}-\sigma}\tau^k da_k - \int_{\mathcal{V}-\sigma} g\,dv = 0, \tag{2.15}$$

$$\frac{d}{dt}\int_{\mathcal{S}-\gamma}\mathbf{q}\cdot d\mathbf{a} - \oint_{\mathcal{C}-\gamma}\mathbf{h}\cdot d\mathbf{x} - \int_{\mathcal{S}-\gamma}\mathbf{r}\cdot d\mathbf{a} = 0 \tag{2.16}$$

Equations (2.15) and (2.16) may be transformed by the generalized Green-Gauss theorem and Stokes' theorem, respectively, into the following forms (see Ref. 15, p. 77):

$$\int_{\mathcal{V}-\sigma}\left[\frac{\partial\phi}{\partial t} + div(\phi\mathbf{v}) - \tau^k{}_{,k} - g\right]dv + \int_\sigma[\phi(v^k - u^k) - \tau^k]n_k da = 0, \tag{2.17}$$

$$\int_{\mathcal{S}-\gamma}(\overset{*}{q}\text{-curl }\mathbf{h} - \mathbf{r})\cdot d\mathbf{a} + \int_\sigma[\mathbf{q}\mathbf{x}(\mathbf{v}-\mathbf{u}) - \mathbf{h}]\cdot\mathbf{k}\,ds = 0, \tag{2.18}$$

where k is the unit tangent vector on γ. A bold face bracket indicates the jump across σ in (2.17) and across γ in (2.18) and

$$\overset{*}{\mathbf{q}} \equiv \frac{\partial\mathbf{q}}{\partial t} + (div\,\mathbf{q})\mathbf{v} + \text{curl}\,(\mathbf{q}\times\mathbf{v}). \tag{2.19}$$

Employing the form (2.17) and (2.18), we transform (2.1), (2.2)-(2.9) to

$$\int_{\mathcal{V}-\sigma}\left[\frac{\partial\rho}{\partial t} + (\rho v^k)_{,k}\right]dv + \int_\sigma[\rho(v^k - u^k)]n_k da = 0, \tag{2.20}$$

$$\int_{\mathcal{V}-\sigma}[\rho(\dot{\mathbf{v}} + \dot{\mathbf{g}}) - \mathbf{t}^k{}_{,k} - \rho\mathbf{f}]dv + \int_{\mathcal{V}-\sigma}(\mathbf{v}+\mathbf{g})\left[\frac{\partial\rho}{\partial t} + (\rho v^k)_{,k}\right]dv + \int_\sigma[\mathbf{v}+\mathbf{g})(v^k - u^k) - \mathbf{t}^k]n_k da = 0, \tag{2.21}$$

$$\begin{aligned}&\int_{\mathcal{V}-\sigma}(\rho\mathbf{v}\times\mathbf{g} - \mathbf{x}_{,k}\times\mathbf{t}^k)dv + \int_{\mathcal{V}-\sigma}\mathbf{x}\times(\mathbf{v}+\mathbf{g})\left[\frac{\partial\rho}{\partial t} + (\rho v^k)_{,k}\right]dv\\ &- \int_{\mathcal{V}-\sigma}\mathbf{x}\times(\mathbf{t}^k{}_{,k} + \rho\mathbf{f} - \rho\dot{\mathbf{v}} - \rho\dot{\mathbf{g}})dv\\ &+ \int_\sigma[\rho\mathbf{x}\times(\mathbf{v}+\mathbf{g})(v^k + u^k) - \rho\mathbf{x}\times\mathbf{t}^k]n_k da = 0,\end{aligned} \tag{2.22}$$

$$\begin{aligned}&\int_{\mathcal{V}-\sigma}(\rho\dot{\epsilon} - \mathbf{t}^k\cdot\mathbf{v}_{,k} - q^k{}_{,k} - \rho\dot{\mathbf{g}}\cdot\mathbf{v} - \rho h)dv + \int_{\mathcal{V}-\sigma}(\epsilon + \tfrac{1}{2}v^2)\left[\frac{\partial\rho}{\partial t} + (\rho v^k)_{,k}\right]dv\\ &- \int_{\mathcal{V}-\sigma}\mathbf{v}\cdot(\mathbf{t}^k{}_{,k} + \rho\mathbf{f} - \rho\dot{\mathbf{v}} - \rho\dot{\mathbf{g}})dv\\ &+ \int_\sigma[\rho(\epsilon + \tfrac{1}{2}v^2)(v^k - u^k) - \mathbf{t}^k\cdot\mathbf{v} - q^k]n_k da = 0,\end{aligned} \tag{2.23}$$

$$\int_{\mathcal{S}-\sigma}(\nabla\times\mathcal{E} + \frac{1}{c}\overset{*}{\mathbf{B}})\cdot d\mathbf{a} + \int_\gamma\left[\mathbf{E} + \frac{1}{c}\mathbf{u}\times\mathbf{B}\right]\cdot\mathbf{k}\,ds = 0, \tag{2.24}$$

$$\int_{\mathcal{S}-\gamma}\left(\nabla\times\mathcal{H} - \frac{1}{c}\overset{*}{\mathbf{D}} - \frac{1}{c}\mathcal{J}\right)\cdot d\mathbf{a} + \int_\gamma[\mathbf{H} - \frac{1}{c}\mathbf{u}\times\mathbf{D}]\cdot\mathbf{k}\,ds = 0, \tag{2.25}$$

$$\int_{\mathcal{S}-\sigma} (\nabla\cdot\mathbf{D}-q)dv + \int_\sigma [\mathbf{D}]\cdot\mathbf{n}da = 0, \tag{2.26}$$

$$\int_{\mathcal{S}-\sigma} \nabla\cdot\mathbf{B}dv + \int_\sigma [\mathbf{B}]\cdot\mathbf{n}da = 0, \tag{2.27}$$

$$\int_{\mathcal{V}-\sigma}\left[\nabla\cdot\mathcal{J} + \frac{\partial q}{\partial t} + \nabla\cdot(q\mathbf{v})\right]dv + \int_\sigma [\mathcal{J}-q\mathbf{u}]\cdot\mathbf{n}da = 0. \tag{2.28}$$

The local balance laws of classical theory are obtained by positing that these integrals be valid for every part of the body. In this case then, the integrands of volume and surface integrals in the foregoing equations are set equal to zero.

When the long range intermolecular forces are strong this is not permissible. However, we can localize these expressions by introducing the *localization residuals* which account for the effects of nonlocal fields. Thus, (2.20) to (2.28) are equivalent to:

Mass:

$$\frac{\partial\rho}{\partial t} + (\rho v^k)_{,k} = \hat{\rho}, \quad \text{in } \mathcal{V}-\sigma,$$
$$[\rho(v^k-u^k)-\hat{\rho}^k]n_k = 0, \quad \text{on } \sigma. \tag{2.29}$$

Momentum:

$$\mathbf{t}^k{}_{,k} + \rho(\mathbf{f}-\dot{\mathbf{v}}-\dot{\mathbf{g}}) = \hat{\rho}(\mathbf{v}+\mathbf{g}) - \rho\hat{\mathbf{f}}, \quad \text{in } \mathcal{V}-\sigma,$$
$$[\mathbf{t}^k - \rho(\mathbf{v}+\mathbf{g})(v^k-u^k) + \hat{\mathbf{f}}^k]n_k = 0, \quad \text{on } \sigma, \tag{2.30}$$

Moment of Momentum:

$$\mathbf{x}_{,k}\times\mathbf{t}^k - \rho\mathbf{v}\times\mathbf{g} = \rho\mathbf{x}\times\hat{\mathbf{f}} - \rho\hat{\boldsymbol{\ell}}, \quad \text{in } \mathcal{V}-\sigma, \tag{2.31a}$$

$$[\rho\mathbf{x}\times[t^k-(\mathbf{v}+\mathbf{g})(v^k-u^k)] + \hat{\boldsymbol{\ell}}^k] = 0, \quad \text{on } \sigma. \tag{2.31b}$$

Energy:

$$\rho\dot{\epsilon} - \rho\dot{\mathbf{g}}\cdot\mathbf{v} - t^k\cdot\mathbf{v}_{,k} - q^k{}_{,k} - \rho h$$
$$= \rho\hat{h} - \rho\mathbf{v}\cdot\hat{\mathbf{f}} - \hat{\rho}[\epsilon - (v^2/2) - \mathbf{v}\cdot\mathbf{g}], \quad \text{in } \mathcal{V}-\sigma, \tag{2.32a}$$

$$[\mathbf{t}^k\cdot\mathbf{v} + q^k - \rho(\epsilon+\tfrac{1}{2}v^2)(v^k-u^k) + \hat{h}^k]n_k = 0, \quad \text{on } \sigma. \tag{2.32b}$$

Faraday's Law:

$$\nabla\times\mathcal{E} + \frac{1}{c}\overset{*}{\mathbf{B}} = \frac{1}{c}\hat{\mathbf{b}}, \quad \text{in } \mathcal{V}-\sigma, \tag{2.33a}$$

$$\left[\mathbf{E} + \frac{1}{c}\mathbf{u}\times\mathbf{B} + \hat{\mathbf{E}}\right]\times\mathbf{n} = 0, \quad \text{on } \sigma. \tag{2.33b}$$

Ampere's Law:

$$\nabla\times\mathcal{H} - \frac{1}{c}\overset{*}{\mathbf{D}} - \frac{1}{c}\mathcal{J} = \frac{1}{c}\hat{\mathcal{J}}, \quad \text{in } \mathcal{V}-\sigma, \tag{2.34a}$$

$$\left[\mathbf{H} - \frac{1}{c}\mathbf{u}\times\mathbf{D} + \hat{\mathbf{H}}\right]\times\mathbf{n} = 0, \quad \text{on } \sigma. \tag{2.34b}$$

Gauss' Law:

$$\nabla\cdot\mathbf{D} - q = \hat{q}, \quad \text{in } \mathcal{V}-\sigma,$$
$$[\mathbf{D}+\hat{\mathbf{D}}]\cdot\mathbf{n} = 0, \quad \text{on } \sigma, \tag{2.35}$$

Magnetic Flux:

$$\nabla\cdot\mathbf{B} = \hat{m}, \quad \text{in } \mathcal{V}-\sigma, \tag{2.36a}$$

$$[\mathbf{B}+\hat{\mathbf{B}}]\cdot\mathbf{n} = 0, \quad \text{on } \sigma. \tag{2.36b}$$

Conservation of Charge:

$$\nabla\cdot\mathcal{J} + \frac{\partial\rho}{\partial t} + \nabla\cdot(q\mathbf{v}) = \hat{\sigma}, \quad \text{in } \mathcal{V}-\sigma, \tag{2.37a}$$

$$[\mathcal{J}+\Sigma]\cdot\mathbf{n} = 0. \tag{2.37b}$$

If we further introduce the nonlocal current, $\hat{\mathbf{J}}$, by

$$\hat{\mathcal{J}} = \hat{\mathbf{J}} - \hat{q}\mathbf{v}, \tag{2.38}$$

then upon taking divergence of (2.34a), and comparing the result with (2.37) we obtain

$$\hat{\sigma} = -\nabla\cdot\hat{\mathbf{J}} - \frac{\partial\hat{q}}{\partial t}, \tag{2.39}$$

so that (2.37a) is equivalent to

$$\nabla\cdot(\mathbf{J}+\hat{\mathbf{J}}) + \frac{\partial}{\partial t}(q+\hat{q}) = 0. \tag{2.40}$$

This is the expression of the law of conservation of nonlocal charge. By taking divergence of (2.33a) and comparing the result with (2.36a) we obtain

$$\frac{\partial\hat{m}}{\partial t} + \nabla\cdot(\hat{m}\mathbf{v}) = \nabla\cdot\hat{\mathbf{b}}. \tag{2.41}$$

This is the expression of the law of balance of *nonlocal pole strength*. It is clear that the nature and existence of $\hat{m}$ is tied with that of $\hat{b}$.

The foregoing equations contain the localization residuals, $\hat{\rho}, \hat{\rho}^k, \hat{\mathbf{f}}, \hat{\mathbf{f}}^k, \hat{\boldsymbol{\ell}}, \hat{\boldsymbol{\ell}}^k, \hat{h}, \hat{h}^k, \hat{\mathbf{b}}, \hat{\mathbf{E}}, \hat{\mathcal{J}}, \hat{\mathbf{H}}, \hat{q}, \hat{\mathbf{D}}, \hat{m}, \hat{\mathbf{B}}, \hat{\sigma}$ and $\hat{\Sigma}$ which are introduced to take into account the effects of fields at all other points of the body on the point at which local balance laws (2.29) to (2.37) are written. Integrals of these residuals over the manifolds of their definitions vanish, i.e.,

$$\int_{\mathcal{V}-\sigma}(\hat{\rho}, \rho\hat{\mathbf{f}}, \rho\hat{\boldsymbol{\ell}}, \rho\hat{h}, \hat{q}, \hat{m}, \hat{\sigma})dv = 0,$$
$$\int_\sigma(\hat{\rho}^k, \hat{\mathbf{f}}^k, \hat{\boldsymbol{\ell}}^k, \hat{h}^k, \hat{D}^k, \hat{B}^k)n_k da = 0, \tag{2.42}$$
$$\int_{\mathcal{S}-\sigma}(\hat{\mathbf{b}}, \hat{\mathcal{J}}, \Sigma)\cdot d\mathbf{a} = 0, \qquad \int_\gamma(\hat{\mathbf{E}}, \hat{\mathbf{H}})\cdot\mathbf{k}ds = 0.$$

Physical significance of these residuals is clear from the equations in which they first occur. For example, $\hat{\rho}$ represents the mass production or destruction at a point $\mathbf{x}$ due to effect of all points of the body. The existence of $\hat{\rho}$ is well known to us from the theories of chemically reacting media. For a nonreacting continuum $\hat{\rho}$ may still exist (even though small) due to the mass carried by the mobile electrons and neutral atoms. Similarly $\hat{\mathbf{f}}$ is the nonlocal body force at $\mathbf{x}$ due to long range intermolecular attractions. The nonlocal residuals are grouped as volume, surface and line residuals as apparent from their domains of integrations shown in (2.42). The mechanical volume residuals $(\hat{\rho}, \hat{\mathbf{f}}, \hat{\boldsymbol{\ell}}, \hat{h})$ and surface residuals $(\hat{\rho}^k, \hat{\mathbf{f}}^k, \hat{\boldsymbol{\ell}}^k, \hat{h}^k)$ were introduced and discussed by us previously, cf. Erigen and Edelen[10]. Electromagnetic residuals $\hat{q}, \hat{m}, \hat{\sigma}, \hat{D}^k, \hat{\mathbf{B}}^K, \hat{\mathbf{b}}, \hat{\mathcal{J}}, \hat{\Sigma}, \hat{\mathbf{E}}$ and $\hat{\mathbf{H}}$ are new and require a few remarks. We recognize $\hat{q}$ as the induced nonlocal charge at a point $\mathbf{x}$ of the body due to charges at all other points. Similarly $\hat{m}$ (if it exists) may be envisioned as the induced magnetic pole strength at $\mathbf{x}$ by the rest of the body. The nonlocal residual $\hat{\mathcal{J}}$ represents the contribution from all other points of the body to the currents at $\mathbf{x}$, and $\hat{\mathbf{b}}$ to the magnetic flux. Corresponding to these surface residuals are indicated by $(\hat{\mathbf{D}}, \hat{\mathbf{B}})$ and line residuals, by $\hat{\mathbf{E}}$ and $\hat{\mathbf{H}}$.

When all residuals vanish at all points of the body the balance laws (2.29) to (2.37) revert back to the local balance laws of electromechanical continua. Thus, the distinguishing feature of nonlocal continuum theories is the presence of the localization residuals. Different nonlocal theories then arise from the different character of the localization residuals. Determination of these residuals is an integral part of the nonlocal continuum physics. The nature and constitution of the body must, therefore, be examined carefully in the light of its response to external effects.

It is well known that in the transmission of high frequency mechanical signals distant intermolecular forces play an important role giving rise to dispersive effects that are not accounted for in the classical field theories. In the E-M theory, importance of action at distance is well accepted. In optical range and beyond (ultraviolet region) dispersive effects become important which cannot be accounted for by use of the Maxwell theory of electromagnetism (Born and Wolf, Ref. 1, Sec. 2.3.4). In this range, optical resonance and absorption play dominant roles. While some of these effects, e.g., spin waves, can be treated by means of polar theories of electromagnetism (e.g., theories involving magnetization gradients), in the region of short wave lengths these theories also fail. Moreover, gradient theories constitute extensions of the classical theory and they are included as special cases in the present theory. We believe that a large class of physical phenomena beyond these theories can be brought within the domain of continuum physics by systematic consideration of nonlocal effects.

3. SECOND LAW OF THERMODYNAMICS

The second law of thermodynamics and ensuing different forms of the Clausius-Duhem inequality were formulated by Grot and Erigen[14] in the context of the special theory of relativity. An appropriate form for nonrelativistic case of the second law[16] is,

$$\frac{d}{dt}\int_{\mathcal{V}-\sigma}\rho\eta dv-\oint_{\mathcal{S}-\sigma}\frac{1}{\theta}(\mathbf{q}-c\mathcal{E}\times\mathcal{H})\cdot d\mathbf{a}-\int_{\mathcal{V}-\sigma}\frac{1}{\theta}(\rho h+\mathcal{J}_0\cdot\mathcal{E})dv\geq 0. \quad (3.1)$$

Here η is the entropy density, $\theta > 0$ is the absolute temperature, $c\mathcal{E}\times\mathcal{H}$ is the poynting vector and $\mathcal{J}_0$ is the external current source.

The localization of (3.1) with the same scheme of Sec. 2, gives

$$\rho\dot{\eta}-\nabla\cdot\left[\frac{1}{\theta}(\mathbf{q}-c\mathcal{E}\times\mathcal{H})\right]-\frac{1}{\theta}(\rho h+\mathcal{J}_0\cdot\mathcal{E}]+\hat{\rho}\eta-\frac{1}{\theta}\rho\hat{s}\geq 0,\quad \text{in } \mathcal{V}-\sigma, \quad (3.2a)$$

$$\left[\rho\eta(v^k-u^k)-\frac{1}{\theta}(\mathbf{q}-c\mathcal{E}\times\mathcal{H})^k-\hat{s}^k\right]n_k\geq 0,\quad \text{on } \sigma, \quad (3.2b)$$

where $\hat{s}$ and $\hat{s}^k$ are entropy localization residuals subject to

$$\int_{\mathcal{V}-\sigma}\frac{1}{\theta}\rho\hat{s}dv=0,\qquad \int_\sigma \hat{s}^k da_k=0. \quad (3.3)$$

Using (2.33a) and (2.34) we write

$$c\nabla\cdot(\mathcal{E}\times\mathcal{H})=-\mathcal{H}\cdot\overset{*}{\mathbf{B}}-\mathcal{E}\cdot\overset{*}{\mathbf{D}}-\mathcal{E}\cdot\mathcal{J}+\mathcal{H}\cdot\hat{\mathbf{b}}-\mathcal{E}\cdot\hat{\mathcal{J}}. \quad (3.4)$$

Substituting this into (3.2a) and eliminating h between (3.2a) and (2.32a), we obtain

$$\rho\left(\dot{\eta}-\frac{\dot{\epsilon}}{\theta}+\frac{\mathbf{v}\cdot\mathbf{g}}{\theta}\right)+\frac{1}{\theta}\mathbf{t}^k\cdot\mathbf{v}_{,k}-\frac{1}{\theta}(\mathbf{q}-c\mathcal{E}\times\mathcal{H})\cdot\nabla\frac{1}{\theta}+\frac{1}{\theta}(\mathcal{E}\cdot\overset{*}{\mathbf{D}}+\mathcal{H}\cdot\overset{*}{\mathbf{B}})$$
$$+\frac{1}{\theta}\mathcal{E}\cdot(\mathcal{J}-\mathcal{J}_0)-\frac{\rho}{\theta}\hat{\mathbf{f}}\cdot\mathbf{v}+\frac{\hat{\rho}}{\theta}\left(\theta\eta-\epsilon+\frac{v^2}{2}+\mathbf{v}\cdot\mathbf{g}\right)$$
$$+\frac{1}{\theta}(\mathcal{E}\cdot\hat{\mathcal{J}}+\mathcal{H}\cdot\hat{\mathbf{b}})+\frac{\rho}{\theta}(\hat{h}-\hat{s})\geq 0,\quad \text{in } \mathcal{V}-\sigma. \quad (3.5)$$

This is the Clausius-Duhem inequality. Inequality (3.5) may be written in other equivalent forms appropriate for bodies with different constitutional properties (e.g., E-M elastic solids, fluids, etc.).

We not postulate that in all thermodynamically admissible processes, the Clausius-Duhem inequality (3.5) *in* $\mathcal{V}-\sigma$ *and* (3.2b) *on* σ *be satisfied. Moreover, the constitution of body must be such as not to violate* (3.5) *for all possible independent states of the body.*

It is important to draw attention to the fact that so far we have not chosen any form for the momentum density g and that the total internal energy includes the kinetic energy arising from motion of the body in E-M fields.

4. ELECTROMAGNETIC MOMENTUM AND ENERGY

For the Maxwell-Lorentz theory of electromagnetism various forms of electromagnetic momentum, energy, and stress tensor have been proposed. The origins of the ideas for these concepts rely heavily on the vacuum or rigid body electromagnetism and therefore are controversial as far as deformable bodies are concerned (cf. remarks made by Dunkin and Eringen[17] and Truesdell and Toupin[12]). A systematic approach was presented by Dixon and Eringen[13] for nonrelativistic deformable media, by Grot and Eringen[14] for relativistic theory and by DeGroot and Suttorp[18] on statistical mechanical basis. While all these different approaches appear to agree in the final forms of these quantities, they require certain primitive postulates in the form of E-M forces, energy or momenta. We do not wish to open this question once again. However, we wish to give a new approach from thermodynamical viewpoint in which no specific forms for stress, energy and momentum are postulated.

Since the question is equally valid for the local theory, to simplify our discussion we examine these concepts in the light of the local theory.

For the local theory, Clausius-Duhem inequality (3.6) may be written in the form

$$\frac{\rho_0}{\theta}(\dot{\overline{\psi}}+\dot{\mathbf{x}}\cdot\mathbf{g}+\dot{\theta}\eta)+\frac{1}{\theta}\mathbf{T}^K\cdot\dot{\mathbf{x}}_{,K}+\frac{1}{\theta^2}Q^K\theta_{,K}+\frac{\rho_0}{\rho_\theta}(\mathcal{E}\cdot\overset{*}{\mathbf{D}}+\mathcal{H}\cdot\overset{*}{\mathbf{B}})+\frac{\rho_0}{\rho\theta}\mathcal{E}\cdot\overline{\mathcal{J}}\geq 0,\quad \text{in } \mathcal{V}-\sigma, \quad (4.1)$$

where we set

$$\overline{\psi}\equiv\psi-\dot{\mathbf{x}}\cdot\mathbf{g},\qquad \psi\equiv\epsilon-\theta\eta,\qquad \overline{\mathcal{J}}\equiv\mathcal{J}-\mathcal{J}_0, \quad (4.2a)$$

$$\mathbf{t}^k\equiv\frac{\rho}{\rho_0}\mathbf{T}^K x^k{}_{,K},\qquad \mathbf{q}-c\mathcal{E}\times\mathcal{H}\equiv\frac{\rho}{\rho_0}Q^K\mathbf{x}_{,K} \quad (4.2b)$$

and use the relation

$$\mathbf{v}_{,k} = \dot{\mathbf{x}}_{,K} X^K{}_{,k}.$$

We must now postulate a constitutive equation for the constitutive variables $\overline{\psi}, \mathbf{T}^K, Q^K, \mathfrak{K}$ and $\mathcal{E}$, and an invariance requirement. However, for the determination of $\mathbf{g}$ the following minimal requirement suffices. For purely mechanical media we have $\mathbf{g} = \mathbf{0}$ and the momentum equations determine all other rates $\ddot{x}, \dddot{x}$, etc. Thus, the constitutive variables cannot depend on rates higher than $\dot{x}$. If they did, then the equation of momentum balance can be used to eliminate all higher rates of $\dot{\mathbf{x}}$.

The *homogeneity of space* requires that these variables must not depend on the origin of the spatial coordinates and hence on $\mathbf{x}$. Thus, these variables shall depend on $\dot{\mathbf{x}}$ and other variables not involving $\dot{\mathbf{x}}$, e.g.,

$$\overline{\psi} = \overline{\psi}(\dot{\mathbf{x}}, \cdots). \tag{4.3}$$

Substitution of (4.3) into (4.1) gives

$$\frac{\rho_0}{\theta}\left(\frac{\partial\overline{\psi}}{\partial\dot{\mathbf{x}}} + \mathbf{g}\right)\cdot\ddot{\mathbf{x}} - \frac{\rho_0}{\theta}(\overline{\psi}' + \dot{\theta}\eta) + \frac{1}{\theta}\mathbf{T}^K\cdot\dot{\mathbf{x}}_{,K} + \frac{1}{\theta^2}Q^K\theta_{,K} + \frac{\rho_0}{\rho\theta}(\mathfrak{K}\cdot\overset{*}{\mathbf{B}} + \mathcal{E}\cdot\overset{*}{\mathbf{D}}) + \frac{\rho_0}{\rho\theta}\mathcal{E}\cdot\mathfrak{J} \geq 0,$$

where $\overline{\psi}'$ denotes material time rate of $\overline{\psi}$ with $\dot{\mathbf{x}} =$ fixed. If this inequality is to be valid for all independent variation of $\ddot{\mathbf{x}}$, then we must have

$$\mathbf{g} = -\frac{\partial\overline{\psi}}{\partial\dot{\mathbf{x}}}. \tag{4.4}$$

Note that this result is valid for any material since we have so far made no assumption on the material constitution. Moreover, we have not yet stated any invariance requirement on $\overline{\psi}$.

From (4.2a) and (4.4) it follows that

$$\frac{\partial\psi}{\partial\mathbf{v}} = v_i\frac{\partial g_i}{\partial\mathbf{v}}. \tag{4.5}$$

Thus, the invariance requirement of ψ is tied to that of g_i. We must investigate this question in detail.

As stated above in the literature several proposals exist for the E-M momenta, energy and stress. We cite three well-known forms of the E-M momentum:

Minkowsky proposal:

$$\mathbf{g} = (\mathbf{D}\times\mathbf{B})/\rho c. \tag{4.6}$$

Abraham's form:

$$\mathbf{g} = (1/\rho c)(\mathbf{E}\times\mathbf{H}). \tag{4.7}$$

Another form:

$$\mathbf{g} = (\mathbf{E}\times\mathbf{B})/\rho c. \tag{4.8}$$

For a discussion of (4.6) and (4.7) see Møller Ref. 19, Sec. 72.

While there exist arguments in favor of one of these forms to others for various types of media (e.g., dielectrics, magnetic materials), none of the arguments can be substantiated either on the basis of physical principles or existing experiments. In fact, the selection of any one of the above forms is consistent with all basic balance laws and thermodynamics. Moreover, they are consistent with the relativistic E-M theory to within an approximation v^2/c^2 as can be seen from the following. Suppose that we select the Minkowski form (4.6). From relativistic consideration it is well known that in a moving frame of reference, to within v^2/c^2, we have

$$\mathfrak{D} = \mathbf{D} + \frac{1}{c}\mathbf{v}\times\mathbf{H}, \quad \mathfrak{B} = \mathbf{B} - \frac{1}{c}\mathbf{v}\times\mathbf{E}, \tag{4.9}$$

where $\mathbf{D}, \mathbf{B}, \mathbf{E}$ and $\mathbf{H}$ are, as usual, the fields in the rest frame. Thus, the E-M momentum in the moving frame of reference is given by

$$\mathcal{G} = (\mathfrak{D}\times\mathfrak{B})/\rho c. \tag{4.10}$$

Upon substituting (4.9) into (4.10) we obtain

$$\mathcal{G}_i = g_i + ({}_Mt_{ij} - h\delta_{ij})(v_j/c) + (1/c^2)v_i\mathbf{H}\cdot\mathbf{v}\times\mathbf{E}, \tag{4.11}$$

where

$${}_Mt_{ij} \equiv E_iD_j + H_iB_j - \tfrac{1}{2}(\mathbf{E}\cdot\mathbf{D} + \mathbf{H}\cdot\mathbf{B})\delta_{ij}, \tag{4.12}$$
$$h \equiv \tfrac{1}{2}(\mathbf{E}\cdot\mathbf{D} + \mathbf{H}\cdot\mathbf{B})$$

are respectively the well-known expressions of the Minkowski E-M stress tensor and energy. We note that to within v^2/c^2 (nonrelativistic limit) $\mathcal{G}_i$ is given by

$$\mathcal{G}_i = g_i + ({}_Mt_{ij} - h\delta_{ij})(v_j/c) + 0(v^2/c^2). \tag{4.13}$$

This result remains valid for other forms (4.7) and (4.8) of the E-M momentum except that ${}_Mt_{ij}$ and h have other forms. For example, in (4.8) they have the forms

$${}_Mt_{ij} = E_iE_j + B_iB_j - \tfrac{1}{2}(E^2 + B^2)\delta_{ij}, \quad h \equiv \tfrac{1}{2}(E^2 + B^2). \tag{4.14}$$

Thus, for the expression of the free energy ψ in moving frame through (4.4), i.e.,

$$\frac{\partial\psi}{\partial\mathbf{v}} = v_i\frac{\partial\mathcal{G}_i}{\partial\mathbf{v}} \tag{4.15}$$

and (4.12) we obtain

$$\Psi = \psi + \frac{1}{2\rho}\frac{v_iv_j}{c^2}({}_Mt_{ij} - h\delta_{ij}), \tag{4.16}$$

where ψ is the value of the free energy in the rest frame. It is now clear that *to within* v^2/c^2, the free energy Ψ is independent of $\mathbf{v}$. We have, therefore, shown that for *all three expressions of the E-M momentum, the free energy ψ is independent of $\mathbf{v}$ to within an approximation excluding the terms of the order v^2/c^2 (or higher)*.

Hence,

$$\overline{\psi} = \psi - \mathbf{v}\cdot\mathbf{g}, \tag{4.17}$$

where ψ is independent of $\mathbf{v}$.

5. CONSTITUTIVE EQUATIONS OF LOCAL E-M ELASTIC SOLIDS

The constitutive equations of E-M elastic solids may be constructed by assuming that $\psi, \mathbf{T}^K, \mathbf{Q}, \mathfrak{K}, \mathcal{E}$ and $\mathfrak{J}$ are functions of $\mathbf{x}_{,K}, \mathbf{D}, \mathbf{B}, \mathfrak{K}$ and $\theta_{,K}$. For simplicity we employ the equivalent set ($\mathbf{x}_{,K}, \bar{D}^K, \bar{B}^K, \theta$ and $\theta_{,K}$), where

$$\bar{D}^K \equiv \frac{\rho_0}{\rho} X^K{}_{,k} D^k, \quad \bar{B}^K \equiv \frac{\rho_0}{\rho} X^K{}_{,k} B^k. \tag{5.1}$$

Thus,

$$\psi = \Psi(\mathbf{x}_{,K}, \bar{D}^K, \bar{B}^K, \theta, \theta_{,k}). \tag{5.2}$$

Similar equations are written for $\mathbf{T}^K, \mathbf{Q}, \mathfrak{K}, \mathcal{E}$ and $\mathfrak{J}$.

First we note the identities

$$\dot{\bar{D}}^K = \frac{\rho_0}{\rho} X^K{}_{,k} \overset{*}{D}{}^k, \quad \dot{\bar{B}}^K \equiv \frac{\rho_0}{\rho} X^K{}_{,k} \overset{*}{B}{}^k. \tag{5.3}$$

Upon substituting (5.2) into (4.1) and using (4.2), (4.4), and (5.3), we obtain

$$\begin{aligned} &\frac{\rho_0}{\theta}\left(\frac{\partial \Psi}{\partial \theta} + \eta\right)\dot{\theta} + \frac{1}{\theta}\left(\mathbf{T}^K - \rho_0 \frac{\partial \Psi}{\partial \mathbf{x}_{,K}}\right)\cdot \dot{\mathbf{x}}_{,K} \\ &+ \frac{\rho_0}{\rho\theta}\left(\mathcal{E}_k - \rho_0 \frac{\partial \Psi}{\partial \bar{D}^K} X^K{}_{,k}\right)\overset{*}{D}{}^k \\ &+ \frac{\rho_0}{\rho\theta}\left(\mathcal{K}_k - \rho_0 \frac{\partial \Psi}{\partial \bar{B}^K} X^K{}_{,k}\right)\overset{*}{B}{}^k + \frac{\partial \Psi}{\partial \theta_{,K}}\dot{\theta}_{,K} \\ &+ \frac{1}{\theta^2} Q^K \theta_{,K} + \frac{\rho_0}{\rho\theta}\mathcal{E}\cdot\bar{\mathfrak{J}} \ge 0. \end{aligned} \tag{5.4}$$

This inequality is valid for all independent variations of $\dot{\theta}, \dot{\mathbf{x}}_{,K}, \overset{*}{D}{}^k, \overset{*}{B}{}^k$ and $\theta_{,K}$. Since it is linear in these quantities it cannot hold unless

$$\begin{aligned} &\eta = -\frac{\partial \Psi}{\partial \theta}, \quad \mathbf{T}^K = \rho_0 \frac{\partial \Psi}{\partial \mathbf{x}_{,K}}, \\ &\mathcal{E}_k = \rho_0 \frac{\partial \Psi}{\partial \bar{D}^K} X^K{}_{,k} \quad \mathcal{K}_K = \rho_0 \frac{\partial \Psi}{\partial \bar{B}^K} X^K{}_{,k} \end{aligned} \tag{5.5}$$

and

$$\frac{\partial \Psi}{\partial \theta_{,K}} = 0, \quad \frac{1}{\theta^2} Q^K \theta_{,K} + \frac{\rho_0}{\rho\theta}\mathcal{E}\cdot\bar{\mathfrak{J}} \ge 0. \tag{5.6}$$

We have therefore proven

Theorem: The constitutive equations of local elastic solids are thermodynamically admissible if and only if they are of the form (5.5) subject to the conditions (5.6) [i.e., Ψ is independent of $\theta_{,K}$, and Q^K and $\mathfrak{J}$ satisfy the inequality of (5.6)], for all independent processes.

Next, we note that the equation of balance of moment of momentum (2.31), with $\mathbf{f} = \boldsymbol{\ell} = 0$, is satisfied if

$$\bar{t}_{kl} \equiv t_{kl} - \rho v_k g_l \tag{5.7}$$

is a symmetric tensor, i.e.,

$$t_{[kl]} - \rho v_{[k} g_{l]} = 0, \tag{5.8}$$

where a square bracket enclosing indices indicates skew-symmetric part and according to $(5.5)_2$ and $(4.2)_4$

$$t_{kl} = \rho \frac{\partial \Psi}{\partial x^l{}_{,K}} x_{k,K} \tag{5.9}$$

Note that if we employ $\bar{t}_{kl} = \bar{t}_{lk}$, given by (5.7), we will need no further reference to the balance of moment of momentum (2.31a).

Remark 1: We note that the selection of the equivalent independent constitutive variables ($\mathbf{x}_{,K}, \bar{D}^K, \bar{B}^K, \theta$ and $\theta_{,K}$) instead of ($\mathbf{x}_{,K}, D^k, B^k, \theta$ and $\theta_{,K}$) eliminates an argument on the invariance requirement to be placed on Ψ. If the second set is selected, the Clausius-Duhem inequality gives a set of partial differential equations whose solution is, as in purely mechanical case, the requirement of invariance under rigid body motions of the spatial frame of reference. This in turn indicates the use of the independent variables used above.

Remark 2: In the dynamical theory of electromagnetism, one must also impose the requirement of invariance under changes of inertial frames. A rigorous treatment of this question requires the relativistic treatment. To within an approximation v^2/c^2 one can show that the variables D^k and B^k can be replaced by their relativistic expressions

$$\mathfrak{D} = \mathbf{D} + \frac{1}{c}\mathbf{u}\times\mathbf{H}, \quad \mathfrak{B} = \mathbf{B} - \frac{1}{c}\mathbf{u}\times\mathbf{D}. \tag{5.10}$$

Remark 3: Different constitutive theories can be constructed employing other sets of E-M variables, e.g., (**E**, **B**), (**D**, **H**), etc. While each form may be more suitable for different types of bodies, (e.g., dielectrics, magnetic materials) they are equivalent to the set (5.5). In fact, they can be transformed into one another by a Legendre transformation applied to the function ψ.

6. NONLOCAL E-M ELASTIC SOLIDS

Determination of the nonlocal constitutive equations of bodies requires the use of the full entropy inequality (3.5). Here we consider nonlocal elastic solids in which the mass production and heat conduction are not appreciable, i.e., we set

$$\hat{\rho} = 0, \quad Q^K = 0, \tag{6.1}$$

so that (3.5) may be written as

$$\begin{aligned} &-\frac{\rho_0}{\theta}(\dot{\bar{\psi}} + \ddot{\mathbf{x}}\cdot\mathbf{g} + \dot{\theta}\eta) + \frac{1}{\theta}\mathbf{T}^K\cdot\dot{\mathbf{x}}_{,K} + \frac{\rho_0}{\rho\theta}(\mathcal{E}\cdot\overset{*}{\mathbf{D}} + \mathfrak{K}\cdot\overset{*}{\mathbf{B}}) \\ &+ \frac{\rho_0}{\rho\theta}\hat{\mathbf{f}}\cdot\mathbf{v} + \frac{\rho_0}{\rho\theta}(\mathcal{E}\cdot\hat{\mathfrak{J}} + \mathfrak{K}\cdot\hat{\mathbf{b}}) + \frac{\rho_0}{\theta}(\hat{h} - \mathfrak{s}) \ge 0, \quad \text{in } \mathcal{V} - \sigma. \end{aligned} \tag{6.2}$$

In accordance with the axiom of causality introduced in[20], we now define a nonlocal E-M elastic solid by the constitutive equations of form

$$\bar{\psi}(\mathbf{X}, t) = \Psi[\mathfrak{K}(\mathbf{X}'), \mathfrak{K}_L(\mathbf{X}'), \mathfrak{D}_L(\mathbf{X}'), \mathfrak{B}_L(\mathbf{X}'); \dot{\mathbf{x}}, \mathbf{x}_{,K}, \bar{\mathbf{D}}, \bar{\mathbf{B}}, \theta, \mathbf{X}], \tag{6.3}$$

where Ψ is a scalar-valued function of $\dot{\mathbf{x}}, \mathbf{x}_{,K}, \bar{\mathbf{D}}, \bar{\mathbf{B}}, \theta, \mathbf{X}$ and a *functional* of the difference functions

$$\begin{aligned} &\mathfrak{K}(\mathbf{X}') \equiv \mathbf{x}(\mathbf{X}', t) - \mathbf{x}(\mathbf{x}, t), \\ &\mathfrak{K}_L(\mathbf{X}') \equiv \mathbf{x}_{,L}(\mathbf{X}', t) - \mathbf{x}_{,L}(\mathbf{X}, t), \\ &\mathfrak{D}(\mathbf{X}') \equiv \bar{\mathbf{D}}(\mathbf{X}', t) - \bar{\mathfrak{D}}(\mathbf{X}, t), \\ &\mathfrak{B}(\mathbf{X}') \equiv \bar{\mathbf{B}}(\mathbf{X}', t) - \bar{\mathbf{B}}(\mathbf{X}, t), \end{aligned} \tag{6.4}$$

defined over *all* material points $\mathbf{X}'$ of the body $\mathfrak{B}$. Similar equations are written for the dependent variables $\eta, T^K, \mathcal{E}, \mathfrak{K}$ and $\bar{\mathfrak{J}}$, except that these are scalar-valued for

η and vector-valued functionals for the remaining dependent variables. We assume that Ψ possesses continuous first order partial derivatives with respect to $\mathbf{x}_{,K}, \overline{\mathbf{D}}, \overline{\mathbf{B}}, \theta$ and the Fréchet partial derivatives[21] with respect to $\mathfrak{X}(\mathbf{X}'), \mathfrak{X}_L(\mathbf{X}'), \mathfrak{D}(\mathbf{X}')$ and $\mathfrak{B}(\mathbf{X}')$, continuous of order zero. From (6.4) it is clear that

$$\mathfrak{X} = \mathfrak{X}_L = \mathfrak{D} = \mathfrak{B} = 0, \quad \text{when } \mathbf{X}' = \mathbf{X} \tag{6.5}$$

since the dependences of Ψ on the arguments $\mathbf{x}_{,K}, \overline{\mathbf{D}}$ and $\overline{\mathbf{B}}$ are separately indicated, without loss in generality, we may also take

$$\Psi[0,0,0,0;\dot{\mathbf{x}},\mathbf{x}_{,K},\overline{\mathbf{D}},\overline{\mathbf{B}},\theta,\mathbf{X}] \equiv \psi_0 \tag{6.6}$$

as the free energy of the local E-M elastic solids. This means that the *functional* gradients of Ψ with respect to $\mathfrak{X}(\mathbf{X}'), \mathfrak{X}_L(\mathbf{X}'), \mathfrak{D}(\mathbf{X}')$ and $\mathfrak{B}(\mathbf{X}')$ at $\mathbf{X}' = \mathbf{X}$ vanish, i.e.,

$$\delta\Psi/\delta\mathfrak{X} = \delta\Psi/\partial\mathfrak{X}_L = \delta\Psi/\partial\mathfrak{D} = \delta\Psi/\delta\mathfrak{B} = 0, \quad \text{at } \mathbf{X}' = \mathbf{X}, \tag{6.7}$$

where δ/δ indicates functional (Fréchet) partial derivative.

We now calculate

$$\begin{aligned}\dot{\Psi} &= \frac{\partial\Psi}{\partial\dot{\mathbf{x}}}\cdot\ddot{\mathbf{x}} + \frac{\partial\Psi}{\partial\theta}\dot{\theta} + \frac{\partial\Psi}{\partial\mathbf{x}_{,K}}\cdot\dot{\mathbf{x}}_{,K} + \frac{\partial\Psi}{\partial\overline{\mathbf{D}}}\cdot\dot{\overline{\mathbf{D}}} + \frac{\partial\Psi}{\partial\overline{\mathbf{B}}}\cdot\dot{\overline{\mathbf{B}}} \\ &+ \int_{\mathcal{V}-\sigma}\frac{\partial\Psi}{\partial\mathfrak{X}}[\mathfrak{X}(\mathbf{X}'),\cdots,\mathbf{X},\boldsymbol{\lambda}]\cdot\dot{\mathfrak{X}}(\boldsymbol{\lambda})dv(\boldsymbol{\lambda}) \\ &+ \int_{\mathcal{V}-\sigma}\frac{\delta\Psi}{\delta\mathfrak{X}_L}[\mathfrak{X}(\mathbf{X}'),\cdots,\mathbf{X},\boldsymbol{\lambda}]\cdot\dot{\mathfrak{X}}_L(\boldsymbol{\lambda})dv(\boldsymbol{\lambda}) \\ &+ \int_{\mathcal{V}-\sigma}\frac{\delta\Psi}{\delta\mathfrak{D}}[\mathfrak{X}(\mathbf{X}'),\cdots,\mathbf{X},\boldsymbol{\lambda}]\cdot\mathfrak{D}(\boldsymbol{\lambda})dv(\boldsymbol{\lambda}) \\ &+ \int_{\mathcal{V}-\sigma}\frac{\delta\Psi}{\delta\mathfrak{B}}[\mathfrak{X}(\mathbf{X}'),\cdots,\mathbf{X},\boldsymbol{\lambda}]\cdot\dot{\mathfrak{B}}(\boldsymbol{\lambda})dv(\boldsymbol{\lambda}).\end{aligned} \tag{6.8}$$

We draw attention to the fact that the functional gradients appearing in the integrals are also functions of a vector $\boldsymbol{\lambda}$.

Substituting (6.8) into (6.2) we obtain

$$\begin{aligned}&-\frac{\rho_0}{\theta}\left(\mathbf{g} + \frac{\partial\Psi}{\partial\dot{\mathbf{x}}}\right)\cdot\ddot{\mathbf{x}} - \frac{\rho_0}{\theta}\left(\eta + \frac{\partial\Psi}{\theta}\right)\dot{\theta} \\ &-\frac{\rho_0}{\theta}\left[\hat{\mathbf{f}} - \int_{\mathcal{V}-\sigma}\frac{\delta\Psi}{\delta\mathfrak{X}}dv(\boldsymbol{\lambda})\right]\cdot\dot{\mathbf{x}} \\ &+\frac{1}{\theta}\left[\mathbf{T}^K - \rho_0\frac{\partial\Psi}{\partial\mathbf{x}_{,K}} + \rho_0\int_{\mathcal{V}-\sigma}\frac{\delta\Psi}{\delta\mathfrak{X}_K}dv(\boldsymbol{\lambda})\right]\cdot\dot{\mathbf{x}}_{,K} \\ &+\frac{\rho_0}{\rho\theta}\left[\mathcal{E}_k - \rho_0\frac{\partial\Psi}{\partial\overline{D}^K}X^K{}_{,k} + \rho_0 X^K{}_{,k}\int_{\mathcal{V}-\sigma}\frac{\delta\Psi}{\delta\mathfrak{D}^K}dv(\boldsymbol{\lambda})\right]\overset{*}{D}{}^k \\ &+\frac{\rho_0}{\rho\theta}\left[\mathcal{H}_k - \rho_0\frac{\partial\Psi}{\partial\overline{B}^K}X^K{}_{,k} + \rho_0 X^K{}_{,k}\int_{\mathcal{V}-\sigma}\frac{\delta\Psi}{\delta\mathfrak{B}^K}dv(\boldsymbol{\lambda})\right]\overset{*}{B}{}^k \\ &+\frac{\rho_0}{\rho\theta}\mathbf{E}\cdot\overline{\mathfrak{J}} - \frac{\rho_0}{\theta}\int_{\mathcal{V}-\sigma}\left[\frac{\delta\Psi}{\delta\mathbf{K}}\cdot\dot{\mathbf{x}}(\boldsymbol{\lambda}) + \frac{\delta\Psi}{\delta\mathfrak{X}_K}\cdot\dot{\mathbf{x}}_{,K}(\boldsymbol{\lambda})\right. \\ &\left.+\frac{\delta\Psi}{\delta\mathfrak{D}^K}X^K{}_{,k}\overset{*}{D}{}^k(\boldsymbol{\lambda}) + \frac{\delta\Psi}{\delta\mathfrak{B}^K}X^K{}_{,k}(\boldsymbol{\lambda})\overset{*}{B}{}^k\right]dv(\boldsymbol{\lambda}) \\ &+\frac{\rho_0}{\rho\theta}(\mathcal{E}\cdot\hat{\mathfrak{J}} + \mathcal{H}\cdot\hat{\mathbf{b}}) + \frac{\rho_0}{\theta}(\hat{h} - \mathfrak{s}) \geq 0, \quad \text{in } \mathcal{V}-\sigma.\end{aligned} \tag{6.9}$$

This inequality is linear in $\ddot{\mathbf{x}}, \dot{\theta}, \mathbf{x}_{,K}, \overset{*}{D}{}^k$ and $\overset{*}{B}{}^k$. If $\hat{\mathfrak{J}}, \hat{\mathbf{b}}$, $\hat{h}$ and $\mathfrak{s}$ are independent of these quantities, then (6.9) cannot be maintained for all possible variation of these independent quantities unless

$$\begin{aligned}&\mathbf{g} = -\frac{\partial\Psi}{\partial\dot{\mathbf{x}}}, \quad \eta = -\frac{\partial\Psi}{\partial\theta}, \\ &\mathbf{T}^K = \rho_0\frac{\partial\Psi}{\partial\mathbf{x}_{,K}} - \rho_0\int_{\mathcal{V}-\sigma}\frac{\delta\Psi}{\delta\mathfrak{X}_K}dv(\boldsymbol{\lambda}), \\ &\mathcal{E}_k = \rho_0\frac{\partial\Psi}{\partial\overline{D}^K}X^K{}_{,k} - \rho_0 X^K{}_{,k}\int_{\mathcal{V}-\sigma}\frac{\delta\Psi}{\delta\mathfrak{D}^K}dv(\boldsymbol{\lambda}), \\ &\mathcal{H}_k = \rho_0\frac{\partial\Psi}{\partial\overline{B}^K}X^K{}_{,k} - \rho_0 X^K{}_{,k}\int_{\mathcal{V}-\sigma}\frac{\delta\Psi}{\delta\mathfrak{B}^K}dv(\boldsymbol{\lambda}), \quad \text{in } \mathcal{V}-\sigma\end{aligned} \tag{6.10}$$

and

$$\begin{aligned}&-\frac{\rho_0}{\theta}\left[\hat{\mathbf{f}} - \int_{\mathcal{V}-\sigma}\frac{\delta\Psi}{\delta\mathfrak{X}}dv(\boldsymbol{\lambda})\right]\cdot\dot{\mathbf{x}} + \frac{\rho_0}{\rho\theta}\mathcal{E}\cdot\overline{\mathfrak{J}} \\ &-\frac{\rho_0}{\theta}\int_{\mathcal{V}-\sigma}\left[\frac{\delta\Psi}{\delta\mathfrak{X}}\cdot\dot{\mathbf{x}}(\boldsymbol{\lambda}) + \frac{\delta\Psi}{\delta\mathfrak{X}_K}\cdot\dot{\mathbf{x}}_{,K}(\boldsymbol{\lambda})\right. \\ &\left.+\frac{\delta\Psi}{\delta\mathfrak{D}^K}X^K{}_{,k}\overset{*}{D}{}^k(\boldsymbol{\lambda}) + \frac{\delta\Psi}{\delta\mathfrak{B}^K}X^K{}_{,k}\overset{*}{B}{}^k(\boldsymbol{\lambda})\right]dv(\boldsymbol{\lambda}) \\ &+\frac{\rho_0}{\rho\theta}(\mathcal{E}\cdot\hat{\mathfrak{J}} + \mathcal{H}\cdot\hat{\mathbf{b}}) + \frac{\rho_0}{\theta}(\hat{h} - \mathfrak{s}) \geq 0, \quad \text{in } \mathcal{V}-\sigma.\end{aligned} \tag{6.11}$$

We have thus proved

Theorem: The constitutive equations of nonlocal E-M elastic solids are thermodynamically admissible if and only if they are of the form (6.11) *for all independent processes.*

To achieve further progress, we need information regarding the nature of the free energy functional Ψ. Motivated with the results of Sec. 4, we assume that $\mathbf{g}$ has the form (4.6) and Ψ may be expressed as

$$\Psi = \psi - \mathbf{v}\cdot\mathbf{g}, \tag{6.12}$$

where ψ is independent of $\mathbf{v}$.

Upon employing (6.12) in (6.11) and recalling the expressions (2.10c) and (2.38) of $\mathfrak{J}$ and $\hat{\mathfrak{J}}$, the argument on the linearity of (6.11) in $\dot{\mathbf{x}}$ ($\hat{h}$ and $\mathfrak{s}$ are assumed to be independent of $\dot{\mathbf{x}}$) leads to

$$\hat{\mathbf{f}} = \frac{1}{\theta}(q + \hat{q})\mathcal{E} + \int_{\mathcal{V}-\sigma}\frac{\delta\Psi}{\delta\mathfrak{X}}dv(\boldsymbol{\lambda}) \tag{6.13}$$

and

$$\begin{aligned}&\frac{\rho_0}{\rho\theta}\mathcal{E}\cdot(\mathbf{J} + \hat{\mathbf{J}} - \mathfrak{J}_0) + \frac{\rho_0}{\rho\theta}\mathcal{H}\cdot\hat{\mathbf{b}} + \frac{\rho_0}{\rho\theta}\mathcal{H}\cdot\hat{\mathbf{b}} + \frac{\rho_0}{\theta}(\hat{h} - \mathfrak{s}) \\ &-\frac{\rho_0}{\theta}\int_{\mathcal{V}-\sigma}\left[\frac{\delta\Psi}{\delta\mathfrak{X}}\cdot\dot{\mathbf{x}}(\boldsymbol{\lambda}) + \frac{\delta\Psi}{\delta\mathfrak{X}_K}\cdot\dot{\mathbf{x}}_{,K}(\boldsymbol{\lambda}) + \frac{\delta\Psi}{\delta\mathfrak{D}^K}X^K{}_{,k}\overset{*}{D}{}^k(\boldsymbol{\lambda})\right. \\ &\left.+\frac{\delta\Psi}{\delta\mathfrak{B}^K}X^K{}_{,k}\overset{*}{B}{}^k(\boldsymbol{\lambda})\right]dv(\boldsymbol{\lambda}) \geq 0, \quad \text{in } \mathcal{V}-\sigma.\end{aligned} \tag{6.14}$$

The constitutive equation for the conduction current $\mathbf{J}_c \equiv \mathbf{J} - \mathfrak{J}_0$ cannot be derived from (6.14), but must be written as a separate equation

$$\mathbf{J}_c = \mathfrak{J}[\mathfrak{X}(\mathbf{X}'), \mathfrak{X}_L(\mathbf{X}'), \mathfrak{D}_L(\mathbf{X}'), \mathfrak{B}_L(\mathbf{X}'); \dot{\mathbf{x}}, \mathbf{x}_{,K}, \overline{\mathbf{D}}, \overline{\mathbf{B}}, \theta, \mathbf{X}] \tag{6.15}$$

subject to the restrictions arising from (6.14).

The constitutive equations (6.10), together with (6.12) and (6.13), are now given by

$$\mathbf{g} = (\mathbf{D} \times \mathbf{B})/\rho c, \quad \eta = -\frac{\partial \Psi}{\partial \theta}, \tag{6.16a}$$

$$\hat{\mathbf{f}} = \frac{1}{\rho}(q + \hat{q})\mathcal{E} + \int_{\mathcal{V}-\sigma} \frac{\delta \Psi}{\delta \mathfrak{K}} dv(\boldsymbol{\lambda}), \tag{6.16b}$$

$$\mathbf{T}^K = \rho_0 \frac{\partial \Psi}{\partial \mathbf{x}_{,K}} - \rho_0 \int_{\mathcal{V}-\sigma} \frac{\delta \Psi}{\delta \mathfrak{K}_K} dv(\boldsymbol{\lambda}), \tag{6.16c}$$

$$\mathcal{E}_k = \rho_0 \frac{\partial \Psi}{\partial \bar{D}^K} X^K{}_{,k} - \rho_0 X^K{}_{,k} \int_{\mathcal{V}-\sigma} \frac{\delta \Psi}{\delta \mathfrak{D}^K} dv(\boldsymbol{\lambda}), \tag{6.16d}$$

$$\mathcal{H}_k = \rho_0 \frac{\partial \Psi}{\partial \bar{B}^K} X^K{}_{,k} - \rho_0 X^K{}_{,k} \int_{\mathcal{V}-\sigma} \frac{\delta \Psi}{\delta \mathfrak{B}^K} dv(\boldsymbol{\lambda}). \tag{6.16e}$$

we now substitute (6.16) into the energy equation (2.32a) in its material form, and employ (4.2), (3.4), and (5.3). This results in

$$\begin{aligned} &\rho_0 \theta \dot{\eta} - \frac{\rho_0}{\rho} \mathcal{E} \cdot (\mathbf{J} + \hat{\mathbf{J}}) - \frac{\rho_0}{\rho} \mathcal{H} \cdot \hat{\mathbf{b}} + \rho_0 \int_{\mathcal{V}-\sigma} \left[\frac{\delta \Psi}{\delta \mathfrak{K}} \cdot \dot{\mathbf{x}}(\boldsymbol{\lambda}) \right. \\ &\left. + \frac{\delta \Psi}{\delta \mathfrak{K}_K} \cdot \dot{\mathbf{x}}_{,K}(\boldsymbol{\lambda}) + \frac{\delta \Psi}{\delta \mathfrak{D}} \cdot \dot{\bar{\mathbf{D}}}(\boldsymbol{\lambda}) + \frac{\delta \Psi}{\delta \mathfrak{B}} \cdot \dot{\bar{\mathfrak{B}}}(\boldsymbol{\lambda}) \right] dv(\boldsymbol{\lambda}) \\ &- \rho_0 (h - \hat{h}) = 0, \quad \text{in } \mathcal{V} - \sigma. \end{aligned} \tag{6.17}$$

This is but another form of the energy equation.

There remains the question of invariance requirements of the constitutive functionals. In this regard we may employ the *Galilean* invariance of the response functionals. This is expressed by the form invariance of response functions under all class of motions of the form

$$\mathbf{x}(\mathbf{X}', t) \to \mathbf{Q}\mathbf{x}(\mathbf{X}', t) + \mathbf{v}_0 t + \mathbf{b}_0, \tag{6.18}$$

when $\mathbf{Q}, \mathbf{v}_0$, and $\mathbf{b}_0$ are independent of time and

$$\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}, \quad \det \mathbf{Q} = 1. \tag{6.19}$$

Mathematically,

$$\begin{aligned} &\Psi[\mathfrak{K}(\mathbf{X}'), \mathfrak{K}_L(\mathbf{X}'), \mathfrak{D}(\mathbf{X}'), \mathfrak{B}(\mathbf{X}'); \mathbf{x}_{,K}, \bar{\mathbf{D}}, \bar{\mathbf{B}}, \theta, \mathbf{X}] \\ &= \Psi[\mathbf{Q}\mathfrak{K}(\mathbf{X}'), \mathbf{Q}\mathfrak{K}_L(\mathbf{X}'), \mathbf{Q}\mathfrak{D}(\mathbf{X}'), \mathbf{Q}\mathfrak{B}(\mathbf{X}'); \mathbf{Q}\mathbf{x}_{,K}, \bar{\mathbf{D}}, \bar{\mathbf{B}}, \theta, \mathbf{X}]. \end{aligned} \tag{6.20}$$

If the relativistic effects are neglected, then the mechanical and E-M balance laws are known to be invariant under (6.18). While for time dependent rotations ($\mathbf{Q} = \mathbf{Q}(t)$) and accelerating frames ($\mathbf{b} = \mathbf{b}(t)$) the invariance requirements under (6.18) fail to apply to E-M quantities, here we impose much milder conditions (constant velocities of rotation and translations). For E-M fluids and memory dependent materials, it is necessary to employ the principle of objectivity in relativistic form. This invariance requirement can be used to restrict Eqs. (6.16) further. Moreover, if the energy equation (6.17) is posited to be invariant under Galilean transformations we can determine the form of $\hat{h}$, similar to Ref. 10. However, since we are not interested in the thermal and dissipative aspects of this problem, we do not pursue this question further.

To complete the constitutive theory, it remains to satisfy the condition

$$\int_{\mathcal{V}-\sigma} \rho \hat{\mathbf{f}} dv = 0. \tag{6.21}$$

This can be achieved by selecting ψ in the form

$$\psi = \tilde{\psi} + \mathbf{A} \cdot \int_{\mathcal{V}-\sigma} \mathfrak{K}(\boldsymbol{\lambda}) dv(\boldsymbol{\lambda}), \tag{6.22}$$

where $\mathbf{A}$ is a constant vector and $\tilde{\psi}$ is a functional of the same type as ψ. Upon substituting (6.22) into (6.16c) and the result into (6.21), we obtain

$$\mathbf{A} = -\frac{1}{M} \int_{\mathcal{V}-\sigma} (q + \hat{q}) \mathcal{E} dv(\mathbf{x}) - \frac{1}{M} \int_{\mathcal{V}-\sigma} \rho dv(\boldsymbol{\lambda}) \int_{\mathcal{V}-\sigma} \frac{\partial \Psi}{\partial \mathfrak{K}} dv(\boldsymbol{\lambda}), \tag{6.23}$$

where M is the total mass of the body. With $\mathbf{A}$ given by (6.23), the condition (6.21) is satisfied. The free energy given by (6.22) can be used in the remaining Eqs. of (6.16). It is clear that the forms of the remaining equations do not change except that ψ is replaced by $\tilde{\psi}$.

A remark on the nature of nonlocal effects, represented by the integrals in (6.16), is in order: It is well known that the intermolecular forces attenuate rapidly with the distance. In fact, the success of the local continuum theories is, primarily, due to this fact and due to the consideration of a limited class of phenomenological effects for which typical size is much larger than atomic and molecular distances composing the materials. In the transmission of waves, for example, when the wave length of the signal becomes comparable with the granular or atomic distances, local theories of continua fail to apply. In this range we must employ nonlocal theories. The strong molecular interactions, however, permit us to consider only a small neighborhood of the point of observation. In a continuum theory this notion may be formalized by certain strong continuity requirements on the constitutive functionals. To this end we have previously introduced the *hypothesis* of *attenuating neighborhood*.[20] This hypothesis will place restrictions on the kernels appearing in the integrals of (6.16) so that the kernels decay rapidly with distance. For a precise statement of this hypothesis we make reference to Ref. 20 and for the application of this idea to nonlocal elasticity to Ref. 10. In a forthcoming paper we shall employ the dispersion of E-M waves to determine the exact forms of some of these kernels.

---

*The research contained in this paper was partially supported by the National Science Foundation

[1]Max Born and Emil Wolf, *Principles of Optics* (Pergamon, New York, 1970), 4th ed.
[2]C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963).
[3]W. F. Brown, *Micromagnetics* (Wiley, New York, 1963).
[4]H. Tiersten, J. Math. Phys. **5**, 1298 (1964); J. Math. Phys. **6**, 779 (1965).
[5]Gérard A. Maugin and A. Cemal Eringen, J. Math. Phys. **13**, 143 (1972).
[6]C. Kittel, Rev. Mod. Phys. **21**, 541 (1949).
[7]J. H. Van Vleck, Ann. Institute Henri Poincaré **10**, 150 (1941).
[8]F. Bloch, Z. Phys. **49**, 295 (1932).
[9]W. Heisenberg, Z. Phys. **49**, 619 (1928).
[9]W. Heisenberg, Z. Phys. **49**, 619 (1928).
[10]A. Cemal Eringen and D. G. B. Edelen, Int. J. Eng. Sci. **10**, 233 (1972).
[11]A. Cemal Eringen, Int. J. Eng. Sci. **10**, 425 (1972).
[12]C. Truesdell and R. Toupin, *Handbuch der Physik III/1* (Springer-Verlag, Berlin, 1960).
[13]R. C. Dixon and A. Cemal Eringen, Int. J. Eng. Sci. **3**, 359 (1965).
[14]R. A. Grot and A. Cemal Eringen, Int. J. Eng. Sci. **4**, 611 (1966); Int J. Eng. Sci. **4**, 639 (1966).
[15]A. C. Eringen, *Mechanics of Continua* (Wiley, New York, 1967).
[16]This form was also used by Dr. R. Grot (private communication).
[17]W. Dunkin and A. Cemal Eringen, Int. J. Eng. Sci. **1**, 461 (1963).
[18]S. R. DeGroot and L. G. Suttorp, Physica (Utr.) **36** and **37**, 77 (1968)
[19]C. Møller, *The Theory of Relativity* (Oxford, U. P., Oxford, 1952).
[20]A. Cemal Eringen, Int. J. Eng. Sci. **4**, 179 (1966).
[21]I. N. Sneddon, *Continuum Physics*, edited by A. C. Eringen (Academic, New York, 1971), Vol. I, p. 447.

Branching rules for $U(N) \supset U(M)$ and the evaluation of outer plethysms

P. H. Butler and R. C. King

The Mathematics Department, The University, Southampton, England
(Received 27 August 1971; final revised manuscript received 29 November 1971)

Littlewood's third method of evaluating plethysms is generalized by noting that plethysms determine the branching rules associated with the subgroup decomposition $U(N) \supset U(M)$, and by making use of the well-known branching rule for $U(M) \supset U(M-1)$. This generalization leads to recurrence formulas which are simpler than those due to Murnaghan and to a new method of evaluating plethysms which is free of the ambiguities inherent in the use of Littlewood's third method.

1. INTRODUCTION

Since Littlewood first defined a new multiplication of S functions[1] which he called the operation of plethysm,[2] many methods have been developed to evaluate such products. Littlewood himself suggested three methods.[3] The third of these proved to be the most effective and forms the basis for the calculations of Ibrahim[4] who derived an array of principal part theorems[5] to eliminate the ambiguities that occur in the calculations.

The plethysm $\{\lambda\} \otimes \{\mu\}$ may be identified[6] with the branching rule for the decomposition of an irreducible representation of $U(N)$ associated with the S function $\{\mu\}$ into irreducible representations of $U(M)$, where the S function $\{\lambda\}$ is associated with an irreducible representation of $U(M)$ of dimension N which defines the embedding of $U(M)$ in $U(N)$. This identification has previously been exploited in the evaulation of plethysms[7] by a technique involving the use of the weight spaces associated with $U(M)$ and $U(N)$. The aim in this paper is to use the same identification to derive, without any use of weights, recurrence formulas for plethysms.

It is, of course, this identification which gives plethysms an important role to play in the study of a number of physical problems. This was first recognised by Elliott[8] who used Ibrahim's tables of plethysms to establish the branching rules for the decomposition $U(N) \supset U(3)$. These results were then used in the study of the $SU(3)$ shell model of nuclei. In atomic spectroscopy plethysms have come to play an increasingly important role both in the resolution of the Kronecker squares of irreducible representations with application to the determination of selection rules,[9] and more generally in the classification of the atomic states of many electron configurations, the analysis and classification of the many particle operators of atomic theory, and the derivation of selection rules for the matrix elements of these operators.[10]

In Sec. 2 the well-known branching rule[11] associated with the embedding of $U(M-1)$ in $U(M)$ is used to generalize Littlewood's third method in a manner appropriate to the removal of ambiguities from the calculation. The application of conjugacy theorems[3] in Sec. 3 then leads to a derivation of two formulas similar to those of Murnaghan[12] which formed the basis of an earlier calculation of plethysms.[13] However, the new formulas are simpler to use.

Finally, in Sec. 4 the general result obtained in Sec. 2 is inverted so as to give a direct method of evaluating plethysms free of any ambiguity.

2. GENERALIZATION OF LITTLEWOOD'S THIRD METHOD

The algebra of S functions (Ref. 2, p. 290; Ref. 3) is well established, but the duality that exists between S functions and the irreducible representations of the classical groups has not been fully exploited. This duality is such that if $\{\lambda\}$ is an S function associated with an irreducible representation of $U(M)$ of dimension N, then an embedding of $U(M)$ in $U(N)$ may be defined by the mapping

$$\{1\} \to \{\lambda\}. \tag{2.1}$$

Under this mapping every S function $\{\mu\}$, associated with an irreducible representation of $U(N)$, decomposes in accordance with the branching rule[6]

$$\{\mu\} \to \{\lambda\} \otimes \{\mu\} = \sum_{\nu} G^{\nu}_{\lambda\mu}\{\nu\}, \tag{2.2}$$

where $\{\lambda\} \otimes \{\mu\}$ is an outer plethysm of S functions and $\{\nu\}$ is an S-function associated with an irreducible representation of $U(M)$. If $\{\lambda\}$, $\{\mu\}$ and $\{\nu\}$ correspond to partitions (λ), (μ) and (ν) of l, m and n into p, q and r parts, respectively, then $n = lm$ and $r \le pm$. It is convenient to call l, m, and n and p, q, and r the degrees and depths of the S functions $\{\lambda\}$, $\{\mu\}$, and $\{\nu\}$.

The determination of the coefficients $G^{\nu}_{\lambda\mu}$ corresponds to the evaluation of the plethysm $\{\lambda\} \otimes \{\mu\}$. These coefficients are independent of M and N even though the branching rule (2.2) for $U(N) \supset U(M)$ may be M-independent in the sense that if $\{\nu\}$ has a depth, r, greater than M, then the corresponding irreducible representation is zero.

The canonical embedding of $U(M-1)$ in $U(M)$ is defined by the mapping

$$\{1\} \to \{1\} + \{0\}, \tag{2.3}$$

which leads through the algebra of plethysm to the branching rule

$$\{\nu\} \to (\{1\} + \{0\}) \otimes \{\nu\} = \sum_{a=0}^{n} \{\nu\}/\{a\}. \tag{2.4}$$

The notation is such that Latin letters are used to denote partitions into one part only, while Greek letters denote general partitions. Thus $\{a\}$ is of degree a and depth 1. The evaluation of the quotient (Ref. 2, p. 110) of S functions $\{\nu\}/\{a\}$ to give a sum of S functions of degree $n-a$ leads directly to the well known branching rule (Ref. 11, p. 391) appropriate to $U(M) \supset U(M-1)$.

The mapping (2.1) and the branching rule (2.4) define an embedding of $U(M-1)$ in $U(N)$ which is such that

$$\{\mu\} \to \left(\sum_{a=0}^{l} \{\lambda\}/\{a\}\right) \otimes \{\mu\}, \tag{2.5}$$

while successive application of (2.2) and (2.4) imply that for this same embedding

$$\{\mu\} \to \sum_{b=0}^{lm} (\{\lambda\} \otimes \{\mu\})/\{b\}. \tag{2.6}$$

Therefore,

$$\sum_{b=0}^{lm} (\{\lambda\} \otimes \{\mu\})/\{b\} = \left(\sum_{a=0}^{l} \{\lambda\}/\{\alpha\}\right) \otimes \{\mu\}. \tag{2.7}$$

This result is an identity involving S functions, and although its derivation depends upon the relationship between S functions and irreducible representations of unitary groups, the result is independent of M and N.

Expanding both the left and right hand sides of (2.7) into terms containing S-functions of a given degree then yields the set of identities

$$\{\lambda\} \otimes \{\mu\} = \{\lambda\} \otimes \{\mu\}, \tag{2.8a}$$

$$(\{\lambda\} \otimes \{\mu\})/\{1\} = \{\lambda\} \otimes (\{\mu\}/\{1\}) \cdot (\{\lambda\}/\{1\}), \tag{2.8b}$$

$$\begin{aligned}(\{\lambda\} \otimes \{\mu\})/\{2\} &= \{\lambda\} \otimes (\{\mu\}/\{1\}) \cdot (\{\lambda\}/\{2\}) \\ &+ \{\lambda\} \otimes (\{\mu\}/\{2\}) \cdot (\{\lambda\}/\{1\}) \otimes \{2\} \\ &+ \{\lambda\} \otimes (\{\mu\}/\{1^2\}) \cdot (\{\lambda\}/\{1\}) \otimes \{1^2\},\end{aligned} \tag{2.8c}$$

etc.

Equation (2.8b) forms the basis of Littlewood's third method of calculating plethysms (Ref. 2, p. 291; Ref. 3) which depends upon the fact that if sufficient plethysms are known to enable the right-hand side of (2.8b) to be evaluated, then it may be possible, from the resulting expression for $(\{\lambda\} \otimes \{\mu\})/\{1\}$, to establish the plethysm $\{\lambda\} \otimes \{\mu\}$. It is this last step which involves a certain amount of trial and error and leads to ambiguities.

For example, assuming that $\{2\} \otimes \{2\} = \{4\} + \{2^2\}$, it follows from (2.8b) that

$$\begin{aligned}(\{2\} \otimes \{3\})/\{1\} &= (\{4\} + \{2^2\}) \cdot \{1\} \\ &= \{5\} + \{41\} + \{32\} + \{2^21\}.\end{aligned} \tag{2.9}$$

However,

$$(\{6\} + \{42\} + \{2^3\})/\{1\} = \{5\} + \{41\} + \{32\} + \{2^21\}$$

and

$$(\{51\} + \{3^2\} + \{2^3\})/\{1\} = \{5\} + \{41\} + \{32\} + \{2^21\},$$

so that $\{2\} \otimes \{3\}$ is not uniquely determined by the use of (2.8b)

In this example, application of (2.8c) gives

$$(\{2\} \otimes \{3\})/\{2\} = 2\{4\} + \{31\} + 2\{2^2\}. \tag{2.10}$$

Since

$$(\{6\} + \{42\} + \{2^3\})/\{2\} = 2\{4\} + \{31\} + 2\{2^2\},$$

while

$$(\{51\} + \{3^2\} + \{2^3\})/\{2\} = \{4\} + 2\{31\} + \{2^2\},$$

it then follows that

$$\{2\} \otimes \{3\} = \{6\} + \{42\} + \{2^3\}. \tag{2.11}$$

It is clear from this example that (2.7) provides a way of generalizing Littlewood's third method of calculating plethysms through the use of the set of identities (2.8).

3. COMPARISON WITH MURNAGHAN'S FORMULAS

Application of the general result (2.7) to the cases for which $\{\lambda\} = \{1^l\}$ and either $\{\mu\} = \{m\}$ or $\{\mu\} = \{1^m\}$ leads to the very simple identities

$$\sum_{a=0}^{lm} (\{1^l\} \otimes \{m\})/\{a\} = (\{1^l\} + \{1^{l-1}\}) \otimes \{m\}, \tag{3.1a}$$

$$\sum_{a=0}^{lm} (\{1^l\} \otimes \{1^m\})/\{a\} = (\{1^l\} + \{1^{l-1}\}) \otimes \{1^m\}. \tag{3.1b}$$

Expanding the right hand sides of these equations then gives, for the terms of degree $lm - a$,

$$(\{1^l\} \otimes \{m\})/\{a\} = \{1^l\} \otimes \{m-a\} \cdot \{1^{l-1}\} \otimes \{a\}, \tag{3.2a}$$

$$(\{1^l\} \otimes \{1^m\})/\{a\} = \{1^l\} \otimes \{1^{m-a}\} \cdot \{1^{l-1}\} \otimes \{1^a\}. \tag{3.2b}$$

It follows from Littlewood's theorem of conjugates[3] that

$$(\{l\} \otimes \{m\})/\{1^a\} = \{l\} \otimes \{m-a\} \cdot \{l-1\} \otimes \{1^a\}, \tag{3.3a}$$

$$(\{l\} \otimes \{1^m\}/\{1^a\} = \{l\} \otimes \{1^{m-a}\} \cdot \{l-1\} \otimes \{a\}. \tag{3.3b}$$

The cases $a > m$ lead immediately to the obvious result that neither $\{l\} \otimes \{m\}$ nor $\{l\} \otimes \{1^m\}$ contain S functions of depth greater than m. Setting $a = m$ in (3.3) then yields the results

$$(\{l\} \otimes \{m\}/\{1^m\} = \{l-1\} \otimes \{1^m\}, \tag{3.4a}$$

$$(\{l\} \otimes \{1^m\}/\{1^m\} = \{l-1\} \otimes \{m\}, \tag{3.4b}$$

which determine uniquely the terms of depth m in the expansion of $\{l\} \otimes \{m\}$ and $\{l\} \otimes \{1^m\}$, since only these terms give any contribution to the left hand sides of (3.4) and these terms have a one-to-one correspondence with those of these quotients.

Setting $a = m - 1$ in (3.3) then yields equations which give the terms of $\{l\} \otimes \{m\}$ and $\{l\} \otimes \{1^m\}$ of depth $m - 1$, the process may clearly be continued. The computation may be simplified by noting that terms of depth greater than $m - c$ may be omitted when solving for those of depth $m - c$.

Using the notation[12]

$$[\{\nu\}]_k = \begin{cases} \{\nu_1 - 1, \nu_2 - 1, \ldots, \nu_k - 1\} & \text{if } r = k \\ 0 & \text{if } r \neq k, \end{cases} \tag{3.5}$$

where r is the depth of $\{\nu\}$, the above procedure gives

$$\begin{aligned}[\{l\} \otimes \{m\}]_{m-c} = \sum_{a=0}^{c} (-1)^a \{l\} \otimes \{c-a\} \\ \cdot \{l-1\} \otimes \{1^{m-c+a}\} \cdot \{1^a\},\end{aligned} \tag{3.6a}$$

$$\begin{aligned}[\{l\} \otimes \{1^m\}]_{m-c} = \sum_{a=0}^{c} (-1)^a \{l\} \otimes \{1^{c-a}\} \\ \cdot \{l-1\} \otimes \{m-c+a\} \cdot \{1^a\},\end{aligned} \tag{3.6b}$$

where on the right-hand side of these equations all S functions of depth greater than $m - c$ may be omitted.

For example,

$$[\{2\} \otimes \{3\}]_3 = \{1\} \otimes \{1^3\} = \{1^3\} + \cdots,$$

$$[\{2\} \otimes \{3\}]_2 = \{2\} \otimes \{1\} \cdot \{1\} \otimes \{1^2\} - \{1\} \otimes \{1^3\} \cdot \{1\} = \{31\} + \cdots,$$

$$[\{2\} \otimes \{3\}]_1 = \{2\} \otimes \{2\} \cdot \{1\} \otimes \{1\} - \{2\} \otimes \{1\} \cdot \{1\} \otimes \{1^2\} \cdot \{1\} + \{1\} \otimes \{1^3\} \cdot \{1^2\} = \{5\} + \cdots, \tag{3.7}$$

where the dots indicates S functions of depth greater than $m - c$. Therefore,

$$\{2\} \otimes \{3\} = \{6\} + \{42\} + \{2^3\} \tag{3.8}$$

in agreement with (2.11).

The Eqs. (3.6) bear a striking resemblance to the very useful results of Murnaghan.[12] However, they are simpler in that fewer terms occur on the right hand side of (3.6) than in the expressions of Murnaghan.

4. A FORMULA FOR PLETHYSM COEFFICIENTS

In deriving the results of Secs. 2 and 3, the key formula is (2.7) which was obtained by a consideration of the branching rules appropriate to the subgroup chain $U(N) \supset U(M) \supset U(M-1)$. To proceed further, it is necessary to consider the last link in this chain in more detail.

Just as (2.3) leads through the algebra of plethysm to the relation (2.4) appropriate to $U(M) \downarrow U(M-1)$, so the inverse of (2.3) given by

$$\{1\} \to \{1\} - \{0\} \tag{4.1}$$

leads to the relation

$$\{\tau\} \to (\{1\} - \{0\}) \otimes \{\tau\} = \sum_{b=0}^{t} (-1)^b \{\tau\}/\{1^b\} \tag{4.2}$$

appropriate to $U(M-1) \uparrow U(M)$. The arrows $\downarrow$ and $\uparrow$ correspond to the processes of subduction and induction respectively.

If

$$\sum_{a=0}^{s} \{\sigma\}/\{a\} = \sum_{\tau} H_\sigma^\tau \{\tau\} \tag{4.3}$$

and

$$\sum_{b=0}^{t} (-1)^b \{\tau\}/\{1^b\} = \sum_{\sigma} K_\tau^\sigma \{\sigma\}, \tag{4.4}$$

it is clear that the matrix K is the inverse of the matrix H. The matrix elements of H and K, derived from (4.3) and (4.4), are given in Tables I and II for $s \le 5, t \le 5$. It is to be noted that $K_\tau^\sigma = (-1)^{s-t} H_{\tilde{\sigma}}^{\tilde{\tau}}$ where $\{\tilde{\sigma}\}$ and $\{\tilde{\tau}\}$ are the S functions conjugate to $\{\sigma\}$ and $\{\tau\}$.

An arbitrary set of S functions, each of weight n, associated with a set of irreducible representations of $U(M)$ corresponds to another set of S functions, each of degree less than or equal to n, associated with the subduced representations of $U(M-1)$. That is, for all coefficients A_ν there exist a set of coefficients B_σ such that under $U(M) \downarrow U(M-1)$

$$\sum_{\nu} A_\nu \{\nu\} \to \sum_{\sigma} B_\sigma \{\sigma\} \tag{4.5}$$

with $s \le n$. In fact, from (2.4) and (4.3)

$$B_\sigma = \sum_{\nu} A_\nu H_\nu^\sigma. \tag{4.6}$$

However, if

$$\{\nu\} = \{\nu_1, \nu_2, \nu_3, \cdots\} = \{n-t, \tau_1, \tau_2, \cdots\}, \tag{4.7}$$

it is easy to see that

$$H_\nu^\sigma = H_\sigma^\tau \tag{4.8}$$

so that, with the notation of (4.7), Table I gives the transpose of the matrix associated with the branching rule

$$\{\nu\} \to \sum_{\sigma} H_\nu^\sigma \{\sigma\}. \tag{4.9}$$

TABLE I. The subduction coefficients, H_σ^τ, associated with $U(M) \downarrow U(M-1)$: $\{\sigma\} \to \sum_\tau H_\sigma^\tau \{\tau\}$ and $\{\nu\} \to \sum_\sigma H_\nu^\sigma \{\sigma\}$.

$\{\nu\}$	$\{n\}$	$\{n-1,1\}$	$\{n-2,2\}$	$\{n-2,1^2\}$	$\{n-3,3\}$	$\{n-3,21\}$	$\{n-3,1^3\}$	$\{n-4,4\}$	$\{n-4,31\}$	$\{n-4,2^2\}$	$\{n-4,21^2\}$	$\{n-4,1^4\}$	$\{n-5,5\}$	$\{n-5,41\}$	$\{n-5,32\}$	$\{n-5,31^2\}$	$\{n-5,2^21\}$	$\{n-5,21^3\}$	$\{n-5,1^5\}$
$\{\sigma\}\backslash\{\tau\}$	$\{0\}$	$\{1\}$	$\{2\}$	$\{1^2\}$	$\{3\}$	$\{21\}$	$\{1^3\}$	$\{4\}$	$\{31\}$	$\{2^2\}$	$\{21^2\}$	$\{1^4\}$	$\{5\}$	$\{41\}$	$\{32\}$	$\{31^2\}$	$\{2^21\}$	$\{21^3\}$	$\{1^5\}$
$\{0\}$	1																		
$\{1\}$	1	1																	
$\{2\}$	1	1	1																
$\{1^2\}$		1		1															
$\{3\}$	1	1	1		1														
$\{21\}$		1	1	1		1													
$\{1^3\}$				1			1												
$\{4\}$	1	1	1		1			1											
$\{31\}$		1	1	1	1	1			1										
$\{2^2\}$			1			1				1									
$\{21^2\}$				1		1	1				1								
$\{1^4\}$							1					1							
$\{5\}$	1	1	1		1			1					1						
$\{41\}$		1	1	1	1	1		1	1					1					
$\{32\}$			1		1	1			1	1					1				
$\{31^2\}$				1		1	1		1		1					1			
$\{2^21\}$						1			1		1						1		
$\{21^3\}$							1				1	1						1	
$\{1^5\}$												1							1

TABLE II. The induction coefficients K_τ^σ associated with $U(M-1) \uparrow U(M)$: $\{\tau\} \to \sum K_\tau^\sigma \{\sigma\}$ and $\sum_\nu A_\nu \{\nu\} = \sum_{\nu\sigma} K_\tau^\sigma B_\sigma \{\nu\} \to \sum_\sigma B_\sigma \{\sigma\}$.

$\{\nu\}$	$\{\tau\}$ \ $\{\sigma\}$	$\{0\}$	$\{1\}$	$\{2\}$	$\{1^2\}$	$\{3\}$	$\{21\}$	$\{1^3\}$	$\{4\}$	$\{31\}$	$\{2^2\}$	$\{21^2\}$	$\{1^4\}$	$\{5\}$	$\{41\}$	$\{32\}$	$\{31^2\}$	$\{2^21\}$	$\{21^3\}$	$\{1^5\}$
$\{n\}$	$\{0\}$	1																		
$\{n-1, 1\}$	$\{1\}$	−1	1																	
$\{n-2, 2\}$	$\{2\}$		−1	1																
$\{n-2, 1^2\}$	$\{1^2\}$	1	−1		1															
$\{n-3, 3\}$	$\{3\}$			−1		1														
$\{n-3, 21\}$	$\{21\}$		1	−1	−1		1													
$\{n-3, 1^3\}$	$\{1^3\}$	−1	1		−1			1												
$\{n-4, 4\}$	$\{4\}$					−1			1											
$\{n-4, 31\}$	$\{31\}$			1		−1	−1			1										
$\{n-4, 2^2\}$	$\{2^2\}$				1		−1				1									
$\{n-4, 21^2\}$	$\{21^2\}$		−1	1	1		−1	−1				1								
$\{n-4, 1^4\}$	$\{1^4\}$	1	−1		1			−1					1							
$\{n-5, 5\}$	$\{5\}$								−1					1						
$\{n-5, 41\}$	$\{41\}$					1			−1	−1					1					
$\{n-5, 32\}$	$\{32\}$						1			−1		−1				1				
$\{n-5, 31^2\}$	$\{31^2\}$			−1		1	1			−1		−1					1			
$\{n-5, 2^21\}$	$\{2^21\}$				−1		1	1			−1	−1						1		
$\{n-5, 21^3\}$	$\{21^3\}$		1	−1	−1		1	1				−1	−1						1	
$\{n-5, 1^5\}$	$\{1^5\}$	−1	1		−1			1					−1							1

The columns of the table are labelled in such a way as to indicate this. For given n the columns have to be suitable extended of course, while for small n some columns may be irrelevant.

Making use of (4.8) in (4.6), together with the fact that $K = H^{-1}$, yields the result

$$A_\nu = \sum_\sigma K_\tau^\sigma B_\sigma. \tag{4.10}$$

It should be stressed that in this equation $\{\nu\}$ and $\{\tau\}$ are related by (4.7) so that $t < n$. This implies that for a given value of n the only information required to determine the left hand side of (4.5) is that given by a knowledge of the coefficients B_σ with $s < n$. The relevant transformation matrix is the matrix K of Table II in which the rows are labelled both by $\{\nu\}$ and by $\{\tau\}$.

Returning to the problem of evaluating plethysms, the result (2.7) implies in the notation of (2.2) that under $U(M) \downarrow U(M-1)$

$$\sum_\nu G_{\lambda\mu}^\nu \{\nu\} \to \left(\sum_{a=0}^{l} \{\lambda\}/\{a\}\right) \otimes \{\mu\} = \sum_\sigma F_{\lambda\mu}^\sigma \{\sigma\}, \tag{4.11}$$

where clearly $F_{\lambda\mu}^\nu = G_{\lambda\mu}^\nu$ and the remaining coefficients $F_{\lambda\mu}^\sigma$ with $s < n$ may be evaluated from a knowledge of plethysms of the type $\{\pi\} \otimes \{\rho\}$ with either $p < l$, $r \le m$ or $p \le l, r < m$, i.e., plethysms of degree lower than $\{\lambda\} \otimes \{\mu\}$.

The branching rule (4.11) is an example of the relation (4.5), and using (4.10) and (4.4) gives the formula

$$G_{\lambda\mu}^\nu = \sum_{b=0}^{t} (-1)^b F_{\lambda\mu}^{\tau/1^b}, \tag{4.12}$$

where $\{\nu\}$ and $\{\tau\}$ are related by (4.7) and $\tau/1^b$ denotes the S function quotient $\{\tau\}/\{1^b\}$.

This formula (4.12) enables any plethysm $\{\lambda\} \otimes \{\mu\}$ to be evaluated unambiguously in terms of plethysms of lower degree. As an example, (4.11) implies that

$$\begin{aligned}\sum_\sigma F_{\{2\}\{3\}}^\sigma \{\sigma\} &= (\{2\} + \{1\} + \{0\}) \otimes \{3\} \\ &= \{2\} \otimes \{3\} + \{5\} + \{41\} + \{32\} + \{2^21\} + 2\{4\} + \{31\} \\ &\quad + 2\{2^2\} + 2\{3\} + \{21\} + 2\{2\} + \{1\} + \{0\},\end{aligned} \tag{4.13}$$

where use has been made of the algebra of plethysms and a knowledge of some plethsyms of degree less than 6. The plethysm coefficients $G_{\{2\}\{3\}}^\nu$ are then found by multiplying by K the column matrix whose elements

$$1\ 1\ 2\ 0\ 2\ 1\ 0\ 2\ 1\ 2\ 0\ 0\ 1\ 1\ 1\ 0\ 1\ 0\ 0 \tag{4.14}$$

are the coefficients of the S function $\{\sigma\}$ appearing in (4.13) with $s \le 5$. It is then clear that

$$\{2\} \otimes \{3\} = \{6\} + \{42\} + \{2^3\} \tag{4.15}$$

in agreement with (2.11) and (3.8).

It is worth noting that the general structure of the matrices H and K is such that $G_{\lambda\mu}^\nu = 0$ if $F_{\lambda\mu}^\tau = 0$, and that the only terms of (4.11) which are relevant are those for which $\sigma_1 \ge n - s$. For example only the last six terms of (4.13) are relevant, so that the list (4.14) need be continued no further than the tenth element. Fruthermore it follows immediately from the absence of the S functions $\{1^2\}, \{1^3\}, \{21^2\}$ and $\{1^4\}$ in (4.13) that the S functions $\{41^2\}, \{31^3\}$ $\{2^21^2\}$ and $\{21^4\}$ are not contained in $\{2\} \otimes \{3\}$.

Clearly, even noting these points, the formula (4.12) does not provide the most rapid method of calculating the plethysm $\{2\} \otimes \{3\}$. However, the great merits of the formula lie in the fact that it is completely general and free of any ambiguity.

ACKNOWLEDGMENT

One of the authors (P.H.B.) wishes to acknowledge the financial support of the New Zealand University Grants Committee.

[1]D. E. Littlewood, J. Lond. Math. Soc. **11**, 49 (1936).
[2]D. E. Littlewood, *The Theory of Group Characters* (Oxford U.P., Oxford, 1950), 2nd ed.
[3]D. E. Littlewood, Philos. Trans. R. Soc. Lond. A **239**, 305 (1943).
[4]E. M. Ibrahim, Proc. Math. Phys. Soc. UAR **5**, 85 (1953); Proc. Math. Phys. Soc. UAR **22**, 137 (1958); "Tables for the Plethysms of S-functions," Royal Society (London), Depository of Unpublished Tables.

[5]E. M. Ibrahim, Oxford Quart. J. Math. **3**, 50 (1952); Proc. Am. Math. Soc. **7**, 199 (1956).
[6]P. H. Butler and B. G. Wybourne, J. Phys. (Paris) **30**, 181 (1969).
[7]R. C. King and S. P. O. Plunkett, J. Phys. (Paris) 33, 145 (1972).
[8]J. P. Elliott, Proc. R. Soc. A **245**, 128 (1958).
[9]P. R. Smith and B. G. Wybourne, J. Math. Phys. **8**, 2434 (1967); P. H. Butler and B. G. Wybourne, J. Phys. (Paris) **30**, 655 (1969).
[10]P. R. Smith and B. G. Wybourne, J. Math. Phys. **9**, 1040 (1968); B. G. Wybourne, *Symmetry Principle and Atomic Spectroscopy*, with an appendix of tables by P. H. Butler (Wiley, New York, 1970).
[11]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Methuen, London, 1931).
[12]F. D. Murnaghan, Proc. Natl. Acad. Sci. USA **37**, 51 (1951); Proc. Natl. Acad. Sci. USA **37**, 439 (1951); Proc. Natl. Acad. Sci. USA **40**, 721 (1954).
[13]P. H. Butler, J. Phys. (Paris) **31**, Suppl. C4, 47 (1970); P. H. Butler and B. G. Wybourne, Atomic Data **3**, 133 (1971).

Oscillator phase states, thermal equilibrium and group representations*

Y. Aharonov†, E. C. Lerner‡, H. W. Huang, and J. M. Knight

Physics Department, University of South Carolina, Columbia, South Carolina 29208

Eigenstates of the annihilation type operator $U = C + iS$, where C and S are the "cosine" and "sine" operators for harmonic oscillator phase, are shown to be closely related to thermal equilibrium states of the oscillator and to provide a new interpretation of the thermal equilibrium density operator. The problem of creating such states is considered and a general theorem is established leading to the construction of interaction Hamiltonians which transform the eigenstates of U among themselves and, in particular, create them from the oscillator ground state. These Hamiltonians lead to representations of the Lie algebras of $O(2,1)$ and $O(3)$. It is suggested that the mathematical technique used, in which generalized U-type operators provide the link between a group and its representations, has its own intrinsic interest for the study of Lie groups.

1. INTRODUCTION

In studying the quantum theory of harmonic oscillator phase, new Hermitian operators C and S were introduced[1-7] whose spectra coincide with the range of values of the trigonometric functions $\cos\phi$ and $\sin\phi$. Since these operators do not commute with one another, one cannot prepare a state in which the phase is arbitrarily sharply defined except in certain limiting cases. However, one might expect that the operator $U = C + iS$, which is the quantum analog of the quantity $e^{i\phi} = \cos\phi + i\sin\phi$, would define states of maximal phase resolution in some reasonable sense.

The eigenstates of U, referred to here as the phase states, have been studied,[7] and not only provide a physically reasonable description of phase, but also possess other interesting physical properties. These properties stem from the close relationship between phase states and the description of an oscillator in thermal equilibrium with its surroundings, a relationship in which the classical concept of oscillator phase plays an important role.

As an example, consider an oscillator of natural frequency ω in thermal equilibrium at a temperature T. Then the statistical average of the oscillator energy can be shown to be equal to a pure quantum expectation value in a suitable phase state. This implies that measurement of the oscillator energy cannot distinguish a phase state from a thermal equilibrium mixture. The expectation value of any other oscillator observable is obtained by uniform averaging of its quantum expectation value in such a state over a single parameter which, in the limit $kT \gg \hbar\omega$, is identifiable as the classical phase of the oscillator.

Therefore, although the density operator formalism makes it clear that thermal equilibrium cannot be described by a pure quantum state, the phase states provide as close a description as one might hope for within the pure state framework. The additional randomness associated with thermal equilibrium is represented by a uniform distribution over the phase parameter associated with the state.

In view of the foregoing remarks, it becomes a matter of considerable interest to examine the possibility of finding a physical model for the creation of a phase state. In this model the oscillator would be part of a well-defined larger dynamical system, the effect of which would be to subject it to an interaction which would take it from its ground state, for example, to a phase state. Such a model would conceivably have the interesting property of exhibiting, within the framework of pure quantum dynamics, a process very closely related to the approach to thermal equilibrium.

An obvious first step in the search for this model is to find interaction Hamiltonians which generate phase states. The formulation and solution of this problem form the bulk of the present paper. In developing the mathematical techniques for this purpose, we find that the desired Hamiltonians form a representation of a Lie algebra. The elements of the group generated by this algebra are identified with transformations of the spectrum of the operator U, which in turn acts in the underlying space of the representation. We believe that this linking of the group with its representation via an operator is of sufficient generality to be of intrinsic interest for the study of group representations.

This paper is therefore one of largely mathematical nature, motivated and guided by physical considerations. The emergence of a possible new approach to the study of Lie groups was quite unforseen at the outset. We therefore do not include here attempts to extend our approach beyond the groups $O(2, 1)$ and $O(3)$ which arise naturally in our treatment of harmonic oscillator dynamics. These further efforts will be the subject of a future paper.

In Sec. 2, the relation indicated above between the statistics of an oscillator in thermal equilibrium and the phase states is established. A brief discussion of the properties of these states and of the operators defining them is included here and in the following section. Further clarification of their significance as phase states is provided by examining them in the classical limit and showing that they have just the interpretation one would expect in terms of an ensemble in phase space.

Section 3 is concerned with the formulation and solution of the problem of finding the Hamiltonians that generate phase states. The formulation is achieved by establishing a general theorem which characterizes those Hamiltonians which transform the eigenstates of an operator of the type U into themselves. This leads to a whole class of operators $A(k)$ parametrized by a real variable k and including the operator U itself. For each value of k, there is a set of allowable Hamiltonians $H(k)$. It follows from the general conditions of our theorem that the $H(k)$ constitute a representation of a Lie algebra, which turns out to be essentially that of $O(2, 1)$. We show also that the familiar destruction operator, whose eigenstates are the coherent states, is a member of our class in the limit $k \to \infty$. We thus refer to the operators $A(k)$ as generalized destruction operators.

In Sec. 4 we explicitly exhibit the general form of the unitary transformations which the $H(k)$ induce on the $A(k)$ eigenstates. It is seen that the eigenvalues of $A(k)$ undergo linear fractional transformations. For a given k, these transformations can be classified into two types. In one, the eigenvalues vary periodically in time; in the other they display the interesting property of approaching a given limit as $t \to \infty$ independently of their initial values.

In the last two sections, we take up the group theoretic aspects of our work. In Sec. 5, we show that an extension of the range of the parameter k allows us to generalize the methods of Sec. 3 to include $O(3)$ as well as $O(2,1)$. By a slight broadening of the concept of eigenstate, we are able to retain the useful transformation properties of the eigenstates of the generalized destruction operator even in the finite dimensional subspaces which support representations of $O(3)$. In the final section, after identifying the relevant representations of $O(3)$ and $O(2,1)$, we give some characterization of the phase states from the point of view of the theory of group representations. The section closes with a discussion of the derivation of the explicit forms of the matrix representation of the group by means of our formalism.

2. PHASE STATES AND THERMAL EQUILIBRIUM

Consider an oscillator which has been brought into thermal contact with a heat bath of temperature $T = (k\beta)^{-1}$ and allowed to reach equilibrium. If the contact is then broken and the oscillator allowed to evolve as an isolated system, its state is represented by the thermal density operator

$$\rho = (1 - e^{-\beta\hbar\omega}) \sum_{n=0}^{\infty} e^{-n\beta\hbar\omega} |n\rangle\langle n|, \tag{2.1}$$

where the $|n\rangle$ are orthonormal eigenstates of the number operator $N = a^+a$.

Formally, we may exhibit a pure quantum state which, for observables diagonal on the oscillator number basis, gives the same expectation values as are obtained from the thermal density operator. This state is

$$|\psi_\beta\rangle = \sqrt{1 - e^{-\beta\hbar\omega}} \sum_{n=0}^{\infty} e^{-n\beta\hbar\omega/2} |n\rangle. \tag{2.2}$$

As an example, the familiar result for the average value of the number operator,

$$\langle N\rangle = 1/(e^{\beta\hbar\omega} - 1) \tag{2.3}$$

is obtained as a pure quantum expectation value. For general observables, we may make use of the time dependent state associated with (2.2),

$$|\psi_\beta(t)\rangle = \sqrt{1 - e^{-\beta\hbar\omega}} \sum_{n=0}^{\infty} e^{-n\beta\hbar\omega/2 - in\omega t} |n\rangle, \tag{2.4}$$

to obtain ρ by time averaging over the oscillator period τ:

$$\rho = (1/\tau) \int_0^\tau dt\, |\psi_\beta(t)\rangle\langle\psi_\beta(t)|. \tag{2.5}$$

These somewhat formal constructs acquire a more tangible significance from the fact that, for the classical oscillator, averaging over a period is equivalent to averaging over a completely random phase. With this in mind, we rewrite Eq. (2.5) as

$$\rho = \int_0^{2\pi} \frac{d\phi}{2\pi} |\psi_\beta(\phi)\rangle\langle\psi_\beta(\phi)|, \tag{2.6}$$

with $|\psi_\beta(\phi)\rangle$ defined by

$$|\psi_\beta(\phi)\rangle = \sqrt{1 - e^{-\beta\hbar\omega}} \sum_{n=0}^{\infty} e^{-n\beta\hbar\omega/2} e^{in\phi} |n\rangle. \tag{2.7}$$

Thus, the time averaging of Eq. (2.5) is replaced by an averaging over a set of time independent states parametrized by the quantity ϕ.

Of course, the resemblance of Eq. (2.6) to classical phase averaging does not guarantee that ϕ is related to the phase. The relation can be established, however, on the basis of a previously published mathematical analysis[7] of oscillator phase operators in which new states of the oscillator, the phase states, were introduced. These are eigenstates of the operator $U = C + iS$, where C and S are the "cosine" and "sine" operators and satisfy

$$[C,N] = iS, \quad [S,N] = -iC, \tag{2.8}$$

or

$$[U,N] = U. \tag{2.9}$$

The simplest choice for U, namely, $U = E$, the unit shift operator, defined by

$$E|n\rangle = (1 - \delta_{on})|n-1\rangle, \tag{2.10}$$

leads to phase states of precisely the form (2.7) with the eigenvalue z identified as

$$z = |z|\, e^{i\phi} = e^{i\phi - \beta\hbar\omega/2} = \sqrt{\frac{\langle N\rangle}{\langle N\rangle + 1}}\, e^{i\phi}. \tag{2.11}$$

Thus the averaging in Eq. (2.6) is carried out with respect to a parameter which, in the limit $\langle N\rangle \to \infty$ (i.e., $\beta\hbar\omega \ll 1$, or $|z| \to 1$), is formally identifiable with the classical oscillator phase.[8]

This formal identification of ϕ with the classical phase can be made intuitive by noting the following behavior, which is established mathematically in the Appendix: for $\phi = 0$, the expectation value of the coordinate $\langle q\rangle$ becomes indefinitely large with large $\langle N\rangle$. Furthermore, the fractional uncertainty $\delta q/\langle q\rangle$ approaches a nonzero constant value less than unity. Therefore, δq also becomes very large, but remains less than $\langle q\rangle$, indicating that the probability distribution covers primarily the positive real axis.[9] At the same time, the expectation value of the momentum $\langle p\rangle$ vanishes, and the uncertainty δp becomes vanishingly small as $\langle N\rangle$ becomes infinite. Thus p is sharply defined about zero. When $\phi = \pi/2$, the roles of p and q are reversed, with the p distribution being smeared out over positive values and q being sharply defined about the origin.

More descriptively, we can picture a classical ensemble of oscillators with different amplitudes but identical phases. Thus, $\phi = 0$ sees them strung out to the right of the origin and at rest, while $\phi = \pi/2$ sees them all located at the origin with different momenta, but moving in the same direction.

The phase states differ in this respect from the well-known coherent states, which are eigenstates of the annihilation operator with eigenvalue α, expressible as

$$\alpha = (1/\sqrt{2})(\langle q\rangle + i\langle p\rangle) = \sqrt{\langle N\rangle}\, e^{i\phi}. \tag{2.12}$$

Here the parameter ϕ also becomes interpretable[3] as the phase of the oscillator in the limit of large $\langle N\rangle$. But in this case, *all* the dynamical quantities, q, p, N, etc., become essentially classical, having fractional uncertainties which vanish in this limit. The classical behavior is evidenced by the fact, shown by Glauber,[10] that the thermal density operator (2.6) can be expressed in terms of the coherent states as[11]

$$\rho = \frac{1}{\pi \mathfrak{N}} \int d^2\alpha\, e^{-|\alpha|^2/\mathfrak{N}}\, |\alpha\rangle\langle\alpha| . \tag{2.13}$$

The distribution function in Eq. (2.13) is Gaussian, as one would expect classically. In the limit of small β, it goes over into the Boltzman function and becomes interpretable as a probability distribution.[10]

In analogous fashion, the uniform distribution over ϕ in Eq. (2.6) becomes interpretable as a uniform probability distribution in phase for large $\langle N\rangle$. The distinguishing property of the phase states is that, for them, only the phase becomes classically definable, the other quantities retaining finite fractional uncertainties. Thus the density operator acquires a new interpretation. The statistics associated with phase independent variables, such as the energy, are purely quantum-mechanical, while the total statistical picture emerges as a result of uniform phase averaging.

In view of this close relationship between phase states and thermal averaging, it becomes a matter of considerable interest to find a model in which these states are generated as a result of some interaction. It is with this aim in mind that we undertake in this paper a general mathematical formulation of the problem of finding Hamiltonians which generate states of this type from the ground state and transform them into one another.

3. GENERAL FORMULATION

We begin by noting[12] that the phase states form a nondegenerate, overcomplete set of eigenstates of the nonunitary shift operator E, whose spectrum consists of the unit circle in the complex plane. An interior point z of the circle corresponds to a phase state

$$|z\rangle = \sqrt{1-|z|^2} \sum_{n=0}^{\infty} z^n |n\rangle . \tag{3.1}$$

These properties bear a strong resemblance to those of the coherent states, where the relevant shift operator is the familiar annihilation operator $E\sqrt{N}$, and the spectrum consists of the entire complex plane.

An important feature of the coherent states is that they retain their character as eigenstates of the annihilation operator under the influence of linear, c-number driving forces, that is, the class of Hamiltonians of the form

$$H = c_1 N + c_2 q + c_3 p \tag{3.2}$$

generates unitary transformations of these states into themselves.

In seeking analogous Hamiltonians for the phases states we were led to the following general formulation of the problem:

(i) Consider an annihilation type operator $A = EF(N)$, so that

$$A|n\rangle = F(n)|n-1\rangle \tag{3.3}$$

with $F(0) = 0$, but $F(n)$ nonvanishing for $n \neq 0$. There is no loss of generality in assuming $F(n)$ to be real and positive.[7] We assume further than $F(n)$ converges to a finite nonzero limit as $n \to \infty$. A solution of the eigenvalue equation

$$A\|\alpha\rangle = \alpha\|\alpha\rangle \tag{3.4}$$

is given by

$$\|\alpha\rangle = |0\rangle + \sum_{n=1}^{\infty} \frac{\alpha^n}{\prod_{m=1}^{n} F(m)} |n\rangle, \tag{3.5}$$

in which we assume that the $F(n)$ are such that $\|\alpha\rangle$ is normalizable for $|\alpha| < R$ where R is some nonzero positive constant. The double bar notation is used to denote states with the normalization (3.5) in which the coefficient of the ground state is unity and the coefficients of the other states are analytic in α. The eigenstate with unit norm is denoted by $|\alpha\rangle$. The spectrum[7] of A then consists of the interior and circumference of a circle of radius R in the complex plane. The states $\|\alpha\rangle$ form a nondegenerate, overcomplete[13] set.

(ii) We now show that a necessary and sufficient condition that a Hamiltonian H generates transformations of the $\|\alpha\rangle$ into themselves is that[14]

$$[A, H] = f(A), \tag{3.6}$$

where the notation $f(A)$ is understood to mean that the eigenstates of A are also eigenstates of $[A, H]$. It will be seen, in fact, that f is an analytic function of its argument.

The proof of necessity proceeds from the fact that if H generates unitary transformations of the $\|\alpha\rangle$ into themselves, it follows that

$$Ae^{-itH}|\alpha\rangle = \lambda e^{-itH}|\alpha\rangle, \tag{3.7}$$

where λ is a number which is dependent on t and α. In infinitesimal form,

$$(1 + i\delta t H)A(1 - i\delta t H)|\alpha\rangle = (\alpha - i\delta t f)|\alpha\rangle, \tag{3.8}$$

where we have written $\lambda = \alpha - i\delta t f$, and f is independent of t. Then,

$$[A, H]|\alpha\rangle = f|\alpha\rangle . \tag{3.9}$$

The dependence of f on α is obtained from the relation

$$\begin{aligned} f \equiv f(\alpha) &= \langle 0|[A,H]\|\alpha\rangle \\ &= \langle 0|[A,H]|0\rangle + \sum_{n=1}^{\infty} \frac{\langle 0|[A,H]|n\rangle}{\prod_{m=1}^{n} F(m)} \alpha^n, \end{aligned} \tag{3.10}$$

which shows[15] that $f(\alpha)$ is analytic in α and therefore that $f(A)$ is well defined.

Sufficiency follows from the fact that Eq. (3.8) is a direct consequence of assuming Eq. (3.6)

(iii) We observe now the important fact that the set of all operators H which satisfy Eq. (3.6) for a given A constitutes a Lie algebra. For, if H_1 and H_2 are members of this set whose commutators with A are $f_1(A)$ and $f_2(A)$, respectively, it follows that

$$[A, \lambda_1 H_1 + \lambda_2 H_2] = \lambda_1 f_1(A) + \lambda_2 f_2(A), \tag{3.11}$$

and

$$[A, [H_1, H_2]] = f_2(A)\frac{df_1(A)}{dA} - f_1(A)\frac{df_2(A)}{dA}, \tag{3.12}$$

which shows that the set is linear and closed with respect to commutation. Eq. (3.12) is deduced from the Jacobi identity and the analyticity of f_1 and f_2.

It is a remarkable fact that, within the framework outlined above, it is possible to deduce the form of all

Hermitian operators H associated in this way with a generalized destruction operator A. Further, all the A operators separate into two classes, one of which leads to Hamiltonians belonging to physically uninteresting Abelian Lie algebras, while the other contains the shift operator E as a special case and the standard annihilation operator as a limiting case. This latter class provides the physically interesting Hamiltonians.

In demonstrating these results it is convenient to work with the states $\| \alpha\rangle$ as defined in Eq. (3.5) because they permit differential representations of operators. Thus we have from Eq. (3.8)

$$(1 - i\delta t H) \| \alpha\rangle = (1 - i\delta t\, g(\alpha, \alpha^*)) \| \alpha - i\delta t f(\alpha)\rangle = \left(1 - i\delta t \left(g(\alpha, \alpha^*) + f(\alpha) \frac{\partial}{\partial \alpha}\right)\right) \| \alpha\rangle . \tag{3.13}$$

The function $g(\alpha, \alpha^*)$, to be determined, serves not only to represent a possible phase factor, but also to preserve the normalization (3.5), and is therefore not necessarily real. Then,

$$H \| \alpha\rangle = [f(\alpha) \frac{\partial}{\partial\alpha} + g(\alpha, \alpha^*)] \| \alpha\rangle. \tag{3.14}$$

We must now apply the requirements of Hermiticity to H, which is done by requiring that

$$\langle\beta \| H \| \alpha\rangle = \langle\alpha \| H \| \beta\rangle^* \tag{3.15}$$

for all α and β. If we write $\| \alpha\rangle$ as

$$\| \alpha\rangle = \sum_{n=0}^{\infty} h_n \alpha^n \,|n\rangle, \tag{3.16}$$

and define the function $\psi(\zeta)$ by

$$\psi(\zeta) = \sum_{n=0}^{\infty} h_n^2 \zeta^n, \tag{3.17}$$

$\psi(\zeta)$ is analytic within the circle $|\zeta| < R^2$, where R is the spectral radius of A. The matrix element in Eq. (3.15) becomes

$$\langle\beta \| H \| \alpha\rangle = [f(\alpha) \frac{\partial}{\partial\alpha} + g(\alpha, \alpha^*)]\psi(\alpha\beta^*). \tag{3.18}$$

Setting $\beta = 0$ in Eq. (3.15) gives

$$g(\alpha, \alpha^*) = f(0)^* h_1^2 \alpha + g(0)^*, \tag{3.19}$$

which shows that g is analytic and, in fact, linear in α and that $g(0)$ is real. Putting Eq. (3.19) into Eq. (3.18), differentiating the latter with respect to β^* and then setting $\beta^* = 0$ yields the following equation for f,

$$f(\alpha) = f_0 + f_1^* \alpha + \frac{f_0^* h_1^2}{k+1} \alpha^2, \tag{3.20}$$

where $f_0 = f(0)$, $f_1 = f'(0)$ and

$$1/(k+1) = (2h_2^2/h_1^4) - 1. \tag{3.21}$$

Differentiating $f(\alpha)$ once with respect to α and setting $\alpha = 0$ shows that f_1 is real. Finally, using Eq. (3.20) with the Hermiticity requirement leads to

$$(f_0\beta^* - f_0^*\alpha)\left[\psi'(\zeta) - \frac{h_1^2}{k+1} \zeta\psi'(\zeta) - h_1^2 \psi(\zeta)\right] = 0, \tag{3.22}$$

with $\zeta = \alpha\beta^*$. A moment's reflection now shows that we must have either

(A) $f_0 = 0$

or

(B) $\psi'(\zeta) - [h_1^2/(k+1)]\zeta\psi'(\zeta) - h_1^2\psi = 0.$

Case (A) may be disposed of quickly by noting that Eq. (3.14) becomes

$$H \| \alpha\rangle = [f_1 \alpha \frac{\partial}{\partial\alpha} + g(0)] \| \alpha\rangle = [f_1 N + g(0)] \| \alpha\rangle . \tag{3.23}$$

Once stated in general operator form, the restriction to the specific normalization defined by $\| \alpha\rangle$ becomes unnecessary and we see that H is a real, linear combination of N and the unit operator I. It is therefore an element of the two-dimensional Abelian Lie algebra $U(1) \times U(1)$. The transformations generated by N and I are physically uninteresting in that they merely represent the unperturbed oscillator.

The differential equation in case (B), along with the condition $\zeta(0) = 1$, has the solution

$$\psi(\zeta) = \{1 - [(h_1^2 \zeta)/(k+1)]\}^{-(k+1)}. \tag{3.24}$$

In order to match this with Eq. (3.17) we note that there is always a finite circle in the ζ-plane within which we can expand Eq. (3.24) in a convergent power series

$$\psi(\zeta) = \sum_{n=0}^{\infty} \frac{\Gamma(k+n+1)}{n!\,\Gamma(k+1)} \left(\frac{h_1^2\zeta}{k+1}\right)^n. \tag{3.25}$$

Note the implicit requirement that $k > -1$ to ensure that the coefficients in the power series all be positive in conformity with the definition of Eq. (3.17). The radius of convergence of the above power series is the square of the spectral radius of A.

Taking the square root of the ratio of successive coefficients in Eq. (3.25) shows the $F(n)$ of Eq. (3.3) to be

$$F(n) = (1/h_1)\sqrt{n(k+1)/(k+n)}. \tag{3.26}$$

The factor $1/h_1$ is a scale factor which plays no essential role in the eigenstates of the resulting A operators. The choice $h_1 = 1$ results in $k = 0$ corresponding to the unit shift operator E, while $k \to \infty$ corresponds, as we shall see, to the annihilation operator. We refer to the resulting A as the generalized destruction operator $A(k)$, i.e.,

$$A(k) = E\sqrt{N(k+1)/(N+k)}. \tag{3.27}$$

It follows from this expression that the spectral radius of $A(k)$ is $\sqrt{k+1}$. Thus, in terms of a single real parameter k, we have the general form of the annihilation type operator associated with Case (B). It is now a simple matter to exhibit the associated eigenstates $|\alpha, k\rangle$ in normalized form. The normalization factor follows from the definition of $\psi(\zeta)$ in Eq. (3.17) with $\zeta = |\alpha|^2$ and from the expression for $\psi(\zeta)$ in Eq. (3.24). The result is

$$|\alpha, k\rangle = \left[1 - \frac{|\alpha|^2}{k+1}\right]^{(k+1)/2} \times \sum_{n=0}^{\infty} \sqrt{\frac{\Gamma(n+k+1)}{n!\,\Gamma(k+1)}} \left(\frac{\alpha}{\sqrt{k+1}}\right)^n |n\rangle. \tag{3.28}$$

It now remains to deduce the form of the associated Hamiltonian and to show that it is an arbitrary linear combination of interaction terms and the free oscillator term. Our starting point is once again Eq. (3.14), into which we substitute the expression in Eq. (3.20). This gives

$$H \| \alpha\rangle = \left[\left(f_0 + f_1\alpha + \frac{f_0^*}{k+1}\alpha^2\right)\frac{\partial}{\partial\alpha} + (f_0^*\alpha + g(0))\right] \| \alpha\rangle . \tag{3.29}$$

Using

$$\frac{\partial}{\partial\alpha}\left\|\alpha\right\rangle = \sqrt{\frac{N(N+k)}{k+1}}\; E^+ \left\|\alpha\right\rangle, \tag{3.30}$$

and

$$\alpha^2 \frac{\partial}{\partial\alpha}\left\|\alpha\right\rangle = \sqrt{\frac{N(N+k)}{k+1}}\, E^+ A^2 \|\alpha\rangle, \tag{3.31}$$

we are again freed of the $\|\alpha\rangle$ normalization and obtain, after some rearrangement,

$$\begin{aligned} H &= (f_0/\sqrt{k+1})H_+ + (f_0^*/\sqrt{k+1})H_- + f_1 N + g(0) \\ &= C_0 + C_1H_1 + C_2H_2 + C_3H_3, \end{aligned} \tag{3.32}$$

where

$$H_1 + iH_2 \equiv H_+ = H_-^\dagger = \sqrt{N(N+k)}\,E^+, \tag{3.33a}$$

$$H_3 = N + (k+1)/2 \tag{3.33b}$$

and C_0, C_1, C_2, C_3 are independent real numbers. The Hamiltonian of Eq. (3.32) is an element of the Lie algebra of $O(2,1) \times U(1)$, since H_1, H_2, and H_3 satisfy the commutation relations of $O(2,1)$. The basic equation (3.6) takes the form

$$\begin{aligned} &[A, H_3] = A, \\ &[A, H_+] = \sqrt{k+1}, \\ &[A, H_-] = A^2/(k+1)^{1/2}. \end{aligned} \tag{3.34}$$

4. DYNAMICS OF THE $|\alpha, k\rangle$ STATES

In this section we turn our attention to the group of transformations induced on the $|\alpha, k\rangle$ states by unitary operators e^{-itH}, where H is a Hamiltonian of the form (3.32). Since the eigenvalue transformations are unaffected by constant terms in the Hamiltonian, we restrict ourselves to the operators H_1, H_2, H_3 by choosing $g(0) = f_1[(k+1)/2]$. The resulting Hamiltonian will be referred to as $H(k)$, i.e.,

$$H(k) = (1/\sqrt{k+1})(f_0 H_+ + f_0^* H_-) + f_1 H_3. \tag{4.1}$$

Using the pertinent results of the previous section, we can rewrite Eq. (3.13) as

$$(1 - i\delta t H(k)) \| \alpha, k\rangle = (1 - i\delta t g(\alpha)) \| \alpha - i\delta t f(\alpha), k\rangle. \tag{4.2}$$

It then follows that

$$\begin{aligned} e^{-itH(k)} \| \alpha, k\rangle &= \lim_{n\to\infty}\left(1 - i\frac{t}{n}H(k)\right)^n \left\| \alpha, k\right\rangle \\ &= \exp\{-i \textstyle\int_0^t dt\, g[\alpha(t)]\} \| \alpha(t), k\rangle, \end{aligned} \tag{4.3}$$

where $\alpha(t)$ satisfies the Riccati type[16] differential equation

$$\frac{d\alpha(t)}{dt} = -if[\alpha(t)] = -i\left[f_0 + f_1\alpha(t) + \frac{f_0^*}{k+1}\alpha^2(t)\right], \tag{4.4}$$

with initial condition $\alpha(0) = \alpha$.

The trivial case $f_0 = 0$ represents the free oscillator with the obvious result that $\alpha(t) = \alpha e^{-if_1 t}$. For $f_0 \neq 0$, we make the substitution

$$\alpha(t) = -i\frac{k+1}{f_0^*}\frac{1}{w(t)}\frac{dw(t)}{dt}. \tag{4.5}$$

The function $w(t)$ need be determined only to within a multiplicative constant. This can be done by noting that it satisfies the second order differential equation

$$\frac{d^2w}{dt^2} + if_1\frac{dw}{dt} - \frac{|f_0|^2}{k+1}w = 0. \tag{4.6}$$

Imposing the condition $\alpha(0) = \alpha$ on the general solution of this equation gives

$$w(t) = e^{-if_1t/2}\left[\cos st + i\left(\frac{f_1}{2} + \frac{f_0^*\alpha}{k+1}\right)\frac{\sin st}{s}\right], \tag{4.7}$$

where

$$s = \begin{cases} \left(\dfrac{f_1^2}{4} - \dfrac{|f_0|^2}{k+1}\right)^{1/2}, & \dfrac{f_1^2}{4} > \dfrac{|f_0|^2}{k+1}, \\[2ex] i\left(\dfrac{|f_0|^2}{k+1} - \dfrac{f_1^2}{4}\right)^{1/2}, & \dfrac{f_1^2}{4} < \dfrac{|f_0|^2}{k+1}. \end{cases} \tag{4.8}$$

This in turn leads to the results that $\alpha/\sqrt{k+1}$ undergoes the linear fractional transformation

$$\alpha(t)/\sqrt{k+1} = \frac{a(\alpha/\sqrt{k+1}) + b}{b^*(\alpha/\sqrt{k+1}) + a^*} \tag{4.9}$$

with

$$a = \cos st - i(f_1/2)(\sin st/s), \tag{4.10a}$$

and

$$b = -(if_0/\sqrt{k+1})(\sin st/s). \tag{4.10b}$$

Thus $|a|^2 - |b|^2 = 1$, so that the transformation has unit determinant.[17] It can be verified also that it maps the interior of the spectral circle $|\alpha| < \sqrt{k+1}$ into itself.

Having established the transformation properties of the eigenvalues within the spectral circle, it now remains to evaluate the multiplier $\exp\{-i\int_0^t dtg[\alpha(t)]\}$ in Eq. (4.3), which, it will be recalled, does not generally have unit magnitude because it is defined with respect to the $\|\alpha, k\rangle$. From Eq. (4.5) and the expression for $g(\alpha)$, it follows immediately that

$$\int_0^t dtg[\alpha(t)] = +\frac{k+1}{2}f_1 t - i(k+1)\int_0^t dt\,\frac{1}{w(t)}\frac{dw(t)}{dt}, \tag{4.11}$$

which leads to the multiplier

$$\begin{aligned} &\exp\{-i\textstyle\int_0^t dtg[\alpha(t)]\} \\ &\quad = \exp\left\{-(k+1)\ln\left[\cos st + i\left(\frac{f_1}{2} + \frac{f_0^*\alpha}{k+1}\right)\frac{\sin st}{s}\right]\right\} \end{aligned} \tag{4.12}$$

The branch of the logarithm in the exponent is determined by taking the principal value zero at $t = 0$, then demanding continuity in t.

The results obtained thus far in this section are summarized in the formula

$$\exp\left\{-it\left(\frac{f_0H_+ + f_0^*H_-}{\sqrt{k+1}} + f_1H_3\right)\right\}\,\|\,\alpha,k\rangle$$
$$= \exp\left\{-(k+1)\ln\left[\cos st + i\left(\frac{f_1}{2}+\frac{f_0^*\alpha}{k+1}\right)\frac{\sin st}{s}\right]\right\}$$
$$\times \left\|\sqrt{k+1}\;\frac{a\alpha + b\sqrt{k+1}}{b^*\alpha + a^*\sqrt{k+1}},k\right\rangle, \tag{4.13}$$

where a, b, and s are given by Eqs. (4.8) and (4.10).

We are now in a position to exhibit unitary transformations which transform the ground state to an arbitrary (normalized) state $|\alpha,k\rangle$. Setting $\alpha(0)\equiv\alpha=0$ in Eq. (4.9), it is apparent from Eq. (4.10) that an arbitrary $\alpha(t)$ is easily attained by making the choice $f_1=0$. For we then have

$$\frac{\alpha(t)}{\sqrt{k+1}} = \frac{b}{a^*} = -\frac{if_0}{|f_0|}\tanh\left(\frac{|f_0|t}{\sqrt{k+1}}\right) \tag{4.14}$$

and

$$\exp\left\{-i\int_0^t dt\, g[\alpha(t)]\right\} = \left[\cosh\left(\frac{|f_0|t}{\sqrt{k+1}}\right)\right]^{-(k+1)}$$
$$= \left[1-\tanh^2\left(\frac{|f_0|t}{\sqrt{k+1}}\right)\right]^{(k+1)/2}. \tag{4.15}$$

Looking at Eq. (3.28) we see that (4.14) provides just the proper normalization factor, as it must, since $\|0\rangle = |0\rangle$. The resulting unitary operator can be put into a particularly simple form which does not depend explicitly on the parameter t by choosing

$$f_0 = ie^{i\phi}, \qquad t = \rho\sqrt{k+1}. \tag{4.16}$$

This leads to the relations

$$D_k(\rho,\phi)\,|0\rangle = |\,e^{i\phi}\sqrt{k+1}\tanh\rho\rangle, \tag{4.17}$$

where the unitary operator $D_k(\rho,\phi)$ is defined by[18]

$$D_k(\rho,\phi) \equiv \exp\{\rho(e^{i\phi}H_+ - e^{-i\phi}H_-)\}. \tag{4.18}$$

The states $|z\rangle$ of Eq. (3.1) are obtained by setting $k=0$, so that

$$D_0(\rho,\phi) = \exp\{\rho(e^{i\phi}NE^+ - e^{-i\phi}EN)\}$$
$$= \exp\{\rho(e^{i\phi}\sqrt{N}a^+ - e^{-i\phi}a\sqrt{N})\}, \tag{4.19}$$

and

$$D_0(\rho,\phi)|0\rangle = |z\rangle, \quad z = e^{i\phi}\tanh\rho. \tag{4.20}$$

These results, and those of the previous section, provide the operators called for by the arguments of Sec. 2. In particular, we note that the interaction Hamiltonians which generate phase states are linear in $a\sqrt{N}$ and $\sqrt{N}a^+$, but not in q and p.

An interesting property of the Hamiltonians (4.1) appears when we consider the eigenvalue transformations of Eq. (4.9) in more detail. The nature of the transformations depends on the value of the parameter s. For each value of k, the Hamiltonians (4.1) fall into two classes according to whether the parameters f_0 and f_1 are such that s is real or pure imaginary as indicated in Eq. (4.8).

The difference between the two classes is best understood if we consider the points which are invariant under the transformation of eigenvalues induced by each Hamiltonian.[19] These can be determined either directly from Eq. (4.9), or by setting the right-hand side of the differential equation (4.4) equal to zero. There are generally two such points, given by

$$\alpha = (k+1)(-f_1 \pm 2s)/2f_0^*. \tag{4.21}$$

It is not difficult to see that for the first class of Hamiltonians (s real), both invariant points lie on the same ray from the center of the spectral circle, one inside the circle and one outside. For the other class (s pure imaginary), both points lie on the circumference of the circle. For the case $s=0$, which properly belongs to the second class, the two invariant points coincide, and lie on the circumference.

The eigenvalue transformations induced by the first class of Hamiltonians are periodic in t with period π/s. The eigenvalue α, starting from any initial value inside the spectral circle, follows a closed trajectory which circumscribes the invariant point within the circle.[20] Thus, the free oscillator mean energy, related to $|\alpha|$, varies periodically with t for these Hamiltonians.

As $s\to 0$, the trajectories are pinched between the two coalescing invariant points, so that in the limit, all trajectories pass through this single point, which now lies on the circumference of the spectral circle. At the same time, the period becomes infinite.

This behavior implies that the asymptotic state of the system for $t\to\pm\infty$ is independent of the initial state, and is determined only by the parameters f_0 and f_1 that specify the Hamiltonian.[21] This final state has infinite mean energy and sharp phase resolution. A Hamiltonian of this class therefore leaves its imprint on the system in the value of the phase of its final state.

If the parameter s is pure imaginary, the trajectories pass through both invariant points on the circumference of the circle $|\alpha| = \sqrt{k+1}$. There are then two asymptotic states of infinite energy, one for $t\to+\infty$ and one for $t\to-\infty$. The behavior is otherwise similar to the case $s=0$.

It is intuitively obvious from Eq. (3.27) that the limit $k\to\infty$ should lead to the well-known coherent states

$$|\alpha\rangle = e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}\,|n\rangle. \tag{4.22}$$

This follows rigorously from the fact that, for a given α, we can choose k sufficiently large so that α is within the spectral radius of $A(k)$ and then compute $\|\,|\alpha\rangle - |\alpha,k\rangle\,\|^2$. Noting that $\langle\alpha\,|\alpha,k\rangle$ is real,

$$\|\,|\alpha\rangle - |\alpha,k\rangle\,\|^2 = 2(1-\langle\alpha\,|\,\alpha,k\rangle), \tag{4.23}$$

where

$$\langle\alpha\,|\,\alpha,k\rangle = e^{-|\alpha|^2/2}\left(1-\frac{|\alpha|^2}{k+1}\right)^{(k+1)/2}$$
$$\times\sum_{n=0}^{\infty}\left(\frac{\Gamma(n+k+1)}{\Gamma(k+1)(k+1)^n}\right)^{1/2}\frac{|\alpha|^{2n}}{n!}. \tag{4.24}$$

Since

$$\frac{\Gamma(n+k+1)}{\Gamma(k+1)(k+1)^n} \ge 1, \tag{4.25}$$

we have

$$e^{|\alpha|^2/2}\left(1-\frac{|\alpha|^2}{k+1}\right)^{(k+1)/2} \le \langle\alpha\,|\,\alpha,k\rangle \le 1, \tag{4.26}$$

so that

$$\langle \alpha \mid \alpha, k\rangle \underset{k\to\infty}{\longrightarrow} 1, \tag{4.27}$$

which proves the point.

Similarly, the behavior of the unitary operator $D_k(\rho,\phi)$ defined by Eq. (4.17) is of interest. Using essentially heuristic arguments, we see[22] that for a fixed

$$\alpha = e^{i\phi}\sqrt{k+1}\,\tanh\rho, \tag{4.28}$$

ρ must become small as $k \to \infty$. Thus $\alpha \simeq e^{i\phi}\sqrt{k}\,\rho$, and from (3.33a),

$$D_k(\rho,\phi) \to \exp\left\{\frac{\alpha}{\sqrt{k}}\sqrt{kN}\,E^+ - \frac{\alpha^*}{\sqrt{k}}\,E\sqrt{kN}\right\} = e^{\alpha a^+ - \alpha^* a}, \tag{4.29}$$

which is just Glauber's[10] $D(\alpha)$ operator.

These results encourage us to seek a resolution of the identity of the form

$$I = \int d^2\alpha\, \rho(\mid \alpha\mid^2, k) \mid \alpha, k\rangle\langle \alpha, k\mid, \tag{4.30}$$

with the weighting function $\rho(\mid\alpha\mid^2,k)$ to be determined, and $d^2\alpha = d(\mathrm{Re}\alpha)d(\mathrm{Im}\alpha)$ with integration over the circle $\mid\alpha\mid < \sqrt{k+1}$. By introducing polar coordinates in the α-plane and making the change of variable $t = \mid\alpha\mid^2/k+1$, the integral in (4.30) can be written as

$$\int d^2\alpha\, \rho(\mid\alpha\mid^2,k)\mid\alpha,k\rangle\langle\alpha,k\mid = \pi\sum_{n=0}^{\infty}\frac{\Gamma(n+k+1)}{\Gamma(n+1)\Gamma(k)}\int_0^1 dt\,\bar\rho(t,k)(1-t)^{k+1}t^n\mid n\rangle\langle n\mid, \tag{4.31}$$

where $\rho(\mid\alpha\mid^2,k) = \bar\rho(t,k)$. A resolution of the identity is obtained if

$$\int_0^1 dt\,\bar\rho(t,k)(1-t)^{k+1}t^n = \frac{1}{\pi}\,\frac{\Gamma(n+1)\Gamma(k)}{\Gamma(n+k+1)}. \tag{4.32}$$

Using

$$\int_0^1 dt\,(1-t)^{q-1}t^{p-1} = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}, \tag{4.33}$$

when the real parts of p and q are positive, we see that

$$\bar\rho(t,k) = (1/\pi)\,[1/(1-t)^2], \tag{4.34a}$$

or,

$$\rho(\mid\alpha\mid^2,k) = (1/\pi)\,[1/(1-\mid\alpha\mid^2/(k+1))^2], \tag{4.34b}$$

with the condition $k>0$. This gives the result

$$\frac{1}{\pi}\int d^2\alpha\left(1-\frac{\mid\alpha\mid^2}{k+1}\right)^{-2}\mid\alpha,k\rangle\langle\alpha,k\mid = I,\ (k>0). \tag{4.35}$$

Note that the weighting function has sufficiently singular behavior for its integral to diverge, a property which is to be expected since $\mathrm{Tr}(I) = \infty$. The condition $k > 0$ indicates that this resolution fails for the phase states (3.1). This is a particularly vexing fact in view of the physical interest attached to these states. The reason for it is most simply seen from Eq. (4.32), which shows that the condition $k = 0$ demands that all of the moments of the function $\bar\rho(t,0)(1-t)$ on the unit interval be equal. This in turn forces $\bar\rho(t,0)$ to have δ-function type behavior at $t = 1$. In a crude sense, this indicates that the unit shift operator E tries very hard to behave like a unitary operator.

5. GENERALIZATIONS AND EXTENSIONS

It should be apparent by now that the formalism developed thus far already exceeds in generality the requirements posed in Sec. 2. Quite aside from relevance to phase states there is thus the mathematically intriguing question of whether the methods developed here are useful for the study of groups.

The essential feature of our procedure is the association of a destruction operator with a unitary group in such a fashion that the eigenstates of the operator transform into themselves under the action of the group. Because eigenstates of destruction operators are of necessity infinite dimensional, this leads to infinite dimensional representations of the group. However, there is a limiting sense in which the association between destruction operator, eigenstate, and group carries over into finite dimensional spaces. In this section we illustrate this by extending our formalism to include representations of $O(3)$ as well as $O(2,1)$.

Our starting point is a consideration of the effect of removing the restriction $-1 < k < \infty$ from the parameter k by extending it to the complex plane. The operators $A(k)$ of Eq. (3.27) and the H of Eq. (3.33) remain well defined. Their matrix elements on the number basis are in fact analytic in the complex k-plane cut from -1 to $-\infty$. Further, the $\mid\alpha,k\rangle$ of Eq. (3.28) continue to be normalized eigenstates of $A(k)$ for $\mid\alpha\mid < \sqrt{\mid k+1\mid}$.

The Hermiticity of the H, which is of course generally lost in this process, can be restored for negative integral k (approached, for example, from the upper half-plane) by a simple modification based on the fact that the H now reduce the oscillator space to two invariant subspaces. Thus, let $k = -p$, where p is a positive integer greater than unity. The Eq. (3.33a) shows that

$$H_+\mid p-1\rangle = 0 \tag{5.1a}$$

and

$$H_-\mid p\rangle = 0. \tag{5.1b}$$

Therefore, a natural division of the number basis into two invariant bases is defined. The vectors $\mid n\rangle$ with $0 \le n \le p-1$ span the p-dimensional subspace X_p, while those with $n \ge p$ span the infinite dimensional subspace X_p^∞. In the latter subspace the H remain Hermitian as defined, and are in fact identical, except for a relabeling of the basic states, with those for the case $k = +p$.

We therefore fix our attention on the subspace X_p, where multiplication of $H_\pm$ by $-i$ achieves the desired result. The re-defined Hamiltonians,

$$H_1 + iH_2 \equiv H_+ = H_-^\dagger = \sqrt{N(p-N)}\,E^+ \tag{5.2a}$$

and

$$H_3 = N - (p-1)/2 \tag{5.2b}$$

satisfy the $O(3)$ commutation rules. We have thus been led naturally to the p-dimensional representation of the $O(3)$ algebra. In conventional notation, $p = 2j+1$ and $J_3 = N - j$.

The question of obvious interest at this point is the status of the $\mid\alpha,k\rangle$ states. Inspection of Eq. (3.28) shows that, as k approaches $-p$ from the upper half of the complex plane, the coefficients of $\mid\alpha,k\rangle$ in X_p^∞ becomes vanishingly small. In the limit, a well-defined p-dimensional state $\mid\alpha,p\rangle$ is obtained:

$$\Big|\alpha,p\Big\rangle = \left(1+\frac{\mid\alpha\mid^2}{p-1}\right)^{-(p-1)/2}\sum_{n=0}^{p-1}\sqrt{\binom{p-1}{n}}\left(\frac{\alpha}{\sqrt{p-1}}\right)^n\Big|n\Big\rangle, \tag{5.3}$$

where

$$\binom{p-1}{n} = \frac{(p-1)!}{n!(p-1-n)!} \tag{5.4}$$

is the standard binomial coefficient. The state $|\alpha,p\rangle$ is now defined for all α, and not merely those within some spectral radius, as it is when k is not a negative integer. Note that $A(k)$, which now becomes

$$A_p \equiv A(-p) = E\sqrt{\frac{N(p-1)}{p-N}}, \tag{5.5}$$

is well defined on X_p. Also, its commutators with the H of (5.2) satisfy the basic relation (3.6). However, the $|\alpha,p\rangle$ of (5.3) are no longer eigenstates of A in the strict sense in view of the fact that a destruction operator cannot have eigenstates in a finite dimensional space. There is, nevertheless, a limiting sense in which the $|\alpha,p\rangle$ are eigenstates of A_p. This can be seen from the equation $A(k)\,|\,\alpha,k\rangle = \alpha\,|\,\alpha,k\rangle$, which holds rigorously for k indefinitely close to $-p$. We therefore have

$$\lim_{k\to -p} A(k)\,|\,\alpha,k\rangle = \alpha\,|\,\alpha,p\rangle, \tag{5.6}$$

which, however, does not imply $A_p\,|\,\alpha,p\rangle = \alpha\,|\,\alpha,p\rangle$. In effect, the singular limiting behavior of $A(k)$ on the state $|\,p\rangle$ combines with the vanishingly small projection of $|\alpha,k\rangle$ on $|\,p\rangle$ to give a finite amplitude on $|\,p-1\rangle$. This feature is not present when A is restricted to X_p.

The above remarks suggest the retention of the concept of eigenstate of $A(k)$ even in the limit, and we thus refer to the $|\,\alpha,p\rangle$ as extended eigenstates of $A(k)$. We now show that these states also retain the property of transforming among themselves under transformations generated by the Hamiltonians of Eq. (5.2).

Our demonstration is based on the fact that the H can be represented as differential operators when acting on the nonnormalized states $\|\,\alpha,p\rangle$. The following equations are easily verified from (5.2) and (5.3):

$$H_3\,\|\,\alpha,p\rangle = \left(\alpha\frac{\partial}{\partial\alpha} - \frac{p-1}{2}\right)\|\,\alpha,p\rangle, \tag{5.7a}$$

$$H_+\,\|\,\alpha,p\rangle = \sqrt{p-1}\,\frac{\partial}{\partial\alpha}\,\|\,\alpha,p\rangle, \tag{5.7b}$$

$$H_-\,\|\,\alpha,p\rangle = \left(\sqrt{p-1}\,\alpha - \frac{1}{\sqrt{p-1}}\,\alpha^2\frac{\partial}{\partial\alpha}\right)\|\,\alpha,p\rangle. \tag{5.7c}$$

The differential relations imply that one can immediately employ the methods of Sec. 4, beginning with Eq. (4.2). In fact, one obtains, "mutatis mutandis", as the analog of Eq. (4.13), the result

$$\begin{aligned}&\exp\{-it[(p-1)^{-1/2}(f_0H_+ + f_0^*H_-) + f_1H_3]\}\,\|\,\alpha,p\rangle\\ &= \exp\left\{(p-1)\ln\left[\cos st + i\left(\frac{f_1}{2} - \frac{f_0^*\alpha}{p-1}\right)\frac{\sin st}{s}\right]\right\}\\ &\quad\times\|\,\sqrt{p-1}\,\frac{a\alpha + ib\sqrt{p-1}}{ib^*\alpha + a^*\sqrt{p-1}}, p\rangle,\end{aligned} \tag{5.8}$$

where

$$s = \left(\frac{f_1^2}{4} + \frac{|f_0|^2}{p-1}\right)^{1/2} > 0, \tag{5.9}$$

and

$$a = \cos st - i\frac{f_1}{2}\frac{\sin st}{s}, \tag{5.10a}$$

$$b = -\frac{f_0}{\sqrt{p-1}}\frac{\sin st}{s}. \tag{5.10b}$$

The analog of Eqs. (4.17) and (4.18) is

$$D_p(\rho,\phi)\,|\,0\rangle = |\,\sqrt{p-1}\,e^{i\phi}\tan\rho, p\rangle, \tag{5.11a}$$

$$D_p(\rho,\phi) = \exp\{\rho e^{i\phi}H_+ - \rho e^{-i\phi}H_-\}. \tag{5.11b}$$

With these formulas, our extension of the formalism to $O(3)$ is complete.

We can characterize the results of this section in group theoretical terms by saying that we pass continuously from unitary representations of $O(2,1)$ to the unitary representations of $O(3)$ via a continuum of nonunitary representations of $O(2,1)$. Thus, representations of the compact group appear at isolated values of the parameter which labels the representations of the noncompact group.

6. REPRESENTATIONS OF $O(3)$ AND $O(2,1)$

We have shown in the preceding sections that the groups underlying the dynamics of phase states are $O(2,1)$ and $O(3)$. These groups are realized in our formalism in two distinct way: (i) as mappings of the eigenvalues of the generalized destruction operator, and (ii) as transformations of the Hilbert space of harmonic oscillator states. The transformations involved are familar.[23-25] In the case of $O(2,1)$, the eigenvalue mappings are the linear fractional transformations obtained by stereographic projection of the unit hyperboloid onto the complex plane. The group of such transformations is known to be isomorphic with $O(2,1)$, and can therefore be identified with this group. A similar situation obtains with respect to $O(3)$ and projections of the unit sphere. The spectrum of the generalized destruction operator can therefore be regarded as the supporting space of the underlying groups $O(2,1)$ and $O(3)$. The transformations of Hilbert space on the other hand are simply the linear irreducible representations of the two groups.

The generalized destruction operator brings the supporting spaces of the group and of its representations into particularly close association. It is an operator which is defined on the representation space and whose spectrum serves as the space on which the group itself acts. Further, the mapping of the spectrum is associated with a mapping of eigenstates by the group, and thereby determines the transformation in Hilbert space associated with a particular group element. It will be shown in a forthcoming paper that these features are general and apply to groups other than $O(3)$ and $O(2,1)$. Here, we content ourselves with showing that some of the familiar results pertaining to representations of the $O(2,1)$ and $O(3)$ groups appear in the present context.

We begin by identifying the representations that have been obtained. In the case of $O(2,1)$, irreducible representations are characterized[25] by the value of the Casimir operator $Q = -H_1^2 - H_2^2 + H_3^2$ and the spectrum of the generator H_3. These are found from Eq. (3.33) in our case:

$$Q = \tfrac{1}{4}(k^2 - 1), \tag{6.1}$$

$$H_3 = n + \tfrac{1}{2}(k+1), \quad n = 0,1,2,\cdots. \tag{6.2}$$

This infinite family of representations, labeled by the real parameter k, coincides with the representations $D^-(\Phi)$ of Ref. (25), with $\Phi = -\frac{1}{2}(k+1)$. The representations $D^+(\Phi)$ are also obtained in our formalism by a different choice of labeling for the generators, which corresponds to the substitutions $H_3 \to -H_3$ and $H_2 \to -H_2$. These are multivalued representations of $O(2,1)$ unless k is an odd integer. Our basic requirement associating a generalized destruction operator with a representation therefore leads to all the unitary representations of $O(2,1)$ in which the spectrum of H_3 is bounded either from above or from below. The remaining representations, in which H_3 has an unbounded spectrum, cannot be supported on the harmonic oscillator basis. They can be obtained, however, by our method if an extension of the basis to negative n is made. The details of this extension will be given in a future paper.

The case of $O(3)$ is simpler. It is evident from the remarks following Eq. (5.2) and from the allowed values of p that we obtain all of the unitary representations except for the trivial one belonging to the eigenvalue zero of the Casimir operator $Q = J^2 = H_1^2 + H_2^2 + H_3^2$. Thus the generalized destruction operator and its eigenstates bring into association the unitary representations of $O(3)$ and the bounded unitary representations of $O(2,1)$.

If we now turn our attention to the eigenstates (3.28) and (5.3), we find that they have a simple group theoretic significance: they are the transforms under the operators of the group of the single state $|0\rangle$. The existence of the operators $D_k(\rho,\phi)$ and $D_p(\rho,\phi)$ of Eqs. (4.16) and (5.11) guarantees that all the eigenstates can be obtained from the ground states in this way, and the basic property (3.6) insures that we obtain only eigenstates.

In the case of the rotation group, the ground state corresponds, for a given representation $j = \frac{1}{2}(p-1)$, to the state with $H_3 = -j$, i.e., the state with "spin down" with respect to the 3-axis. The rotation which transforms this state into another extended eigenstate of the generalized destruction operator simply rotates this spin to some other direction in space. We may therefore characterize the eigenstates as those states having the minimum value of the component of angular momentum in some definite direction.

A similar characterization is possible for $O(2,1)$ if we interpret H_1 and H_2 as the generators of pure Lorentz transformations in two orthogonal spatial directions, and H_3 as the generator of rotations in the plane of these directions. Here again, although the eigenvalues of H_3 are no longer restricted to integers or half-integers, the state $|0\rangle$ corresponds to the minimum eigenvalue. Therefore the eigenstates of the generalized destruction operator are just those states which have minimum eigenvalue of H_3 in some definite Lorentz frame.

Another interesting description of the eigenstates is obtained by considering their components $\langle n|\alpha\rangle$ with respect to the number basis. The above remarks indicate that these components can be written in the form $\langle n|D|0\rangle$ for some operator D belonging to the representation. Furthermore, the quantities $\langle n|D|0\rangle$ for an arbitrary operator of the representation form the components of some eigenstate of the generalized destruction operator. This shows that the first columns of the matrix representatives of all the operators of the representation comprise the class of eigenvectors of the generalized destruction operator.

The full matrix for an operator D of exponential form is also calculable from the eigenstates of the generalized destruction operator. The results are not new and the derivation is similar to treatments found in the literature,[24] but the method given here offers a considerable conceptual simplification. Beginning from Eqs. (4.13) and (5.8), we obtain an equation

$$\langle n|D\|\alpha\rangle = \exp\{-i\int_0^t dt\, g[\alpha(t)]\}\langle n\|\alpha(t)\rangle \tag{6.3}$$

in which the left-hand side is expressible as a convergent power series in α with coefficients proportional to $\langle n|D|n'\rangle$, and the right-hand side is a known analytic function of α for a given D. We can therefore obtain $\langle n|D|n'\rangle$ by comparison of coefficients on either side of Eq. (6.3). In this way we can obtain, for example, the familiar result[26] for the matrix elements of $\exp(-i\beta H_2)$ for $O(3)$, and its analog for the bounded representations $D^-(\Phi)$ of $O(2,1)$:

$$\begin{aligned}\langle n|e^{-i\psi H_2(k)}|n'\rangle &= \left(\frac{n'!\,\Gamma(n+k+1)}{n!\,\Gamma(n'+k+1)}\right)^{1/2}\\ &\times\sum_r \frac{n!}{(n'-r)!(n-n'+r)!}\\ &\times(-1)^{n-n'+r}\frac{\Gamma(n+k+r+1)}{r!\,\Gamma(n+k+1)}\\ &\times(\cosh\tfrac{1}{2}\psi)^{n-n'-k-1-2r}(\sinh\tfrac{1}{2}\psi)^{n-n'+2r}.\end{aligned} \tag{6.4}$$

The summation over the integer r runs over a finite range which is determined by the factorials in the denominator.

Note added in proof: Since submitting this paper it has come to our attention that some properties of the states of Eq. (5.3) have been discussed in other contexts by various authors.[27]

APPENDIX

We show here that when an oscillator is in a state $|z\rangle$ of the form (3.1) with z positive real, i.e., $\phi = 0$ in Eq. (2.11), then $\delta q/\langle q\rangle$ approaches a finite number less than unity and δp becomes vanishingly small as $\langle N\rangle \to \infty$.

It is convenient to normalize units so that

$$q = (2)^{-1/2}(a + a^+), \qquad p = [-i(2)^{-1/2}](a - a^+). \tag{A1}$$

In terms of the unit shift operator E of Eq. (2.10) we have

$$a = E\sqrt{N} = \sqrt{N+1}\,E \tag{A2}$$

and

$$a^2 = \sqrt{N+1}\,E\sqrt{N+1}\,E = \sqrt{(N+1)(N+2)}\,E^2. \tag{A3}$$

Using the fact that $E|z\rangle = z|z\rangle$ and z is real gives

$$\frac{\langle q^2\rangle}{\langle q\rangle^2} = \frac{1}{2\langle\sqrt{N+1}\rangle^2}\times\left(\frac{(\langle N\rangle+\frac{1}{2})(\langle N\rangle+1)}{\langle N\rangle} + \langle\sqrt{(N+1)(N+2)}\rangle\right), \tag{A4}$$

where $\langle N\rangle$ is related to z by Eq. (2.11). We shall see that the behavior of this expression depends strongly on the behavior of the expectation value

$$\langle\sqrt{N+1}\rangle = (1-z^2)\sum_{n=0}^{\infty}\sqrt{n+1}\,z^{2n}, \tag{A5}$$

which we now examine.

Let

$$z^2 = \frac{\langle N\rangle}{\langle N\rangle + 1} = \zeta, \tag{A6}$$

so that

$$\langle \sqrt{N+1}\rangle = \sum_{n=0}^{\infty} (\sqrt{n+1} - \sqrt{n})\zeta^n. \tag{A7}$$

Noting that

$$\sqrt{n+1} - \sqrt{n} = \frac{1}{2}\int_0^1 \frac{dr}{\sqrt{n+r}}, \tag{A8}$$

enables us to write

$$\langle \sqrt{N+1}\rangle = \frac{1}{2\sqrt{\pi}}\int_{-\infty}^{\infty} \frac{dx}{x^2}\,\frac{1-e^{-x^2}}{1-\zeta e^{-x^2}}, \tag{A9}$$

where we have used

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} dx\, e^{-(n+r)x^2} = \frac{1}{\sqrt{n+r}}. \tag{A10}$$

Expressing ζ in terms of $\langle N\rangle$ and making the change of integration variable to $y = x\sqrt{\langle N\rangle}$ puts (A9) into the form

$$\langle \sqrt{N+1}\rangle = \frac{\sqrt{\langle N\rangle}\,(1+\langle N\rangle)}{\sqrt{\pi}}\int_0^{\infty} \frac{dy}{y^2}\,\frac{1-e^{-y^2/\langle N\rangle}}{1+\langle N\rangle(1-e^{-y^2/\langle N\rangle})}. \tag{A11}$$

For very large $\langle N\rangle$ the integrand in (A11) goes as $[\langle N\rangle(1+y^2)]^{-1}$. This suggests writing $\langle\sqrt{N+1}$ as

$$\langle \sqrt{N+1}\rangle = \frac{1+\langle N\rangle}{\sqrt{\pi\langle N\rangle}}\int_0^{\infty} \frac{dy}{1+y^2}\, G\left(y^2, \frac{1}{\langle N\rangle}\right), \tag{A12}$$

where the function $G(x,\epsilon)$ is defined by

$$G(x,\epsilon) = [1 + (1/x)]/[1 + (1/f(x,\epsilon))] \tag{A13}$$

and

$$f(x,\epsilon) = (1 - e^{-\epsilon x})/\epsilon. \tag{A14}$$

It may be verified by straightforward calculation that $G(x,\epsilon)$ decreases monotonically from the value $G(0,\epsilon) = 1$ to $G(\infty,\epsilon) = 1/(1+\epsilon)$. This then gives

$$\tfrac{1}{2}\sqrt{\pi\langle N\rangle} \le \langle\sqrt{N+1}\rangle \le \tfrac{1}{2}\sqrt{\pi\langle N\rangle}\,(1 + 1/\langle N\rangle). \tag{A15}$$

Going back to Eq. (A4) and using

$$\langle N\rangle + \tfrac{1}{2} < \langle\sqrt{(N+1)(N+2)}\rangle < \langle N\rangle + \tfrac{3}{2}, \tag{A16}$$

which follows directly from the fact that

$$\langle\sqrt{(N+1)(N+2)}\rangle = (1-\zeta)\sum_{n=0}^{\infty}\sqrt{(n+1)(n+2)}\,\zeta^n, \tag{A17}$$

with the appropriate inequalities holding term by term, we get

$$\frac{4}{\pi}\left(\frac{1+(1/2\langle N\rangle)}{1+(1/\langle N\rangle)}\right)^2 < \frac{\langle q^2\rangle}{\langle q\rangle^2} < \frac{4}{\pi}\left(1 + \frac{3}{2\langle N\rangle} + \frac{1}{4\langle N\rangle^2}\right). \tag{A18}$$

Thus as $\langle N\rangle$ becomes indefinitely large,

$$\frac{(\delta q)^2}{\langle q\rangle^2} = \frac{\langle q^2\rangle}{\langle q\rangle^2} - 1 \sim \frac{4}{\pi} - 1, \tag{A19}$$

or

$$\frac{\delta q}{\langle q\rangle} \sim 0.52. \tag{A20}$$

Now, since $\langle p\rangle = 0$ for a state $|z\rangle$ with $\phi = 0$, the calculation of δp involves only the calculation of $\langle p^2\rangle$, which from Eqs. (A1) and (A3) is given by

$$\langle p^2\rangle = \langle N\rangle + \tfrac{1}{2} - \left(\frac{\langle N\rangle}{\langle N\rangle + 1}\right)\langle\sqrt{(N+1)(N+2)}\rangle \tag{A21}$$

The demonstration that this expression becomes vanishingly small for large $\langle N\rangle$ requires a somewhat tighter inequality than that of Eq. (A16). Such an inequality is

$$\langle\sqrt{(N+1)(N+2)}\rangle > \langle N\rangle + \tfrac{3}{2} - \tfrac{1}{4}\left\langle\frac{1}{N+\frac{3}{2}}\right\rangle, \tag{A22}$$

which again follows from the fact that it holds term by term if one writes out the expectation value as in Eq. (A17). Thus,

$$\langle p^2\rangle < \langle N\rangle + \tfrac{1}{2} - \left(1 + \frac{1}{\langle N\rangle}\right)^{-1}\left(\langle N\rangle + \tfrac{3}{2} - \tfrac{1}{4}\left\langle\frac{1}{N+\frac{3}{2}}\right\rangle\right). \tag{A23}$$

Using the obvious fact that the expectation value $\langle 1/(N+\frac{3}{2})\rangle$ is bounded from above by unity, we see that for $\langle N\rangle \gg 1$

$$\langle p^2\rangle < \frac{1}{4}\left\langle\frac{1}{N+\frac{3}{2}}\right\rangle + 0\left(\frac{1}{\langle N\rangle}\right). \tag{A24}$$

The relevant expectation value is

$$\left\langle\frac{1}{N+\frac{3}{2}}\right\rangle = (1-\zeta)\sum_{n=0}^{\infty}\frac{\zeta^n}{n+\frac{3}{2}} < \frac{1-\zeta}{\zeta}\sum_{n=0}^{\infty}\frac{\zeta^{n+1}}{n+1}, \tag{A25}$$

so that

$$\left\langle\frac{1}{N+\frac{3}{2}}\right\rangle < -\frac{1-\zeta}{\zeta}\ln(1-\zeta) = \frac{\ln(\langle N\rangle + 1)}{\langle N\rangle} \xrightarrow[\langle N\rangle\to\infty]{} 0. \tag{A26}$$

We see, then, that $\langle p^2\rangle$, and thus δp, becomes vanishingly small as $\langle N\rangle$ becomes indefinitely large.

---

*A preliminary report of this work and its possible group theoretic implications is contained in Y. Aharonov, H. W. Huang, J. M. Knight, and E. C. Lerner, Nuovo Cimento Lett. **2**, 1317 (1971). Parts of this work are contained in a doctoral dissertation presented by one of us (HWH) to the faculty of the University of South Carolina.

†Permanent Address: Physics Department, Belfer Graduate School, Yeshiva University, New York, and Physics Department, Tel-Aviv University, Tel-Aviv.

‡Parts of this author's contribution were developed while he was on leave at Birkbeck College, University of London. He would particularly like to thank Professor David Bohm for his very valuable comments.

[1]W. H. Louisell, Phys. Lett. **7**, 60 (1963).
[2]L. Susskind and J. Glogower, Physics **1**, 49 (1964).
[3]P. Carruthers and M. M. Nieto, Phys. Rev. Lett. **14**, 387 (1965).
[4]E. C. Lerner, "Harmonic Oscillator Phase Operators," AFCRL 66-624, Brandeis University (1966).
[5]E. C. Lerner, Nuovo Cimento B **56**, 183 (1968); Nuovo Cimento B **57**, 251 (1968).
[6]For a review of the subject of oscillator phase in quantum theory and an exhaustive list of references, see P. Carruthers and M. M. Nieto, Rev. Mod. Phys. **40**, 411 (1968).
[7]E. C. Lerner, H. W. Huang, and G. E. Walters, J. Math. Phys. **11**, 1679 (1970).
[8]It is shown in Ref. 7 that, for states of the form (2.7), the uncertainties δC and δS both approach zero for $\langle N\rangle \to \infty$, thus giving a well-defined phase.

[9]We cannot provide a rigorous proof of the stronger statement that the probability that q is negative vanishes as $\langle N\rangle \to \infty$.

[10]R. J. Glauber, Phys. Rev. **131**, 2766 (1963).

[11]$\mathfrak{N}$ denotes the statistical average of N, as distinguished from the quantum expectation value $\langle N\rangle = |\alpha|^2$ in a coherent state $|\alpha\rangle$.

[12]These properties are proved in Ref. 7.

[13]The argument for overcompleteness is essentially the one given by K. E. Cahill, Phys. Rev. B **138**, 1566 (1965).

[14]A similar equation has been written for coherent states by C. L. Mehta and E. C. G. Sudarshan, Phys. Lett. **22**, 574 (1966).

[15]We exclude the possibility that the radius of convergence of (3.10) may be zero, since it will appear in the sequel that the series actually terminates.

[16]See, for example, E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1944).

[17]An elementary discussion of the pertinent linear fractional transformations is contained in M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, London, 1964).

[18]Note that this operator is formally identical with the corresponding operator of Ref. 10 for coherent states.

[19]A more thorough discussion of these transformations is found in H. W. Huang, thesis, University of South Carolina (unpublished).

[20]These closed trajectories are themselves circles, as the following argument shows: For the Hamiltonian H_3, it is easily verified that the trajectories are circles centered at the invariant point, which is located at the origin. The trajectories of other Hamiltonians of this class can be obtained by applying the unitary transformation of the form (4.18) which takes the origin to the invariant point of the new Hamiltonian. Under this transformation, the old trajectories are mapped into the new ones by a linear fractional transformation. The result now follows from the elementary property that these transformations map circles into circles. The argument can be extended to Hamiltonians of the second class.

[21]It is possible to interpret these results in terms of $O(2,1)$ transformations of a particle of velocity $|\alpha|c/\sqrt{k+1}$ in the direction corresponding to the phase of α. The circumference of the circle then corresponds to particles traveling with the velocity of light.

[22]This limit does not appear to be a contraction in the sense of E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39**, 510 (1953). It is identical, however, to the limiting process discussed by J. D. Talman in *Special Functions* (Benjamin, New York, 1968), Chap. 13.

[23]A similar method has been applied to the Lorentz group by I. Bars and F. Gürsey, J. Math. Phys. **13**, 131 (1972).

[24]V. Bargman, Ann. Math. **48**, 568 (1947); I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Integral Geometry and Representation Theory* (Academic, New York, 1966); I. M. Gel'fand, M. I. Graev, and I. I. Pyatetskii-Shapiro, *Representation Theory and Automorphic Functions* (W. B. Saunders Co., Philadelphia, 1969).

[25]A. O. Barut and C. Fronsdal, Proc. R. Soc. A **287**, 532 (1965).

[26]See, for example, A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, 1960).

[27]P. W. Atkins and J. C. Dobson, Proc. R. Soc. A **321**, 321 (1971); J. M. Radcliffe, J. Phys. A **4**, 313 (1971); F. T. Arecchi, E. Courteus, R. Gilmore, and H. Thomas, Phys. Rev. A **6**, 2211 (1972); C. M. Bowden, Int. J. Quantum Chem. **6**, 133 (1972).

Stochastic particle trajectories in turbulent flow

C. P. Tsokos

Department of Mathematics, University of South Florida, Tampa, Florida 33620

D. P. Telionis

Department of Engineering Science and Mechanics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

(Received 6 December 1972)

The random position of a particle in turbulent flow $r(t;\omega)$ is a vector random function of time $t \geq 0$, $\omega \in \Omega$, where Ω is the supporting set of the underlying probability space (Ω, A, μ). If the statistics of a Lagrangian velocity field is known, the random position vector $r(t;\omega)$ is the solution of a nonlinear stochastic integral equation. Depending on the particular application, this equation may be homogeneous or inhomogeneous and with deterministic or random limits. A theorem is presented on the existence and uniqueness of the solution of this integral equation. The proof is based on Banach's fixed point theorem.

1. INTRODUCTION

The study of the path of a single particle in turbulent flow has recently attracted more attention, because although such theories give little information about mean flow properties, they find applications in other interesting problems of our times. As such we may mention here the study of trajectories of nondispersive particle pollutants, that may be assumed to follow fluid particle paths in the atmosphere or in rivers.

One of the main difficulties of such problems is that experimental means of measurements usually provide the Eulerian velocity field $u(x, t)$ where x a fixed point in space, while our integral equations involve the Lagrangian field $u(r, t)$ with r the instantaneous position vector of a particle. Lumley[1] suggests discretizing the time axis for integration while Wandel and Kofoed-Hansen[2] propose expressions for the transformation of autocorrelations and power spectra obtained by Eulerian and Lagrangian probing.

If the Lagrangian velocity field $u[r(t;\omega), t, \omega]$ or any general function ϕ of the position vector $r(t;\omega)$ is considered known then the problem can be formulated as a nonlinear stochastic integral equation. Chao[3] was able to relate the stochastic particle path of a particle to the statistics of the fluid motion by linearizing this equation to an integro-differential equation with a stationary forcing function. For a fully developed turbulent flow the integral equation is homogeneous with deterministic limits of integration and was considered by Lumley[1] and later by Padgett and Tsokos[4] who were concerned with its existence and uniqueness. If the particle trajectory is initiated in the laminar portion of the flow, then the instant $t_0(\omega)$ at which the particle enters the turbulent regime is a random quantity and the integral equation becomes nonhomogeneous with random limits. The conditions for existence and uniqueness of a random solution to this equation is the subject of the present paper.

2. PRELIMINARIES

Let us first consider the motion of a particle which follows a path within the laminar portion of the flow, passes through the region of transition and enters the turbulent regime. Let the time that elapses until the particle enters the turbulent regime be $t_0(\omega)$. Then the position vector of the particle is the solution of the integral equation

$$r(t;\omega) = h(t;\omega) + \int_{t_0(\omega)}^{t} \phi[r(\tau;\omega), t, \tau; \omega] d\tau, \qquad (2.1)$$

where (i) $\omega \in \Omega$, the supporting set of the probability measure space (Ω, A, μ), Ω being the sample space, A the σ-algebra of subsets of Ω and μ a complete probability measure on A; (ii) $t_0(\omega)$ is a random variable such that $0 \leq t_0(\omega) < T_2$, $\omega \in \Omega$, $t_0 \in R^+$. We shall let $T_1 = \inf t_0(\omega)$, $\omega \in \Omega$ and denote by I the closed interval $[T_1, T_2]$; (iii) $r(t,\omega)$ is the unknown random position vector; (iv) $\phi[r(\tau;\omega), t, \tau; \omega]$ is a vector function of the position vector and time and in our specific case the random velocity field $u(r, t;\omega)$; (v) $h(t;\omega)$ is a random vector function defined for $t \in I$ and $\omega \in \Omega$. Note that $h(t;\omega)$ represents the integration through the laminar portion of the flow and hence can be expressed in terms of a deterministic integral with a random limit.

$$h(t;\omega) = \int_0^{t_0(\omega)} \phi(r(\tau), t, \tau) d\tau. \qquad (2.2)$$

We will show in this paper that the above stochastic integral equation possesses a unique random solution, a second order stochastic process, which satisfies the equation with probability one. However, the results can be easily extended so that our solution will be any stochastic process of finite order.

Let $L_2^*(\Omega, A, \mu)$ denote the set of all three-dimensional random vectors of the form $r(t;\omega) = [r_i(t;\omega)]$, $i = 1, 2, 3$, where for each i, $r_i(t;\omega)$ is an element of $L_2^*(\Omega, A, \mu)$.

Lemma 2.1: The space $L^*(\Omega, A, \mu)$ is a normed linear space over the reals with the usual definition of componentwise addition and scalar multiplication where the norm in $L_2^*(\Omega, A, \mu)$ is given by

$$|r(t;\omega)|_{L_2^*(\Omega, A, \mu)} = \max_i \| r_i(t;\omega)\|.$$

The proof of the above lemma is straightforward and we proceed with the following definition.

Definition 2.1: Let $C_T(I, L_2^*(\Omega, A, \mu))$ be the set of all continuous functions from I into $L_2^*(\Omega, A, \mu)$. This definition says that $t \to (r(t;\omega))$ is continuous and that for each $t \in I$ and each i, $i = 1, 2, 3$, $r_i(t;\omega) \in L_2^*(\Omega, A, \mu)$. Thus for fixed $t \in I$

$$\|r(t;\omega)\|_{L_2^*(\Omega, A, \mu)} = \max_i \|r_i(t;\omega)\|.$$

We shall define the norm of the space $C_T(I, L_2^*(\Omega, A, \mu))$ as follows:

$$\|r(t;\omega)\|_{C_T(I, L_2^*(\Omega, A, \mu))} = \max_i \left\{ \sup_{T_1 \leq t \leq T_2} \|r(t;\omega)\|_{L_2^*(\Omega, A, \mu)} \right\}.$$

We shall use the concept of admissibility and some techniques similar to those developed by Tsokos.[5] We also refer the reader to a previous publication[4] for proof and comments of the following Lemma. Also, see Bharucha-Reid.[6]

Lemma 2.2: The space $C_T(I, L_2^*(\Omega, A, \mu))$ is a linear space over the reals with the usual definitions of addition and scalar multiplication for continuous functions.

Finally let us restate the well-known[7] "Banach's fixed-point theorem".

Theorem 2.1: If F is a contraction mapping from a subset W of a Banach space D into itself, then there exists a unique point x in W such that $F(x) = x$, that is a unique fixed point of the operator exists in W.

3. MAIN RESULTS

For the operator U defined by $(Ur)(t;\omega) = \int_{t_0(\omega)}^{t} \phi[r(\tau;\omega), t, \tau;\omega]d\tau$ we state the following lemma.

Lemma 3.1: We shall show that $UC_T(I, L_2^*(\Omega, A, \mu)) \subset C_T(I, L_2^*(\Omega, A, \mu))$. For $r(t;\omega) \in C_T(I, L_2^*(\Omega, A, \mu))$, we can write

$$\begin{aligned}\|r(t_1;\omega) - r(t_2;\omega)\|_{L_2^*(\Omega,A,\mu)} &= \Big\| \int_{t_0(\omega)}^{t_1} \phi[r(\tau;\omega), t_1, \tau;\omega]d\tau \\ &\quad - \int_{t_0(\omega)}^{t_2} \phi[r(\tau;\omega), t_2, \tau;\omega]d\tau \Big\|_{L_2^*(\Omega,A,\mu)}.\end{aligned}$$

The above expression can be written as follows:

$$\begin{aligned}\|r(t_1;\omega) - r(t_2;\omega)\|_{L_2^*(\Omega,A,\mu)} &= \Big\| \int_{t_0(\omega)}^{t_1} \{\phi[r(\tau;\omega), t_1, \tau;\omega] \\ &\quad - \phi[r(\tau;\omega), t_2, \tau;\omega]\} d\tau \Big\|_{L_2^*(\Omega,A,\mu)} \\ &\quad + \int_{t_1}^{t_2} \phi[r(\tau;\omega), t_2, \tau;\omega]d\tau \Big\|_{L_2^*(\Omega,A,\mu)},\end{aligned}$$

$$\begin{aligned}\|r(t_1;\omega) - r(t_2;\omega)\|_{L_2^*(\Omega,A,\mu)} &\le \int_{t_0(\omega)}^{t_1} \|\phi[r(\tau;\omega), t_1, \tau;\omega] \\ &\quad - \phi[r(\tau;\omega), t_2, \tau;\omega]\|_{L_2^*(\Omega,A,\mu)} d\tau \\ &\quad + \int_{t_1}^{t_2} \|\phi[r(\tau;\omega), t_2, \tau;\omega]\|_{L_2^*(\Omega,A,\mu)} d\tau.\end{aligned}$$

The continuity of $\|\phi[r(\tau;\omega), t_1, \tau;\omega]\|$ implies that the first term in the above inequality goes to zero with $|t_1 - t_2|$. Also, since $t_1, t_2 \in I$ and $\|\phi[r(\tau;\omega), t_2, \tau;\omega]\|$ is continuous, the second term goes to zero as $|t_1 - t_2| \to 0$, which proves the desired result.

Theorem 3.1: Consider the above random integral equation (2.1) under the following conditions:

(i) $\phi[r(t;\omega), t, \tau;\omega]$ is a continuous function from $R \times I \times I \times \Omega$ into $L_2^*(\Omega, A, \mu)$;

(ii) $h(t;\omega)$ is a continuous function from I into $L_2^*(\Omega, A, \mu)$;

(iii) $$|\phi[r(t;\omega), t, \tau;\omega] - \phi[r^*(t;\omega), t, \tau;\omega]|^2 \le \tau |r(t;\omega) - r^*(t;\omega)|^2 \quad (3.1)$$

for $r, r^* \in L_2^*(\Omega, A, \mu)$, $t, \tau \in R^+$, where $\tau \in R^+$. Then there exists a unique random solution $r(t;\omega)$ of (2.1) in the space $C_T(I, L_2^*(\Omega, A, \mu))$.

Proof: Define the operator U as

$$(Ur)(t;\omega) = \int_{t_0(\omega)}^{t} \phi[r(\tau;\omega), t, \tau;\omega]d\tau. \quad (3.2)$$

From Lemma 3.1 we know that $UC_T(I, L_2^*(\Omega, A, \mu)) \in C_T(I, L_2^*(\Omega, A, \mu))$. We shall prove that the operator U^n is a contraction operator for sufficiently large n. Let $r^*(t;\omega) \in C_T(I, L_2^*(\Omega, A, \mu))$. Thus, we may write

$$(Ur^*)(t;\omega) = \int_{t_0(\omega)}^{t} \phi[r^*(\tau;\omega), t, \tau;\omega]d\tau. \quad (3.3)$$

Subtracting Eq. (3.3) from Eq. (3.2), since the difference of elements of the space $C_T(I, L_2^*(\Omega, A, \mu))$ is in the space $C_T(I, L_2^*(\Omega, A, \mu))$, we have

$$Ur - Ur^* = \int_{t_0(\omega)}^{t} \{\phi[r(\tau;\omega), t, \tau;\omega] - \phi[r^*(\tau;\omega), t, \tau;\omega]\} d\tau.$$

Taking the expected value of the square of the absolute difference of the above expression, we have

$$\begin{aligned}E|Ur - Ur^*|^2 &= \int_\Omega d\mu \Big\{\int_{t_0(\omega)}^{t} \{\phi[r(\tau;\omega), t, \tau;\omega] - \phi[r^*(\tau;\omega), t, \tau;\omega] d\tau\Big\}^2 \\ &\le \int_\Omega d\mu \Big\{\int_T^t |\phi[r(\tau;\omega), t, \tau;\omega] - \phi[r^*(\tau;\omega), t, \tau;\omega] d\tau\Big\}^2.\end{aligned}$$

We now use Schwartz's inequality to write the above expression as

$$\begin{aligned}E|Ur - Ur^*|^2 \le (t - T_1)\int_\Omega d\mu \int_{T_1}^t |\phi[r(\tau,\omega), t, \tau;\omega] \\ - \phi[r^*(\tau;\omega), t, \tau;\omega]|^2 d\tau. \quad (3.4)\end{aligned}$$

Interchanging the order of integration in Eq. (3.4), we have

$$\begin{aligned}E|Ur - Ur^*|^2 \le (t - T_1)\int_{T_1}^t \int_\Omega |\phi[r(\tau;\omega), t, \tau;\omega] \\ - \phi[r^*(\tau;\omega), t, \tau;\omega]|^2 d\mu d\tau,\end{aligned}$$

which by virtue of the condition (3.1) can be put in the form

$$\begin{aligned}E|Ur - Ur^*|^2 &\le \lambda(t - T_1)\int_{T_1}^t E|r(\tau;\omega) - r^*(\tau;\omega)|^2 d\tau, \\ &- c\int_{T_1}^t E|r(\tau;\omega) - r^*(\tau;\omega)|^2 d\tau,\end{aligned}$$

where $c = \lambda(t - T_1)$. Successive integration of the above expression now yields

$$E|U^n r - U^n r^*|^2 \le \frac{c^n t^n}{n!} \|r(t;\omega) - r^*(t;\omega)\|^2,$$

which implies that

$$\|U^n r - U^n r^*\|^2 \le \frac{\lambda^n t^n}{n!} \|r(t;\omega) - r^*(t;\omega)\|^2.$$

Therefore, for sufficiently large n, the operator U^n is a contraction operator and existence and uniqueness of a random solution, $r(t;\omega)$, follows from Theorem 2.1.

[1] J. L. Lumley, J. Math. Phys. **3**, 309 (1962).
[2] C. F. Wandel and O. Kofoed-Hansen, J. Geophys. Res. **67**, 3089 (1962).
[3] B. T. Chao, Osterreichisches Ing.-Archiv **18**, 7 (1964).
[4] W. J. Padgett and C. P. Tsokos, J. Math. Phys. **12**, 210 (1971).
[5] C. P. Tsokos, Math. Systems Theory **3**, 222 (1969).
[6] A. T. Bharucha-Reid, *Random Integral Equations* (Academic, New York, 1972).
[7] P. M. Anselone, *Nonlinear Integral Equations* (U. of Wisconsin Press, Madison, 1964).

Projective representations of $SL(3,C)$ in the Z-operator formalism*

Itzhak Bars

Department of Physics, University of California, Berkeley, California 94720

(Received 6 December 1972)

An operator formalism, previously developed to discuss the Gel'fand–Naimark z-basis for the homogeneous Lorentz group, is now generalized to treat the projective representations of $SL(3,C)$. We find the projective "position" operators Z_1, Z_2, Z_3, as well as their canonical conjugate "momenta" Π_1, Π_2, Π_3 which form the building blocks of the generators of $SL(3,C)$. The z-representation, the states in which the "position" operators Z_i are diagonal, has very simple global transformation properties. This representation also leads to a Hilbert space endowed with an affine metric G, which relates the "covariant" and "contravariant" states. All unitary representations are unified by means of a single scalar product in which the matrix elements of G play the role of an intertwining operator.

1. INTRODUCTION

Recently some noncompact groups have played a major role in the understanding and development of important physical ideas. Among these, the most striking one is the development of the dual Veneziano-type models, which relied heavily on the unitary representations of $SU(1,1)$[1] and its extension to the conformal Virasoro algebra. Similarly, $SL(2,C)$ has been used in this connection for the group theoretical treatment of the Virasoro–Shapiro model.[2] The unitary representations of the above groups, relevant for dual models, are given by Bargmann,[3] Gel'fand and Naimark,[4] Bars and Gursey,[5] and are best understood in the so-called z-basis,[5,6,7]

In view of such considerations, we are motivated to study further the structure of the groups that fall into the same class with the hope that they may be useful in the further development of dual models.

In this paper, we study the group of 3×3 complex matrices with determinant one, called $SL(3,C)$. This group can be thought of also as complexified $SU(3)$ or complexified $SU(2,1)$, etc. It has 16 generators, eight of them form a maximal compact subgroup of $SU(3)$, and the rest are noncompact. Thus, unitary representations of this group are infinite dimensional, and can be labelled either by a set of continuous indices or, equivalently, by a discrete set of indices with an infinite number of values.

Our method of investigation is the z-operator formalism, previously developed to discuss the Gel'fand–Naimark z-basis for $SL(2,C)$,[5,6] as well as the z-basis for $SU(1,1)$ and $SL(2,R)$.[6,7] Here, we generalize this method to the more complicated case of $SL(3,C)$ which is isomorphic to $SU(3)_L \times SU(3)_R$. The group and its Lie algebra are defined in Sec. 2. In Sec. 3 (and Appendix A), we find for $SU(3)_L$ and $SU(3)_R$, the three commuting *projective "position" operators* Z_1, Z_2, Z_3, and their canonical conjugate "momenta" Π_1, Π_2, Π_3. These form the building blocks of an operator representation of the generators. The construction of the generators is given in Sec. 4. We then define the irreducible *z-basis*, in Sec. 5, and find the finite global transformations of these states, which turn out to be rather simple. In Sec. 6, we discuss the *covariant* and *contravariant* states which are related by a *metric* operator G constructed from Π_3. We also find the bilinear invariant functionals with the help of the matrix elements of G. In Sec. 7 we impose unitarity and find the necessary constraints on the Casimir operators. Finally, in Sec. 8, we construct a Hermitian, positive definite scalar product. We do this in a unified formulation, which applies to the principal series, supplementary series, and the two kinds of integer-point representations. The principal and supplementary series have also been discussed by Gel'fand and Naimark in a completely different approach.[8] We arrive at the same conclusions as Ref. 8, for these two cases.

2. GENERATORS, COMMUTATION RELATIONS, AND TRANSFORMATION PROPERTIES

We consider the group $SL(3,C)$ of 3×3 complex matrices of determinant one. The generators of the infinitesimal transformations are denoted by J_α and K_α, $\alpha = 1, 2, \cdots 8$. They have the following commutation relations:

$$[J_\alpha, J_\beta] = if_{\alpha\beta\lambda}J_\lambda, \tag{2.1a}$$

$$[J_\alpha, K_\beta] = if_{\alpha\beta\lambda}K_\lambda, \tag{2.1b}$$

$$[K_\alpha, K_\beta] = -if_{\alpha\beta\lambda}J_\lambda, \tag{2.1c}$$

where the structure constants $f_{\alpha\beta\lambda}$ are the usual $SU(3)$ structure constants given in Ref. 9.

We define the left-handed and right-handed operators X^L_α and X^R_α which correspond to an $SU(3)_L \times SU(3)_R$ decomposition of $SL(3,C)$

$$X^L_\alpha = \tfrac{1}{2}(J_\alpha + iK_\alpha), \tag{2.2a}$$

$$X^R_\alpha = \tfrac{1}{2}(J_\alpha - iK_\alpha) \tag{2.2b}$$

with the commutation relations

$$[X^L_\alpha, X^L_\beta] = if_{\alpha\beta\lambda}X^L_\lambda, \tag{2.3a}$$

$$[X^R_\alpha, X^R_\beta] = if_{\alpha\beta\lambda}X^R_\lambda, \tag{2.3b}$$

$$[X^L_\alpha, X^R_\beta] = 0. \tag{2.3c}$$

The 3×3 representation of the left- and right-handed $SU(3)$ groups is denoted by $\lambda_\alpha/2$ (or $-\lambda^*_\alpha/2$), where the 3×3 traceless matrices $\lambda_\alpha/2$ satisfy the same commutation relations as (2.3a) or (2.3b) and are given explicitly by Gell-Mann.[9]

With the help of the matrices λ_α and their complex conjugate λ^*_α we define two matrices A^L_{ij} and B^R_{ij} with *operator entries*

$$A^L_{ij} = \sum_{\alpha=1}^{\infty} X^\alpha_L \lambda^\alpha_{ij} \text{ with } \mathrm{Tr}A^L = 0, \tag{2.4a}$$

$$A^R_{ij} = \sum_{\alpha=1}^{\infty} X^\alpha_R \lambda^{*\alpha}_{ij} \text{ with } \mathrm{Tr}A^R = 0. \tag{2.4b}$$

Notice that for a unitary representation we have $X_R^\dagger = X_L$, therefore also $A_{ij}^{R\dagger} = A_{ij}^L$.

The commutation relation satisfied by A_{ij}^L and A_{ij}^R can be deduced from (2.3) and from the relations

$$4if_{\alpha\beta\gamma} = \mathrm{Tr}(\lambda^\gamma[\lambda^\alpha, \lambda^\beta]), \tag{2.5a}$$

$$(\delta_{ad}\delta_{cb} - 2/3\,\delta_{ab}\delta_{cd}) = \sum_{\alpha=1}^{8} \lambda^\alpha_{ab}\lambda^\alpha_{cd} \tag{2.5b}$$

we obtain

$$[A_{ij}^L, A_{kl}^L] = \delta_{il} A_{kj}^L - \delta_{kj} A_{il}^L, \tag{2.6a}$$

$$[A_{ij}^R, A_{kl}^R] = -\delta_{il} A_{kj}^R + \delta_{kj} A_{il}^R, \tag{2.6c}$$

$$[A_{ij}^R, A_{kl}^L] = 0.$$

Then we can easily see that the Casimir operators for the group $SL(3,C)$ are given as

$$C_{IL} = \mathrm{Tr}A_L^2, C_{2_L} = \mathrm{Tr}A_L^3, C_{IR} = \mathrm{Tr}A_R^2, C_{2_R} = \mathrm{Tr}A_R^3. \tag{2.7}$$

The operators $C_i^{L,R}$ commute with all A_{ij}^L and A_{ij}^R as well as all J_α and K_α. Thus, a general representation of $SL(3,C)$ will be denoted by the set of 4 numbers $(C_{IL}, C_{2L}, C_{1R}, C_{2R})$.

A general finite transformation with the parameters ω_α, ν_α, $\alpha = 1,\cdots 8$, is denoted by U,

$$U = e^{i(\omega_\alpha J_\alpha + \nu_\alpha K_\alpha)} = e^{ia_\alpha^L X_\alpha^L} e^{ia_\alpha^R X_\alpha^R}, \tag{2.8a}$$

where

$$a_\alpha^L = \omega_\alpha - i\nu_\alpha, \quad a_\alpha^R = \omega_\alpha + i\nu_\alpha. \tag{2.8b}$$

We assign the following representations to the left- and right-handed generators: The 3×3 *left*-handed representation is obtained by letting $J_\alpha \to 1/2\lambda_\alpha$, $K_\alpha \to -i/2\lambda_\alpha$, or equivalently $X_\alpha^L \to \lambda_\alpha/2$, $X_\alpha^R \to 0$, and similarly the 3×3 *right*-handed representation is obtained by letting $J_\alpha \to -1/2\lambda_\alpha^*$, $K_\alpha \to -i/2\lambda_\alpha^*$ or equivalently $X_\alpha^L \to 0$, $X_\alpha^R \to -1/2\lambda_\alpha^*$. Thus, corresponding to the general transformation U we can write the 3×3 left-handed representation Λ (letting $a_L = a_R^* \equiv a$)

$$\Lambda(a) = e^{ia\cdot\lambda/2} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}. \tag{2.9}$$

The 3×3 right-handed representation is then, according to the above prescription, simply Λ^*. Clearly, by construction we have $\det\Lambda = 1$, $(\mathrm{Tr}\lambda_\alpha = 0)$, and thus Λ corresponds to a general $SL(3,C)$ representation.

We would like to find the transformation properties of A_{ij}^L and A_{ij}^R under a general transformation U:

$$U^{-1}\mathbf{A}^L U = \boldsymbol{\lambda}_\alpha e^{-ia^L.X^L} X_\alpha^L e^{+ia^L.X^L} \tag{2.10a}$$

$$= \boldsymbol{\lambda}_\alpha (S_{\alpha\beta} X_\beta) \tag{2.10b}$$

$$= (S_{\alpha\beta}^{-1}\boldsymbol{\lambda}_\beta) X_\alpha^L \tag{2.10c}$$

$$= e^{ia^L.\lambda/2}\boldsymbol{\lambda}_\alpha e^{-ia^L.\lambda/2} X_\alpha^L \tag{2.10d}$$

$$= \boldsymbol{\Lambda}\mathbf{A}^L\boldsymbol{\Lambda}^{-1}. \tag{2.10e}$$

In (2.10a) we used $[X^L, X^R] = 0$. In (2.10b) we used the fact that X_α^L form the adjoint representation of $SU(3)_L$, thus the expression in (2.10a) induces an 8×8 linear transformation on X_α^L which we denoted by $S_{\alpha\beta}$ in 2.10b. In (2.10c) we used the fact that the product $\lambda_\alpha X_\alpha^L$ is an invariant when λ_α and X_α^L are transformed in the same way; this can be seen immediately by analogy to the Casimir operator $C_{IL} = X_\alpha^L X_\alpha^L$. In 2.10d we used the same reasoning as in passing from (2.10a) to (2.10b).

Similarly we find

$$U^{-1}A^R U = \Lambda^* A^R \Lambda^{*-1}. \tag{2.11}$$

The expressions in (2.10) and (2.11) can be also checked for an infinitesimal transformation directly by using the commutation relations (2.3) as well as those for λ_α.

Finally, we note some identities which will be useful shortly. They can be derived as given in Ref. 10 and using Eqs. (2.6):

$$A_L^3 + 3A_L^2 + (2 - 1/2\,\mathrm{Tr}A_L^2)A_L - (1/3\,\mathrm{Tr}A_L^3 + 1/2\,\mathrm{Tr}A_L^2) = 0, \tag{2.12a}$$

$$A_R^3 - 3A_R^2 + (2 - 1/2\,\mathrm{Tr}A_R^2)A_R - (1/3\,\mathrm{Tr}A_R^3 - 3/2\,\mathrm{Tr}A_R^2) = 0 \tag{2.12b}$$

3. THE Z-OPERATORS OF $SL(3, C)$ AND THEIR CANONICAL CONJUGATES

A. Definition and transformation properties of the Z-operators

In this section we will find the Z_i-operators and their canonical conjugates Π_i for the left-handed $SU(3)_L$ characterized by A_{ij}^L or X_α^L. At the end of the section we will simply state the results for the right-handed $SU(3)_R$ since everything works just in the same way. From here on we drop the index L or R. Thus, we consider the operators A_{ij} arranged as a 3×3 matrix and having the commutation relation (2.6). We consider the *operator* eigenvalue equation

$$\mathbf{A}^L\boldsymbol{\Psi}^L = \lambda^L\boldsymbol{\Psi}^L; \quad (\mathbf{A}^R\boldsymbol{\Psi}^R = -\lambda^R\boldsymbol{\Psi}^R), \tag{3.1}$$

where Ψ is a 3-dimensional column with *operator* entries, and λ is an operator proportional to unity. Applying (2.12a) to Ψ we find that the eigenvalue λ satisfies the cubic equation

$$\lambda^3 + 3\lambda^2 + (2 - 1/2\,\mathrm{Tr}\mathbf{A}^2)\lambda - (1/3\,\mathrm{Tr}\mathbf{A}^3 + 1/2\,\mathrm{Tr}A^2) = 0. \tag{3.2}$$

The roots of this equation $\lambda_1, \lambda_2, \lambda_3$ satisfy

$$\lambda_1 + \lambda_2 + \lambda_3 = -3, \tag{3.3a}$$

$$\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1 = 2 - 1/2\,\mathrm{Tr}\mathbf{A}^2, \tag{3.3b}$$

$$\lambda_1\lambda_2\lambda_3 = 1/3\,\mathrm{Tr}\mathbf{A}^3 + 1/2\,\mathrm{Tr}\mathbf{A}^2. \tag{3.3c}$$

Since $\mathrm{Tr}\mathbf{A}^2$ and $\mathrm{Tr}\mathbf{A}^3$ are Casimir operators, the roots $\lambda_1, \lambda_2, \lambda_3$ are functions of *only* Casimir operators, thus they commute with each other and with *all* X_α^L and X_α^R. Therefore, any representation of the left-handed $SU(3)_L$ can be characterised by two roots (λ_1, λ_2) or (λ_1, λ_3), with $\lambda_3 = -3 - \lambda_1 - \lambda_2$.

We would like to find also the transformation properties of the operators Ψ; to do so, we apply U^{-1}, U to both sides of (3.1), and use the fact that $[\lambda, U] = 0$:

$$U^{-1}\mathbf{A}UU^{-1}\mathbf{\Psi}U = \lambda U^{-1}\mathbf{\Psi}U. \tag{3.4a}$$

Using Eqs. (2.10), we write

$$\Lambda\mathbf{A}\Lambda^{-1}\, U^{-1}\mathbf{\Psi}U = \lambda U^{-1}\mathbf{\Psi}U. \tag{3.4b}$$

Applying Λ^{-1} from the left we obtain

$$\mathbf{A}(\Lambda^{-1}U^{-1}\Psi U) = \lambda(\Lambda^{-1}U^{-1}\Psi U). \tag{3.4c}$$

Comparing (3.4c) to (3.1) and taking the most general $\lambda_1, \lambda_2, \lambda_3$ so that there is no degeneracy, we find

$$\mathbf{\Psi} = (\text{const}) \times \Lambda^{-1}U^{-1}\mathbf{\Psi}U \tag{3.4d}$$

or

$$U^{-1}\mathbf{\Psi}U = (\text{const})\Lambda\Psi. \tag{3.4e}$$

Now we define two Z operators Z_1 and Z_2 as

$$Z_1 = \Psi_1\Psi_3{}^{-1}, \qquad Z_2 = \Psi_2\Psi_3{}^{-1}, \tag{3.5}$$

where $\Psi_{1,2,3}$ are the 3 components of the operator column matrix Ψ.

From (3.4e) we immediately deduce the transformation properties of Z_1 and Z_2. We find that they transform according to a generalized *projective* transformation

$$U^{-1}Z_1U = \frac{\Lambda_{11}Z_1 + \Lambda_{12}Z_2 + \Lambda_{13}}{\Lambda_{31}Z_1 + \Lambda_{32}Z_2 + \Lambda_{33}}, \tag{3.6a}$$

$$U^{-1}Z_2U = \frac{\Lambda_{21}Z_1 + \Lambda_{22}Z_2 + \Lambda_{23}}{\Lambda_{31}Z_1 + \Lambda_{32}Z_2 + \Lambda_{33}}. \tag{3.6b}$$

We will show shortly that $[Z_1, Z_2] = 0$, so we are justified in writing (3.6) like rational functions, even though Z_i are operators. The Λ_{ij} have been defined in Eq. (2.9).

In addition to Z_1 and Z_2 we can define two more Z operators Z_3 and Z_4. For this purpose we write another eigenvalue equation in analogy to (3.1).

$$\mathbf{\Phi}^T\mathbf{A} = \eta\mathbf{\Phi}^T. \tag{3.7}$$

Notice that now $\mathbf{A}$ operates from the right on the row matrix $\mathbf{\Phi}^T$. Applying again (2.12a) on $\mathbf{\Phi}^T$ from the right we find that η satisfies the same cubic equation as λ, i.e., Eq. (3.2). Therefore, η is equal to one of the roots $\lambda_1, \lambda_2, \lambda_3$ and is a Casimir operator. Going through similar steps to (3.4) we find

$$U^{-1}\mathbf{\Phi}^TU = (\text{const})\mathbf{\Phi}^T\Lambda^{-1}, \tag{3.8}$$

where Λ^{-1} is the inverse of the matrix Λ ($\Lambda\Lambda^{-1} = \Lambda^{-1}\Lambda = 1$).

Now we define Z_3 and Z_4 as

$$Z_3 = \Phi_1{}^{-1}\Phi_2, \qquad Z_4 = \Phi_1{}^{-1}\Phi_3. \tag{3.9}$$

From Eq. 3.8 we obtain the transformation properties of Z_3 and Z_4.

$$U^{-1}Z_3U = \frac{V_{12} + V_{22}Z_3 + V_{32}Z_4}{V_{11} + V_{21}Z_3 + V_{31}Z_4}, \tag{3.10a}$$

$$U^{-1}Z_4U = \frac{V_{13} + V_{23}Z_3 + V_{33}Z_4}{V_{11} + V_{21}Z_3 + V_{31}Z_4}, \tag{3.10b}$$

where we have defined

$$\Lambda^{-1} \equiv V. \tag{3.11}$$

Again, we will show shortly that $[Z_3, Z_4] = 0$, so that it is legitimate to write the rational expressions of (3.10).

We remark that Z_1, Z_2, Z_3, Z_4 are not all independent, there is a nonlinear relation among them, which is obtained as follows. Using (3.1) and (3.7) we can write

$$\mathbf{\Phi}^T\mathbf{A}\mathbf{\Psi} = \lambda\mathbf{\Phi}^T\mathbf{\Psi} = \eta\mathbf{\Phi}^T\mathbf{\Psi}. \tag{3.12}$$

Therefore, taking $\lambda \neq \eta$ as the most general case we obtain

$$\mathbf{\Phi}^T\Psi = 0 \equiv \Phi_1\Psi_1 + \Phi_2\Psi_2 + \Phi_3\Psi_3 \tag{3.13}$$

multiplying (3.13) by Φ_1^{-1} from left and Ψ_3^{-1} from the right

$$0 = \Psi_1\Psi_3^{-1} + \Phi^{-1}\Phi_2\Psi_2\Psi_3^{-1} + \Phi_1^{-1}\Phi_3 \tag{3.13$'$}$$

or

$$0 = Z_1 + Z_3Z_2 + Z_4. \tag{3.13$''$}$$

Thus, there are *only 3 independent Z-operators*. Therefore, in the transformation equations (3.6) and (3.10) we should eliminate one of them, (say Z_4), in favor of the others. In the following, for simplicity, we shall continue to write our expressions in terms of Z_1, Z_2, Z_3, Z_4, but we should bear in mind that Z_4 is not independent and should be replaced everywhere by $Z_4 = -Z_1 - Z_3Z_2$.

B. Explicit construction of the Z-operators

Now we construct Z_i explicitly as functions of the generators X_α or equivalently A_{ij}. The details of the construction are demonstrated in Appendix A. The result is

$$Z_1 = K_{3n}^{-1}(\lambda)K_{1n}(\lambda) = K_{1m}(\lambda+1)K_{3m}^{-1}(\lambda+1), \tag{3.14a}$$

$$Z_2 = K_{3n}^{-1}(\lambda)K_{2n}(\lambda) = K_{2m}(\lambda+1)K_{3m}^{-1}(\lambda+1), \tag{3.14b}$$

$$Z_3 = K_{n2}(\eta)K_{n1}^{-1}(\eta) = K_{m1}^{-1}(\eta+1)K_{m3}(\eta+1), \tag{3.14c}$$

$$Z_4 = K_{n3}(\eta)K_{n1}^{-1}(\eta) = K_{m1}^{-1}(\eta+1)K_{m3}(\eta+1), \tag{3.14d}$$

$m, n = 1, 2, 3$, *no sum* over n or m, the expressions are *equal* for any value of n or m. The operators $K_{ij}(\lambda)$ or $K_{ij}(\eta)$ are given as

$$K_{ij}(\lambda) = A_{ij}(A_{kk} - \lambda - 1) - A_{ik}A_{kj}, \qquad i \neq j, \tag{3.15a}$$

$$K_{ii}(\lambda) = -(A_{jj} - \lambda)(A_{kk} - \lambda - 1) + A_{jk}A_{kj}, \tag{3.15b}$$

$i, j, k = 1, 2, 3$ and *cyclic* or *anticyclic, no sum* over repeated indices.

In Appendix A we give some interesting properties of the operators $K_{ij}(\lambda)$, and we prove that

$$[Z_i, Z_j] = 0 \quad \text{for any } i, j = 1, 2, 3, 4. \tag{3.16a}$$

C. Canonical conjugates of the Z-operators, and the quantum mechanical analog.

We have obtained three Z^L-operators (Z_4^L is dependent) which commute with each other and which transform projectively as in Eqs. (3.6) and (3.10). We now wish to find their canonical conjugates $\Pi_1^L, \Pi_2^L, \Pi_3^L$ as functions of the A_{ij}^L, which satisfy:

$$[\Pi_i^L, \Pi_j^L] = 0, \tag{3.16b}$$

$$[\Pi_i^L, Z_j^L] = -\delta_{ij}. \tag{3.16c}$$

According to Eq. (A13) in the appendix it appears that the candidates which satisfy (3.16c) are almost $\Pi_1^L = A_{31}^L$, $\Pi_2^L = A_{32}^L$, $\Pi_3^L = -A_{21}^L$ except that $[A_{21}^L, Z_1^L] \neq 0$. To correct this, we have to take

$$\Pi_1^L = A_{31}^L, \tag{3.17a}$$

$$\Pi_2^L = A_{32}^L, \tag{3.17b}$$

$$\Pi_3^L = -A_{21}^L + A_{31}^L Z_2^L. \tag{3.17c}$$

With this choice, we see immediately that Eq. (3.16b) is also satisfied. This asymmetry can be traced also to the elimination of Z_4 as a dependent operator.

At this point it is interesting to note the similarity of Z_i and Π_i to the quantum mechanical operators of position and momenta. In fact, our procedure of finding the representations for $SL(3,C)$ will be to define states which diagonalize Z_i^L, just as the position states in quantum mechanics, on which Π_i act as derivatives $\partial/\partial Z_i^L$ (just like momenta). This is why we went through all the trouble of finding commuting Z_i^L operators (so that they can be diagonalized simultaneously) and their canonical conjugates Π_i^L.

D. Correspondence principle for $SU(3)_L \leftrightarrow SU(3)_R$

We finally mention the result in the case of $SU(3)_R$. The only difference is that the A_R^{ij} operators have commutation relations with an extra minus sign Eq. (2.6). We can see that an easy way to deduce the results for $SU(3)_R$ from the ones of $SU(3)_L$, is to consider a correspondence principle $A_R^{ij} \leftrightarrow A_L^{ij\dagger}$ (this is an equality in the case of unitary representations). Since $A_L^{ij\dagger}$ has the same commutation relations as A_R^{ij}, we can simply take the Hermitian conjugation of the $SU(3)_L$ results and identify them with $SU(3)_R$. (If any rearrangement is needed, one should remember the commutation relations. In this connection compare also Eqs. (2.12a and 2.13b.) Thus, we obtain

$$[Z_i^R, Z_j^R] = 0 = [\Pi_i^R, \Pi_j^R], \tag{3.18a}$$

$$[\Pi_i^R, Z_j^R] = +\delta_{ij}, \tag{3.18b}$$

with

$$\Pi_1^R = A_{31}^R, \tag{3.19a}$$

$$\Pi_2^R = A_{32}^R, \tag{3.19b}$$

$$\Pi_3^R = -A_{21}^R + A_{31}^R Z_2^R. \tag{3.19c}$$

Notice the difference between (3.16c) and (3.18b). In general, also, we should take Casimir operators for $SU(3)_R$ different than those for $SU(3)_L$. It turns out that the correct correspondence principle is that at Hermitian conjugation one should replace

$$\lambda_L^* + 2 \to -\lambda_R, \quad \eta_L^* + 2 \to -\eta_R. \tag{3.20}$$

This will then be consistent with $\lambda_{L,R}$ and $\eta_{L,R}$ defined according to Eqs. 3.1 and 3.7.

4. CONSTRUCTION OF THE $SU(3)_{L,R}$ INFINITESIMAL GENERATORS IN TERMS OF THE CANONICALLY CONJUGATE OPERATORS $Z^i_{L,R}$ AND $\Pi^i_{L,R}$

A. Representations of the generators

Using only the commutation properties of the A_{ij}^L and A_{ij}^R operators, we have constructed, in Appendix A and Sec. 3, the canonically conjugate operators $Z_i^{L,R}$ and $\Pi_i^{L,R}$, $i = 1, 2, 3$:

$$[Z_i^L, Z_j^L] = 0 = [\Pi_i^L, \Pi_j^L], \tag{4.1a}$$

$$[\Pi_i^L, Z_j^L] = -\delta_{ij} \text{ etc.} \tag{4.1b}$$

In this section we consider the inverse problem: Given the simple commutation relations (4.1), how do we construct the generators A_{ij}^L, A_{ij}^R in terms of the canonically conjugate operators $Z_i^{L,R}$ and $\Pi_i^{L,R}$? Again, first we consider the problem only for $SU(3)_L$. Taking the Hermitian conjugation at the end, we obtain the analogous results for $SU(3)_R$ by the correspondence principle of Sec. 3. From here on we drop the index L,R.

In general, we find two classes of representations for the generators A_{ij} in terms of Z, Π. The first class (class A) depends only on $Z_{1,2}$, and $\Pi_{1,2}$ and only on one Casimir operator λ. The second class (class B) depends in addition to Z_3, Π_3 and a second Casimir operator η.

Class A: We notice that, if we completely ignore Z_3, Π_3, then from the Appendix Eq. A1, we can solve for A_{12}, A_{23}, A_{33}. We already have $A_{31} = \Pi_1, A_{32} = \Pi_2$. Thus, the only unknowns are $A_{11}, A_{12}, A_{21}, A_{22}$, which must be constructed from Z_i, Π_i such as to satisfy the proper commutation relations [Eqs. A13 (e–f) and the trace condition $\mathrm{Tr}A = 0$. We find the solution

$$\begin{aligned} A_{11} &= \Pi_1 Z_1 - \lambda/2, \quad A_{12} = \Pi_2 Z_1, \\ A_{21} &= \Pi_1 Z_2, \quad A_{22} = \Pi_2 Z_2 - \lambda/2. \end{aligned} \tag{4.2a}$$

Then, from Eq. A1 we find

$$\begin{aligned} A_{13} &= -\Pi_1 Z_1^2 - \Pi_2 Z_1 Z_2 + (3/2)\lambda Z_1, \quad A_{31} = \Pi_1, \\ A_{23} &= -\Pi_2 Z_2^2 - \Pi_1 Z_1 Z_2 + (3/2)\lambda Z_2, \quad A_{32} = \Pi_2, \\ A_{33} &= \lambda - \Pi_1 Z_1 - \Pi_2 Z_2. \end{aligned} \tag{4.2b}$$

These satisfy explicitly the commutation relations (2.6a) and equations such as (2.12a), etc., as they should do.

We obtain the $SU(3)_R$ representations from the above by the Hermitian conjugation correspondence principle: For example,

$$A_{11}^{L\dagger} = (Z_1^{L\dagger}\Pi_1^{L\dagger} - \lambda_L^*/2) \to A_{11}^R = \left(Z_1^R \Pi_1^R + \frac{\lambda_R + 2}{2}\right).$$

Using Eq. (3.18b) we obtain

$$A_{11}^R = \Pi_1^R Z_1^R + \lambda_R/2 \tag{4.3}$$

and similarly for the rest. It turns out that the form is exactly the same as (4.2) except that λ_L is replaced by $-\lambda_R$.

Class B: When Z_3 and Π_3 are also included in the representation, then the Eqs. A1 and A2 of the appendix and $\mathrm{Tr}A = 0$ give a complete solution of the A_{ij} in terms of Π_i, Z_i. With a little rearrangement we find

$$\begin{aligned} A_{11} &= \Pi_1 Z_1 + \Pi_3 Z_3 + \eta + 2, \\ A_{12} &= \Pi_2 Z_1 + \Pi_3 Z_3^2 + (2\eta + \lambda + 4) Z_3, \\ A_{13} &= -\Pi_1 Z_1^2 - \Pi_2 Z_1 Z_2 - \Pi_3 Z_3 (Z_1 + Z_2 Z_3) \\ &\quad + (\lambda - \eta - 2) Z_1 - (2\eta + \lambda + 4) Z_2 Z_3, \end{aligned} \tag{4.4}$$

$$\begin{aligned} A_{21} &= -\Pi_3 + \Pi_1 Z_2, \\ A_{22} &= \Pi_2 Z_2 - \Pi_3 Z_3 - (\eta + \lambda + 2), \end{aligned}$$

$$A_{23} = -\Pi_2 Z_2^2 - \Pi_1 Z_1 Z_2 + \Pi_3(Z_1 + Z_2 Z_3) + (\eta + 2\lambda + 2)Z_2,$$
$$A_{31} = \Pi_1,$$
$$A_{32} = \Pi_2,$$
$$A_{33} = \lambda - \Pi_1 Z_1 - \Pi_2 Z_2.$$

These again satisfy the commutation relations and Eqs. such as (2.12), etc.

To obtain the representation for $SU(3)_R$, again we appeal to the correspondence principle and Eq. 3.20. The result is that the form is the same (with $\Pi_L \to \Pi_R, Z_L \to Z_R$), except that $\eta_L + 2 \to -(\eta_R + 2)$, $\lambda_L \to -\lambda_R$. For example,

$$A_{11}^R = \Pi_1^R Z_1^R + \Pi_3^R Z_3^R - (\eta_R + 2), \tag{4.5}$$

etc.

B. Equivalent representations

We now make an important remark, that will lead later to the construction of the "metric" in the z-representation. According to Eqs. (3.2) and (3.3), there are three possible roots λ_i corresponding to only two Casimir operators. In the representation of Eq. 4.4 we have used the two roots, say, $\lambda_1 = \lambda$ and $\lambda_2 = \eta$. Thus, this representation is labeled by the roots (λ_1, λ_2). Clearly we are naturally led to expect that the representation labeled by (λ_1, λ_3), where $\lambda_3 = -3 - \lambda_1 - \lambda_2$ (Eq. 3.3), should be equivalent to the first one, since they both correspond to the *same* values of the Casimir operators $C_1 = \mathrm{Tr}A^2$ and $C_2 = \mathrm{Tr}A^3$. Indeed, we have been able to find a similarity transformation in operator form which can take us from the (λ_1, λ_2) representation to (λ_1, λ_3).

By explicit commutation, we verify that

$$\Pi_3^{(-\lambda_2+\lambda_3)} A_{ij}(\lambda_1, \lambda_2) \Pi_3^{(\lambda_2-\lambda_3)} = A_{ij}(\lambda_1, \lambda_3). \tag{4.6}$$

Thus, the operator $\Pi_3^{-\lambda_2+\lambda_3}$ is analogous to a "metric" operator (not yet positive definite and Hermitian), which takes us from "covariant" representations (λ_1, λ_2) to "contravariant" ones (λ_1, λ_3). The analogous operator for $SU(3)_R$ is: $\Pi_{3_R}^{-\lambda_{2R}+\lambda_{3R}}$, obtained again by the correspondence principle. As it will be seen below, for unitary representations the product operator $[\Pi_{3L}^{-\lambda_{2L}+\lambda_{3L}} \Pi_{3R}^{-\lambda_{2R}+\lambda_{3R}}]$ is indeed interpretable as a positive definite Hermitian metric operator, just as it was possible for $SL(2,C)$.[5]

5. THE IRREDUCIBLE Z-BASIS AND FINITE TRANSFORMATIONS

In the previous section, we gave an operator representation for the infinitesimal generators of $SU(3)_L \otimes SU(3)_R$. In this section, we find the realization on the irreducible z-basis, which will then lead to a very simple representation for finite global transformations.

In analogy to Ref. 5, we define the z-basis as the simultaneous eigenstates of the commuting operators $C_{1L,R}$, $C_{2L,R}$, $Z_{iL,R}$. This is a complete specification for an irreducible representation of $SL(3,C)$. It is also quite analogous to the "position-space" commonly used in quantum mechanics, if we remember that Z_i are like position operators, and Π_i like momentum operators.

There will be two classes of states: Class A and class B depending on whether Z_3 and Π_3 are included or not.

We denote the z-basis simply as $|z\rangle$. It actually stands for $|\lambda_L, \lambda_R; \eta_L, \eta_R; z_{1_L}, z_{2_L}, z_{3_L}; z_{1_R}, z_{2_R}, z_{3_R}\rangle$. The dependence on $z_{3_{L,R}}$ is removed for the representations of class A. The operators Z_i and Π_i are applied just like position and momentum:

$$Z_{iL}|z\rangle = z_{iL}|z\rangle; \; Z_{iR}|z\rangle = z_{iR}|z\rangle, \tag{5.1a}$$

$$\Pi_{iL}|z\rangle = \partial/\partial z_{iL}|z\rangle; \; \Pi_{iR}|z\rangle = -\partial/\partial z_{iR}|z\rangle. \tag{5.1b}$$

These are consistent with the canonical commutation relations (3.16) and (3.18). (Notice the sign difference for $\Pi_{L,R}$.) With the help of (5.1) we can apply the infinitesimal generators (4.2), (4.3), (4.4) and (4.5) on the states $|z\rangle$.

Thus, for the representations of class A ($|z\rangle_A$) we find

$$A_{11}^L |z\rangle_A = (z_{1_L}\partial/\partial z_{1_L} - \lambda_L/2)|z\rangle_A, \tag{5.2a}$$

$$A_{12}^L |z\rangle_A = z_{1_L}\partial/\partial z_{2_L}|z\rangle_A, \text{ etc}, \tag{5.2b}$$

and for the representations of class B we find

$$A_{11}^L |z\rangle_B = (z_{1_L}\partial/\partial z_{1_L} + z_{3_L}\partial/\partial z_{3_L} + \eta_L + 2)|z\rangle_B, \text{ etc.} \tag{5.3}$$

Now we are ready to calculate the effect of a finite transformation $U = e^{ia\cdot X_L} e^{ia^*\cdot X_R}$ on the states $|z\rangle$. We recall that the 3×3 representation of U is Λ, as given by (2.9).

Let us first notice that with the help of (3.6) and (3.10) we can write

$$Z_{1_L}U|z\rangle = UU^{-1}Z_{1_L}U|z\rangle = U\,\frac{aZ_{1L} + bZ_{2L} + c}{gZ_{1L} + hZ_{2L} + k}\,|z\rangle$$
$$= \frac{az_{iL} + bz_{2L} + c}{gz_{1L} + hz_{2L} + k}\,U|z\rangle. \tag{5.4}$$

Similarly for the other $Z_{iL,R}$. Therefore, the state $U|z\rangle$ is an eigenstate of the operators $Z_{iL,R}$ with the eigenvalues $z'_{1L} = (az_{1L} + bz_{2L} + c)\,(gz_{1L} + hz_{2L} + k)^{-1}$, $z'_{1R} = (a^*z_{1R} + b^*z_{2R} + c^*)(g^*z_{1R} + h^*z_{2R} + k^*)^{-1}$ etc. These transformation properties could only be consistent if we let $z_R = z_L^*$, and we must write in general

$$U(\Lambda)|z\rangle = \Omega(\Lambda, z)|z'\rangle, \tag{5.5}$$

where $|z'\rangle$ stands for a state with transformed labels z'_i as indicated by (3.6) and (3.10) [the right-handed $z_{iR} = z_{iL}^*$ (transforming with Λ^*)], and $\Omega(\Lambda, z)$ is just a multiplicative functions of the complex z_i and the transformation parameters Λ. Clearly we must have $\Omega(1, z) = 1$.

To find the finite transformation of the states we simply have to specify $\Omega(\Lambda, z)$. We first find the infinitesimal form of Ω, by applying an infinitesimal transformation: $U(\Lambda) = 1 + ia_{ij}A_{ij}^L + ia_{ij}^*A_{ij}^R$, with a_{ij} small. Then, the left-handed 3×3 representation is given by replacing the 3×3 representation of $A_{ij}^L \to (|j\rangle\langle i| - 1/3\,\mathbf{1})$ and letting $A_{ij}^R \to 0$. [This representation of A_{ij}^L is, of course, consistent with the commutation relations (2.6a)].

$$\Lambda = \begin{bmatrix} 1 + i[2/3a_{11} - 1/3(a_{22} + a_{33})] & ia_{21} & ia_{31} \\ ia_{12} & 1 + i[2/3a_{22} - 1/3\,(a_{11} + a_{33})] & ia_{32} \\ ia_{13} & ia_{23} & 1 + i[2/3\,a_{33} - 1/3\,(a_{11} + a_{22})] \end{bmatrix}. \tag{5.6}$$

The inverse of Λ, denoted by V, is obtained by changing the sign of a_{ij}. Then we can write

$$z_1' = z_1 + i\{(a_{11} - a_{33})z_1 + a_{21}z_2 - a_{13}z_1^2 - a_{23}z_1z_2 + a_{31}\}, \quad (5.7\text{a})$$

$$z_2' = z_2 + i\{a_{12}z_1 + (a_{22} - a_{33})z_2 + a_{32} - a_{13}z_1z_2 - a_{23}z_2^2\}, \quad (5.7\text{b})$$

$$z_3' = z_3 + i\{(a_{11} - a_{22})z_3 + a_{23}(z_1 + z_2z_3) - a_{21} + a_{12}z_3^2 - a_{13}z_3(z_1 + z_2z_3)\}, \quad (5.7\text{c})$$

etc. Also,

$$\Omega(\Lambda, z) = [1 + a_{ij}(\partial\Omega/\partial a_{ij})|_{a_{ij}=0} + a_{ij}^*(\partial\Omega/\partial a_{ij}^*)|_{a_{ij}^*=0}], \quad (5.8)$$

$$|z'\rangle = |z\rangle + a_{ij}\frac{\partial z'_k}{\partial a_{ij}}\frac{\partial}{\partial z_k}|z\rangle + a_{ij}^*\frac{\partial z'^*_k}{\partial a_{ij}^*}\frac{\partial}{\partial z_k^*}|z\rangle. \quad (5.9)$$

Combining (5.7–9), we can write

$$U(\Lambda)|z\rangle = (1 + ia_{ij}A_{ij}^L + ia_{ij}^*A_{ij}^R)|z\rangle \quad (5.10\text{a})$$

$$= \left\{1 + a_{ij}\left[\frac{\partial\Omega}{\partial a_{ij}}\Big|_{a_{ij}=0} + \frac{\partial z'_k}{\partial a_{ij}}\frac{\partial}{\partial z_k}\right] + a_{ij}^*\left[\frac{\partial\Omega}{a_{ij}^*}\Big|_{a_{ij}^*=0} + \frac{\partial z'^*_k}{\partial a_{ij}^*}\frac{\partial}{\partial z_k^*}\right]\right\}|z\rangle. \quad (5.10\text{b})$$

Comparing 5.10a and 5.10b, we finally obtain

$$\left(\frac{\partial\Omega}{\partial a_{ij}}\Big|_0 + \frac{\partial z'_k}{\partial a_{ij}}\frac{\partial}{\partial z_k}\right)|z\rangle = iA_{ij}^L|z\rangle, \quad (5.11\text{a})$$

$$\left(\frac{\partial\Omega}{\partial a_{ij}^*}\Big|_0 + \frac{\partial z'^*_k}{\partial a_{ij}^*}\frac{\partial}{\partial z^*_k}\right)|z\rangle = iA_{ij}^R|z\rangle. \quad (5.11\text{b})$$

The right-hand side of (5.11) is known from (5.2–3); thus, we get information about the derivatives of Ω. In fact, we find that with infinitesimal Λ (and V), Ω is given for class A, B as

$$\Omega_A = (\Lambda_{31}z_1 + \Lambda_{32}z_2 + \Lambda_{33})^{3/2\lambda_L} \times (\Lambda_{31}^*z_1^* + \Lambda_{32}^*z_2^* + \Lambda_{33}^*)^{3/2\lambda_R}, \quad (5.12\text{a})$$

$$\begin{aligned}\Omega_B = {} & (\Lambda_{31}z_1 + \Lambda_{32}z_2 + \Lambda_{33})^{\eta_L+2\lambda_L+2} \\ & \times (\Lambda_{31}^*z_1^* + \Lambda_{32}^*z_2^* + \Lambda_{33}^*)^{\eta_R+2\lambda_R+2} \\ & \times (V_{31}z_4 + V_{21}z_3 + V_{11})^{-2\eta_L-\lambda_L-4} \\ & \times (V_{31}^*z_4^* + V_{21}^*z_3^* + V_{11}^*)^{-2\eta_R-\lambda_R-4}\end{aligned} \quad (5.12\text{b})$$

We can check that this form of Ω has the group property, as it should, namely:

$$U(\Lambda_2)U(\Lambda_1)|z\rangle = U(\Lambda_2)\Omega(\Lambda_1, z)|\Lambda_1 z\rangle \quad (5.13\text{a})$$

$$= \Omega(\Lambda_1, z)\Omega(\Lambda_2, \Lambda_1 z)|(\Lambda_2\Lambda_1)z\rangle \quad (5.13\text{b})$$

$$\begin{aligned}&= \Omega(\Lambda_2\Lambda_1, z)|(\Lambda_2\Lambda_1)z\rangle \\ &= U(\Lambda_2\Lambda_1)|z\rangle.\end{aligned} \quad (5.13\text{c})$$

Therefore, by applying successive infinitesimal transformations, we can integrate (5.12) to obtain the same formula as above, but now with finite Λ. Thus (5.5) and (5.12) give a very simple expression for all representations of $SL(3,C)$.

6. COVARIANT AND CONTRAVARIANT KET AND BRA VECTORS AND INVARIANT BILINEAR FUNCTIONALS

The states $|z\rangle$ labeled with the two sets of roots $\lambda_1^{L,R} = \lambda^{L,R}$, $\lambda_2^{L,R} = \eta^{L,R}$ of Eq. (2.12), (3.2), (3.3), and transforming as specified in (5.5) and (5.12) will be called *covariant kets*. Thus a covariant ket is characterized by the set of four numbers $(\lambda_L, \eta_L; \lambda_R, \eta_R)$. We now introduce the *contravariant kets*, as the states $|z)$, labeled by $\lambda_1^{L,R} = \lambda^{L,R}$ and $\lambda_3^{L,R} = \rho^{L,R}$ (the third root), which transforms like $|z\rangle$, but with $\rho^{L,R}$ replacing $\eta^{L,R}$. Notice that $\rho^{L,R}$ are *not* independent, and must satisfy

$$\rho^L = -3 - \lambda^L - \eta^L, \qquad \rho^R = -3 - \lambda^R - \eta^R. \quad (6.1)$$

Thus, the contravariant ket is characterized by $(\lambda_L, \rho_L; \lambda_R, \rho_R)$.

We define a *metric* operator G, which transforms the covariant kets to the contravariant ones:

$$G = \Pi_{3L}^{\rho_L-\eta_L}\Pi_{3R}^{\rho_R-\eta_R}, \quad (6.2)$$

$$|z) = G|z\rangle. \quad (6.3)$$

Of course, $G = 1$ for class A, thus, for this class there is no distinction between covariant and contravariant states.

Indeed, we find, by virtue of Eq. (4.6), that $|z)$ differs from $|z\rangle$ in its transformation properties only by the change $\eta^{L,R} \leftrightarrow \rho^{L,R}$.

Now, we define *covariant bras* by the normalization condition (it should be emphasized that we do *not* identify a bra state with the naive Hermitian conjugate of a ket state):

$$\langle\omega|z\rangle = \delta(\omega - z)\delta(\omega^* - z^*). \quad (6.4)$$

The transformation properties of $\langle\omega|$ must be consistent with Eq. (6.4). Defining

$$\langle\omega|U^{-1}(\Lambda) = \tilde{\Omega}(\Lambda, \omega)\langle\omega'|, \quad (6.5)$$

we find

$$\langle\omega|z\rangle = \langle\omega|U^{-1}(\Lambda)U(\Lambda)|z\rangle \quad (6.6\text{a})$$

$$= \tilde{\Omega}(\Lambda, \omega)\Omega(\Lambda, z)\langle\omega'|z'\rangle, \quad (6.6\text{b})$$

which gives

$$\delta(\omega - z)\delta(\omega^* - z^*) = \tilde{\Omega}(\Lambda, \omega)\Omega(\Lambda, z)\delta(\omega' - z')\delta(\omega'^* - z'^*). \quad (6.7)$$

By considering Eqs. (3.6) and (3.10) we find, for class A

$$\delta^{(2)}(z' - \omega') = \delta^{(2)}(z - \omega)(\Lambda_{31}z_1 + \Lambda_{32}z_2 + \Lambda_{33})^3 \quad (6.8\text{a})$$

and for class B

$$\delta^{(3)}(z' - \omega') = \delta^{(3)}(z - \omega)(\Lambda_{31}z_1 + \Lambda_{32}z_2 + \Lambda_{33})^2 \times (V_{11} + V_{21}z_3 + V_{31}z_4)^2. \quad (6.8\text{b})$$

Thus, we solve for $\tilde{\Omega}(\Lambda, \omega)$:

$$\tilde{\Omega}_A = (\Lambda_{31}\omega_1 + \Lambda_{32}\omega_2 + \Lambda_{33})^{-3/2\lambda_L-3} \times (\Lambda_{31}^*\omega_1^* + \Lambda_{32}^*\omega_2^* + \Lambda_{33}^*)^{-3/2\lambda_R-3}, \quad (6.9\text{a})$$

$$\begin{aligned}\tilde{\Omega}_B &= (\Lambda_{31}\omega_1 + \Lambda_{32}\omega_2 + \Lambda_{33})^{-\eta_L-2\lambda_L-4} \\ &\times (\Lambda^*_{31}\omega^*_1 + \Lambda^*_{32}\omega^*_1 + \Lambda^*_{33})^{-\eta_R-2\lambda_R-4} \\ &\times (V_{11} + V_{21}z_3 + V_{31}z_4)^{2\eta_L+\lambda_L+2} \\ &\times (V^*_{11} + V^*_{21}z^*_3 + V^*_{31}z^*_4)^{2\eta_R+\lambda_R+2}. \end{aligned} \tag{6.9b}$$

We have thus obtained the transformation properties of the bra vector $\langle\omega|$. We see that, if the ket vector transforms like $(\lambda_L, \eta_L; \lambda_R, \eta_R)$, then the bra vector transforms like $(-\lambda_L - 2, -\eta_L - 2; -\lambda_R - 2, -\eta_R - 2)$. While the naive Hermitian conjugate of the ket vector denoted by $(|z\rangle)^+$ would transform according to $(\lambda^*_R, \eta^*_R, \lambda^*_L, \eta^*_L)$, which in general is different than the transformation properties of the bra vector $\langle z|$. This is due to the presence of a nontrivial metric operator G.

Similarly, we can define contravariant bras $(\omega|$

$$(\omega|z) = \delta(\omega - z)\delta(\omega^* - z^*), \tag{6.10}$$

which transform like $(-\lambda_L - 2, -\rho_L - 2; -\lambda_R - 2, -\rho_R - 2)$.

The covariant and contravariant bras are related to each other by

$$\langle\omega|G^{-1} = (\omega|. \tag{6.11}$$

Thus, we find that the matrix elements of G between covariant states are equal to the overlap functions between covariant bras and contravariant kets, etc.:

$$\langle\omega|z) = \langle\omega|G|z\rangle = (\omega|G|z) \equiv G(\omega, z), \tag{6.12a}$$

$$(\omega|z = \langle\omega|G^{-1}|z\rangle = (\omega|G^{-1}|z) \equiv \tilde{G}(\omega, z). \tag{6.12b}$$

We will explicitly evaluate these matrix elements in Sec. 8.

We finally note that the normalization conditions (6.4) and (6.10) imply the completeness relations

$$1 = \int(d^2z_i)|z\rangle\langle z| = \int(d^2\omega_i)|\omega)(\omega| \tag{6.13a}$$

$$= \int(d^2z_i)(d^2\omega_i)|z\rangle\langle z|\omega)(\omega| = \int(d^2z_i)(d^2\omega_i)|z\rangle G(z,\omega)(\omega| \tag{6.13b}$$

$$= \int(d^2z_i)(d^2\omega_i)|\omega)(\omega|z\rangle\langle z| = \int(d^2z_i)(d^2\omega_i)|\omega)\tilde{G}(z,\omega)\langle z|. \tag{6.13c}$$

These can be used to define invariant bilinear products in the group. For example, we consider the "wave-packets"

$$\langle\Psi| = \int(d^2z_i)\Psi(z)\langle z|, \tag{6.14a}$$

$$|\phi\rangle = \int(d^2z_i)\phi(z)|z), \tag{6.14b}$$

$$|\chi\rangle = \int(d^2z_i)\chi(z)|z\rangle,$$

which transform as

$$\langle\Psi'| = \langle\Psi|U^{-1}(\Lambda), \qquad |\phi'\rangle = U(\Lambda)|\phi\rangle, \qquad |\chi'\rangle = U(\Lambda)|\chi\rangle. \tag{6.15}$$

From the transformation properties of the states $\langle z|$, $|z)$ and $|z\rangle$ we deduce those for the "vector components" $\psi(z)$ and $\phi(\omega)$:

$$\chi(z) \to \chi'(z) = \tilde{\Omega}^{\lambda,\eta}(\Lambda, z)\chi(z'),$$

$$\psi(z) \to \psi'(z) = \Omega^{\lambda,\eta}(\Lambda, z)\psi(z'), \tag{6.16a}$$

$$\phi(z) \to \phi'(z) = \tilde{\Omega}^{\lambda,\rho}(\Lambda, z)\phi(z'), \tag{6.16b}$$

where we have indicated also the dependence on the roots λ, η, ρ, etc., and, $\Omega, \tilde{\Omega}$ are given in (5.12) and (6.9). That is, $\psi(z)$ transforms like $(\lambda_L, \eta_L; \eta_R, \lambda_R)$, $\phi(z)$ transforms like $(-\lambda_L - 2, -\rho_L - 2; -\lambda_R - 2, -\rho_R - 2)$ and $\chi(z)$ like $(-\lambda_L - 2, -\eta_L - 2; -\lambda_R - 2, -\eta_L - 2)$. For functions transforming according to (6.16), we can easily write invariant bilinear functionals $\langle\psi|\phi\rangle$ and $\langle\psi|\chi\rangle$ by using (6.13):

$$\langle\psi|\phi\rangle = \int(d^2z_i)(d^2\omega_i)\langle\psi|z\rangle G(z,\omega)(\omega|\phi\rangle \tag{6.17a}$$

$$= \int(d^2z_i)(d^2\omega_i)\psi(z)G(z,\omega)\phi(\omega), \tag{6.17b}$$

$$\langle\psi|\chi\rangle = \int d^2z\langle\psi|z\rangle\langle z|\chi\rangle \tag{6.18a}$$

$$= \int d^2z\psi(z)\chi(z). \tag{6.18b}$$

7. UNITARITY CONDITIONS

To impose unitarity, it is enough to demand that the generators J_α, K_α of Sec. 2 be Hermitian

$$G^{-1}J^+_\alpha G = J_\alpha, \qquad G^{-1}K^+_\alpha G = K_\alpha. \tag{7.1}$$

Where, in general, G is the metric of the representation. This means that $G^{-1}X^{R+}_\alpha G = X^L_\alpha$, or from Eq. (2.4)

$$G^{-1}A^{L+}_{ij}G = A^R_{ij}. \tag{7.2}$$

We further apply the condition (7.2) to the Casimir operators, and obtain

$$C^*_{1L} = G^{-1}\,(\mathrm{Tr}A^2_L)^+G = \mathrm{Tr}A^2_R = C_{1R}, \tag{7.3a}$$

$$C^*_{2L} = G^{-1}\,(\mathrm{Tr}A^3_L)^+G = \mathrm{Tr}A^3_R - 3\,\mathrm{Tr}A^2_R = C_{2R} - 3C_{1R}. \tag{7.3b}$$

Now considering Eqs. (2.12b), and (3.1), we can write

$$\lambda^3_{Ri} + 3\lambda^2_{Ri} + (2 - 1/2C^*_{1L})\lambda_{Ri} + (1/3C^*_{2L} - 1/2C^*_{1L}) = 0 \tag{7.4}$$

Replacing from (3.3), we obtain the cubic equation

$$\lambda^3_{Ri} + 3\lambda^2_{Ri} + (\lambda^*_{1L}\lambda^*_{2L} + \lambda^*_{2L}\lambda^*_{3L} + \lambda^*_{3L}\lambda^*_{1L})\lambda_{Ri} + \lambda^*_{1L}\lambda^*_{2L}\lambda^*_{3L} + 2(\lambda^*_{1L}\lambda^*_{2L} + \lambda^*_{1L}\lambda^*_{3L} + \lambda^*_{2L}\lambda^*_{3L}) - 4 = 0. \tag{7.5}$$

Remembering (3.3a), this can be written in the form

$$(\lambda_{Ri} + \lambda^*_{1L} + 2)(\lambda_{Ri} + \lambda^*_{2L} + 2)(\lambda_{Ri} - \lambda^*_{1L} - \lambda^*_{2L} - 1) = 0. \tag{7.6}$$

Therefore, according to the unitarity condition (7.1), a representation labeled by the numbers $(\lambda_L, \eta_L; \lambda_R, \eta_R)$ cannot have arbitrary complex numbers as its labels, but they must satisfy *one* of the following sets of conditions (7a–f):

$$\lambda_R = -\lambda^*_L - 2, \qquad \eta_R = -\eta^*_L - 2, \tag{7.7a}$$

$$\lambda_R = -\lambda^*_L - 2, \qquad \eta_R = +\lambda^*_L + \eta^*_L + 1, \tag{7.7b}$$

$$\lambda_R = -\eta^*_L - 2, \qquad \eta_R = -\lambda^*_L - 2, \tag{7.7c}$$

$$\lambda_R = -\eta^*_L - 2, \qquad \eta_R = \lambda^*_L + \eta^*_L + 1, \tag{7.7d}$$

$$\lambda_R = \eta^*_L + \lambda^*_L + 1, \qquad \eta_R = -\lambda^*_L - 2, \tag{7.7e}$$

$$\lambda_R = \eta^*_L + \lambda^*_L + 1, \qquad \eta_R = -\eta^*_L - 2. \tag{7.7f}$$

Further restrictions are obtained if we realize that the generators J_α form an $SU(3)$ compact subgroup. Therefore, the third components of its isospin and U-spin subgroups, J_3 and $-1/2J_3+\sqrt{3}/2J_8$, have half-integer eigenvalues; while its hypercharge $Y=2/\sqrt{3}J_8$ has 1/3 integer eigenvalues (e.g., quarks). These operators are given as

$$J_3 = 1/2(A^L_{11}-A^L_{22}+A^R_{11}-A^R_{22}) = 1/2(\Pi_{1L}Z_{1L}+\Pi_{1R}Z_{1R}-\Pi_{2L}Z_{2L}-\Pi_{2R}Z_{2R}) + (\Pi_{3L}Z_{3L}+\Pi_{3R}Z_{3R})+(\eta_L-\eta_R)+1/2(\lambda_L-\lambda_R), \tag{7.8a}$$

$$-1/2J_3+\sqrt{3}/2J_8 = 1/2(A^L_{11}-A^L_{33}+A^R_{11}-A^R_{33}) = (\Pi_{1L}Z_{1L}+\Pi_{1R}Z_{2R}) + 1/2(\Pi_{2L}Z_{2L}+\Pi_{2R}Z_{2R}+\Pi_{3L}Z_{3L}+\Pi_{3R}Z_{3R}) + 1/2\,(\eta_L-\eta_R)-1/2\,(\lambda_L-\lambda_R), \tag{7.8b}$$

$$Y=-(A^L_{33}+A^R_{33}) = (\Pi_{1L}Z_{1L}+\Pi_{1R}Z_{2R}+\Pi_{2L}Z_{2L}+\Pi_{2R}Z_{2R})-(\lambda_L-\lambda_R). \tag{7.8c}$$

We see now easily that the state $|z=0\rangle$ is an eigenstate of these operators. Thus, we must take $(\eta_L-\eta_R)+1/2(\lambda_L-\lambda_R)$ and $1/2(\eta_L-\eta_R)-1/2(\lambda_L-\lambda_R)$ half integers, and $(\lambda_L-\lambda_R)$, 1/3 integer. The most general solution for class B is then

$$\lambda_L-\lambda_R=\tfrac{1}{3}(n-2m), \tag{7.9a}$$

$$\eta_L-\eta_R=\tfrac{1}{3}(n+m),\qquad m,n=\text{integers}. \tag{7.9b}$$

For the representation of class A, the condition is simpler:

$$\lambda_L-\lambda_R=-2m/3. \tag{7.10}$$

The most general classes of solutions compatible with 7.7 and 7.9 or 7.10 for class B are

$$\lambda_L=\tfrac{1}{6}(n-2m)-1+ia,\quad \eta_L=\tfrac{1}{6}(n+m)+1+ib,\quad \lambda_R=-\lambda_L^*-2,\quad \eta_R=-\eta_L^*-2, \tag{7.11a}$$

$$\lambda_L=-m/3-1-ia,\quad \eta_L=m/6-\tfrac{1}{2}(2+c)+ia/2,\quad \lambda_R=-\lambda_L^*-2,\quad \eta_R=\lambda_L^*+\eta_L^*+1, \tag{7.11b}$$

$$\lambda_L=-n/6-\tfrac{1}{2}(2+c)-ia/2,\quad \eta_L=n/3-1+ia,\quad \lambda_R=\lambda_L^*+\eta_L^*+1,\quad \eta_R=-\eta_L-2 \tag{7.11c}$$

with a,b,c arbitrary real numbers. For class A, we simply ignore η_L, and put $n=0$. We have thus obtained some necessary conditions for the representations of the type (4.2), (4.4) or (5.12) to be unitary. Further restrictions will come from demanding a positive definite Hermitian scalar product, as discussed in the next section.

8. HERMITIAN, POSITIVE DEFINITE SCALAR PRODUCT

For a unitary representation, we must specify a Hermitian, positive definite scalar product. Thus, we consider two functions $\Psi_1(z)$ and $\Psi_2(z)$, transforming according to the general representation $(\lambda_L,\eta_L;\lambda_R,\eta_R)$, like a covariant ket, as in Eq. (6.16a). The complex conjugate of, say, the first function, $\Psi_1(z)^*$, then transforms as $(\lambda_R^*,\eta_R^*;\lambda_L^*,\eta_L^*)$, [see Eq. (5.12)]. Using Eqs. (7.11a–c) and (6.1), we find three cases for class B):

1. $$(\lambda_R^*,\eta_R^*;\lambda_L^*,\eta_L^*) = (-\lambda_L-2,-\eta_L-2;-\lambda_R-2,-\eta_R-2), \tag{8.1a}$$

2. $$(\lambda_R^*,\eta_R^*;\lambda_L^*,\eta_L^*) = (-\lambda_L-2,-\rho_L-2;-\lambda_R-2;-\rho_R-2), \tag{8.1b}$$

3. $$(\lambda_R^*,\eta_R^*;\lambda_L^*,\eta_L^*) = (-\rho_L-2,-\eta_L-2;-\rho_R-2;-\eta_R-2). \tag{8.1c}$$

We call case 1 the principal series and case 2 the supplementary series and integer points. Comparing the definition of covariant and contravariant kets and bras, we see that, in a unitary representation, $\Psi_1^*(z)$ transforms like a covariant bra in case 1, and like a contravariant bra in case 2. We have not defined a name for case 3. Thus, according to Eqs. (6.17) and (6.18), we can write the invariant scalar products, provided the integrals converge

$$(\Psi_1,\Psi_2)=\int d^2z_i d^2\omega_i \Psi_1^*(z)G(z,\omega)\Psi_2(\omega) \tag{8.2}$$

where, for case 1 we simply take $G(z,\omega)=\delta(z-\omega)$ and for case 2, the operator G is now given as

$$G=\Pi_{3L}^c\Pi_{3R}^c. \tag{8.3a}$$

According to (7.2), we can write $\Pi_{3R}=\Pi_{3L}{}^+$, therefore

$$G=(\Pi_{3L}\Pi_{3L}{}^+)^c \tag{8.3b}$$

is a positive definite operator! Thus, the function $G(z,\omega)=\langle z|G|\omega\rangle$ is positive definite, and therefore the scalar product of (8.2) is positive definite and Hermitian, for the principal series, supplementary series and integer points of cases 1 and 2. For case 3 we have not found a metric operator analogous to G, thus we cannot write a unitary scalar product for this case.

It is not hard to evaluate the matrix elements of the operator G. As in Ref. 5, we define the Π_3-representation, in which Π_3 rather than Z_3 is diagonalized. The Π_3 and Z_3 representations are related by a Fourier transformation, which allow the evaluation of $G(z,\omega)$:

$$G(z,\omega)=\langle z|(\Pi_{3L}\Pi_{3L}^+)^c|\omega\rangle=\frac{\Gamma(1+c)}{\pi\Gamma(-c)}|z_3-\omega_3|^{-2c-2}. \tag{8.4}$$

This expression now allows us to discuss the different explicit forms of the scalar product defined in Eq. (8.2), by taking appropriate limits of the Casimir operators.

Class A: $\Psi_1(z)$ and $\Psi_2(z)$ transform as in Eq. (5.12), with

$$\lambda_L=-\tfrac{1}{3}m-1+ia,\qquad \lambda_R=\tfrac{1}{3}m-1+ia. \tag{8.5a}$$

We can write the scalar product

$$(\Psi_1,\Psi_2)=\int d^2z_1d^2z_2\,\Psi_1^*(z)\Psi_2(z). \tag{8.5b}$$

For convergence, we demand that Ψ_1,Ψ_2 are $L^{(2)}$ functions.

Class B: $\Psi_1(z)$ and $\Psi_2(z)$ transform as in Eq. (5.12) with $(\lambda_L,\eta_L;\lambda_R,\eta_R)$ specified as below for the various representations.

1. Principal series.

$$\lambda_L=\tfrac{1}{6}(n-2m)-1+ia,\qquad \eta_L=\tfrac{1}{6}(n+m)-1+ib, \tag{8.6a}$$

$$\lambda_R = -\tfrac{1}{6}(n-2m) - 1 + ia, \quad \eta_R = -\tfrac{1}{6}(n+m) - 1 + ib, \tag{8.6b}$$

$$(\psi_1, \psi_2) = \int d^2z_1 d^2z_2 d^2z_3 \psi_1^*(z)\psi_2(z). \tag{8.6c}$$

For convergence, we must have $\psi_1, \psi_2, L^{(2)}$ functions.

2. Supplementary series.

$$\lambda_L = -\frac{m}{3} - 1 - ia, \quad \eta_L = \frac{m}{6} - \frac{1}{2}(2+c) + i\frac{a}{2}, \tag{8.7a}$$

$$\lambda_R = \frac{m}{3} - 1 - ia, \quad \eta_R = -\frac{m}{6} - \frac{1}{2}(2+c) + i\frac{a}{2}, \tag{8.7b}$$

$$(\psi_1, \psi_2) = \frac{\Gamma(1+c)}{\pi\Gamma(-c)} \int d^2z_1 d^2z_2 d^2z_3 d^2\omega_3 \psi_1^*(z_1, z_2, z_3) \times |z_3 - \omega_3|^{-2c-2}\psi_2(z_1, z_2, \omega_3). \tag{8.7c}$$

If ψ_1, ψ_2 are $L^{(2)}$ functions, then the integral converges only for

$$0 < |c| < 1. \tag{8.7d}$$

3. Integer points. For an analogous situation in $SL(2,c)$ and a discussion of such representations we refer to ref. 5. Here we only mention the result. λ_L, η_L, etc., are as in (8.7) with $c =$ integer.

Case a: $c = k, k$ being a positive integer or zero.

$$(\psi_1, \psi_2) = (-1)^k \int d^2z_1 d^2z_2 d^2z_3 \psi_1^*(z) \frac{\partial^{2k}}{\partial z_3^k \partial z_3^{*k}} \psi_2(z). \tag{8.8a}$$

The functional space is restricted to functions such that

$$\frac{\partial^{2k}}{\partial z_3^k \partial z_3^{*k}} \psi_{1,2}(z) \neq 0. \tag{8.8b}$$

Case b: $c = -k, k$ being a positive integer.

$$(\psi_1, \psi_2) = \{2(-1)^k / \pi[\Gamma(k)]^2\} \int d^2z_1 d^2z_2 d^2z_3 d^2\omega_3 \times |z_3 - \omega_3|^{2k-2} \log|z_3 - \omega_3| \psi_1^*(z_1, z_2, z_3)\psi(z_1, z_2, \omega_3). \tag{8.9a}$$

The functional space is restricted to functions such that

$$\int d^2z_3 z_3^r z_3^{*s} \psi_{1,2}(z) = 0 \tag{8.9b}$$

for $r, s \leqslant k-1$.

We have thus classified the unitary representations of $SL(3,C)$, and have given the explicit transformation properties and the Hermitian positive definite scalar products. Our operational method yielded the form of the scalar product for the principal and supplementary series as given by Gel'fand and Naimark.[2] In addition our method has given consistent normalization factors due to the normalization condition of Eq. (6.4). Further, it has unified all scalar products, for all representations, in a single closed expression (8.2) and (8.4), from which the various scalar products are obtained by taking the appropriate limits.

APPENDIX

In this appendix we construct the operators Z_i and show that they commute with each other. We also give some interesting properties of the operators K_{ij}. With the help of these we also find the commutation relations of A_{ij} with Z_k.

In the following, we shall need to write inverses of certain operators. Throughout, we shall assume that the inverses do exist, by interpreting them as operating on a complete set of states for which the inverse is defined.

Starting with Eqs. (3.1) and (3.7) and multiplying the first one from the right by Ψ_3^{-1} and the second from the left by Φ_3^{-1}, we obtain explicitly

$$\begin{pmatrix} A_{11}-\lambda & A_{12} & A_{13} \\ A_{21} & A_{22}-\lambda & A_{23} \\ A_{31} & A_{32} & A_{33}-\lambda \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \\ 1 \end{pmatrix} = 0, \tag{A1a}$$

$$(1\ Z_3\ Z_4) \begin{pmatrix} A_{11}-\eta & A_{12} & A_{13} \\ A_{21} & A_{22}-\eta & A_{23} \\ A_{31} & A_{32} & A_{33}-\eta \end{pmatrix} = 0. \tag{A1b}$$

Since we are dealing with operators, we wish to detail our procedure carefully. We have six equations satisfied by Z_1, Z_2, Z_3, Z_4. Using the set (A1a) we can solve for Z_1 by eliminating Z_2 from any two equations, and similarly, we can solve for Z_2. Of course, the solutions obtained in this way for Z_1 must be compatible no matter which set of two equations is used to eliminate Z_2.

In eliminating Z_2 one must be careful that A_{ij} are operators which do not commute. For example, supposing we want to eliminate Z_2 by using the first and second rows in Eq. A1a):

$$(A_{11}-\lambda)Z_1 + A_{12}Z_2 + A_{13} = 0, \tag{A2a}$$

$$A_{21}Z_1 + (A_{22}-\lambda)Z_2 + A_{23} = 0. \tag{A2b}$$

We multiply (A2a) from the left by $(A_{22}-\lambda-1)$ and (A2b) from the left by $-A_{12}$, and add the resulting equations. Since $(A_{22}-\lambda-1)A_{12} = A_{12}(A_{22}-\lambda)$ the Z_2 terms cancel and we get

$$[(A_{22}-\lambda-1)(A_{11}-\lambda) - A_{12}A_{21}]Z_1 + [(A_{22}-\lambda-1)A_{13} - A_{12}A_{23}] = 0, \tag{A3a}$$

from which we obtain Z_1 in terms of A_{ij}. Thus,

$$Z_1 = [-(A_{11}-\lambda)(A_{22}-\lambda-1) + A_{12}A_{21}]^{-1} \times [(A_{22}-\lambda-1)A_{13} - A_{12}A_{23}]. \tag{A3b}$$

Similarly, we could have used the first and third rows, or the second and third rows, which would have given different expressions for Z_1. All of these solutions must be equal to each other. Here we give the result in a compact form. We define some new operators K_{ij}:

$$K_{ij}(\lambda) = A_{ij}(A_{kk}-\lambda-1) - A_{ik}A_{kj}, \quad i \neq j, \tag{A4a}$$

$$K_{ii}(\lambda) = -(A_{jj}-\lambda)(A_{kk}-\lambda-1) + A_{jk}A_{kj}, \tag{A4b}$$

$i, j, k = 1, 2, 3$ and *cyclic* or *anticyclic, no sum* over repeated indices. We notice that $K_{ij}(\lambda)$ is almost the cofactor matrix transposed of the matrix in (A1a), but not quite. This is because A_{ij} are noncommuting operators rather than numbers.

In terms of K_{ij} we can write Z_i in a compact form as

$$Z_1 = K_{3n}^{-1}(\lambda)K_{1n}(\lambda) = K_{1m}(\lambda+1)K_{3m}^{-1}(\lambda+1), \tag{A5a}$$

$$Z_2 = K_{3n}^{-1}(\lambda)K_{2n}(\lambda) = K_{2m}(\lambda+1)K_{3m}^{-1}(\lambda+1), \tag{A5b}$$

$$Z_3 = K_{n2}(\eta)K_{n1}^{-1}(\eta) = K_{m1}^{-1}(\eta+1)K_{m2}(\eta+1), \tag{A5c}$$

$$Z_4 = K_{n3}(\eta)K_{n1}^{-1}(\eta) = K_{m1}^{-1}(\eta+1)K_{m3}(\eta+1), \tag{A5d}$$

$n, m = 1, 2, 3$, no sum over repeated indices.

The three values of $n, m = 1, 2, 3$ correspond to taking different sets of equations in order to solve for Z_i. They are all equal to each other as will be shown below. The second expression for each Z_i is obtained due to certain properties of the K_{ij}. We will also show that $[Z_i, Z_j] = 0$ for any i, j. All of these claims will be proven below.

The operators $K_{ij}(\lambda)$ have the following interesting properties. They can be easily checked by explicitly using the commutation properties of A_{ij} [Eq. (2.6)].

$$[A_{ij}, K_{lm}(\lambda)] = \delta_{im}K_{lj}(\lambda) - \delta_{lj}K_{im}(\lambda), \tag{A6a}$$

$$K_{ij}(\lambda)K_{lm}(\lambda+1) = K_{lj}(\lambda)K_{im}(\lambda+1), \tag{A6b}$$

$$K_{ij}(\eta+1)K_{lm}(\eta) = K_{im}(\eta+1)K_{lj}(\eta), \tag{A6c}$$

$$\sum_{i=1}^{3} K_{in}(\lambda)K_{mi}(\eta) = 0, \tag{A6d}$$

$$\sum_{i=1}^{3} K_{mi}(\eta+1)K_{in}(\lambda+1) = 0, \tag{A6e}$$

$$\sum_{j=1}^{3} A_{ij}K_{jl}(\lambda+1) = \lambda K_{il}(\lambda+1), \tag{A6f}$$

$$\sum_{j=1}^{3} K_{ij}(\eta+1)A_{jl} = \eta K_{il}(\eta+1). \tag{A6g}$$

No sum over repeated indices unless shown.

We obtained (A6a–c) by direct commution of the A_{ij}. These three equations are enough to show that Z_i are equal for each n, and also that $[Z_1, Z_2] = 0 = [Z_3, Z_4]$. We rewrite (A6b) as

$$K_{lj}(\lambda)^{-1}K_{ij}(\lambda) = K_{im}(\lambda+1)K_{lm}^{-1}(\lambda+1). \tag{A7a}$$

Taking $l = 3$, $i = 1, 2$, we see that the second expressions for Z_1 and Z_2 are the same as the first for arbitrary j, m. Now we can write Z_1 as given by Eq. (A5a):

$$\begin{aligned} Z_1 &= K_{33}^{-1}(\lambda)K_{13}(\lambda) \\ &= K_{13}(\lambda+1)K_{33}^{-1}(\lambda+1) \\ &= K_{31}^{-1}(\lambda)K_{11}(\lambda). \end{aligned} \tag{A7b}$$

In writing the second line we used (A7a) with $l = 3$, $j = 3$, $i = 1$, $m = 3$, and in writing the third line we used (A7a) once more with $l = 3$, $j = 1$, $i = 1$, $m = 3$. Therefore, (A7b) shows that the expressions for Z_1 with $n = 3$ and $n = 1$ are equal to each other. Similarly for Z_2, and for all values of n, m. Now we calculate the commutator

$$\begin{aligned} [Z_1, Z_2] &= Z_1Z_2 - Z_2Z_1 \\ &= K_{3n}^{-1}(\lambda)K_{1n}(\lambda)K_{2m}(\lambda+1)K_{3m}^{-1}(\lambda+1) \\ &\quad - K_{3n}^{-1}(\lambda)K_{2n}(\lambda)K_{1m}(\lambda+1)K_{3m}^{-1}(\lambda+1) \\ &= K_{3n}^{-1}(\lambda)[K_{1n}(\lambda)K_{2m}(\lambda+1) \\ &\quad - K_{2n}(\lambda)K_{1m}(\lambda+1)]K_{3m}^{-1}(\lambda+1). \end{aligned} \tag{A8a}$$

Using (A6b) we find

$$[Z_1, Z_2] = 0. \tag{A8b}$$

The same things can be proven for Z_3 and Z_4 by using Eq. (A6c) instead of (A6b).

Equation (A6e) is obtained by replacing in (3.13c) the second expressions in (A5). Similarly, (A6f) and (A6g) are obtained by replacing (A5) in (A1).

To obtain Eq. (A6d), we first write the identity

$$K_{lm}(\lambda+1) = K_{lm}(\eta) - (\lambda+1-\eta)[A_{lm} + \delta_{lm}(\lambda+\eta+2)] \tag{A9a}$$

and replace it in (A6b), to get

$$\begin{aligned} K_{ij}(\lambda)K_{lm}(\eta) &= K_{lj}(\lambda)K_{im}(\lambda+1) \\ &\quad + (\lambda+1-\eta)K_{ij}(\lambda)[A_{lm} + \delta_{lm}(\lambda+\eta+2)]. \end{aligned} \tag{A9b}$$

We then let $i = m$, sum over $i = 1, 2, 3$, and replace in the resulting equation

$$\mathrm{Tr}K(\lambda+1) = (\eta - \lambda - 1)(2\lambda + \eta + 4), \tag{A10}$$

which is evaluated with the help of (A9a), (A4b), (3.3a–b), and the fact that λ and η are roots of Eq. (3.2). With all that, Eq. (A9b) becomes

$$\sum_i K_{ij}(\lambda)K_{li}(\eta) = (\lambda+1-\eta)\{-K_{lj}(\lambda)(\lambda+2) + K_{ij}(\lambda)A_{li}\}. \tag{A11}$$

The expression in the brackets on the right hand side is equal to zero as can be seen by direct calculation. Thus, we obtain Eq. (a6d),

$$\sum_i K_{ij}(\lambda)K_{li}(\eta) = 0.$$

At this point we are ready to prove that Z_1, Z_2, and Z_3, Z_4 all commute with each other. We already have $[Z_1, Z_2] = 0 = [Z_3, Z_4]$. To prove $[Z_2, Z_3] = 0$, we start by rewriting Eq. (A6e) as

$$\begin{aligned} K_{m2}(\eta+1)K_{2n}(\lambda+1) &= -K_{m1}(\eta+1)K_{1n}(\lambda+1) \\ &\quad - K_{m3}(\eta+1)K_{3n}(\lambda+1). \end{aligned} \tag{A12a}$$

Then using (A7a), we find

$$\begin{aligned} &= -K_{m1}(\eta+1)K_{3n}^{-1}(\lambda)K_{1n}(\lambda)K_{3n}(\lambda+1) \\ &\quad - K_{m1}(\eta+1)K_{m3}(\eta)K_{m1}^{-1}(\eta)K_{3n}(\lambda+1), \end{aligned} \tag{A12b}$$

$$\begin{aligned} &= -K_{m1}(\eta+1)K_{3n}^{-1}(\lambda)\{K_{1n}(\lambda)K_{m1}(\eta) \\ &\quad + K_{3n}(\lambda)K_{m3}(\eta)\}K_{m1}^{-1}(\eta)K_{3n}(\lambda+1) \end{aligned} \tag{A12c}$$

and using (A6d) we find

$$= K_{m1}(\eta+1)K_{3n}^{-1}(\lambda)K_{2n}(\lambda)K_{m2}(\eta)K_{m1}^{-1}(\eta)K_{3n}(\lambda+1). \tag{A12d}$$

Multiplying from the left by $K_{m1}^{-1}(\eta+1)$ and from the right by $K_{3n}^{-1}(\lambda+1)$ and using (A5), we obtain

$$Z_3Z_2 = Z_2Z_3, \tag{A12e}$$

which gives

$$[Z_2, Z_3] = 0. \tag{A12f}$$

By similar operations, we obtain $[Z_1, Z_3] = 0$, and then using (3.13c) we obtain

$$[Z_1, Z_4] = 0 = [Z_2, Z_4], \tag{A12g}$$

which is the desired result.

Now we calculate the commutation relations of A_{ij} with Z_k. Using (A5) and Eq. (A6a), we find

$$A_{12}, A_{22}, A_{32} \text{ commute with } Z_1, \tag{A13a}$$

$$A_{11}, A_{21}, A_{31} \text{ commute with } Z_2, \tag{A13b}$$

$$A_{31}, A_{32}, A_{33} \text{ commute with } Z_3, \tag{A13c}$$

$$A_{21}, A_{22}, A_{23} \text{ commute with } Z_4, \tag{A13d}$$

$$[A_{11}, Z_1] = -Z_1, [A_{21}, Z_1] = -Z_2, [A_{31}, Z_1] = -1, \quad [A_{13}, Z_1] = Z_1^2, [A_{23}, Z_1] = Z_1 Z_2, \tag{A13e}$$

$$[A_{12}, Z_2] = -Z_1, [A_{22}, Z_2] = -Z_2, [A_{32}, Z_2] = -1, \quad [A_{13}, Z_2] = Z_1 Z_2, [A_{23}, Z_2] = Z_2^2, \tag{A13f}$$

$$[A_{21}, Z_3] = +1, [A_{22}, Z_3] = +Z_3, [A_{23}, Z_3] = +Z_4, \quad [A_{12}, Z_3] = -Z_3^2, [A_{13}, Z_3] = -Z_3 Z_4 \tag{A13g}$$

$$[A_{31}, Z_4] = +1, [A_{32}, Z_4] = +Z_3, [A_{33}, Z_4] = +Z_4, \quad [A_{12}, Z_4] = -Z_3 Z_4, [A_{13}, Z_4] = -Z_4^2. \tag{A13h}$$

*Research supported by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under contract number F44620-70-C-0028. This document has been approved for public release and sale; its distribution is unlimited.

[1]S. Fubini and G. Veneziano, Nuovo Cimento **67**, 29 (1970). L. Clavelli and Ramond, Phys. Rev. D **2**, 973 (1970). For further literature on dual models the reader is referred to the following review papers: V. A. Allessandrini and D. Amati, "Dual Models: Their Group Theoretic Structure. Lectures delivered at the International School of Physics, Enrico Fermi," Varenna, August 1971, CERN Preprint, Ref. TH 1425; P. Ramond, "Group Theoretical Properties of Dual Resonance Models," in *Lectures in Theoretical Physics,* edited by W. Brittin (University of Colorado Press, Boulder, 1971).
[2]I. Bars and F. Gursey, Phys. Rev. D **4**, 1769 (1971).
[3]V. Bargmann, Ann. Math. **48**, 568 (1947).
[4]I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5, Chap. 3.
[5]I. Bars and F. Gursey, J. Math. Phys. **13**, 131 (1972).
[6]F. Gursey, "Representations of Some Noncompact Groups Related to the Poincaré Group, Yale Lecture Notes," compiled by Murat Gunaydin, Yale preprint (1971)(unpublished).
[7]S. Orphanidis, Yale thesis (1972), Appendix, (unpublished).
[8]I. M. Gel'fand and M. A. Naimark, *Unitäre Darstellungen der Klassischen Gruppen* (Akademie-Verlag, Berlin, 1957).
[9]M. Gell-mann, Phys. Rev. **125**, 1067 (1962).
[10]S. Okubo, Prog. Theor. Phys. **27**, 949 (1962), Appendix.

Finite-contour dispersive inequalities

Tarik Celik*

Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

David R. Palmer[†‡]

Department of Physics, University of Toronto, Toronto 181, Ontario, Canada
(Received 13 December 1972)

Finite-contour dispersive inequalities are derived for a function $f(\xi)$ which is analytic in the ξ plane except for a right-hand cut. Inequalities are also derived for the derivative of the function. These inequalities are rigorous and the sharpest ones that can be derived using only analyticity.

1. INTRODUCTION

Dispersion relations were originally introduced as general restrictions imposed by causality on an amplitude. To exploit these restrictions by using the dispersion relations in a computational sense, one must know the absorptive part of the amplitude on the cut and its behavior at infinity. If, however, only a limited amount of information about the amplitude is available then in lieu of dispersion relations, i.e., dispersive equalities, one can study dispersive inequalities. These inequalities usually require only knowledge of an upper bound for the modulus of the form factor on the cut. In several cases these inequalities are stringent enough to place meaningful constraints on theoretical constructs.[1]

In this paper we derive a set of finite-contour dispersive inequalities. Specifically, we will consider a function $f(t)$, a form factor for example, which is the boundary value of a function $f(\xi)$ real-analytic in the disc $|\xi - t_0| \leqslant R$ except for a branch cut on the positive real axis from $\xi = t_0$ to $\xi = t_0 + R$. Upper bounds for $|f(s)|$ and $|df(s)/ds|$ $(s < t_0)$ will be found in terms of weighted integrals of an upper bound for $|f(t)|$ along the cut and of the modulus of $f(\xi)$ on the ring $|\xi - t_0| = R$. The construction of this type of inequality was motivated by the development of finite-energy sum rules[2] and the techniques we use are very similar to Okubo's.[3]

Finite-contour dispersive inequalities have several nice properties. One does not need to worry about the asymtotic behavior of $f(\xi)$ nor its singularities for $|\xi - t_0| > R$. Furthermore, they should serve as a useful tool for correlating the low-energy and high-energy properties of form factors. This is an important consideration since a good asymptotic description for a form factor does not always lead to the correct couplings and models which give the proper couplings do not always have a suitable asymptotic behavior. This is the sort of trouble one encounters when constructing Veneziano type models for form factors.[4] Since our inequalities are rigorous, every model must satisfy them.

The paper is organized as follows: In Sec. 2 inequalities for the analytic function itself are derived. Inequalities involving the derivative of the function are then constructed in Sec. 3. A certain type of integral encountered in the analysis is evaluated in the Appendix. Our principle results are contained in Eqs. (2.24), (2.25), (3.11), and (3.12).

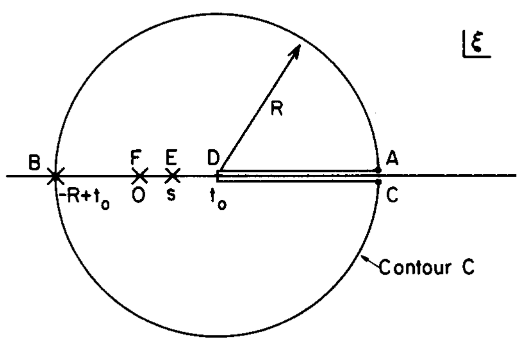

FIG. 1. Closed contour C in the complex ξ plane for the dispersion integral in Eq. (2.1)

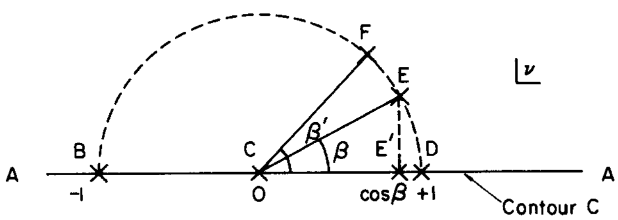

FIG. 2. Complex v plane.

2. INEQUALITIES FOR THE ANALYTIC FUNCTION

Our function $f(\xi)$, analytic in $|\xi - t_0| \leqslant R$ except for a branch cut on the positive real axis from $\xi = t_0$ to $\xi = t_0 + R$, is assumed to be "real" so that $f(\xi^*) = f^*(\xi)$ in this cut disc. For $-R + t_0 < s < t_0$ we have the representation

$$f(s) = \frac{1}{2\pi i} \int_C d\xi \frac{f(\xi)\phi(\xi)}{\xi - s} \tag{2.1}$$

where the closed contour C is indicated in Fig. 1 and $\phi(\xi)$ is any function analytic in the cut disk $|\xi - t_0| \leqslant R$ satisfying the requirements that the integral in Eq. (2.1) exists and that $\phi(s) = 1$. We assume further that the inequality

$$\rho(t) \geqslant w(t)|f(t)|^2 \tag{2.2}$$

is satisfied for $t \geqslant t_0$, where $\rho(t)$ is some positive spectral function and $w(t)$ is a known (kinematic), positive-definite factor which we take to be of the form

$$w(t) = k[(t - t_0)^{a/2}/t^{b/2}], \tag{2.3}$$

where k, a, and b are constants. This form for $w(t)$ is sufficiently general to apply to most situations; it will be apparent how the analysis can be altered to include other forms. We now map the cut-disk $|\xi - t_0| \leqslant R$ onto the upper half-plane $\mathrm{Im}\, v \geqslant 0$ by means of the conformal transformation

$$v = \left[\frac{R^{1/2} + (\xi - t_0)^{1/2}}{R^{1/2} - (\xi - t_0)^{1/2}}\right]^2 \tag{2.4}$$

We note the following properties of this transformation (see Fig. 2):

(i) The contour C is mapped onto the real v axis such that the points $\xi = R + t_0 + i0$, $-R + t_0$, $R + t_0 - i0$, and t_0 correspond to the points $v = \pm\infty$, -1, 0, $+1$ respectively

(ii) The points $\xi = s$ and $\xi = 0$ are mapped onto the points $v = \exp(i\beta)$, and $v = \exp(i\beta')$ respectively, where

$$\tan^2(\beta/4) = (t_0 - s)/R, \tag{2.5a}$$

$$\tan^2(\beta'/4) = t_0/R. \tag{2.5b}$$

The integral Eq. (2.1) can be rewritten as

$$f(s) = \frac{1}{2\pi i}\int_{-\infty}^{+\infty} dv F(v)\phi(v)\mu^{(1)}(v;t_0 - s) \tag{2.6}$$

with $F(v) = f(\xi(v))$, $\Phi(v) = \phi(\xi(v))$ and, for convenience, we have introduced the function

$$\mu^{(n)}(v;x) = \frac{2R(v^{1/2}-1)}{v^{1/2}(v^{1/2}+1)^3}\left|x + R\left(\frac{v^{1/2}-1}{v^{1/2}+1}\right)^2\right|^{-n}. \tag{2.7}$$

Applying the Schwarz inequality to Eq. (2.6) gives the relation

$$|f(s)|^2 \leqslant \frac{1}{2\pi}\int_{-\infty}^{\infty} dv |F(v)\mu^{(1)}(v;t_0 - s)|^2 g(v), \tag{2.8}$$

where

$$J = \frac{1}{2\pi}\int_{-\infty}^{\infty} dv \frac{|\Phi(v)|^2}{g(v)} \tag{2.9}$$

and $g(v)$ is a positive function to be specified.

Using the transformation Eq. (2.4), one can show

$$\int_{t_0}^{t_0+R} dt\,\frac{\rho(t)}{t^n} = \int_1^{\infty} dv\rho(t(v))\mu^{(n)}(v;t_0). \tag{2.10}$$

Eqs. (2.2), (2.3), and (2.10) give the inequality

$$\int_{t_0}^{t_0+R} dt\,\frac{\rho(t)}{t^n} \geqslant kR^{a/2}\int_1^{\infty} dv\left(\frac{v^{1/2}-1}{v^{1/2}+1}\right)^a |F(v)|^2\mu^{(n+b/2)}(v;t_0) \tag{2.11}$$

The reality condition for $f(\xi)$ implies

$$F(v^*) = F^*(v^{-1}), \tag{2.12}$$

which can be used to rewrite the right-hand side of Eq. (2.11) as an integral over positive values of v:

$$\int_{t_0}^{t_0+R} dt\,\frac{\rho(t)}{t^n} \geqslant \frac{kR^{a/2}}{2}\int_0^{\infty} dv\left|\left(\frac{v^{1/2}-1}{v^{1/2}+1}\right)^a\mu^{(n+b/2)}(v;t_0)\right| |F(v)|^2. \tag{2.13}$$

If we let

$$g(v) = \left|\frac{v^{1/2}-1}{v^{1/2}+1}\right|^a \frac{|\mu^{(n+b/2)}(v;t_0)|}{|\mu^{(1)}(v;t_0 - s)|^2} \tag{2.14}$$

and break the integral in Eq. (2.8) up into an integral from $v = -\infty$ to 0 and one from $v = 0$ to $+\infty$ we obtain, with the aid of Eq. (2.13), the expression

$$|f(s)|^2 \leqslant \frac{J}{\pi k R^{a/2}}\left\{\int_{t_0}^{t_0+R} dt\,\frac{\rho(t)}{t^n} + \frac{kR^{1+a/2}}{2}\int_0^{2\pi} d\varphi \frac{|f(t_0 + Re^{i\varphi})|^2}{|t_0 + Re^{i\varphi}|^{n+b/2}}\right\} \tag{2.15}$$

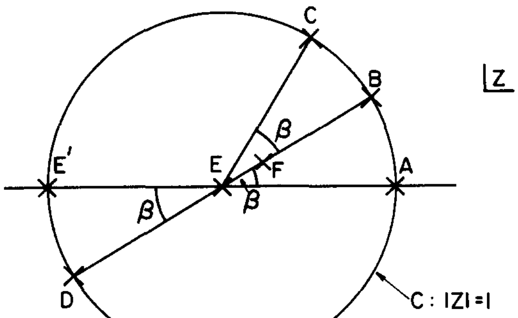

FIG. 3. Complex z plane.

where we have mapped the integral from $v = -\infty$ to 0 back onto the ξ plane.

We now evaluate the least-upper-bound of J over the set of all admissible functions $\Phi(v)$. From Eqs. (2.9) and (2.14) we have

$$J = \frac{1}{2\pi}\int_{-\infty}^{\infty} dv \left|\frac{v^{1/2}+1}{v^{1/2}-1}\right|^a \frac{|\mu^{(1)}(v;t_0 - s)\Phi(v)|^2}{|\mu^{(n+b/2)}(v;t_0)|} \tag{2.16}$$

We now map the half-plane $\mathrm{Im}\,v \geqslant 0$ onto the unit disk $|z| \leqslant 1$ by means of the transformation

$$z = (v - e^{i\beta})/(v - e^{-i\beta}) \tag{2.17}$$

This transformation has the following properties (see Fig. 3):

(i) The real v axis is mapped onto the circle $|z| = 1$ such that the points $v = \pm\infty, -1, 0, \cos\beta$, and 1 are mapped onto the points $\arg z = 0, \beta, 2\beta, \pi, \pi + \beta$ respectively.

(ii) The arc $|v| = 1$ is mapped onto the line $\arg z = \beta$ in such a way that $v = \exp(i\beta)$ (i.e., $\xi = s$) and $v = \exp(i\beta')$ (i.e., $\xi = 0$) correspond to $z = 0$ and $z = [\sin\frac{1}{2}(\beta'-\beta)/\sin\frac{1}{2}(\beta'+\beta)]\exp(i\beta)$ respectively.

(iii) For real v one has the relation

$$v = \sin\left(\frac{\theta}{2} - \beta\right)/\sin\frac{\theta}{2}; \quad \arg z = \theta. \tag{2.18}$$

with this transformation the integral J becomes

$$J = \frac{\sin\beta}{\pi}\int_0^{2\pi} d\theta |h(e^{i\theta})||\Phi(v(e^{i\theta}))|^2, \tag{2.19}$$

where

$$h(z) = \frac{1}{(1-z)^2}\left(\frac{v^{1/2}(z)+1}{v^{1/2}(z)-1}\right)^a \frac{[\mu^{(1)}(v(z);t_0 - s)]^2}{\mu^{(n+b/2)}(v(z);t_0)}.$$

With the aid of Eqs. (2.4), (2.7), and (2.17) we have

$$h(z) = \Sigma(z)\frac{(\xi(z))^{n+b/2}}{(\xi(z)-s)^2}(1-z)^{-1/2}(1-e^{-2i\beta}z)^{-1/2} \times (1-e^{-i(\beta+\pi)}z)^{1-a}, \tag{2.20a}$$

$$\Sigma(z) = \frac{2iRe^{-i(\beta-\pi)/2}}{[2\sin(\beta/2)]^{a-1}} \times [e^{i\beta/2}(1-e^{-2i\beta}z)^{1/2} + (1-z)^{1/2}]^{2a-4} \tag{2.20b}$$

Equations (2.20) show that $h(z)$ has singularities on the unit circle at $\arg z = 0, 2\beta, \pi + \beta$. We therefore define

$$\Phi(v(z)) = (1-z)^{\gamma_1}(1-e^{-2i\beta}z)^{\gamma_2}(1-e^{-i(\beta+\pi)}z)^{\gamma_3}\hat{\Phi}(z), \tag{2.21}$$

where the constants γ_i are to be picked so that the absolute value of

$$\tilde{h}(z) = (1-z)^{2\gamma_1}(1-e^{-2i\beta}z)^{2\gamma_2}(1-e^{-i(\beta+\pi)}z)^{2\gamma_3}h(z) \tag{2.22}$$

is integrable on the unit circle, i.e., $\gamma_1 \geq \frac{1}{4}, \gamma_2 \geq \frac{1}{4}$ $\gamma_3 \geq (a-1)/2$. (Zeros of $\tilde{h}$ on the unit circle are not particularly troublesome for one can consider the integral J to be a limit as $|z| \to 1^-$.)

We are now in a position to apply the Szegö theorem[5] to bound J from above:

$$J \leq J_M \equiv 2 \sin\beta \exp\left(\frac{1}{2\pi}\int_0^{2\pi} d\theta \log|\tilde{h}(e^{i\theta})|\right). \tag{2.23}$$

Substituting Eq. (2.23) with the value for J_M obtained in the Appendix into Eq. (2.15) gives

$$\begin{aligned}|f(s)|^2 \leq &\frac{1}{4\pi k}\left(\frac{R+t_0-s}{R-t_0+s}\right)(t_0-s)^{n+1+(b-a)/2} \\ &\times \frac{\{1+[(R-t_0+s)/(R-t_0)]\,[t_0^{1/2}/(t_0-s)^{1/2}]\}^{2n+b}}{[1+t_0(2R-2t_0+s)/(R-t_0)^2]^{n+b/2}} \\ &\times \left(\int_{t_0}^{t_0+R} dt \frac{\rho(t)}{t^n} + \frac{kR^{1+a/2}}{2}\int_0^{2\pi} d\varphi \frac{|f(t_0+Re^{i\varphi})|^2}{|t_0+Re^{i\varphi}|^{n+b/2}}\right).\end{aligned} \tag{2.24}$$

At $s=0$, Eq. (2.24) reduces to

$$\begin{aligned}|f(0)|^2 \leq &\frac{2^a}{\pi k}(4t_0)^{n-1+(b-a)/2}\left(\frac{R-t_0}{R+t_0}\right)^{n-1+b/2} \\ &\times \left(\int_{t_0}^{t_0+R} dt \frac{\rho(t)}{t^n} + \frac{kR^{1+a/2}}{2}\int_0^{2\pi} d\varphi \frac{|f(t_0+Re^{i\varphi})|^2}{|t_0+Re^{i\varphi}|^{n+b/2}}\right).\end{aligned} \tag{2.25}$$

It is worth noting that the inequalities Eqs. (2.24) and (2.25) are the sharpest ones that can be derived given only Eq. (2.2) and the analyticity of $f(\xi)$. Let $f_{\max}(R)$ be the maximum of $|f(\xi)|$ on the circle $|\xi - t_0| = R$. From Eq. (2.25), we derive

$$|f_{\max}(R)|^2 \geq \frac{2}{kR^{1+a/2}\delta}\left[\frac{|f(0)|^2 k\pi}{2^a(4t_0)^{n-1+(b-a)/2}}\left(\frac{R+t_0}{R-t_0}\right)^{n-1+b/2} - \int_{t_0}^{t_0+R} dt \frac{\rho(t)}{t^n}\right], \tag{2.26}$$

where

$$\delta = \int_0^{2\pi} \frac{d\varphi}{|t_0+Re^{i\varphi}|^{n+b/2}}.$$

Equation (2.26) gives a lower bound for $|f_{\max}(R)|$ throughout the cut plane in terms of $|f(0)|^2$ and $|f(t)|^2$ for $t_0 \leq t \leq t_0+R$. In other words, Eq. (2.26) gives the constraints on the behavior of $f(\xi)$ (e.g., an amplitude, a form factor) in the whole cut-plane in terms of the low energy behavior of $f(t)$.

3. INEQUALITIES FOR THE DERIVATIVE OF THE FUNCTION

To find an inequality involving the derivative of $f(\xi)$, we first differentiate both sides of Eq. (2.1) with respect to s,

$$\frac{df(s)}{ds} - \gamma f(s) = \frac{1}{2\pi i}\int_c d\xi \frac{f(\xi)\phi(\xi)}{(\xi-s)^2}, \tag{3.1a}$$

where

$$\gamma \equiv \frac{d\phi(s)}{ds}. \tag{3.1b}$$

By repeating the same procedure as before, one obtains

$$\left|\frac{df(s)}{ds} + \gamma f(s)\right|^2 \leq \frac{J'}{\pi k R^{a/2}}\left(\int_{t_0}^{t_0+R} dt \frac{\rho(t)}{t^n} + \frac{kR^{1+a/2}}{2}\int_0^{2\pi} d\varphi \frac{|f(t_0+Re^{i\varphi})|^2}{|t_0+Re^{i\varphi}|^{n+b/2}}\right), \tag{3.2}$$

where J' is given by Eq. (2.9) with $g(v)$ replaced by $g'(v)$ where

$$g'(v) = \left|\frac{v^{1/2}-1}{v^{1/2}+1}\right|^a \frac{|\mu^{(n+b/2)}(v;t_0)|}{|\mu^{(2)}(v;t_0-s)|^2}. \tag{3.3}$$

Besides the condition $\Phi(v(z=0)) = 1$, the function $\Phi(v(z))$ must now satisfy the constraint

$$\left(\frac{d\Phi(v(z))}{dz}\right)_{z=0} = -4(t_0-s)\frac{R-t_0+s}{R+t_0-s}e^{i\beta}\gamma. \tag{3.4}$$

Because of this new constraint, we cannot directly apply Szegö's theorem. We can nevertheless find an upper bound J'_M for J' by using a slightly different method.[3] We define a new function $\Psi(z)$ by the relation

$$\Phi(v(z)) = \left[1 - 4(t_0-s)\frac{R-t_0+s}{R+t_0-s}e^{i\beta}\gamma z\right]\Psi(z). \tag{3.5}$$

Now Ψ is an analytic function arbitrary to the extent that

$$\Psi(z=0) = 1, \quad \frac{d\Psi(z=0)}{dz} = 0. \tag{3.6}$$

The expression for J' becomes

$$J' = \frac{\sin\beta}{\pi}\int_0^{2\pi} d\theta\, |H(e^{i\theta})|\, |\Psi(e^{i\theta})|^2, \tag{3.7}$$

where

$$H(z) = \frac{h(z)}{(\xi(z)-s)^2}\left|1 - 4(t_0-s)\frac{R-t_0+s}{R+t_0-s}\gamma e^{-i\beta}z\right|^2 \tag{3.8}$$

with $h(z)$ given by Eqs. (2.20). Since $|H(z)|$ is a positive definite, summable function on the unit circle, one can apply the generalized Szegö theorem[6] to obtain

$$J' \leq J'_M = 2\sin\beta\left(|\eta(z)|^2 + \left|\frac{d\eta(z)}{dz}\right|^2\right)_{z=0} \tag{3.9}$$

where

$$\eta(z) = \exp\left(\frac{1}{4\pi}\int_0^{2\pi} d\theta \frac{e^{i\theta}+z}{e^{i\theta}-z}\log\tilde{H}(e^{i\theta})\right). \tag{3.10}$$

Here $\tilde{H}(z)$ is given by Eq. (3.8) with $h(z)$ replaced by $\tilde{h}(z)$ [see Eq. (2.22)].

By using the same method employed in the Appendix and the elementary formula

$$\int_0^{2\pi} d\theta \log(1 + a\sin\theta + b\cos\theta) = 2\pi\log\frac{1+\sqrt{1-a^2-b^2}}{2}, \quad (a^2+b^2<1),$$

we find

$$\left|\frac{df(s)}{ds} + \gamma f(s)\right|^2 \leq \frac{1}{64\pi k}\left(\frac{R+t_0-s}{R-t_0+s}\right)^3 (t_0-s)^{n-3+(b-a)/2}$$

$$\times \frac{\{1+[(R-t_0+s)/(R-t_0)][t_0^{1/2}/(t_0-s)^{1/2}]\}^{2n+b}}{[1+t_0(2R-2t_0+s)/(R-t_0)^2]^{n+b/2}}$$

$$\times\Big\{1+\left[\left|\frac{C_1-C_2}{C_1+C_2}+8\frac{\lambda}{\lambda-1}-(2n+b)\frac{\lambda'}{\lambda'-1}\right| + |(2a-4)C_3|\right]^2$$

$$\times\left(\int_{t_0}^{t_0+R} dt\,\frac{\rho(t)}{t^n}+\frac{kR^{1+a/2}}{2}\int_0^{2\pi} d\varphi\,\frac{|f(t_0-Re^{i\varphi})|^2}{|t_0+Re^{i\varphi}|^{n+b/2}}\right), \tag{3.11a}$$

where

$$\lambda=\frac{4R^2}{(R-t_0+s)^2},\quad C_1=\left|1+4(t_0-s)\frac{R-t_0+s}{R+t_0-s}\gamma\right|$$

$$\lambda'=\frac{4R^2}{(R-t_0+s)^2}\,\frac{t_0-s}{t_0},$$

$$C_2=\left|1-4\,(t_0-s)\frac{R-t_0+s}{R+t_0-s}\gamma\right| \tag{3.11b}$$

and for convenience we have written c_3 in the integral form

$$C_3=\frac{1}{2\pi}\int_0^{2\pi} e^{-i\theta}d\theta\,\log|e^{i\beta/2}(1-e^{-2i\beta}z)^{1/2}+(1-z)^{1/2}| \tag{3.11c}$$

In Eqs. (3.11), γ is a free parameter. Bounds on the derivative of the function which do not involve the value of the function itself may be obtained by setting $\gamma=0$. In applications it may be useful to give some other value to γ; the inequality (3.11a) is valid for all values of γ. For $s=0$ and $\gamma=0$, Eq. (3.11) becomes

$$\left|\frac{df(0)}{ds}\right|^2\leqslant\frac{2^a}{\pi k}(4t_0)^{n-3+(b-a)/2}\left(\frac{R-t_0}{R+t_0}\right)^{n-3+b/2}$$

$$\times\left[1+\left(\left|\frac{(8-2n-b)\,4R^2}{(R+t_0)(3R-t_0)}\right|+|(2a-4)C_3|\right)^2\right]$$

$$\times\left[\int_{t_0}^{t_0+R}dt\,\frac{\rho(t)}{t^n}+\frac{kR^{1+a/2}}{2}\int_0^{2\pi}d\varphi\,\frac{|f(t_0+Re^{i\varphi})|^2}{|t_0+Re^{i\varphi}|^{n+b/2}}\right]. \tag{3.12}$$

One need not stop here. Dispersive inequalities for higher derivatives can also be constructed by using the methods reported in this article.

ACKNOWLEDGMENTS

The authors would like to thank Professor S. Okubo and Dr. D. Levin for informative discussions.

APPENDIX

In this appendix we evaluate the integral

$$J_M=2\sin\beta\,\exp\left(\frac{1}{2\pi}\int_0^{2\pi}d\theta\,\log|\tilde{h}(e^{i\theta})|\right), \tag{A1}$$

with $\tilde{h}(z)$ given by Eqs. (2.20) and (2.22). We first write

$$J_M=(2\sin\beta)\exp[I^{(1)}+(n+b/2)\,I^{(2)}-2I^{(3)}], \tag{A2}$$

where

$$I^{(1)}=\frac{1}{2\pi}\int_0^{2\pi}d\theta\,\log|\Sigma(e^{i\theta})|,$$

$$I^{(2)}=\frac{1}{2\pi}\int_0^{2\pi}d\theta\,\log|\xi(e^{i\theta})|,$$

$$I^{(3)}=\frac{1}{2\pi}\int_0^{2\pi}d\theta\,\log|\xi(e^{i\theta})-s|, \tag{A3}$$

and we have used

$$\int_0^{2\pi}d\theta\,\log|e^{i\theta}-e^{i\theta_0}|=0,\qquad \text{all } \theta_0. \tag{A4}$$

The integrals $I^{(i)}$ are most simply evaluated by means of Jensen's well-known formula:

$$\log\frac{|\zeta|}{\prod_{i=1}^{n}|z_i|}=\frac{1}{2\pi}\int_0^{2\pi}d\theta\,\log|f(e^{i\theta})|, \tag{A5}$$

where $f(z)$ is an analytic function for $|z|\leqslant 1$ having zeros at $z_1,\dots,z_n$ in the interior $0\neq|z|<1$ (multiple zeros being repeated) and the behavior $f(z)\to\zeta z^{\tau}(\tau\geqslant 0)$ as $z\to 0$. Referring to Eq. (2.20b), we see $\Sigma(z)$ is nonvanishing for $|z|<1$, hence

$$I^{(1)}=\log|\Sigma(0)|=\log\left(\frac{R}{2\sin^{a-1}(\beta/4)\cos^{3-a}(\beta/4)}\right). \tag{A6}$$

By construction, $\xi(z)$ has a simple zero at $z=[\sin\frac{1}{2}(\beta'-\beta)/\sin\frac{1}{2}(\beta'+\beta)]\exp(i\beta)$ and $\xi(z)-s$ has a simple zero at $z=0$. Obviously $\xi(0)=s$ and using Eqs. (2.4), (2,5a), and (2.17), one can show

$$[(\xi(z)-s)/z]_{z=0}=-4(t_0-s)\cos(\beta/2)e^{-i\beta}, \tag{A7}$$

therefore,

$$I^{(2)}=\log\left(s\frac{\sin\frac{1}{2}(\beta'+\beta)}{\sin\frac{1}{2}(\beta'-\beta)}\right), \tag{A8}$$

$$I^{(3)}=\log[4(t_0-s)\cos(\beta/2)]. \tag{A9}$$

Substituting Eq. (A6), (A8), and (A9) into (A2) gives

$$J_M=\frac{R}{4(t_0-s)^2}\,\frac{\tan^{2-a}(\beta/4)}{\cos(\beta/2)}\left(s\frac{\sin\frac{1}{2}(\beta'+\beta)}{\sin\frac{1}{2}(\beta'-\beta)}\right)^{n+b/2} \tag{A10}$$

or, with the definitions Eqs. (2.5),

$$J_M=\frac{R^{a/2}}{4}\left(\frac{R+t_0-s}{R-t_0+s}\right)(t_0-s)^{n-1(b-a)/2}$$

$$\times\frac{\{1+[(R-t_0+s)/(R-t_0)]\,[(t_0^{1/2}/(t_0-s)^{1/2}]^{2n+b}}{[1+t_0(2R-2t_0+s)/(R-t_0)^2]^{n+b/2}}. \tag{A11}$$

---

*Supported in part by the U. S. Atomic Energy Commission.
†Institute of Particle Physics Fellow
‡Supported in part by the National Research Council of Canada.
[1]For a summary see S. Okubo, University of Rochester preprint COO-3065-1 (1972) (to be published in *Proceedings of the 1972 Coral Gables Conference on Fundamental Interactions at High Energy*).
[2]A. Logunov, L. Soloviev, and A. Tavkhelidge, Phys. Lett. B **24B**, 181 (1967); K. Igi and S. Matsuda, Phys. Rev. Lett. **18**, 625 (1967); R. Dolen, D. Horn, and C. Schmid, Phys. Rev. **166**, 1768 (1968).
[3]S. Okubo, Phys. Rev. D **3**, 2807 (1971).
[4]J. Cleymans and B. Rodenberg, Nuovo Cimento A **5A**, 297 (1971).
[5]See. for example, K. Hoffman, *Banach Spaces of Analytic Functions* (Prentice-Hall, Englewood CLiffs, N. J., 1962).
[6]H. Helson, *Lectures on Invariant Subspaces* (Academic, New York, 1964).

Complex coordinate transformations and the Schwarzschild–Kerr metrics*

E. T. Newman

Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

(Received 24 October 1972)

Utilizing the fact that the Schwarzschild and Kerr geometries have an intrinsically defined Minkowski space associated with them, we show that these Minkowski spaces (as the Kerr parameter varies) can be viewed as "real slices" in a complexified Minkowski space. The complex Weyl tensor of each member of the family can then be viewed as a single complex field on the complex Minkowski space. Further, the degenerate principle null vectors associated with each geometry can be considered as projections into the "real slices" of a complex null vector field in the complex Minkowski space. These results may be considered as clarifying earlier work on obtaining the Kerr metric from the Schwarzschild metric by a complex coordinate transformation.

1. INTRODUCTION

Several years ago we published two papers[1,2] in which the Kerr[3] and the charged Kerr metrics were "derived" from the Schwarzschild and charged Schwarzschild (Reissner–Nordstrom) metrics by a complex coordinate transformation. "Derivation" was originally put in quotation marks because there was no simple, clear reason for the series of operations performed on the Schwarzschild metric to yield a new solution of the Einstein equations. Until recently, aside from the essentially trivial remark that the field equations sanction these operations,[4,5] there has been little progress in giving a geometric interpretation to the complex transformations. Adler et al.[6] have just shown that the Kerr and Schwarzschild metrics can be obtained from a common generating function by just a complex displacement of the origin. The present author has recently shown[7] that the Maxwell equations and the linearized Einstein equations (for the Weyl tensor) may be complexified and considered as field equations in complex Minkowski space. From this point of view the Weyl tensor of the linearized Kerr (charged Kerr) and linearized Schwarzschild (charged Schwarzschild) may be looked upon as the same field but viewed in different "real slices" of complex Minkowski space. (Each value of the Kerr parameter a yields a different slice; $a = 0$ yields the Schwarzschild slice.)

It is the purpose of the present paper to show that these results for the linearized Schwarzschild–Kerr metrics can be extended to their exact form.

Of fundamental importance in the analysis is the fact that for each member of the Kerr family there is a covariantly defined flat-space defined from

$$ds^2 = ds^2_{\text{Flat}} + \lambda(l^*_\mu dx^\mu)^2, \tag{1.1}$$

where λ is a scalar function and l^*_μ is a (degenerate) principle null vector. What we wish to show is that these flat-spaces may be considered as a family of flat-spaces all imbedded in complex Minkowski space. The Weyl tensor [or more precisely the "complex" self-dual Weyl tensor $\frac{1}{2}(C^{\alpha\beta\gamma\delta} + iC^{\alpha\beta\overset{*}{\gamma}\delta})$] will then be considered as a function on the complex Minkowski space. On any of the real slices the real Weyl tensor can be reconstructed. Furthermore, the principle null vectors of the complex Weyl tensor are degenerate in the sense of being arbitrary in a null complex two-plane and in general complex (they are not degenerate in this sense for the real Weyl tensor). The degeneracy, however, can be removed on each real slice by demanding a real (tangent to the slice) principle null vector. This yields the vector l^*_μ of (1.1). Finally, the scalar function λ can be written as the real part of a function on the complex Minkowski space.

In Sec. 2 we discuss some properties of complex Minkowski space and define "real slices". After complex null polar coordinates are introduced (the real coordinate values yield a real slice) a basic complex coordinate transformation (similar to that of Ref. 1) is presented. The real new coordinate values yield a Kerr-type of coordinate system on a new slice. In Sec. 3 complex tetrad transformations and their relation to the principle (complex) null vectors of the complex Weyl tensor are described. It is shown that the complex Weyl tensor of each member of the Kerr–Schwarzschild family can be viewed as a single complex function in complex Minkowski space.

We wish to point out that we make no use here of the Einstein equations to find the relation between the different members of the Kerr–Schwarzschild family. We simply take the known metrics and show that a certain unity exists between them when viewed in complex Minkowski space.

2. COMPLEX MINKOWSKI SPACE

We consider the complexification of Minkowski space by taking a four-dimensional complex manifold[8] (eight real dimensions) covered by a single complex chart, the coordinates labeled by $z^\mu = x^\mu + iy^\mu$ (x^μ and y^μ real) endowed with a complex metric

$$ds^2 = \eta_{\mu\nu} dz^\mu dz^\mu. \tag{2.1}$$

The only transformations considered are holomorphic transformations of the coordinates. The group which preserves the form of the line element is the ten (complex) parameter Poincare group[9]

$$z'^\mu = a^\mu_\nu z^\nu + b^\mu, \tag{2.2}$$

b^μ being a constant complex vector and a^μ_ν being a constant complex matrix satisfying

$$a^\mu_\alpha a^\nu_\beta \eta_{\mu\nu} = \eta_{\alpha\beta}, \tag{2.3}$$

By a "real slice" of this space we mean a four (real) dimensional subspace such that the metric induced on it by (2.1) is real.

It obviously follows from (2.2) that there exists, at least, a ten (real) parameter family of these slices arising from the imaginary Poincare transformations, four from the imaginary translations and six from the imaginary homogeneous transformations. (It is unknown to us whether other "real slices" exist. Presumably if they did, the induced metrics would not be flat.) Although in the remainder of this paper we will confine ourselves to the "real slices" obtained by the imaginary translations (parallel slices), we would like to conjecture that new solutions of the Einstein equations can be obtained by applying the ideas of this paper to the complex boosts. The solution, presumably, would be similar to the solution of Maxwell's equations obtained from the Coulomb solution by a complex boost.[7]

We now wish to consider a generic parallel slice which can be considered, with no loss of generality, to arise from a translation of real Minkowski space in the imaginary z^3 direction, i.e., it is defined by

$$z^\mu = x'^\mu - ia\delta^\mu_3, \qquad x'^\mu \text{ real.} \tag{2.4}$$

We digress for a moment to introduce complex null polar coordinates by

$$z^\mu = (1/\sqrt{2})(2u + r, r\cos\theta, r\sin\theta\cos\phi, r\sin\theta\sin\phi) \tag{2.5}$$

with u, r, θ and ϕ complex. The line-element (2.1) becomes

$$ds^2 = 2du^2 + 2du dr - \frac{r^2}{2}(d\theta^2 + \sin^2\theta d\phi^2). \tag{2.6}$$

Note that real values of u, r, θ, ϕ defines the same real slice as $z^\mu = x^\mu, x^\mu$ real.

If we now perform the complex transformation

$$u = u' + ia\cos\theta', \qquad r = r' - 2ia\cos\theta',$$
$$\cos\theta = \frac{r'\cos\theta' - 2ia}{r' - 2ia\cos\theta'}, \qquad \cos 2(\phi - \phi') = \frac{r'^2 - 4a^2}{r'^2 + 4a^2}, \tag{2.7}$$

(2.6) becomes

$$ds^2 = 2du'^2 + 2du'dr' - 2a\sin^2\theta' dr' d\phi' - \tfrac{1}{2}(r'^2 + 4a^2\cos^2\theta')(d\theta'^2 + \sin^2\theta' d\phi'^2) - 2a^2\sin^4\theta' d\phi'^2. \tag{2.8}$$

If u', r', θ' and ϕ' are restricted to real values, it can be shown that this is equivalent to choosing the real slice (2.4). Equation (2.8) is then the real Minkowski metric in Kerr-type coordinates.[1,2] One could consider that the transformation (2.7), taking (2.6) into (2.8), offers a partial explanation of the algorithm of Refs. 1 and 2 leading from the Schwarzschild to the Kerr metric.

[We wish to point out that knowing (2.8) and (2.6) it would be possible to derive (2.7) by solving a set of partial differential equations. We however found it far easier to use the Penrose theory of twistors[10,11] where a series of relatively simple algebraic steps yielded (2.7). It is however unnecessary for the purposes of this paper to enter into this question.]

3. THE COMPLEX WEYL TENSOR

Associated with the complex null polar coordinate system of (2.6) is a null tetrad system

$$l^\mu = \delta^\mu_1, \quad n^\mu = \delta^\mu_0 - \delta^\mu_1, \quad m^\mu = \frac{1}{r}\left(\delta^\mu_2 + \frac{i}{\sin\theta}\delta^\mu_3\right),$$
$$\overline{m}^\mu = \frac{1}{r}\left(\delta^\mu_3 - \frac{i}{\sin\theta}\delta^\mu_3\right), \tag{3.1}$$

where (0, 1, 2, 3) refers to (u, r, θ, ϕ). Only on the real slice, $z^\mu = x^\mu$, should these be viewed as a "normal" null tetrad system with l and n real and m and $\overline{m}$ complex conjugates of each other.

A second complex null tetrad system introduced by

$$l^* = l - Qm, \quad n^* = n + Qm, \quad m^* = Sm,$$
$$\overline{m}^* = S^{-1}(\overline{m} + Ql - Qn - Q^2 m), \tag{3.2}$$

with $Q = 2ia\sin\theta'/r'^2 + 4a^2)^{1/2}$ and $S = (r' - 2ia\cos\theta')/(r'^2 + 4a^2)^{1/2}$, has the form [when expressed in the coordinate system of (2.8)]

$$l^* = \delta^\mu_1, \quad n^{*\mu} = \delta^\mu_0 - \delta^\mu_1,$$
$$m^{*\mu} = \frac{1}{(r' + 2ia\cos\theta')}\left[ia\sin\theta'(\delta^\mu_0 - 2\delta^\mu_1) + \delta^\mu_2 + \frac{i}{\sin\theta'}\delta^\mu_3\right], \tag{3.3}$$
$$\overline{m}^{*\mu} = \frac{1}{(r' - 2ia\cos\theta')}\left[-ia\sin\theta'(\delta^\mu_0 - 2\delta^\mu_1) + \delta^\mu_2 - \frac{i}{\sin\theta'}\delta^\mu_3\right].$$

On the real slice (2.4) this is a "normal" null tetrad system,[1,2] l^* and n^* real, m^* and $\overline{m}^*$ complex conjugates of each other, with l^* shear-free and having complex divergence $\rho \equiv l^*_{\mu;\nu} m^{*\mu}\overline{m}^{*\nu} = -1/(r' - 2ia\cos\theta')$.

Starting with a tensor $C_{\alpha\beta\gamma\delta}$ whose algebraic symmetries are those of the Weyl tensor we consider its complexification

$$W_{\alpha\beta\gamma\delta} = \tfrac{1}{2}(C_{\alpha\beta\gamma\delta} + iC^*_{\alpha\beta\gamma\delta}) \tag{3.4}$$

with * denoting the dual. $W_{\alpha\beta\gamma\delta}$ is now to be considered a field on the complex Minkowski space. Its only independent nonvanishing tetrad components are

$$\begin{aligned}
\psi_0 &= -W_{\alpha\beta\gamma\delta} l^\alpha m^\beta l^\gamma m^\delta,\\
\psi_1 &= -W_{\alpha\beta\gamma\delta} l^\alpha n^\beta l^\gamma m^\delta,\\
\psi_2 &= -W_{\alpha\beta\gamma\delta} \overline{m}^\alpha n^\beta l^\gamma m^\delta,\\
\psi_3 &= -W_{\alpha\beta\gamma\delta} \overline{m}^\alpha n^\beta l^\gamma n^\delta,\\
\psi_4 &= -W_{\alpha\beta\gamma\delta} \overline{m}^\alpha n^\beta \overline{m}^\gamma n^\delta.
\end{aligned} \tag{3.5}$$

(Note that $W_{\alpha\beta\gamma\delta} l^\gamma \overline{m}^\delta = W_{\alpha\beta\gamma\delta} \overline{m}^\gamma n^\delta = 0$.)

If we give the particular field, expressed in the complex null polar coordinates with the tetrad (3.1) as

$$\psi_0 = \psi_1 = \psi_3 = \psi_4 = 0, \qquad \psi_2 = -2\sqrt{2}m/r^3, \tag{3.5'}$$

it is easily seen that on the "real slice" $z^\mu = x^\mu$, it represents the Schwarzschild Weyl tensor with l^μ a degenerate principle null vector. [l is a null vector of both Schwarzschild space and its associated Minkowski space. The same is not true for n. See (1.1).]

The same field but now expressed in the coordinates of (2.8), using the tetrad (3.3), has the form

$$\psi^*_3 = 0, \qquad \psi^*_1 = 0, \qquad \psi^*_2 = \frac{-2\sqrt{2}m}{(r' - 2ia\cos\theta')^3},$$
$$\psi^*_3 = \frac{-6\sqrt{2}\,S^{-1}Qm}{(r' - 2ia\cos\theta')^3} = \frac{12\sqrt{2}\,ima\sin\theta'}{(r' - 2ia\cos\theta')^4},$$
$$\psi^*_4 = \frac{-12\sqrt{2}\,S^{-2}Q^2 m}{(r' - 2ia\cos\theta')^3} = \frac{48\sqrt{2}\,ma^2\sin^2\theta'}{(r' - 2ia\cos\theta')^5}.$$

This can be shown to be [on the real slice (2.4)] equivalent to the Kerr Weyl tensor.

Finally we mention that the λ of (1.1) can be chosen as the real part of the scalar function $4\sqrt{2}\, m/r$ on complex Minkowski space, thus

$$\lambda = 2\sqrt{2}\, m\left(\frac{1}{r} + \frac{1}{\bar{r}}\right) = \frac{4\sqrt{2}\, mr'}{r'^2 + 4a^2 \cos^2\theta'}.$$

DISCUSSION

We have shown that there exists a certain unity in the Kerr-Schwarzschild class of metrics (possibly including the conjectured new metrics arising from the imaginary boosts) when viewed from complex Minkowski space. We have little idea whether other new metrics can be obtained by similar techniques though it appears conceivable that similar unities between other members of the Kerr-Schwarzschild class [Eq. (1.1)] could exist.

There does however appear to be some evidence that the work described here is part of a larger structure. Recent work[11,12] on asymptotically flat spaces indicates that null infinity possesses a natural complex structure and that the invariance group (suitably defined) is the complex Poincaré group. In fact, transformations similar to Eq. (2.7) in the limit $r \to \infty$ arise naturally.

*This research has been supported by NSF Grant No. GP 22789.

[1] E. T. Newman and A. I. Janis, J. Math. Phys. **6**, 914 (1965).
[2] E. T. Newman *et al.*, J. Math. Phys. **6**, 918 (1965).
[3] R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
[4] R. P. Kerr, private communication.
[5] C. J. Talbot, Commun. Math. Phys. **13**, 45 (1969).
[6] R. J. Adler *et al.*, preprint.
[7] E. T. Newman, J. Math. Phys. **14**, 102 (1973).
[8] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1969), Vol. II.
[9] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
[10] R. Penrose, J. Math. Phys. **8**, 345 (1967).
[11] B. Aronson *et al.*, J. Math. Phys. **12**, 2462 (1971).
[12] Unpublished results of author.

Central decomposition of invariant states: Applications to the Galilei group

G. Loupias

Equipe de Recherche Associée au C.N.R.S., Département de Mathématiques, Université des Sciences et Techniques du Languedoc, 34060 Montpellier CEDEX, France

M. Mebkhout

Université d'Aix-Marseille, U.E.R. de Luminy, 13009-Marseille, France

Centre de Physique Théorique-C.N.R.S., Marseille, France
(Received 24 October 1972)

We follow the investigation initiated in an earlier paper. Here we treat the case of Galilei group $\mathcal{G}$. The complete description of all closed subgroups H of $\mathcal{G}$ is given. Then among them we describe those such that the corresponding homogeneous spaces $\mathcal{G}/H$ bear a bounded invariant measure.

INTRODUCTION

Many studies of ergodic (i.e., extremal invariant) states of a C^*-algebra acted upon in an asymptotically abelian way by a group G have revealed that a wide class of ergodic states are obtained as integrals of states with lesser symmetry, i.e., invariant only under a subgroup of G.

The case of crystal states led to seek for a principle yielding decompositions of ergodic states into states which retain the symmetry of normal as well as non-normal subgroups running through a conjugacy class of subgroups of G. This principle was shown in Ref. 1 to be the central decomposition of ergodic states in the sense of Sakaï.[2] In the case of transitive states, the description of such components "with broken symmetry" amounts to the classification of all homogeneous spaces of G which carry a bounded invariant measure.

In Ref. 1, this work was done for the Euclidean group, giving rise to the cristallographic groups in 1, 2, and 3 dimensions, and some "helicoïdal" groups. Our purpose is to give an equivalent classification for the Galilei group $\mathcal{G}$, up to a conjugation in $\mathcal{G}$.

Section 1 gives a strategical description of the method; Sec. 2 recalls some essential properties of $\mathcal{G}$; Secs. 3 and 4 contain a complete description of closed subgroups of $\mathcal{G}$; Sec. 5 retains among them those which are an answer to our problem.

1. PRELIMINARIES

We refer to Ref. 1 for a listing of some mathematical results concerning quotients, and recall the strategical remarks upon which our method rests.

Let H be a solution of our problem for $\mathcal{G}$, i.e., H is a closed subgroup of $\mathcal{G}$ such that the homogeneous space $\mathcal{G}/H$ carries a bounded $\mathcal{G}$-invariant measure. Then:

(1) All closed subgroups H' of $\mathcal{G}$ such that $H \subset H' \subset \mathcal{G}$ are solutions of the same problem;

(2) All subgroups $H'' \subset H$ solutions of the problem for H are solutions of the problem for $\mathcal{G}$;

(3) Conversely if K is not a solution of the problem for H, it is not a solution of the problem for $\mathcal{G}$.

Moreover, Mostow has shown[3]:

Theorem: If G is a solvable Lie group, then, for any closed subgroup A of G such that G/A carries an invariant measure m_A, the finiteness of m_A is equivalent to the compactness of G/A.

This is also true when G is the extension of a compact Lie group by a solvable group (Mostow, private communication). As this is the case for the Galilei group, our work verifies that it satisfies this extension of Mostow's theorem.

2. THE GALILEI GROUP

Let us recall that the Galilei group $\mathcal{G}$ is a ten-parameter real Lie group, the elements of which will be written

$$(x, v, t, r), \tag{1}$$

where

$$\begin{array}{ll} x \in R^3_x, & \text{translations in three-dimensional Euclidean space,} \\ v \in R^3_v, & \text{pure Galilei transformations,} \\ t \in R_t, & \text{time translations,} \\ r \in SO(3), & \text{three-dimensional connected real orthogonal group}^4, \end{array} \tag{2}$$

the group law being:

$$\begin{aligned} &(x', v', t', r')(x, v, t, r) \\ &\quad = (r'x + x' + tv', r'v + v', t' + t, r'r), \\ &(0, 0, 0, 1) = 1_{\mathcal{G}}, \\ &(x, v, t, r)^{-1} = (-r^{-1}(x - tv), -r^{-1}v, -t, r^{-1}). \end{aligned} \tag{3}$$

There exist several decompositions of $\mathcal{G}$ as a topological semidirect product. We shall retain the following one, which we will use in the sequel.

Let K be the subgroup of $\mathcal{G}$ of elements of the form

$$(x, v, t, 1). \tag{4}$$

Because of

$$\begin{aligned} &(x', v', t', r')^{-1}(x, v, t, 1)(x', v', t', r') \\ &\qquad = (r'^{-1}(x + t'v - tv'), r'^{-1}v, t, 1) \end{aligned} \tag{5}$$

one sees that K is invariant and we can define the quotient group

$$\mathcal{G}/K \approx SO(3). \tag{6}$$

The application

$$\mathbf{r}:(x, v, t, 1) \to \mathbf{r}(x, v, t, 1) = (rx, rv, t, 1), \tag{7}$$

where rx and rv are the transformations of x and v under the orthogonal transformation $r \in SO(3)$, defines a continuous homomorphism from $SO(3)$ into aut$K$ and we can write $\mathcal{G}$ as the topological semidirect product of K and $SO(3)$, which we denote

$$\mathcal{G} = K \,\Box\, SO(3). \tag{8}$$

Let us now call $R_x^3 \oplus R_v^3$ the subgroup of K of elements of the form

$$(x, v, 0, 1). \tag{9}$$

Because of

$$(x', v', t', 1)^{-1}(x, v, 0, 1)(x', v', t', 1) = (x + t'v, v, 0, 1) \tag{10}$$

one sees that $R_x^3 \oplus R_v^3$ is invariant in K and we can define the quotient group

$$K/(R_x^3 \oplus R_v^3) \approx R_t. \tag{11}$$

The application

$$\mathbf{t}:(x, v, 0, 1) \to \mathbf{t}(x, v, 0, 1) = (x + tv, 0, 1), \tag{12}$$

where tv is the multiplication of $v \in R_v^3$ by the scalar $t \in R_t$, defines a continuous homomorphism from R_t into aut$(R_x^3 \oplus R_v^3)$, and we can write K as the topological semidirect product of $R_x^3 \oplus R_v^3$ and R_t, which we denote

$$K = (R_x^3 \oplus R_v^3) \,\Box\, R_t. \tag{13}$$

Finally, we get for $\mathcal{G}$ the decomposition

$$\mathcal{G} = (R_x^3 \oplus R_v^3) \,\Box\, R_t \,\Box\, SO(3) \tag{14}$$

with the law

$$\begin{aligned}
&[\{(x', v'), t'\}, r'][\{(x, v), t\}, r] \\
&\quad = [\{x', v'), t'\}\cdot\mathbf{r}'\{(x, v), t\}, r'r] \\
&\quad = [\{(x', v'), t'\}\{(r'x, r'v), t\}, r'r] \\
&\quad = [\{\mathbf{t}(x', v')(r'x, r'v), t' + t\}, r'r] \\
&\quad = [\{(x' + tv', v')(r'x, r'v), t' + t\}, r'r] \\
&\quad = [\{(r'x + x' + tv', r'v + v'), t' + t\}, r'r].
\end{aligned} \tag{15}$$

It must be noticed that the composition law in (13) is written in an unusual (t acting on the left) but, nevertheless, coherent way, thanks to the Abelian character of R_t.

Finally, (7) and (12) being continuous automorphisms of $R_x^3 \oplus R_v^3$, and $\Delta_{\mathcal{G}}$ denoting the modular function of $\mathcal{G}$, one sees that

$$\begin{tikzcd}
N_K(F) \arrow[r] & M_F \arrow[r] & 1 \\
K_H \arrow[u] \arrow[r] & F_H \arrow[u] \arrow[r] & 1 \\
1 \arrow[u] & 1 \arrow[u] &
\end{tikzcd}$$

Fig. 1

$$\Delta_{\mathcal{G}} = \Delta_K \cdot \det r \cdot \Delta_{SO(3)} = \Delta_{R_x^3} \cdot \Delta_{R_x^3} \cdot \det t \cdot \Delta_{R_t} \cdot \det r \cdot \Delta_{SO(3)} = 1$$

and that $\mathcal{G}$ is unimodular. Hence, the subgroups H giving rise to an invariant measure over $\mathcal{G}/H$ have to be unimodular.[5]

3. THE CLOSED SUBGROUPS OF K

This section will be devoted to the search for the closed subgroups of K. Let K_H be such a closed subgroup (the choice of this notation will become clear in the next section) and let us define

$$F = K_H \cap (R_x^3 \oplus R_v^3). \tag{16}$$

It is a closed invariant Abelian subgroup of K_H, which allows us to introduce the quotient group

$$F_H = K_H/F. \tag{17}$$

If $N_K(F)$ denotes the normalizer of F in K, it contains K_H as a closed subgroup, F and $R_x^3 \oplus R_v^3$ as closed invariant Abelian subgroups, and we get the two quotient groups

$$M_F = N_K(F)/(R_x^3 \oplus R_v^3) \tag{18}$$

$$N_F = N_K(F)/F. \tag{19}$$

Thanks to the explicit definition of $N_K(F)$,

$$\begin{aligned}
N_K(F) = \{(x, v, t, 1) \in K :& (x, v, t, 1)^{-1}(x', v', 0, 1)(x, v, t, 1) \\
&= (x' + tv', v', 0, 1) \\
&= \mathbf{t}(x', v', 0, 1) \in F, (x', v', 0, 1) \in F\},
\end{aligned} \tag{20}$$

it is evident that M_F is the subgroup of R_t under the action of which F is left stable. Moreover, $N_K(F)$ can be written as the semidirect product of M_F and $R_x^3 \oplus R_v^3$ with the law induced by the one of K:

$$N_K(F) = (R_x^3 \oplus R_v^3) \,\Box\, M_F. \tag{21}$$

On the other hand, the continuous homomorphisms

$$K_H \to F_H \quad \text{and} \quad K_H \to M_F, \tag{22}$$

having the same kernel, there exists a continuous injective homomorphism λ

$$\lambda: F_H \to M_F, \quad \lambda(F_H) = [K_H \cdot (R_x^3 \oplus R_v^3)]/(R_x^3 \oplus R_v^3) \tag{23}$$

such that the diagram in Fig. 1 is commutative.

It is important to notice that, λ having no closed range in general, $\lambda(F_H)$ is an algebraic, but not necessarily topological, subgroup of M_F.

Introducing the quotient

$$T_F = (R_x^3 \oplus R_v^3)/F, \tag{24}$$

we are able to consider $R_x^3 \oplus R_v^3$ as an extension of F by T_F characterized by some 2-cocycle ω:

$$R_x^3 \oplus R_v^3 = F \overset{\omega}{\Box} T_F. \tag{25}$$

Any element $(x, v, t, 1) \in K$ can then be written

$$\begin{aligned}
(x, v, t, 1) &= \{[(x_F, v_F), (\chi_x, \chi_v)], t, 1\} \\
&= \{[0, (\chi_x, \chi_v)], t, 1\}\{[(x_F, v_F), 0], 0, 1\},
\end{aligned} \tag{26}$$

where $(x_F, v_F) \in F$ and $(\chi_x, \chi_v) \in T_F$. In particular, the elements of F_H are of the form $\{[0, (\chi_x, \chi_v)], t, 1\}$; as

$$\lambda\{[0, (\chi_x, \chi_v)], t, 1\} = t \tag{27}$$

and λ is injective, to any $t \in \lambda(F_H)$ there corresponds a unique element $(\chi_x(t), \chi_v(t)) \in T_F$, we can then define the application from $\lambda(F_H)$ into T_F according to

$$t \xrightarrow{(\chi_x, \chi_v)} (\chi_x(t), \chi_v(t)) \tag{28}$$

and write any element of K_H in the form

$$\{[(x_F, v_F), (\chi_x(t), \chi_v(t))], t, 1\}. \tag{29}$$

If, on the other hand, we remark that

$$N_F = [(R_x^3 \oplus R_v^3) \,\Box\, M_F]/F = T_F \,\Box\, M_F, \tag{30}$$

where the action of M_F into T_F is deduced from the action of M_F onto $R_x^3 \oplus R_v^3$ through the quotient (24), we then discover that F_H is a subgroup of N_F with the law

$$\begin{aligned} &\{[0, (\chi_x(t'), \chi_v(t'))], t', 1\}\{[0, (\chi_x(t), \chi_v(t))], t, 1\} \\ &= \{[0, (\chi_x(t') + \chi_x(t) + t\chi_v(t'), \chi_v(t') + \chi_v(t))], t' + t, 1\} \\ &= \{[0, (\chi_x(t + t'), \chi_v(t + t'))], t' + t, 1\}, \end{aligned} \tag{31}$$

and, moreover, that the following relations are true in T_F:

$$\begin{aligned} \chi_x(t' + t) &= \chi_x(t) + \chi_x(t') + t\chi_v(t') \\ &= \chi_x(t) + \chi_x(t') + t'\chi_v(t), \\ \chi_v(t + t') &= \chi_v(t) + \chi_v(t'), \\ \chi_x(-t) &= t\chi_v(t) - \chi_x(t), \\ \chi_v(-t) &= -\chi_v(t), \\ \chi_x(0) &= 0, \\ \chi_v(0) &= 0. \end{aligned} \tag{32}$$

In these conditions

$$\begin{aligned} N_K(F) &= (R_x^3 \oplus R_v^3) \,\Box\, M_F = (F \overset{\omega}{\Box} T_F) \,\Box\, M_F \\ &= F \overset{\omega'}{\Box} (T_F \,\Box\, M_F) = F \overset{\omega'}{\Box} N_F \end{aligned} \tag{33}$$

and in particular

$$K_H = F \overset{\omega'}{\Box} F_H \tag{34}$$

where the law $\overset{\omega'}{\Box}$ is defined according to

$$\begin{aligned} &\{(x'_F, v'_F), [(\chi_x(t'), \chi_v(t')), t', 1]\} \\ &\quad \times \{(x_F, v_F), [(\chi_x(t), \chi_v(t)), t, 1]\} \\ &= \{(x'_F + tv'_F + x_F, v'_F + v_F) \\ &\quad + \omega[(\chi_x(t') + t\chi_v(t') + \chi_x(t), \chi_v(t')), (\chi_x(t), \chi_v(t))], \\ &\quad [(\chi_x(t') + t\chi_v(t') + \chi_x(t), \chi_v(t') + \chi_v(t)), t' + t, 1]\} \\ &= \{(x'_F + tv'_F + x_F, v'_F + v_F) + \omega'(t', t), \\ &\quad [(\chi_x(t' + t), \chi_v(t' + t)), t' + t, 1]\} \\ &= \{(x'_F + tv'_F + \omega'_x(t', t), v'_F + v_F + \omega'_v(t', t)), \\ &\quad [(\chi_x(t' + t), \chi_v(t' + t)), t' + t, 1]\}, \end{aligned} \tag{35}$$

where we denote

$$\begin{aligned} &\omega[(\chi_x(t') + t\chi_v(t'), \chi_v(t')), (\chi_x(t), \chi_v(t))] \\ &\qquad = \omega'(t', t) = \{\omega'_x(t', t), \omega'_v(t', t)\} \in F. \end{aligned} \tag{36}$$

Finally, K_H having to be closed into $N_K(F)$, then F_H has to be closed into N_F.

We have the following procedure to get any closed subgroup K_H of K: A choice of F among the closed subgroups $R_x^3 \oplus R_v^3$ determines in a unique way T_F, M_F, N_F, ω, and ω'. The subgroups F_H are then defined as closed sections of N_F with the help of a suitable application (χ_x, χ_v) defined in a (nonnecessarily closed) subgroup of M_F with values in T_F. K_H is then the ω'-extension of F and F_H. The situation can be described, in terms of exact sequences, according to the scheme in Fig. 2.

The closed subgroups F of $R_x^3 \oplus R_v^3$ are of the form[6]

$$F = (R_x^m \oplus Z_x^p) \oplus (R_v^n \oplus Z_v^q), \quad \begin{cases} 0 \le m + p \le 3, \\ 0 \le n + q \le 3. \end{cases} \tag{37}$$

There are one hundred of them, up to an automorphism. According to the definition of M_F,

$$\begin{aligned} t \in M_F \Leftrightarrow t(F) = \{&[(R_x^m \oplus Z_x^p) \\ &+ t(R_v^n \oplus Z_v^q)] \oplus (R_v^n \oplus Z_v^q)\} \subset F, \end{aligned} \tag{38}$$

where the sum $(R_x^m \oplus Z_x^p) + t(R_v^n \oplus Z_v^q)$ is defined through a (nonunique) imbedding of $R_v^n \oplus Z_v^q$ into $R_x^m \oplus Z_x^p$. Then, in particular, $M_F = \{0\}$ if and only if $n + q > m + p$ or $n > m$. M_F being a closed subgroup of R_t, that is to say, $M_F = \{0\}$, $M_F = Z$, or $M_F = R$, we can classify the subgroups F into three families $F_{\{0\}}, F_Z, F_R$ according to the type of the corresponding M_F. As, in the sequel, we will eliminate the $F_{\{0\}}$, we give in Tables I and II the list of the F_Z and F_R, with the corresponding T_{F_Z} and T_{F_R}.

One must notice that some groups are in both tables, according to the way into which $R_v^n \oplus Z_v^q$ is imbedded into $R_x^m \oplus Z_x^p$.

On the other hand, we know that

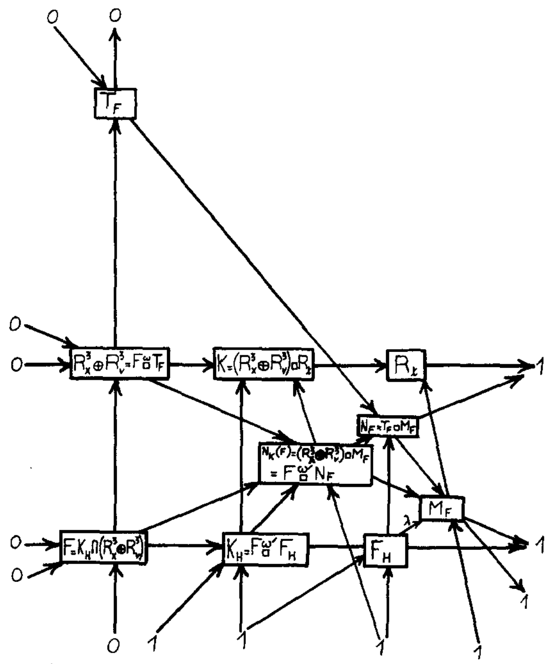

Fig. 2

$$
\begin{aligned}
&\text{if } M_F = \{0\}, && \text{then } \lambda(F_H) = 0,\\
&\text{if } M_F = Z, && \text{then } \lambda(F_H) = \{0\} \text{ or } Z,\\
&\text{if } M_F = R, && \text{then } \lambda(F_H) = \{0\}, Z, R \text{ or a dense subgroup of } R.
\end{aligned}
\tag{39}
$$

TABLE I.

	F_Z	
$Z_x \oplus Z_v$	$(Z_x^2 \oplus R_x) \oplus Z_v$	$Z_x^3 \oplus Z_v^3$
$(Z_x \oplus R_x) \oplus Z_v$	$(Z_x \oplus R_x^2) \oplus Z_v$	$(Z_x^2 \oplus R_x) \oplus Z_v^3$
$Z_x^2 \oplus Z_v$	$Z_x^3 \oplus Z_v^2$	$(Z_x^2 \oplus R_x) \oplus (Z_v^2 \oplus R_v)$
$(Z_x \oplus R_x) \oplus (Z_v \oplus R_v)$	$(Z_x^2 \oplus R_x) \oplus Z_v^2$	$(Z_x \oplus R_x^2) \oplus Z_v^3$
$(Z_x \oplus R_x) \oplus Z_v^2$	$(Z_x^2 \oplus R_x) \oplus (Z_v \oplus R_v)$	$(Z_x \oplus R_x^2) \oplus (Z_v^2 \oplus R_v)$
$Z_x^2 \oplus Z_v^2$	$(Z_x \oplus R_x^2) \oplus (Z_v \oplus R_v)$	$(Z_x \oplus R_x^2) \oplus (Z_v \oplus R_v^2)$
$Z_x^3 \oplus Z_v$	$(Z_x \oplus R_x^2) \oplus Z_v^2$	
	T_{F_Z}	
$(R_x^2 \oplus T_x) \oplus (R_v^2 \oplus T_v)$	$T_x^2 \oplus (R_v^2 \oplus T_v)$	$T_x^3 \oplus T_v^3$
$(R_x \oplus T_x) \oplus (R_v^2 \oplus T_v)$	$T_x \oplus (R_v^2 \oplus T_v)$	$T_x^2 \oplus T_v^3$
$(R_x \oplus T_x^2) \oplus (R_v^2 \oplus T_v)$	$T_x^3 \oplus (R_v \oplus T_v^2)$	$T_x^2 \oplus T_v^2$
$(R_x \oplus T_x) \oplus (R_x \oplus T_v)$	$T_x^2 \oplus (R_v \oplus T_v^2)$	$T_x \oplus T_v^3$
$(R_x \oplus T_x) \oplus (R_v \oplus T_v^2)$	$T_x^2 \oplus (R_v \oplus T_v)$	$T_x \oplus T_v^2$
$(R_x \oplus T_x^2) \oplus (R_v \oplus T_v^2)$	$T_x \oplus (R_v \oplus T_v)$	$T_x \oplus T_v$
$T_x^3 \oplus (R_v^2 \oplus T_v)$	$T_x \oplus (R_v \oplus T_v^2)$	

TABLE II.

	F_R	
$O_x \oplus O_v$	$R_x^3 \oplus O_v$	$(Z_x \oplus R_x^2) \oplus Z_v$
$R_x \oplus O_v$	$Z_x^3 \oplus O_v$	$R_x^3 \oplus R_x^2$
$Z_x \oplus O_v$	$(Z_x^2 \oplus R_x) \oplus O_v$	$R_x^3 \oplus (Z_v \oplus R_v)$
$R_x \oplus R_v$	$(Z_x \oplus R_x^2) \oplus O_v$	$R_x^3 \oplus Z_v^2$
$R_x \oplus Z_v$	$R_x^2 \oplus R_x^2$	$(Z_x \oplus R_x^2) \oplus R_v^2$
$R_x^2 \oplus O_v$	$R_x^2 \oplus (Z_v \oplus R_v)$	$(Z_x \oplus R_x^2) \oplus (Z_v \oplus R_v)$
$(Z_x \oplus R_x) \oplus O_v$	$R_x^2 \oplus Z_v^2$	$(Z_x \oplus R_x^2) \oplus Z_v^2$
$Z_x^2 \oplus O_v$	$R_x^3 \oplus R_v$	$R_x^3 \oplus R_v^3$
$R_x^2 \oplus R_v$	$R_x^3 \oplus Z_v$	$R_x^3 \oplus Z_v^3$
$R_x^2 \oplus Z_v$	$(Z_x^2 \oplus R_x) \oplus R_v$	$R_x^3 \oplus (Z_v^2 \oplus R_v)$
$(Z_x \oplus R_x) \oplus R_v$	$(Z_x^2 \oplus R_x) \oplus Z_v$	$R_x^3 \oplus (Z_v \oplus R_v^2)$
$(Z_x \oplus R_x) \oplus Z_v$	$(Z_x \oplus R_x^2) \oplus R_v$	
	T_{F_R}	
$R_x^3 \oplus R_v^3$	R_v^3	$T_x \oplus (R_v^2 \oplus T_v)$
$R_x^2 \oplus R_v^3$	$T_x^3 \oplus R_x^3$	$O_x \oplus R_v$
$(R_x^2 \oplus T_x) \oplus R_v^3$	$T_x^2 \oplus R_x^3$	$O_x \oplus (R_v \oplus T_v)$
$R_x^2 \oplus R_v^2$	$T_x \oplus R_v^3$	$O_x \oplus (R_v \oplus T_v^2)$
$R_x^2 \oplus (R_v^2 \oplus T_v)$	$R_x \oplus R_v$	$T_x \oplus R_v$
$R_x \oplus R_v^3$	$R_x \oplus (R_v \oplus T_v)$	$T_x \oplus (R_v \oplus T_v)$
$(R_x \oplus T_x) \oplus R_v^3$	$R_x \oplus (R_v \oplus T_v^2)$	$T_x \oplus (R_v \oplus T_v^2)$
$(R_x \oplus T_x^2) \oplus R_v^3$	$O_x \oplus R_v^2$	$O_x \oplus O_v$
$R_x \oplus R_v^2$	$O_x \oplus (R_v^2 \oplus T_v)$	$O_x \oplus T_v^3$
$R_x \oplus (R_v^2 \oplus T_v)$	$T_x^2 \oplus R_v^2$	$O_x \oplus T_v^2$
$(R_x \oplus T_x) \oplus R_v^2$	$T_x^2 \oplus (R_v^2 \oplus T_v)$	$O_x \oplus T_v$
$(R_x \oplus T_x) \oplus (R_v^2 \oplus T_v)$	$T_x \oplus R_v^2$	

We are now going to explicit the form of the application (χ_x, χ_v) and show that the case where $\lambda(F_H)$ would be a nonclosed subgroup of R is not relevant for our problem. From (32) we get immediately that

$$\chi_v(t) = t\,\chi_v(t_0)/t_0,$$

where t_0 is any element of $\lambda(F_H)$ different from zero,

$$\chi_x(nt) = n\chi_x(t) + \tfrac{1}{2}n(n-1)t\,\chi_v(t), \qquad n \in Z, t \in R,$$

$$\chi_x((p/q)t) = (p/q)\chi_x(t) + [p(p-q)/2q^2]t\,\chi_v(t), \tag{40}$$

$$p, q \in Z,\ t \in R.$$

These formulas completely fix the form of χ_x and χ_v when $\lambda(F_H) = 0$ or Z, but only on the dense subgroup of rationals when $\lambda(F_H) = R$. But as F_H, that is to say, the graph of (28), has to be closed into $N_F = T_{F_R} \square R$, we have then to extend these formulas by continuity, and we get

$$\left.\begin{aligned} &\chi_v(t) = t\,\chi_v(1)\\ &\chi_x(t) = \tfrac{1}{2}t^2\chi_v(1) + t(\chi_v(1) - \chi_v(1)/2)\ , \end{aligned}\right\}\ t \in R, \tag{41}$$

where $\chi_x(1)$ and $\chi_v(1)$ are arbitrarily chosen constants in T_{F_R} and T_{F_v} respectively. Let us end with the case where $\lambda(F_H)$ would be a dense nonclosed subgroup of R. We will show in the sequel that the only interesting cases for our problem will be the case where $m = 3$ whenever $\lambda(F_H) = R$ or a dense subgroup.

Hence $\chi_x \equiv 0$ and the graph of the application

$$t \in \lambda(F_H) \to \chi_v(t) = t\,\chi_v(t_0)/t_0 \in T_{F_v} = R_v^{3-n-q} \oplus T_v^q \tag{42}$$

cannot be closed if $\lambda(F_H)$ is not.

Finally we see that the closed subgroups of K interesting for us are all of the type

$$
\begin{aligned}
K_H &= F_{\{0\}},\\
K_H &= F_Z,\\
K_H &= F_Z \overset{\omega'}{\square} \{(n^2\chi_v(1)/2 + n(\chi_x(1) - \chi_v(1)/2), n\chi_v(1)), n\},\\
&\qquad\qquad n \in Z,\\
K_H &= F_R,\\
K_H &= F_R \overset{\omega'}{\square} \{(n^2\chi_v(1)/2 + n(\chi_x(1) - \chi_v(1)/2), n\chi_v(1)), n\},\\
&\qquad\qquad n \in Z,\\
K_H &= F_R \overset{\omega'}{\square} \{(t^2\chi_v(1)/2 + t(\chi_x(1) - \chi_v(1)/2, t\chi_v(1)), t\},\\
&\qquad\qquad t \in R.
\end{aligned}
\tag{43}
$$

4. THE CLOSED SUBGROUPS OF $\mathcal{G}$

We are now reaching the next stage of this work, i.e., the description of closed subgroups of $\mathcal{G}$ itself. The method will be similar to the one used in the preceding section, but more complicated because the subgroup K is not Abelian.

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

$$K_H = K \cap H. \tag{44}$$

Then K_H is a closed subgroup (see the beginning of Sec. 3), invariant in H, but in general not invariant in K. Let us introduce the normalizer $N_{\mathcal{G}}(K_H)$ of K_H in $\mathcal{G}$: H is a closed subgroup of $N_{\mathcal{G}}(K_H)$ but in general K is not. This leads us to consider the normalizer $N_K(K_H)$ of K_H in K. We have then the following relations between closed subgroups:

$$K_H \subset N_K(K_H) \subset K, \qquad N_K(K_H) \subset N_{\mathcal{G}}(K_H) \tag{45}$$

and consequently,

$$N_K(K_H) \subset N_{\mathcal{G}}(K_H) \cap K. \tag{46}$$

But, conversely, K_H being invariant in $N_{\mathcal{G}}(K_H)$, is also invariant into $N_{\mathcal{G}}(K_H) \cap K$, which gives the opposite inclusion, and, finally,

$$N_K(K_H) = N_{\mathcal{G}}(K_H) \cap K. \tag{47}$$

Moreover, $N_K(K_H)$ is closed invariant in $N_{\mathcal{G}}(K_H)$: If $g \in N_{\mathcal{G}}(K_H)$, then $gN_K(K_H)g^{-1} \subset K$ because of the invariance of K into $\mathcal{G}$, and $gN_K(K_H)g^{-1} \subset N_{\mathcal{G}}(K_H)$ because of the inclusion $N_K(K_H) \subset N_{\mathcal{G}}(K_H)$.

We are able to introduce the quotient groups

$$P_H = H/K_H \tag{48}$$

and

$$Q_H = N_{\mathcal{G}}(K_H)/N_K(K_H) \tag{49}$$

and, by an argument analoguous to the one used in the preceding section, we can assert the existence of a continuous, injective (but in general with nonclosed range) homomorphism φ from P_H into Q_H.

On the other hand, let

$$N_H = N_{\mathcal{G}}(K_H)/K_H \tag{50}$$

and

$$T_H = N_K(K_H)/K_H. \tag{51}$$

Then P_H is a subgroup of N_H and

$$Q_H = N_{\mathcal{G}}(K_H)/N_K(K_H) = N_{\mathcal{G}}(K_H)/K_H \big/ N_K(K_H)/K_H = N_H/T_H \tag{52}$$

is a quotient group.

From the injectivity of φ, we conclude that the elements of P_H are of the form

$$\{f(r), r\}, \quad r \in \varphi(P_H), f(r) \in T_H, \tag{53}$$

where $\varphi(P_H)$ is a (nonnecessarily closed) subgroup of Q_H.

Let us go on with the investigation of the structure of $N_K(K_H)$. We know that

$$N_K(K_H) = \{(x', v', t') \in K:(x', v', t')^{-1}(x, v, t)(x', v', t') = (x + t'v - tv', v, t) \in K_H, (x, v, t) \in K_H\}. \tag{54}$$

Hence, using the notation of (26), we have that

$$\text{if } t = 0, \quad (x_F + t'v_F, v_F, 0) \in K_H \quad \text{which means that } t' \in M_F,$$
$$\text{if } v = 0, \quad (x - tv', 0, t) \in K_H \quad \text{which means that } tv' \in (K_H)_x. \tag{55}$$

Let us then write

$$tv' = ((tv')_{F_x}, (tv')_{T_{F_x}}). \tag{56}$$

We now have

$$\{(x_F + t'v_F - (tv')_{F_x}, v_F), [(\chi_x(t) + t'\chi_v(t) - (tv')_{T_{F_x}}, \chi_v(t)), t]\} \in K_H, \tag{57}$$

which implies that

$$(tv')_{F_x} \in F_x = R_x^m \oplus Z_x^p \tag{58}$$

and then

$$v_F' = v' \in R_v^3 \quad \text{if } t \in \lambda(F_H) = \{0\}, \tag{59a}$$

$$v_F' \in R_v^m \oplus Z_v^p \quad \text{if } t \in \lambda(F_H) = Z, \tag{59b}$$

$$v_F' \in R_v^m \quad \text{if } t \in \lambda(F_H) = R \text{ or dense.} \tag{59c}$$

$$\chi_x(t) + t'\chi_v(t) - (tv')_{T_{F_x}} = \chi_x(t) \text{ in } T_{F_x}, \tag{60}$$
$$\text{or else } t(v'_{T_{F_v}} - \chi_v(t')) = 0 \text{ in } T_{F_x},$$

and then

$$v'_{T_{F_v}} = 0 \quad \text{if } t \in \lambda(F_H) = \{0\}, \tag{61a}$$

$$v'_{T_{F_v}} = \chi_v(t') \in R_v^{3-m-p} \oplus T_v^p \quad \text{if } t \in \lambda(F_H) = Z, \quad t' \in M_F, \tag{61b}$$

$$v'_{T_{F_v}} = \chi_v(t') \in R^{3-m} \quad \text{if } t \in \lambda(F_H) = R \text{ or dense}, \quad t' \in M_F. \tag{61c}$$

In other words, if $F_{M_F} = (R_x^m \oplus Z_x^p) \oplus (R_v^n \oplus Z_v^q)$, then

$$N_K(K_H) = (R_x^3 \oplus R_v^3) \,\square\, M_F \quad \text{if } \lambda(F_H) = \{0\},$$
$$N_K(K_H) = [R_x^3 \oplus (R_v^m \oplus Z_v^p)] \overset{\omega'}{\square} \{\chi_v(t'), t'\} \quad \text{if } \lambda(F_H) = Z, \quad t' \in M_F, \tag{62}$$
$$N_K(K_H) = (R_x^3 \oplus R_v^m) \,\square\, \{\chi_v(t'), t'\} \quad \text{if } \lambda(F_H) = R \text{ or dense}, \quad t' \in M_F.$$

Moreover, the fact that the x-component of $N_K(K_H)$ is the whole of R_x^3 shows that $N_K(K_H)$ is invariant in K. In the same way, we will determine the structure of $N_{\mathcal{G}}(K_H)$. We know that

$$N_{\mathcal{G}}(K_H) = \{(x', v', t', r') \in \mathcal{G}:(x', v', t', r')^{-1}(x, v, t, 1)(x', v', t', r') = (r'^{-1}(x + t'v - tv'), r'^{-1}v, t, 1) \in K_H, (x, v, t, 1) \in K_H\}. \tag{63}$$

Hence, with the same notations,

$$\text{if } t = 0, \quad (r'^{-1}(x_F + t'v_F), r'^{-1}v_F, 0, 1) \in K_H, \tag{64}$$

which means that r' has to leave F_v stable;

$$\text{if } t = 0 \text{ and } v = 0, \quad (r'^{-1}x_F, 0, 0, 1) \in K_H, \tag{65}$$

which means that r' has to leave F_x stable. So, if $t = 0$, we have

$$(r'^{-1}x_F + t'r'^{-1}v_F, r'^{-1}v_F, 0, 1) \in K_H, \tag{66}$$

which means that $t' \in M_F$, and we conclude that, through the quotient, r' leaves T_{F_x} and T_{F_v} separately stable.

$$\text{if } v = 0, \quad (r'^{-1}(x - tv'), 0, t, 1) \in K_H, \tag{67}$$

which means, according to the preceding remark, that $tv' \in K_{H_x}$.

In other words, if we compare with the structure of $N_K(K_H)$, we have

$$N_{\mathcal{G}}(K_H) = N_K(K_H) \,\square\, Q_H, \tag{68}$$

with

$$r'^{-1}\chi_x(t) = \chi_x(t) \quad \text{and} \quad r'^{-1}\chi_v(t) = \chi_v(t), t \in \lambda(F_H), \tag{69}$$

i.e., where Q_H is the closed subgroup of $SO(3)$ such that

$$Q_H \text{ leaves } F_x = R_x^m \oplus Z_m^p \text{ stable,}$$
$$Q_H \text{ leaves } F_v = R_v^n \oplus Z_v^q \text{ stable,} \tag{70}$$
$$Q_H \text{ leaves } \chi_x(t) \text{ and } \chi_v(t_0)/t_0 \text{ fixed.}$$

Moreover,

$$N_H = N_{\mathcal{G}}(K_H)/K_H = (N_K(K_H) \,\square\, Q_H)/K_H = T_H \,\square\, Q_H. \tag{71}$$

We can also determine the form of T_H:

$$\text{if } M_F = \{0\}, \quad \lambda(F_H) = \{0\},$$
$$T_H = (R_x^{3-m-p} \oplus T_x^p) \oplus (R_v^{3-n-q} \oplus T_v^q),$$

$$\begin{aligned}
&\text{if } M_F = Z, \quad \lambda(F_H) = \{0\},\\
&\qquad T_H = [(R_x^{3-m-p} \oplus T_x^p) \oplus (R_v^{3-n-q} \oplus T_v^q)] \,\square\, Z,\\
&\text{if } M_F = Z, \quad \lambda(F_H) = Z,\\
&\qquad T_H = [(R_x^{3-m-p} \oplus T_x{}^p) \oplus ((R_v^m \oplus Z_v^p)/(R_v^n \oplus Z_v^q))],\\
&\text{if } M_F = R, \quad \lambda(F_H) = \{0\},\\
&\qquad T_H = [(R_x^{3-m-p} \oplus T_x^p) \oplus (R_v^{3-n-q} \oplus T_v^q)] \,\square\, R,\\
&\text{if } M_F = R, \quad \lambda(F_H) = Z,\\
&\qquad T_H = [(R^{3-m-p} \oplus T^p) \oplus ((R_v^m \oplus Z_v^p)/(R_v^n \oplus Z_v^q))] \,\square\, T,\\
&\text{if } M_F = R, \quad \lambda(F_H) = R,\\
&\qquad T_H = [(R_x^{3-m-p} \oplus T_x^p) \oplus (R_v^{m-n-q} \oplus T_v^q)].
\end{aligned} \tag{72}$$

Then formula (53) gives

$$f(r) = \{f_x(r), f_v(r), f_t(r)\} \in T_H \tag{73}$$

and

$$\begin{aligned}
f(r'r) = \{f_x(r'r), f_v(r'r), f_t(r'r)\} &= f(r')\cdot r'f(r)\\
&= \{f_x(r') + r'f_x(r) + f_t(r)f_v(r'), f_v(r') + r'f_v(r),\\
&\qquad f_t(r') + f_t(r)\}.
\end{aligned} \tag{74}$$

If Ω is now the 2-cocycle defining the extension

$$N_K(K_H) = K_H \overset{\Omega}{\square} T_H, \tag{75}$$

then

$$N_{\mathcal{G}}(K_H) = (K_H \overset{\Omega}{\square} T_H) \,\square\, Q_H = K_H \overset{\Omega'}{\square} (T_H \,\square\, Q_H) = K_H \overset{\Omega'}{\square} N_H \tag{76}$$

and in particular

$$H = K_H \overset{\Omega'}{\square} P_H = (F \overset{\omega'}{\square} F_H) \overset{\Omega'}{\square} P_H \tag{77}$$

where the law $\overset{\Omega'}{\square}$ is defined according to

$$\begin{aligned}
&[(x', v', t'), (f(r'), r')][(x, v, t), (f(r), r)]\\
&= [(x', v', t')(r'x, r'v, t)\Omega(f(r'), r'f(r)), (f(r')\cdot r'f(r), r'r)]\\
&= [(x', v', t')(r'x, r'v, t)\Omega'(r', r), (f(r'), r'r)],
\end{aligned} \tag{78}$$

where we denote

$$\begin{aligned}
\Omega'(r'r) &= \Omega(f(r'), r'f(r))\\
&= \{\Omega'_{F_x}(r', r), \Omega'_{F_v}(r', r), \Omega'_t(r', r)\} \in K_H.
\end{aligned} \tag{79}$$

Hence, more precisely,

$$\begin{aligned}
&\{[(x'_F, v'_F), (\chi_x(t'), \chi_v(t'), t')], (f(r'), r')\}\{[(x_F, v_F), (\chi_x(t), \chi_v(t), t)], (f(r), r)\}\\
&= \{[(x'_F + r'x_F + tv'_F + \omega'_x(t', t) + \Omega'_{F_x}(r', r), v'_F + r'v_F + \omega'_v(t', t) + \Omega'_{F_v}(r', r)),\\
&\qquad (\chi_x(t' + t), \chi_v(t' + t), t' + t)\Omega'_t(r', r)], ((f_x(r'r), f_v(r', r), f_t(r'r)), r'r)\}
\end{aligned} \tag{80}$$

or else, if we put

$$\lambda(\Omega'_t(r', r)) = \tau(r', r), \tag{81}$$

$$\begin{aligned}
&\{[(x'_F, v'_F), (\chi_x(t'), \chi_v(t'), t')], (f(r'), r')\}\{[(x_F, v_F), (\chi_x(t), \chi_v(t), t)], (f(r), r)\}\\
&= \{[(x'_F + r'x_F + tv'_F + \omega'_x(t', t) + \Omega'_{F_x}(r', r), v'_F + r'v_F + \omega'_v(t', t) + \Omega'_{F_v}(r', r)),\\
&\qquad (\chi_x(t' + t + \tau(r', r)), \chi_v(t' + t + \tau(r', r)), t' + t + \tau(r', r))], ((f_x(r'r), f_v(r'r), f_t(r'r), r'r)\}.
\end{aligned} \tag{82}$$

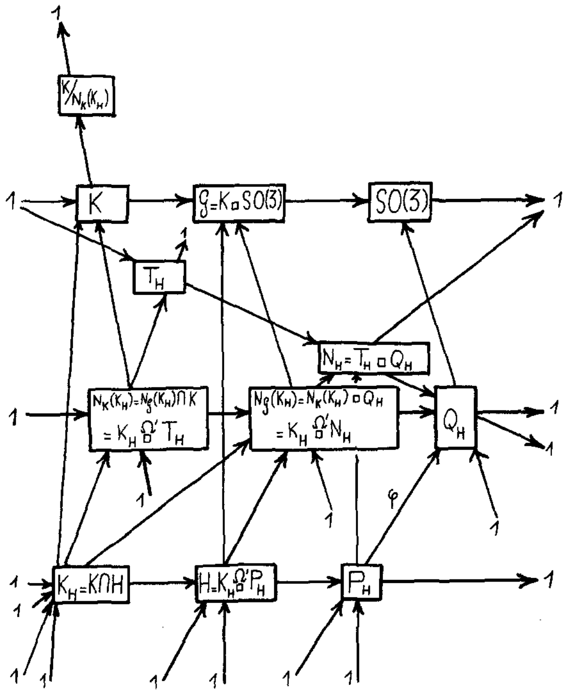

Fig. 3

Finally, H having to be closed into $N_{\mathcal{G}}(K_H)$, P_H has to be closed into N_H.

The whole of this procedure can be described, in terms of exact sequences, in Fig. 3.

Our last task is to define the structure of the application $r \to f(r)$.

We have first

$$f_t(r'r) = f_t(r') + f_t(r). \tag{83}$$

Secondly, we have

$$f_v(r'r) = f_v(r') + r'f_v(r') \tag{84}$$

i.e., f_v is a 1-cocycle onto a subgroup G_0 of $SO(3)$ with value into T_{H_v}. This subgroup can be considered as the connected component of a subgroup G of $O(3)$ by adjunction of the symmetry -1. Then f_v can be considered as the restriction to G_0 of the 1-cocycle $\bar{f}_v$ onto G obtained by defining arbitrarily $\bar{f}_v(-1) = a \in T_{H_v}$, and, if $r \in G_0$, and then $-r \in G$,

$$\bar{f}_v(-r) = f_v(r) + r\bar{f}_v(-1) = \bar{f}_v(-1) - f_v(r) \tag{85}$$

and

$$2f_v(r) = (1 - r)\bar{f}_v(-1). \tag{86}$$

If we call b one of the solutions of the equation $2b = a$ into T_H, we get

$$f_v(r) = (1-r)b, \tag{87}$$

where b is an arbitrary element of T_H. In fact we showed that any 1-cocycle onto G_0 is a 1-coboundary.

Finally we have

$$f_x(r'r) = f_x(r') + r'f_x(r) + f_t(r)f_v(r'). \tag{88}$$

Using the same method as above, and defining arbitrarily $\bar{f}_x(-1) = \alpha$, we have

$$\begin{aligned}\bar{f}_x(-r) &= f_x(r) + r\bar{f}_x(-1) + \bar{f}_t(-1)f_v(r)\\ &= \bar{f}_x(-1) - f_x(r) + f_t(r)\bar{f}_v(-1),\end{aligned} \tag{89}$$

and then

$$2f_x(r) = (1-r)\bar{f}_x(-1) + f_t(r)a - \bar{f}_t(-1)(1-r)b. \tag{90}$$

If we put

$$2\beta = \alpha, \quad 2\epsilon = \bar{f}_t(-1), \tag{91}$$

then

$$f_x(r) = (1-r)\{\beta - \epsilon b\} + f_t(r)b. \tag{92}$$

5. THE HOMOGENEOUS SPACES WITH BOUNDED INVARIANT MEASURE

Now that we have described all closed subgroups of $\mathcal{G}$, we are ready to filter out among them those which give a positive answer to our problem.

We are going to prove a few theorems which among all those subgroups, will retain or reject some ones; the remaining ones, for which no theorems are available, will have to be studied separately.

Theorem 1: For any closed subgroup H of $\mathcal{G}$ containing, as a closed invariant subgroup, $K_H = F \overset{\omega'}{\square} \{\chi_x(t), \chi_v(t), t\}$ with $Z^6 \subset F$ and $t \in Z$ or R, the homogeneous space $\mathcal{G}/H$ carries a bounded invariant measure.

An easy computation shows that K_H is unimodular; let us first show that $\mathcal{G}/K_H$ is compact, i.e., that the theorem is true for $H = K_H$. We have

$$\mathcal{G}/K_H = \{(R^6 \square R) \square SO(3)\}/K_H \simeq \{(R^6 \square R)/K_H\} \cdot SO(3), \tag{93}$$

and it is sufficient to show that the first factor is compact. Let (X,t) be an element of $R^6 \square R$: We can decompose it in an unique way according to

$$(X,t) = [(X_1, X_2), (t_1, t_2)], \tag{94}$$

where $X_1 \in F$, $X_2 \in R^6/F = T_F$, $t_1 \in Z$ or R, $t_2 \in T$ or $\{0\}$. Then there exists a unique $(Y_1, Y_2) \in R^6 = F \overset{\omega}{\square} T_F$ such that

$$\begin{aligned} X_2 &= t_1 Y_2 + (\chi_x(t_1), \chi_v(t_1)),\\ X_1 &= Y_1 + \omega(t_1 Y_2, (\chi_x(t_1), \chi_v(t_1))). \end{aligned} \tag{95}$$

When (X_1, X_2) runs through $R^6 = F \overset{\omega}{\square} T_F$, (Y_1, Y_2) do the same and, for each t, the application $(X_1, X_2) \to (Y_1, Y_2)$ is an homeomorphism. Hence,

$$\begin{aligned}(X,t) &= \{[Y_1 + \omega'(t_1 Y_2, (\chi_x(t_1), \chi_v(t_1))), t_1 Y_2\\ &\quad + (\chi_x(t_1), \chi_v(t_1))], (t_1, t_2)\}\\ &= \{(0, Y_2)(0, t_2)\}\{(Y_1, (\chi_x(t_1), \chi_v(t_1))), (t_1, 0)\},\end{aligned} \tag{96}$$

and we conclude that $(R^6 \square R)/K_H$ is homeomorphic to $R^6/F \times T$ or $R^6/F \times \{0\}$, and compact as soon as $Z^6 \subset F$. But as H contains K_H as a closed subgroup, the theorem is true thanks to our first strategical remark.

Theorem 2: H is not a solution as soon as:

(i) $M_F = \{0\}$,

(ii) $M_F = Z$ or R, $\lambda(F_H) = Z$, and $m + p < 3$,

(iii) $M_F = \lambda(F_H) = R$ and $m < 3$, or $M_F = R$ and $\lambda(F_H)$ dense,

(iv) $M_F = Z$ or R, $\lambda(F_H) = \{0\}$, $\varphi(P_H)$ finite,

(v) $M_F = Z$ or R, $\lambda(F_H) = Z$, $m + p = 3$, $n + q < 3$, and $\varphi(P_H)$ closed,

(vi) $M_F = \lambda(F_H) = R$, $m = 3$, $n + q < 3$, and $\varphi(P_H)$ closed.

Combining (ii) and (iv), we conclude also that

(iv$^{\text{bis}}$) $M_F = Z$ or R, $\lambda(F_H) = \{0\}$, and $m + p < 3$.

Let us first study our problem when $H = N_{\mathcal{G}}(K_H)$. Then

$$\begin{aligned}\mathcal{G}/N_{\mathcal{G}}(K_H) &= \{(R^6 \square R) \square SO(3)\}/(N_K(K_H) \square Q_H)\\ &\simeq (R^6 \square R)/N_K(K_H) \times SO(3)/Q_H,\end{aligned} \tag{97}$$

and we are left with the same problem with the first factor.

But we know that if $K_H = F \overset{\omega'}{\square} F_H$ is such that

$\lambda(F_H) = \{0\}$,

$$\text{then } (R^6 \square R)/N_K(K_H) = (R^6 \square R)/(R^6 \square M_F) \approx R/M_F, \quad M_F = \{0\}, Z, \text{ or } R, \tag{98}$$

$\lambda(F_H) = Z$,

$$\begin{aligned}&\text{then } (R^6 \square R)/N_K(K_H)\\ &= (R^6 \square R)/[R_x^3 \oplus (R_v^m \oplus Z_v^p)] \overset{\omega'}{\square} \{\chi_v(t'), t'\},\\ &\qquad t' \in M_F = Z \text{ or } R,\end{aligned} \tag{99}$$

and, as $R_x^3 \oplus (R_v^m \oplus Z_v^p)$ can be seen to be invariant in $R^6 \square R$,

$$(R^6 \square R)/N_K(K_H) \approx [R_v^3/(R_v^m \oplus Z_v^p)] \times R/(t'\chi_v(t_0)/t_0, t'), \tag{100}$$

$\lambda(F_H) = R$ or dense,

$$\begin{aligned}&\text{then } (R^6 \square R)/N_K(K_H)\\ &= (R^6 \square R)/[(R_x^3 \oplus R_v^m) \square \{\chi_v(t'), t'\}],\\ &\qquad t' \in M_F = R,\end{aligned} \tag{101}$$

and, as $R_x^3 \oplus R_v^m$ can be seen to be invariant in $R^6 \square R$,

$$R^6 \square R/N_K(K_H) \approx (R_v^3/R_v^m) \times R/(t'\chi_v(t_0)/t_0, t'). \tag{102}$$

We then conclude that $H = N_{\mathcal{G}}(K_H)$ is a solution every time that

$$\begin{aligned}
&\text{(i)}\ M_F = Z \text{ or } R \quad \text{and} \quad \lambda(F_H) = \{0\},\\
&\text{(ii)}\ M_F = Z \text{ or } R, \quad \lambda(F_H) = Z, \quad \text{and} \quad m + p = 3,\\
&\text{(iii)}\ M_F = R, \lambda(F_H) = R \text{ or dense} \quad \text{and} \quad m = 3,
\end{aligned} \qquad (103)$$

[which in fact reduces to $\lambda(F_H) = R$ as here $\chi_x \equiv 0$].

Those results might have been given by the Theorem 1. Conversely, if $(R^6 \square R)/N_K(K_H)$ is noncompact, then it is isomorphic the quotient of two Abelian groups and its measure, here the Lebesgue measure, is nonbounded. It is the case when

$$\begin{aligned}
&\text{(i)}\ M_F = \{0\},\\
&\text{(ii)}\ M_F = Z \text{ or } R, \quad \lambda(F_H) = Z, \quad \text{and} \quad m + p < 3,\\
&\text{(iii)}\ M_F = R, \lambda(F_H) = R \text{ or dense} \quad \text{and} \quad m < 3.
\end{aligned} \qquad (104)$$

Let us now study the same problem for any H but replacing $\mathcal{G}$ by the corresponding $N_{\mathcal{G}}(K_H)$ in the case where it is itself a solution for $\mathcal{G}$. Then

$$\begin{aligned}
N_{\mathcal{G}}(K_H)/H &= (N_K(K_H) \square Q_H)/(K_H \overset{\Omega'}{\square} P_H)\\
&\simeq (T_H \square Q_H)/P_H = N_H/P_H.
\end{aligned}$$

We are going to select some cases where P_H is compact.

In fact, the closed subgroups of $SO(3)$ are $SO(3)$, $SO(2)$, $O(2)$, or finite: $\{0\}$, cyclic of order n (C_n), dihedral (C_{nk}), tetrahedral, octohedral, icosahedral. In the case where $\varphi(P_H)$ is finite, P_H is also finite and then compact.

If $M_F = Z$ or R, $\lambda(F_H) = Z$, and $m + p = 3$, and $\varphi(P_H) = SO(3)$, $SO(2)$, or $O(2)$, then T_H is compact in its x and t components and $f_v(r) = (1 - r)b$ is continuous. Hence P_H is closed in the compact:

$\varphi(P_H) \times \{(T_H)_x \times f_v(\varphi(P_H)) \times (T_H)_t\}$ and then compact.

If $M_F = \lambda(F_H) = R$ and $m = 3$, and $\varphi(P_H) = SO(3)$, $SO(2)$, or $O(2)$, then $f(r) = f_v(r) = (1 - r)b$ is continuous. Hence P_H is closed in the compact $\varphi(P_H) \times f(\varphi(P_H))$ and then is compact.

But, when P_H is compact, N_H/P_H is noncompact as soon as N_H is not, i.e., as soon as T_H is not. It is so

$$\begin{aligned}
&\text{if } M_F = Z \text{ or } R \quad \text{and} \quad \lambda(F_H) = \{0\}, \ \varphi(P_H) \text{ finite},\\
&\text{if } M_F = Z \text{ or } R, \quad \lambda(F_H) = Z, \quad m + p = 3,\\
&\qquad\qquad \text{and } n + q < 3, \ \varphi(P_H) \text{ closed},\\
&\text{if } M_F = \lambda(F_H) = R, \quad m = 3,\\
&\qquad\qquad \text{and } n + q < 3, \ \varphi(P_H) \text{ closed}.
\end{aligned} \qquad (105)$$

The theorem then comes from the fact that if H is a solution, then

$$m(\mathcal{G}/H) = m(\mathcal{G}/N_{\mathcal{G}}(K_H)) \cdot m(N_{\mathcal{G}}(K_H)/H). \qquad (106)$$

Remarks:

(1) It is this theorem which allows us to neglect the study of subgroups constructed from $F_{\{0\}}$ or with $\lambda(F_H)$ dense.

(2) In the case of Theorem 1, $\varphi(P_H)$ is necessarily closed. In fact T_H is then compact, as is also N_H. Then P_H, closed in N_H, is also compact, as is also $\varphi(P_H)$ as φ is continuous.

(3) If a group is not retained by Theorem 1, nor rejected by Theorem 2, then $\mathcal{G}/N_{\mathcal{G}}(K_H)$ is compact and it is sufficient to study $N_{\mathcal{G}}(K_H)/H \simeq N_H/P_H$.

We are first going to list the groups that are retained by Theorem 1, what we might call the "Galilean cristallographic groups":

$$\left.\begin{aligned}
&[(Z^3_x \oplus Z^3_x) \overset{\omega'}{\square} \{(n^2\chi_v(1)/2 + n(\chi_x(1) - \chi_v(1)/2), n\chi_v(1)), n\}] \overset{\Omega'}{\square} \{(1-r)\beta, r\},\\
&[((Z^2_x \oplus R_x) \oplus Z^3_v) \overset{\omega'}{\square} \{(n^2\chi_v(1)/2 + n(\chi_x(1) - \chi_v(1)/2), n\chi_v(1)), n\}]\\
&\overset{\Omega'}{\square} \{((1-r)\beta, (1-r)b), r\},\\
&[((Z^2_x \oplus R_x) \oplus (Z^2_x \oplus R_x)) \overset{\omega'}{\square} \{(n^2\chi_v(1)/2 + n(\chi_x(1) - \chi_v(1)/2)n\chi_v(1)), n\}]\\
&\overset{\Omega'}{\square} \{(1-r)\beta, r\},\\
&[((Z_x \oplus R^2_x) \oplus Z^3_v) \overset{\omega'}{\square} \{(n^2\chi_v(1)/2 + n(\chi_x(1) - \chi_v(1)/2), n\chi_v(1)), n\}]\\
&\overset{\Omega'}{\square} \{((1-r)\beta, (1-r)b), r\},\\
&[((Z_x \oplus R^2_x) \oplus (Z^2_x \oplus R_x)) \overset{\omega'}{\square} \{(n^2\chi_v(1)/2 + n(\chi_x(1) - \chi_v(1)/2), n\chi_v(1)), n\}]\\
&\overset{\Omega'}{\square} \{((1-r)\beta, (1-r)b), r\},\\
&[((Z_x \oplus R^2_x) \oplus (Z_x \oplus R^2_v)) \overset{\omega'}{\square} \{(n^2\chi_v(1)/2 + n(\chi_x(1) - n\chi_v(1)/2)), n\chi_v(1)), n\}]\\
&\overset{\Omega'}{\square} \{(1-r)\beta, r\};
\end{aligned}\right\} \qquad (107)$$

$$\left.\begin{aligned}
&[(R^3_x \oplus Z^3_v) \overset{\omega'}{\square} \{n\chi_v(1), n\}] \overset{\Omega'}{\square} \{((1-r)\beta, f_t(r)), r\}\\
&[(R^3_x \oplus Z^3_v) \overset{\omega'}{\square} \{t\chi_v(1), t\}] \overset{\Omega'}{\square} \{(1-r)\beta, r\}\\
&[(R^3_x \oplus (R_v \oplus Z^2_v)) \overset{\omega'}{\square} \{n\chi_v(1), n\}] \overset{\Omega'}{\square} \{((1-r)\beta, f_t(r)), r\},\\
&[(R^3_x \oplus (R_v \oplus Z^2_v)) \overset{\omega'}{\square} \{t\chi_v(1), t\}] \overset{\Omega'}{\square} \{(1-r)\beta, r\},\\
&[(R^3_x \oplus (R^2_v \oplus Z_v)) \overset{\omega'}{\square} \{n\chi_v(1), n\}] \overset{\Omega'}{\square} \{((1-r)\beta, f_t(r)), r\},\\
&[(R^3_x \oplus (R^2_v \oplus Z_v)) \overset{\omega'}{\square} \{t\chi_v(1), t\}] \overset{\Omega'}{\square} \{(1-r)\beta, r\};
\end{aligned}\right\} \qquad (107')$$

$$[(R^3_x \oplus R^3_v) \,\square\, Z] \overset{\Omega'}{\square} \{f_t(r), r\}; \tag{107''}$$

$$[(R^3_x \oplus R^3_v) \,\square\, R] \,\square\, \{r\}. \tag{107'''}$$

In brace (107) $r \in$ finite group, in brace (107′) $r \in$ closed subgroup of $O(2) \times Z_2$ or $O(2)$, and in (107″) and (107‴) $r \in$ closed subgroup of $O(3)$.

Concerning the remaining groups, which are to be studied separately, i.e., for which we have to search for a subgroup P_H of N_H such that $P_H \cap T_H = \{0\}$ and N_H/P_H compact, we are going to classify them according to the form of $N_H = T_H \,\square\, Q_H$, each family giving rise to a certain type of proof.

1st family:

$$\left.\begin{array}{l}
\{[T_x \oplus (R^2_v \oplus T_v)] \,\square\, Z\} \,\square\, (O(2) \times Z_2),\\
\{T_x \oplus R^2_v\} \,\square\, (O(2) \times Z_2 \text{ or } O(2)),\\
\{[T_x \oplus R^3_v] \,\square\, R\} \,\square\, (O(2) \times Z_2),\\
\{[T_x \oplus (R^2_v \oplus Z_v)] \,\square\, T\} \,\square\, (O(2) \times Z_2 \text{ or } O(2)),\\
\{[R^2_v \oplus T_v] \times R\} \,\square\, (O(2) \times Z_2),\\
\{[R^2_v \oplus T_v] \times T\} \,\square\, (O(2) \times Z_2),\\
\{R^2_v \oplus T_v\} \,\square\, (O(2) \times Z_2 \text{ or } O(2)),\\
\{R^2_v \times R\} \,\square\, (O(2) \times Z_2),\\
\{R^2_v \times T\} \,\square\, (O(2) \times Z_2),\\
\{R^2_v\} \,\square\, (O(2) \times Z_2 \text{ or } O(2)),\ \varphi(P_H) \text{ nonfinite or nonclosed.}
\end{array}\right. \tag{108}$$

We are going to eliminate this family: as each of those N_H contains, as a closed subgroup, the last one, it is sufficient to eliminate this case. But this has been done in Ref. 1 [5.4b2-β), 5.4b2-γ)] and we refer to it for the proof.

2nd family:

$$\left.\begin{array}{l}
\{R^3_v \times R\} \,\square\, O(3),\\
\{R^3_v \times T\} \,\square\, O(3) \quad [\text{or } O(2)\text{: see 1st family}],\\
R^3_v \,\square\, O(3) \quad [\text{or } O(2)\text{: see 1st family}],\\
\varphi(P_H) \text{ nonfinite or nonclosed.}
\end{array}\right. \tag{109}$$

As each of those N_H contains, as a closed subgroup, the last one, it is sufficient to eliminate this case. But this has been done in Ref. 1 (5.4c), and we refer to it for the proof.

3rd family:

$$\left.\begin{array}{l}
\{[T_x \oplus T_v] \,\square\, Z\} \times (O(2) \times Z_2),\\
\{[T_x \oplus Z_v] \,\square\, T\} \times (O(2) \times Z_2 \text{ or } O(2)),\\
\{R_v \times T\} \times (O(2) \times Z_2),\\
R_v \times (O(2)),\\
\{T_v \times R\} \times (O(2) \times Z_2), \qquad \varphi(P_H) \text{ nonfinite or nonclosed.}
\end{array}\right. \tag{110}$$

If we project on the noncompact component of T_H, we get groups of the type $Z \times O(2)$ or $R \times O(2)$, and the projection of P_H is closed too. Hence for these projections the problem has been solved in Ref. 1 [5.4b2-α)] and we refer to it for the proof, at least for the second case: The generators are of the form

$$\left.\begin{array}{l}
(\lambda \mathbf{u}, \theta_1), \quad \theta_1/2\pi \text{ irrational},\\
\qquad \lambda \in R, \quad \lambda \neq 0, \quad \mathbf{u} \text{ unitary vector of } R \text{ in } R \times O(2),\\
\theta_2, \quad \theta_2/2\pi \text{ rational.}
\end{array}\right. \tag{111}$$

In the first case, an analoguous proof would show that the generators are of the form

$$\left.\begin{array}{l}
(na, \theta_1), \quad \theta_1/2\pi \text{ irrational},\\
\qquad a \text{ the fundamental length of } Z, \quad n \neq 0,\\
(0, \theta_2), \quad \theta_2/2\pi \text{ rational.}
\end{array}\right. \tag{112}$$

We get then

$$\left.\begin{array}{l}
[(Z_x \oplus R^2_x) \oplus (Z_v \oplus R^2_v)] \overset{\Omega'}{\square} \{(f_x(\theta_1), f_v(\theta_1), nt_0, \theta_1), (f_x'(\theta_2), f_v'(\theta_2), 0, \theta_2)\},\\
\qquad\qquad \theta_1/2\pi \text{ irrational}, \quad \theta_2/2\pi \text{ rational}, \quad n \neq 0, \quad t_0 \text{ the fundamental length of } Z_t = M_F,\\
[((R^2_x \oplus Z_x) \oplus R^2_v) \overset{\omega'}{\square} \{(n^2\chi_v(1)/2 + n(\chi_x(1) - \chi_v(1)/2), n\chi_v(1)), n\}]\\
\quad \overset{\Omega'}{\square} \{(f_x(\theta_1), pv_0, f_t(\theta_1), \theta_1), (f_x'(\theta_2), 0, f_t'(\theta_2), \theta_2)\},\\
\qquad\qquad \theta_1/2\pi \text{ irrational}, \quad \theta_2/2\pi \text{ rational}, \quad p \neq 0, \quad v_0 \text{ the fundamental length of } Z_v = (Z_x)_v,\\
[(R^3_x \oplus R^2_v) \,\square\, \{n\chi_v(1), n\}] \overset{\Omega'}{\square} \{(\lambda \mathbf{v}_0, f_t(\theta_1), \theta_1), (0, f_t'(\theta_2), \theta_2)\},\\
\qquad\qquad \theta_1/2\pi \text{ irrational}, \quad \theta_2/2\pi \text{ rational}, \quad \lambda \neq 0, \quad \mathbf{v}_0 \text{ unitary vector of } R_v,\\
[(R^3_x \oplus R^2_v) \,\square\, \{t\chi_v(1), t\}] \overset{\Omega'}{\square} \{(\lambda \mathbf{v}_0, \theta_1), (0, \theta_1)\},\\
\qquad\qquad \theta_1/2\pi \text{ irrational}, \quad \theta_2/2\pi \text{ rational}, \quad \lambda \neq 0, \quad \mathbf{v}_0 \text{ unitary vector of } R_v,\\
(R^3_x \oplus (R^2_v \oplus Z_v)) \overset{\Omega'}{\square} \{(f_v(\theta_1), \lambda \mathbf{t}_0, \theta_1), (f_v'(\theta_2), 0, \theta_2)\},\\
\qquad\qquad \theta_2/2\pi \text{ irrational}, \quad \theta_1/2\pi \text{ rational}, \quad \lambda \neq 0, \quad t_0 \text{ unitary vector of } R_t = M_F.
\end{array}\right\} \tag{113}$$

4th family:

$$\{[T_x \oplus R_v] \,\square\, R\} \times (O(2) \times Z_2),$$
$$\{R_v \times R\} \times (O(2) \times Z_2). \tag{114}$$

By projection, we have to study the second case. The problem is analoguous to the preceding one, treated in Ref. 1 [5. 4b2-α)], with one dimension more. We get then

$$\left.\begin{array}{l} ((R_x^2 \oplus Z_x) \oplus R_v^2) \overset{\Omega'}{\square} \{(f_x(\theta_1), \lambda \mathbf{v}_0, 0, \theta_1), \\ \qquad (f_x'(\theta_2), 0, \mu \mathbf{t}_0, \theta_2), (f_x''(\theta_3), 0, 0, \theta_3)\}, \\ \quad \theta_1/2\pi \text{ irrational}, \quad \theta_2/2\pi \text{ irrational}, \\ \quad \theta_1/\theta_2 \text{ irrational}, \\ \quad \theta_3/2\pi \text{ rational}, \quad \lambda \neq 0, \quad \mu \neq 0, \\ \quad \mathbf{v}_0 \text{ unitary vector of } R_v, \\ \quad t_0 \text{ unitary vector of } R_t = M_F \\ (R_x^3 \oplus R_v^2) \overset{\Omega'}{\square} \{(\lambda \mathbf{v}_0, 0, \theta_1), (0, \mu \mathbf{t}_0, \theta_2), (0, 0, \theta_3)\} \end{array}\right\} \tag{115}$$

with the same notations.

5th family:

$$R \times O(3). \tag{116}$$

The application $(f_t(r), r) \overset{\varphi}{\to} r$ is here an algebraic isomorphism, which becomes topological if and only if the subgroup we are searching for is compact. But in that case the quotient would not be compact. Hence we have to search for subgroups P_H such that their projection into $SO(3)$ is nonclosed. But then the closure of this projection can only be $SO(2)$, $O(2)$, or $SO(3)$ itself.

(i) *The closure of* $\varphi(P_H)$ *is* $SO(2)$: We are led to the problem for the group $R \times SO(2)$, the solution of which is in Ref. 1 [5. 4b2-α)], and the answer is positive.

(ii) *The closure of* $\varphi(P_H)$ *is* $O(2)$: one must add to the preceding case a symmetry through the origin, and the answer is positive too.

(iii) *The closure of* $\varphi(P_H)$ *is* $SO(3)$:

(α) Let us first show that P_H cannot be discrete. Let $B_\epsilon \subset SO(3)$ the ball $B_\epsilon = \{r \in SO(3), \|r - 1\| < \epsilon\}$. For ϵ sufficiently small, B_ϵ is such that if $s, t \in B_\epsilon$ and $[s,[s,t]] = 1$, then $[s,t] = 1$; if $s, t \in B_\epsilon$, then $[s,t]$, $[s,[s,t]], [s,[s,[s,t]]], \cdots$ is a sequence in B_ϵ which converges to 1.[7] (Here $[s,t] = sts^{-1}t^{-1}$).

Let $\gamma_i = (f_t(r_i), r_i)$, $i = 1, 2$ and $r_i \in B_\epsilon$, and $\gamma_{i+1} = [\gamma_1, \gamma_i]$ $(0, [r_1, r_i]) = (0, r_{i+1})$ for $i \geq 2$. Then $r_i \in B_\epsilon$ and $\{r_i\}$ or $\{\gamma_i\}$ are sequences tending to 1. As P_H is discrete, $\gamma_i = 1$ for i large enough. In particular for $i \geq 4$, $\gamma_i = 1$ implies $\gamma_{i-1} = 1$ and then $\gamma_3 = 1$. But then $[\gamma_1, \gamma_2] = 1$ and the set of γ_i is Abelian. Then $\varphi(P_H) \cap B_\epsilon$ is Abelian and generates $\overline{\varphi(P_H)} = SO(3)$ which should be Abelian.

(β) So P_{H_0}, the connected component of the identity in P_H, cannot be reduced to a point. Since φ is here a continuous injective group isomorphism, $\overline{\varphi(P_{H_0})}$ is a connected closed subgroup of $SO(3)$: i.e., either $SO(2)$ or $SO(3)$ itself. But it has been shown in Ref. 1 [5. 4c-γ)] that in this case $\varphi(P_H) = SO(3)$, which is contradictory to our hypothesis. (We refer to Ref. 1 for the proof).

We get then finally

$$\begin{array}{l} (R_x^3 \oplus R_v^3) \times ((\lambda t_0, \theta_1), (0, \theta_2)) \\ \quad \theta_1/2\pi \text{ irrational}, \quad \theta_2/2\pi \text{ rational}, \quad \lambda \neq 0, \\ \quad t_0 \text{ the unitary vector of } R_t = M_F. \end{array} \tag{117}$$

CONCLUSION

We have then listed all the symmetry groups we were aiming at. As in the case of the Euclidean group, we have to notice that they are defined up to a conjugation in $\mathcal{G}$, while the symmetry groups are usually classified up to a conjugation in the general linear group.

The families we have obtained can be considered as "Galilean extensions" of the ones obtained in the case of the Euclidean group, in the sense that we do not obtain fundamentaly new groups. In fact, we get

(1) one-, two- and three-dimensional cristallographic groups in x, v and t (107),

(2) one-, two- and three-dimensional cristallographic groups in v and (or) t only (107′),

(3) a cristallographic group in t alone (107″),

(4) a group of the form $[(R_x^3 \oplus R_v^3) \,\square\, R) \,\square\, K$, where K is a closed subgroup of $SO(3)$ (107‴),

(5) some "helicoïdal groups" in t or (and) v direction (113, 115, 117).

ACKNOWLEDGMENTS

We give warm thanks to Daniel Kastler and Louis Michel. This work has been achieved thanks to their ideas, their help, and their encouragements.

---

[1] D. Kastler, G. Loupias, M. Mebkhout, and L. Michel, Commun. Math. Phys. 27, 195 (1972), (1972).

[2] S. Sakai, Trans. Am. Math. Soc. **118**, 406 (1965)..

[3] G. D. Mostow, Ann. Math. **75**, 17 (1962).

[4] The case of the physically relevant full Galilei group, corresponding to $O(3)$, can be easily adapted from our results.

[5] N. Bourbaki, *Livre VI — Integration* (Hermann, Paris, 1963), Chap. VII.

[6] N. Bourbaki, *Livre III — Topologie générale* (Hermann, Paris, 1967), Chap. VII.

[7] J. A. Wolf, *Spaces of Constant Curvature* (McGraw-Hill, New York, 1966), Lemmas 3.2.4 and 3.2.5.

On Killing tensors and constants of motion

Paul Sommers*

Center for Relativity Theory, The University of Texas at Austin, Austin, Texas 78712
(Received 29 November 1972)

Some general properties belonging to constants of motion for geodesics and charged particle orbits are derived. A constant of motion for geodesics is seen to be a function on the cotangent bundle which has vanishing Poisson bracket with the "energy function" determined by the metric tensor. The resulting algebraic structure on the set of constants of motion is closely related to the Lie algebra of Killing tensors. Each constant of motion is shown to provide a family of mappings of geodesics into geodesics. Constants of motion for charged particles also possess a Lie algebra structure. The relationship of Killing tensors to charged particle constants of motion is derived. The linear and quadratic constants of motion for charged particle orbits in the charged Kerr metric illustrate the results. Examples of valence 2 Killing tensors are given in an appendix.

1. INTRODUCTION

The analysis of physical processes in curved spacetimes requires a knowledge of the trajectories of test particles. The calculation of these trajectories is a formidable task in general, but it is feasible if constants of motion are known. For geodesics[1] or charged particle orbits, the norm of the momentum is conserved. If three other constants of motion are known, the momentum may be computed algebraically at each point of the trajectory, thereby reducing the orbit problem to first order differential equations. A constant of motion which is linear in the momentum exists whenever the metric admits a one-parameter group of isometries. The origins of higher degree constants of motion are more obscure. Some examples are given in Appendix B.

The Kerr metric provides a physically important example of a quadratic constant of motion. Because this metric may be identified as the field of a rotating black hole, calculations in black hole physics rely heavily on the properties of geodesics and charged particle orbits in the Kerr metric. The discovery of the quadratic constant of motion[2] has made it practicable, for example, to construct astrophysical models of matter accretion onto black holes and to calculate radiation patterns from test particles near black holes, without restricting consideration to the equatorial plane. This quadratic constant of motion arises from a Killing tensor of valence 2.[3]

Motivated by this example, the present paper discusses some properties which are common to constants of motion of all degrees. Section 2 reviews some aspects of Killing tensors, including the generalization of the Lie algebra of Killing vectors to a Lie algebra of Killing tensors of all valences. In Sec. 3 a constant of motion for geodesics is regarded as a function on the cotangent bundle which is constant along the integral curves of the geodesic spray. This leads immediately to a Lie algebra of the constants of motion for geodesics. In addition, each constant of motion is seen to determine a vector field on the cotangent bundle which commutes with the geodesic spray, thereby providing a one-parameter family of transformations of geodesics into geodesics. Section 4 shows that constants of motion for charged particle orbits enjoy the same properties. Charged particle constants of motion for the charged Kerr metric are discussed.

2. KILLING TENSORS

In a Riemannian or pseudo-Riemannian space, a *Killing tensor* is a completely symmetric tensor $K^{ab\cdots c}$ which satisfies the Killing equation

$$\nabla^{(m}K^{ab\cdots c)} = 0, \tag{2.1}$$

where ∇_a denotes covariant differentiation. A constant of motion for geodesics is associated with any such Killing tensor. To be explicit, suppose p_a is the covariant form of the tangent to a geodesic congruence, so that $p_m\nabla^m p_a = 0$. Then $K^{ab\cdots c}p_a p_b \cdots p_c$ is constant along the geodesics:

$$p_m\nabla^m(K^{ab\cdots c}p_a p_b\cdots p_c) = p_m p_a p_b\cdots p_c\nabla^{(m}K^{ab\cdots c)} = 0. \tag{2.2}$$

The metric tensor trivially satisfies the Killing equation, in consequence of which the norm of the tangent to a geodesic is conserved. The symmetrized outer product of Killing vectors also satisfies the Killing equation, the constant of motion being simply the product of those associated with the Killing vectors individually. To distinguish these trivial cases, a Killing tensor is said to be *reducible* if it can be written as a fixed sum of symmetrized outer products of lower valence Killing tensors and the metric tensor.[3] Otherwise it is *irreducible*.

The set of Killing tensors on a space enjoys a Lie algebra structure which is a generalization of the Lie algebra of Killing vectors using the Lie bracket as multiplication. One first introduces a Lie algebra on the set of completely symmetric contravariant tensors of all valences. Let $S^{ab\cdots c}$ and $T^{ab\cdots d}$ be symmetric tensors of valence m and n, respectively. Their skew product, $P^{b\cdots f}$, is a symmetric tensor of valence $m+n-1$ and is given by[4]

$$P^{b\cdots f} = mS^{r(b\cdots c}\partial_r T^{d\cdots f)} - nT^{r(b\cdots d}\partial_r S^{e\cdots f)}. \tag{2.3}$$

Suppressing indices, this will be written

$$P = [S,T]. \tag{2.4}$$

The product, so defined, is antisymmetric in S and T, is linear in each slot, is unchanged if ordinary partial derivatives are replaced by covariant derivatives, reduces to the Lie derivative of T along S if S is a vector field, satisfies the Jacobi identity and the following Leibnitz rule: If $T \cap V$ denotes the symmetrized outer product of tensors T and V, then

$$[S, T\cap V] = [S,T]\cap V + [S,V]\cap T. \tag{2.5}$$

Geroch[5] has pointed out that a Lie algebra of Killing tensors can then be defined as the subalgebra of symmetric tensors $K^{ab\cdots c}$ which commute with the metric tensor g^{ab},

$$[K,g]=0. \tag{2.6}$$

This equation is identical to the Killing equation when covariant differentiation is used in the product definition. One easily verifies that this subset is closed under the operations $\cap$, $[\ ,\ ]$, and addition of two tensors of equal valence.

3. CONSTANTS OF MOTION FOR GEODESICS

A function F, which depends both on position in a manifold M and on a vector p_a in the cotangent space at the point of M, should properly be regarded as a function on the cotangent bundle, T^*M, of M. Since the cotangent bundle to any manifold is endowed with a canonical Hamiltonian symplectic structure $\Omega_{\alpha\beta}$ [cf. Appendix A], indices[6] of tensors on T^*M will be lowered via $\Omega_{\alpha\beta}$ and raised via its inverse, $\Omega^{\alpha\beta}$. In particular, the gradient of a differentiable function F may be regarded as a vector field,

$$F^\alpha=\Omega^{\alpha\beta}\partial_\beta F. \tag{3.1}$$

This field may then act on another differentiable function G,

$$F^\alpha\partial_\alpha G=\Omega^{\alpha\beta}(\partial_\beta F)(\partial_\alpha G)\equiv[F,G]_P, \tag{3.2}$$

the result being the Poisson bracket of F and G. On T^*M, then, scalar functions may be combined by addition, multiplication, or Poisson bracket multiplication.

A constant of motion for geodesics is a scalar function K on T^*M which is constant along the curves of T^*M obtained by lifting geodesics from M. Because a unique geodesic is associated with a particular covariant vector at a particular point in M, the lifted geodesics form a simple congruence of curves on T^*M. The geodesic spray, g^α is the vector field tangent to this congruence. As shown in Appendix A, g^α is given by

$$g^\alpha=\Omega^{\alpha\beta}\partial_\beta g, \tag{3.3}$$

where g is the *energy function*[7]

$$g:=\tfrac{1}{2}g^{ab}p_ap_b. \tag{3.4}$$

The condition that K be a constant of motion for geodesics is therefore

$$0=g^\alpha\partial_\alpha K\equiv[g,K]_P. \tag{3.5}$$

Constants of motion are the functions which commute with the energy function g. From the properties of the Poisson bracket it follows that the sum, product, or Poisson bracket of two constants of motion is again a constant of motion.

The bracket operation for symmetric tensors, as defined in Sec. 1, is closely related to the Poisson bracket. A function on T^*M,

$$F:=F^{ab\cdots c}p_ap_b\cdots p_c, \tag{3.6}$$

may be associated with any contravariant symmetric tensor $F^{ab\cdots c}$ on M. If F and G are the functions associated, respectively, with $F^{ab\cdots c}$ and $G^{ab\cdots d}$, then the tensor bracket operation yields a new symmetric tensor whose associated scalar function is $[F,G]_P$. In the set of constants of motion, those which have the simple form $K^{ab\cdots c}p_ap_b\cdots p_c$ thereby constitute a subset which is closed under the Poisson bracket operation.

Another important property of constants of motion follows immediately from these considerations. If K is a constant of motion, then *the geodesic spray is Lie dragged along the gradient vector field K^α*:

$$[K,g]^\alpha\equiv\mathcal{L}_K g^\alpha=0. \tag{3.7}$$

This follows from the fact that, since $\Omega_{\alpha\beta}g^\beta$ is a gradient and $\partial_{[\gamma}\Omega_{\alpha\beta]}=0$, then $\mathcal{L}_g\Omega_{\alpha\beta}=0$. Using this, together with the hypothesis that $g^\alpha\partial_\alpha K\equiv\mathcal{L}_g K=0$, one obtains

$$0=\partial_\beta(\mathcal{L}_g K)=\mathcal{L}_g\partial_\beta K=\mathcal{L}_g(\Omega_{\alpha\beta}K^\alpha)=\Omega_{\alpha\beta}\mathcal{L}_g K^\alpha. \tag{3.8}$$

Because $\Omega_{\alpha\beta}$ is of maximal rank, this implies $\mathcal{L}_g K^\alpha=-\mathcal{L}_K g^\alpha=0$. In this way, *a constant of motion K moves geodesics onto geodesics.*

In a four-dimensional space, suppose three independent commuting constants of motion, K,L,M, are known. On T^*M, since $g^\alpha\partial_\alpha K, K^\alpha\partial_\alpha g, K^\alpha\partial_\alpha L$, etc., all vanish, the vector fields $g^\alpha, K^\alpha, L^\alpha, M^\alpha$ are tangent to the 4-surfaces of constant g,K,L,M. By the argument of the preceding paragraph, $g^\alpha, K^\alpha, L^\alpha, M^\alpha$ all commute as vector fields and, therefore, may serve as a basis on these 4-surfaces. Functions k,l,m may be chosen such that $K^\alpha\partial_\alpha=\partial/\partial k$, $L^\alpha\partial_\alpha=\partial/\partial l$, $M^\alpha\partial_\alpha=\partial/\partial m$, and such that the points of a lifted geodesic are given by specifying constant values for g,K,L,M,k,l,m.

4. CHARGED PARTICLE ORBITS

The discussion of the preceding section generalizes to the case of charged particles moving in a Maxwell field. In four-dimensional space-times with physical Maxwell field, the charged particle orbits are determined by the Lorentz equation of motion:

$$p_m\nabla^m p_a=q\,F_a{}^m p_m, \tag{4.1}$$

where q is the charge of the particle and F_{ab} is the Maxwell field. The orbits satisfying the Lorentz equation for a fixed value of q constitute a single congruence of curves when lifted to T^*M. The vector field tangent to the congruence is given by

$$g^\alpha[q]=\Omega^{\alpha\beta}[q]\partial_\beta g. \tag{4.2}$$

In this equation, g is the energy function constructed from the metric as before, and $\Omega^{\alpha\beta}[q]$ is the symplectic structure appropriate to the charge q [cf. Appendix A]. Since all the properties of the canonical Hamiltonian structure which were used in the discussion of geodesics are also properties of $\Omega^{\alpha\beta}[q]$, it follows that constants of motion for particles of charge q possess the same algebraic structure.

The relationship of Killing tensors to constants of motion for charged particle orbits is slightly more involved than in the case of geodesics. Suppose a function K on T^*M is of the form

$$K=\overset{0}{K}+\overset{1}{K}{}^a p_a+\overset{2}{K}{}^{ab}p_ap_b+\cdots \tag{4.3}$$

for symmetric tensor fields $\overset{0}{K}, \overset{1}{K}{}^a, \overset{2}{K}{}^{ab}, \cdots$ on M. When the geodesic spray g^α [in the form given by Eq. (A2) of Appendix A] acts on K, term by term, each resulting term is of a different degree in p_a and must therefore vanish by itself if K is to be a constant of motion for all geodesics. The individual terms then show that each of the tensors $\overset{0}{K}, \overset{1}{K}{}^a, \overset{2}{K}{}^{ab}, \ldots$, must be a Killing tensor. When $g^\alpha[q]$ [in the form given by Eq. (A3) of Appendix A] is applied to K, however, the result vanishes for arbitrary p_a if and only if,[8] for each value of n,

$$\nabla^{(a}\overset{n}{K}{}^{b\ldots c)} + (n+1)qF_m{}^{(a}\overset{(n+1)}{K}{}^{b\ldots c)m} = 0. \tag{4.4}$$

Two special cases, which occur in the charged Kerr metric, deserve mention. Suppose K is linear in p_a so that

$$K = \overset{0}{K} + \overset{1}{K}{}^a p_a. \tag{4.5}$$

The conditions that $g^{\alpha}[q]\partial_\alpha K = 0$, in this case, are

$$\nabla^{(a}\overset{1}{K}{}^{b)} = 0. \tag{4.6}$$

and

$$\nabla^a \overset{0}{K} + qF_m{}^a \overset{1}{K}{}^m = 0. \tag{4.7}$$

Therefore $\overset{1}{K}{}^a$ must be a Killing vector, and, using the Maxwell equation $\partial_{[a}F_{bc]} = 0$, the Eq. (4.7) is seen to be equivalent[9] to the condition

$$\mathcal{L}_{\overset{1}{K}} F_{ab} = 0. \tag{4.8}$$

The two Killing vectors of the charged Kerr metric in this way yield constants of motion for particles of arbitrary charge.

The Kerr spacetime also admits a quadratic constant of motion for charged particles,[2,10]

$$K = \overset{2}{K}{}^{ab}p_a p_b, \tag{4.9}$$

where $\overset{2}{K}{}^{ab}$ is the irreducible Killing tensor.[3] The conditions (4.4) in this case yield the two relations

$$\nabla^{(a}\overset{2}{K}{}^{bc)} = 0 \tag{4.10}$$

and

$$F_m{}^{(a}\overset{2}{K}{}^{b)m} = 0. \tag{4.11}$$

The first equation is the Killing equation. Exploiting the antisymmetry of F_{ab}, the second relation is easily verified by writing the Killing tensor of the Kerr metric in the form[10]

$$\overset{2}{K}{}^{ab} = \alpha F^a{}_m F^{mb} + \beta g^{ab}. \tag{4.12}$$

APPENDIX A: HAMILTONIAN STRUCTURES AND LIFTED ORBITS ON T*M

The canonical Hamiltonian structure on the cotangent bundle, T^*M, may be defined invariantly by means of certain bundle projection maps. Let ξ^α be a vector at a point of T^*M, so ξ^α may be regarded as a point in TT^*M, the tangent bundle to T^*M. The bundle projection h from TT^*M to T^*M maps ξ^α into the point of T^*M to which it is attached. This point may be regarded as a covector p_a at the point of M under the fiber to which ξ^α is attached. At the same time, the projection π from T^*M to M determines a differential map π_* from TT^*M to TM which maps ξ^α into a contravariant vector ξ^a at the point of M under the fiber to which ξ^α is attached. The scalar $\xi^a p_a$ is naturally determined in this way from ξ^α. Doing this for each vector ξ^α on T^*M defines a canonical 1-form θ_α on T^*M with values $\theta_\alpha \xi^\alpha = \xi^a p_a$. The curl of θ_α is the canonical Hamiltonian symplectic structure:[11]

$$\Omega_{\alpha\beta} = 2\partial_{[\alpha}\theta_{\beta]} \tag{A1}$$

The symplectic structures $\Omega_{\alpha\beta}[q]$ may be obtained in the same way. Let V_a be a potential for a Maxwell field F_{ab}. The projection h from TT^*M to T^*M may be changed to $h[q]$ so that, instead of the covector p_a at a point in M, the projection yields $p_a + qV_a$ at the same point in M. Using the maps $h[q]$ and π_* a 1-form $\theta_\alpha[q]$ is defined as above. The curl of $\theta_\alpha[q]$ is $\Omega_{\alpha\beta}[q]$. The gauge freedom in V_a is of no consequence since it does not enter $\Omega_{\alpha\beta}[g]$.

Regarding each point of T^*M as a point of M together with a covariant vector at that point, it is natural to adopt coordinates (x^a, p_a) on T^*M such that x^a are coordinate of the point in M and p_a are the components of the covector in the same coordinate basis of M. Then the components of $\Omega^{\alpha\beta}$ are constants and $g^\alpha\partial_\alpha$ takes the form

$$g^\alpha\partial_\alpha \equiv g^{\alpha\beta}(\partial_\beta g)\partial_\alpha = \frac{\partial g}{\partial p_a}\frac{\partial}{\partial x^a} - \frac{\partial g}{\partial x^a}\frac{\partial}{\partial p_a}$$
$$= g^{ab}p_b\frac{\partial}{\partial x^a} - \frac{1}{2}(\partial_a g^{bc})p_b p_c\frac{\partial}{\partial p_a} \tag{A2}$$

If ρ is a parameter along the integral curves of g^α, then $dx^a/d\rho = g^{ab}p_b$ and $dp_a/d\rho = -\frac{1}{2}(\partial_a g^{bc})p_b p_c$. The first equation identifies p_a as the covariant tangent vector to the curve projected onto M. The second equation verifies that the projected curves are, in fact, the geodesics of M.

In a special coordinate system of this type on T^*M, it is straightforward to check that

$$g^\alpha[q]\partial_\alpha \equiv \Omega^{\alpha\beta}[q](\partial_\beta g)\partial_\alpha = g^\alpha\partial_\alpha + qF_a{}^b p_b\frac{\partial}{\partial p_a}. \tag{A3}$$

If now the integral curves of $g^\alpha[q]$ are parametrized by ρ, then

$$\frac{dp_a}{d\rho} = -\tfrac{1}{2}(\partial_a g^{bc})p_b p_c + qF_a{}^b p_b, \tag{A4}$$

or

$$p_m\nabla^m p_a = qF_a{}^b p_b, \tag{A5}$$

which is the Lorentz equation of motion.

APPENDIX B: SOME EXAMPLES OF KILLING TENSORS OF VALENCE 2

(1) Any covariantly constant symmetric tensor is a particularly simple solution of the Killing equation. If such a tensor is nondegenerate, it may be regarded as a second metric tensor compatible with the original affine connection.

Covariantly constant tensors with zero determinant occur whenever the space is decomposable into complementary, orthogonal surfaces with the induced metric on each surface depending only on the coordinates of that surface.[12]

(2) If two (pseudo-)Riemannian spaces with metrics g_{ab} and g'_{ab} have the same geodesic paths, then they are projectively related. The affine connections have the relationship

$$\Gamma'^a_{bc} = \Gamma^a_{bc} + 2\delta^a_{(b}\partial_{c)}\psi \tag{B1}$$

for some scalar field ψ. It follows that

$$K_{ab} := e^{-4\psi}g'_{ab} \tag{B2}$$

is a Killing tensor in the space of g_{ab}.[12]

(3) An affine collineation is a vector field ξ^a satisfying $\mathcal{L}_\xi \Gamma^a_{bc} = 0$. Using the identity

$$\mathcal{L}_\xi \Gamma^a_{bc} \equiv \nabla_b \nabla_c \xi^a - R^a{}_{cbm}\xi^m, \tag{B3}$$

it is seen that $\nabla_{(b}\nabla_c \xi_{a)} = 0$ if ξ^a is an affine collineation. In that case $\nabla_{(a}\xi_{b)}$ is a Killing tensor.

A projective collineation, η^a, satisfies

$$\mathcal{L}_\eta \Gamma^a_{bc} = 2\delta^a_{(b}\partial_{c)}\phi \tag{B4}$$

for some scalar field ϕ. The same identity (B3) then shows that the tensor

$$K_{ab} := \nabla_{(a}\eta_{b)} - 2\phi g_{ab} \tag{B5}$$

satisfies the Killing equation when η^a is a projective collineation.[13]

(4) In four-dimensional space-time, each vacuum solution of the Einstein field equation for which the Weyl tensor is of type $\{2, 2\}$ admits a trace-free *conformal Killing tensor*[3] P_{ab} which satisfies the equation

$$\nabla_{(a}P_{bc)} = \tfrac{1}{3} g_{(ab}\nabla^m P_{c)m}. \tag{B6}$$

Except in the cases of the C-metric and its rotating generalization due to Kinnersley,[14] the divergence $\frac{1}{3}\nabla^m P_{cm}$ is the gradient of a scalar α.[15] Then the tensor

$$K_{ab} := P_{ab} - \alpha g_{ab} \tag{B7}$$

is a Killing tensor. The result holds equally well for the charged versions of these metrics.

ACKNOWLEDGMENTS

I thank R. Geroch, L. Hughston, R. Matzner, J. Rosenman, A. Schild, L. Smarr, M. Walker, and J. York for helpful discussions. I thank The University of Texas at Dallas and Max Planck Institute for their hospitality.

*This work supported in part by National Science Foundation Grant GP-34639X.

[1] In this paper "geodesic" will mean "affinely parametrized geodesic."

[2] B. Carter, Phys. Rev. **174**, 1559 (1968).

[3] M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).

[4] A. Nijenhuis, Idag. Math. **17**, 390 (1955).

[5] R. Geroch, J. Math. Phys. **11**, 1955 (1970).

[6] Abstract indices for tensors on T^*M will be chosen from the Greek alphabet to distinguish tensors on T^*M from tensors on M.

[7] Since the energy function is a function on T^*M, it should rigorously be given by $g := (1/2)(g^{ab}\circ\pi)p_a p_b$ where π is the bundle projection. The projection map will be omitted from such expressions for the sake of compact equations since there is little danger of confusion.

[8] The relationship (4.4) was deduced previously by L. Hughston (private communication) using a different approach.

[9] M. Walker, Ph.D. thesis, Birkbeck College, 1969.

[10] L. Hughston, R. Penrose, P. Sommers, and M. Walker, Commun. Math. Phys. **27**, 303 (1972).

[11] R. Bishop and S. Goldberg, *Tensor Analysis on Manifolds* (MacMillan, New York, 1968).

[12] J. Schouten, *Ricci Calculus* (Springer-Verlag, Berlin, 1954).

[13] W. Davis, M. Moss, and J. York, Nuovo Cimento **B 65**, 19 (1970).

[14] W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).

[15] L. Hughston and P. Sommers, "Spacetimes with Killing tensors", Commun. Math. Phys. (to appear).

Hypercontractivity for fermions

Ivan F. Wilde*

Seminar for Theoretical Physics, E.T.H., 8049 Zürich, Switzerland

(Received 3 January 1973)

If T is a fermion one-particle operator, with $\|T\| < 1$, then T induces a bounded map from $L^2(\mathcal{C}) \to L^p(\mathcal{C})$, for some $p > 2$, where $\mathcal{C}$ is the fermion von Neumann algebra. The proof is an adaptation to the fermion case of the corresponding proof for bosons given by Nelson. This leads to a generalization of a theorem of Gross, inasmuch as T is not required to be self-adjoint.

Recently, Nelson[1] has given a quick proof of a hypercontractive bound for free bosons using stochastic integrals. For fermions, the relevant algebra is noncommutative, and so cannot be regarded as an algebra of random variables in the conventional sense. Stochastic integration, therefore, is not immediately applicable. However, all one needs are special properties of Gaussian stochastic integrals which are equivalent to properties of Wick products as given by Wick's theorem. It is the latter which carry over immediately to the fermion case.

Let $\mathcal{K}$ be a complex Hilbert space (we may assume that $\dim \mathcal{K} = \infty$) and let J be a conjugation on $\mathcal{K}$ (i.e., J is antilinear, antiunitary, and $J^2 = 1$). Let $\mathfrak{F} = \oplus_{n=0}^{\infty} \mathfrak{F}_n$ be the antisymmetric Fock space over $\mathcal{K}$, and let $C(z)$, for $z \in \mathcal{K}$, be the creation operator on $\mathfrak{F}$.[2]

Put $A(z) = C(z)^*$, and define the field $B(z)$ by $B(z) = C(z) + A(Jz)$, $z \in \mathcal{K}$.

Let $\mathcal{C}$ denote the smallest von Neumann algebra containing all the $B(z)$, and let $m(\cdot)$ denote the vacuum expectation value; $m(u) = (\Omega, u\Omega)$, $\forall u \in \mathcal{C}$. Then[3] m is a faithful, central state on $\mathcal{C}$, and $(\mathfrak{F}, \mathcal{C}, m)$ is a regular probability gage space in the sense of Segal.[4]

For $1 \le p < \infty$, one defines $L^p(\mathcal{C})$ to be the completion (modulo null elements) of $\mathcal{C}$ with respect to the norm $\|u\|_p = m(|u|^p)^{1/p} = (\Omega, (u^*u)^{p/2}\Omega)^{1/p}$. $L^\infty(\mathcal{C})$ is, by definition, $\mathcal{C}$ with respect to its operators norm.[5]

$L^p(\mathcal{C})$ is the fermion analog of the "Q-space" function spaces $L^p(Q)$, constructed from the free boson field. Indeed, $L^p(Q)$ is just the corresponding completion of $\mathfrak{M}$, the maximal Abelian algebra generated by the time-zero free boson fields, with respect to the boson Fock vacuum.[6]

One can prove the following.

Theorem[3,7]: The map $u \mapsto u\Omega$, from $\mathcal{C} \to \mathfrak{F}$, extends to a unitary operator $D: L^2(\mathcal{C}) \to \mathfrak{F}$, and the action of $\mathcal{C}$ on $L^2(\mathcal{C})$, given via D, is left multiplication; i.e., if $u \in \mathcal{C}$ and $v \in L^2(\mathcal{C})$, then $D^{-1}uDv = uv$.[8]

If T is an operator on $\mathcal{K}$, then by "tensoring"[1,3,9] one obtains an operator, denoted $\Gamma(T)$, in $\mathfrak{F}$. [$\Gamma(T)$ acts on $\mathfrak{F}_n$ like $T \otimes \cdots \otimes T$ (n factors)].

We are now in a position to state the following.

Theorem: With the notation as above, we suppose that $\|T\| < 1$. Then $D^{-1}\Gamma(T)D$ is bounded from $L^2(\mathcal{C}) \to L^p(\mathcal{C})$ for some $p > 2$.

Proof: We may, and shall, assume that $\mathcal{K} = L^2(\mathbb{R}, dk)$. This is possible because unitary equivalence between one-particle spaces induces an equivalence between the corresponding Fock spaces and the operators thereon. Since m is given by a vector state, it is evident that the L^p-norms are unchanged by such equivalences. We may further suppose that J is given by complex conjugation in $L^2(\mathbb{R}, dk)$.

Thus, for $f \in L^2(\mathbb{R}, dk)$, we can write

$$B(f) = \int f(k)[b^*(k) + b(k)]\, dk,$$

where $b^*(k)$ and $b(k)$ are the usual fermion creation and annihilation forms and satisfy the usual anticommutation relations.

Let $\Psi(k_1, \dots k_n) \in \mathfrak{F}_n$, be an n-particle vector of the form

$$\Psi = \sum_{i=1}^{N} b^*(f_1^i) \cdots b^*(f_n^i)\Omega$$

for some N, and $f_j^i \in L^2(\mathbb{R}, dk)$, $1 \le i \le N, 1 \le j \le n$.

Such vectors are dense in $\mathfrak{F}_n$. Ψ can be written as $\Psi = \sum_{i=1}^N : B(f_1^i) \dots B(f_n^i) : \Omega$, where : : denotes the Wick product. Since all the pairings are finite, we can "undo" the Wick product to obtain a polynomial in the fields $B(f_j^i)$. Thus, Ψ has the form $\Psi = W\Omega$, with $W \in \mathcal{C}$, and where W can be written as

$$W = \int w(k_1, \dots, k_n) : B(k_1) \cdots B(k_n) : dk_1 \cdots dk_n,$$

for some antisymmetric square-integrable function w. Now,

$$\|D^{-1}\Psi\|_2 = \|D^{-1}W\Omega\|_2 = \|D^{-1}WD1\|_2.$$

But $D^{-1}WD$ acts on $L^2(\mathcal{C})$ by left multiplication, and so $D^{-1}WD1 = D^{-1}WD$. Hence, writing $\hat{\Psi} = D^{-1}\Psi$,

$$\begin{aligned}\|\hat{\Psi}\|_2^2 &= \|D^{-1}WD\|_2^2 = (\Omega, W^*W\Omega)\\ &= (\Omega, \textstyle\int \overline{w}(l_1, \dots, l_n) w(k_1, \dots, k_n) : B(l_n) \cdots\\ &\quad \cdots B(l_1) :: B(k_1) \cdots B(k_n) : dl_1 \cdots dk_n \Omega)\\ &= n!\, \|w\|_2^2 \qquad \text{(by applying Wick's theorem).}\end{aligned}$$

Similarly, one obtains

$$\begin{aligned}\|\hat{\Psi}\|_{2j}^{2j} &= (\Omega, (W^*W)^j\Omega)\\ &= \sum \textstyle\int \overline{w}(k_1', \dots, k_n') \cdots \overline{w}(\cdots, k_{nj}')\\ &\quad \times w(k_{nj+1}', \cdots) \cdots w(\cdots, k_{2nj}')dk\end{aligned}$$

where $(k_1', \dots, k_{2nj}')$ is a permutation of $(k_1, k_2, \dots, k_{nj}, k_1, k_2, \dots, k_{nj})$ such that no two variables in a given w-factor coincide (this is just the statement that Wick pairings inside a Wick monomial do not contribute) and where the sum is over all such permutations. The proof now proceeds as in Ref. 1. That is, one notes that each summand is bounded by $\|w\|_2^{2j}$ (by repeated use of the Schwarz inequality) and that the number of allowed permutations is smaller than the number of all permutations, viz. $(2nj - 1)(2nj - 3) \cdots 3.1$.

Hence,

$$\|\hat{\Psi}\|_{2j}^{2j} \le (2nj-1)(2nj-3)\cdots 3.1\|w\|_2^{2j} = \frac{(2nj-1)\cdots 3.1}{(n!)^j}\|\hat{\Psi}\|_2^{2j} \le (2j)^{nj}\|\hat{\Psi}\|_2^{2j},$$

i.e., for Ψ of the assumed form, $\|\hat{\Psi}\|_{2j} \le (2j)^{n/2}\|\hat{\Psi}\|_2$. But, as already noted, such Ψ are dense in $\mathfrak{F}_n$, and these estimates imply that their image under D^{-1} is dense in $L^{2j}(\mathcal{C}) \cap D^{-1}\mathfrak{F}_n$ (in the L^{2j}-norm).[10] Therefore, the above inequalities remain valid for all $\Psi \in \mathfrak{F}_n$.

An application of the Riesz–Thorin–Kunze theorem[11] (interpolating between $j = 1$ and $j = 2$) yields

$$\|\hat{\Psi}\|_p \le 4^{(1-2/p)n}\|\hat{\Psi}\|_2, \quad \text{for } 2 \le p \le 4, \quad \Psi \in \mathfrak{F}_n.$$

For $u \in L^2(\mathcal{C})$, we can write $u = \sum_{n=0}^{\infty} u_n$, with $Du_n \in \mathfrak{F}_n$. Then

$$\begin{aligned}\|D^{-1}\Gamma(T)Du\|_p &\le \sum_n \|D^{-1}\Gamma(T)Du_n\|_p \\ &\le \sum_n 4^{(1-2/p)n}\|D^{-1}\Gamma(T)Du_n\|_2 \\ &\le \sum_n 4^{(1-2/p)n}\|\Gamma(T)Du_n\| \\ &\le \sum_n (4^{(1-2/p)}\|T\|)^n\|u\|_2.\end{aligned}$$

Since $\|T\| < 1$, the geometric series converges for $p - 2$ sufficiently small. QED

Corollary: $D^{-1}\Gamma(T^k)D$ is a contraction from $L^2(\mathcal{C}) \to L^4(\mathcal{C})$ provided k is sufficiently large.

Proof[12]*:* For $u \in L^2(\mathcal{C})$, write $u = \alpha\mathbf{1} + u'$, where $(Du', \Omega) = 0$. Then, if $\alpha = 0$, we have, as in the proof of the theorem,

$$\begin{aligned}\|D^{-1}\Gamma(T^k)Du\|_4 &\le \sum_n (2\|T\|^k)^n\|Du_n\| \\ &\le \sum_{n=1}^{\infty} (2\|T\|^k)^n\|Du'\| \\ &\le \delta\|u'\|_2 = \delta\|u\|_2\end{aligned}$$

for arbitrary $\delta > 0$, provided k is sufficiently large. We can now employ the technique of J. Glimm[13] to complete the proof.

Similarly, one can show that $D^{-1}\Gamma(T^k)D$ is bounded from $L^2(\mathcal{C}) \to L^p(\mathcal{C})$ for any $2 \le p < \infty$ for large enough k (depending on p).

If $T = e^{-X}$ is self-adjoint, with $X \ge c\mathbf{1}$, $c > 0$, then the corollary reduces to a theorem of Gross[3] except that we require that $ck \ge \log 2(1 + \sqrt{5})$ whereas Gross requires only $ck \ge (\log 3)/2$.

In conclusion, we remark that interpolation and duality[11] yield a similar boundedness result from $L^p(\mathcal{C}) \to L^q(\mathcal{C})$ for $1 < p, q < \infty$.

---

*Supported by the Swiss National Foundation.

[1] E. Nelson, J. Functional Anal. **12**, 211 (1973).
[2] See, e.g., Ref. 3.
[3] L. Gross, J. Functional Anal. **10**, 52 (1972).
[4] I. E. Segal, Ann. Math. **57**, 401 (1953); **58**, 595 (1953).
[5] For further details, we refer the reader to Refs. 3, 4, 11, and J. Dixmier, Bull. Soc. Math. France **81**, 9 (1953).
[6] See, e.g., I. E. Segal, Bull. Amer. Math. Soc. **71**, 419 (1965), and B. Simon and R. Høegh-Krohn, J. Functional Anal. **9**, 121 (1972).
[7] I. E. Segal, Ann. Math. **63**, 160 (1956).
[8] $\mathcal{C}$ is identified as a (dense) subset of $L^2(\mathcal{C})$.
[9] I. E. Segal, Trans. Amer. Math. Soc. **81**, 106 (1956).
[10] 2j can be replaced by any $p \geq 1$, since $\|\cdot\|_p$ increases with p.
[11] R. Kunze, Trans. Amer. Math. Soc. **89**, 519 (1958).
[12] We would like to thank Roger J. Plymen for pointing out an error in our original proof.
[13] See page 19 of J. Glimm, Commun. Math. Phys. **8**, 12 (1968).

The characteristic development of trapped surfaces*

Paul N. Demmie

University of Pittsburgh at Johnstown, Johnstown, Pennsylvania 15904

Allen I. Janis

University of Pittsburgh, Pittsburgh, Pennsylvania 15213

(Received 23 October 1972)

Conditions are found that are sufficient to insure the development of trapped surfaces in space-times whose metrics satisfy the Einstein field equations in vacuum or in the presence of a massless scalar field. These conditions involve a topological requirement that a certain two-surface be compact and inequalities that must be satisfied by certain pieces of the characteristic data determining these space-times. It is shown that a particular piece of data playing an important role in these inequalities is related to angular momentum.

1. INTRODUCTION

The concept of a trapped surface—a compact, spacelike two-surface having the property that all null geodesics meeting it orthogonally converge locally to the future—was introduced by Penrose[1] as a characterization of gravitational collapse that has proceeded beyond the point of no return. Under rather general conditions, if an object collapses sufficiently far that trapped surfaces develop in the region surrounding this object, then the spacetime containing the object must be singular.[2] Therefore, an important problem in the theory of gravitational collapse is the determination of the conditions under which trapped surfaces develop.

An important step toward solving this problem in the case of empty space-times was taken by Pajerski and Newman[3] (PN). Exploiting the property of the Schwarzschild space-time that the region containing trapped surfaces is separated from those that do not contain trapped surfaces by a nondiverging null hypersurface, they generalized the Schwarzschild space-time by considering a class of space-times each containing a nondiverging null hypersurface and determining the restrictions on the characteristic data for which trapped surfaces develop. The present work generalizes their work not only to Einstein-scalar space-times, but also to a larger class of empty space-times.

In Sec. 2 the formalism used in this investigation will be presented. This formalism was found to be particularly useful since it provides for a convenient characterization of trapped surfaces. In Sec. 3 the formalism presented in Sec. 2 will be used to determine all Einstein-scalar space-times containing a nondiverging null hypersurface. That these space-times are more general than those obtained in PN follows not only from their being Einstein-scalar space-times rather than empty space-times, but also from their dependence on an arbitrary function that was required to vanish in PN. Evidence suggesting that this function is related to angular momentum will be presented in Sec. 3. Also the characteristic data for these space-times will be determined, examined for restrictions placed on them in order that trapped surfaces develop, and discussed there. In Sec. 4 the results of Sec. 3 will be generalized and it will be established that there exist space-times more general than those containing a nondiverging null hypersurface that also contain trapped surfaces. Finally, in Sec. 5 the results of Secs. 3 and 4 will be summarized and discussed.

2. THE FORMALISM

The Newman–Penrose (NP) formalism[4] was found to be particularly useful for this investigation of the characteristic development of trapped surfaces. This formalism requires introducing into the tangent space at each point of the spacetime a null tetrad system,[5]

$$\left\{D = \frac{\ell^\mu \partial}{\partial x^\mu}, \quad \Delta = \frac{n^\mu \partial}{\partial x^\mu}, \quad \delta = \frac{m^\mu \partial}{\partial x^\mu}, \quad \bar{\delta} = \frac{\bar{m}^\mu \partial}{\partial x^\mu}\right\}, \tag{2.1}$$

consisting of two real null vectors, D and Δ, and a pair of complex null vectors, δ and $\bar{\delta}$, formed from two real, orthonormal, spacelike vectors, s_1 and s_2, as

$$\delta = (s_1 + is_2)/\sqrt{2},$$

and satisfying the orthonormality conditions

$$\begin{aligned} &\ell_\mu n^\mu = -m_\mu \bar{m}^\mu = 1, \\ &\ell_\mu \ell^\mu = n_\mu n^\mu = m_\mu m^\mu = \bar{m}_\mu \bar{m}^\mu = 0, \\ &\ell_\mu m^\mu = \ell_\mu n^\mu = n_\mu m^\mu = n_\mu \bar{m}^\mu = 0. \end{aligned} \tag{2.2}$$

The components $g^{\mu\nu}$ of the contravariant metric are found from (2.2) to be[6]

$$g^{\mu\nu} = 2\ell^{(\mu} n^{\nu)} - 2m^{(\mu}\bar{m}^{\nu)}. \tag{2.3}$$

The formalism then provides a set of partial differential equations equivalent to the Einstein field equations for the determination of the $g^{\mu\nu}$. These equations are given in terms of the five independent physical components of the Weyl tensor $C_{\mu\nu\rho\sigma}$[7]

$$\begin{aligned} \Psi_0 &= -C_{\mu\nu\rho\sigma}\ell^\mu m^\nu \ell^\rho m^\sigma, \\ \Psi_1 &= -C_{\mu\nu\rho\sigma}\ell^\mu n^\nu \ell^\rho m^\sigma, \\ \Psi_2 &= -C_{\mu\nu\rho\sigma}\bar{m}^\mu n^\nu \ell^\rho m^\sigma, \\ \Psi_3 &= -C_{\mu\nu\rho\sigma}\bar{m}^\mu n^\nu \ell^\rho n^\sigma, \\ \Psi_4 &= -C_{\mu\nu\rho\sigma}\bar{m}^\mu n^\nu \bar{m}^\rho n^\sigma, \end{aligned} \tag{2.4}$$

the six independent physical components of the trace-free Ricci tensor $R_{\mu\nu}$,

$$\begin{aligned} \Phi_{00} &= -\tfrac{1}{2}R_{\mu\nu}\ell^\mu \ell^\nu = \bar{\Phi}_{00}, \\ \Phi_{01} &= -\tfrac{1}{2}R_{\mu\nu}\ell^\mu m^\nu = \bar{\Phi}_{10}, \\ \Phi_{02} &= -\tfrac{1}{2}R_{\mu\nu}m^\mu m^\nu = \bar{\Phi}_{20}, \\ \Phi_{11} &= -\tfrac{1}{4}R_{\mu\nu}(\ell^\mu n^\nu + m^\mu \bar{m}^\nu) = \bar{\Phi}_{11}, \\ \Phi_{12} &= -\tfrac{1}{2}R_{\mu\nu}n^\mu m^\nu = \bar{\Phi}_{21}, \\ \Phi_{22} &= -\tfrac{1}{2}R_{\mu\nu}n^\mu n^\nu = \bar{\Phi}_{22}, \end{aligned} \tag{2.5}$$

the Riemann scalar R, and the twelve spin coefficients,[8]

$$\begin{aligned}
&\kappa = \ell_{\mu;\nu} m^\mu \ell^\nu, \qquad \nu = -n_{\mu;\nu}\bar{m}^\mu n^\nu,\\
&\rho = \ell_{\mu;\nu} m^\mu \bar{m}^\nu, \qquad \mu = -n_{\mu;\nu}\bar{m}^\mu m^\nu,\\
&\sigma = \ell_{\mu;\nu} m^\mu m^\nu, \qquad \lambda = -n_{\mu;\nu}\bar{m}^\mu \bar{m}^\nu,\\
&\tau = \ell_{\mu;\nu} m^\mu n^\nu, \qquad \pi = -n_{\mu;\nu}\bar{m}^\mu \ell^\nu,\\
&\alpha = \tfrac{1}{2}(\ell_{\mu;\nu} n^\mu \bar{m}^\nu - m_{\mu;\nu}\bar{m}^\mu \bar{m}^\nu),\\
&\beta = \tfrac{1}{2}(\ell_{\mu;\nu} n^\mu m^\nu - m_{\mu;\nu}\bar{m}^\mu m^\nu),\\
&\gamma = \tfrac{1}{2}(\ell_{\mu;\nu} n^\mu n^\nu - m_{\mu;\nu}\bar{m}^\mu n^\nu),\\
&\epsilon = \tfrac{1}{2}(\ell_{\mu;\nu} n^\mu \ell^\nu - m_{\mu;\nu}\bar{m}^\mu \ell^\nu).
\end{aligned} \tag{2.6}$$

Before exhibiting the NP equations, a class of null tetrad systems appropriate for this investigation will be given. This class of null tetrad systems, which consists of those systems associated in a particular way with a class of null coordinate systems, will simplify these equations somewhat.

In a space–time it is possible to introduce at least locally a family of null hypersurfaces given by $u =$ const, where u is a scalar function satisfying

$$g^{\mu\nu} u_{,\mu} u_{,\nu} = 0. \tag{2.7}$$

Let the x^0 coordinate be u. Then choose the first member of the null tetrad system, D, so that

$$\ell_\mu = u_{,\mu}. \tag{2.8}$$

That D is null and hypersurface orthogonal follows from (2.7) and (2.8), respectively. Therefore D is tangent to a family of null geodesics. Let the x^1 coordinate be r where r is an affine parameter for D. Then

$$D = \frac{dx^\mu}{dr}\frac{\partial}{\partial x^\mu} = \frac{\partial}{\partial r}.$$

These properties of D imply that

$$\kappa = 0 = (\epsilon + \bar{\epsilon}), \qquad \rho = \bar{\rho}, \qquad \tau = \bar{\alpha} + \beta. \tag{2.9}$$

Finally let the x^m coordinates label the null geodesics in the $u =$ constant hypersurface. In this manner a null coordinate system[9]

$$\{u, r, x^m\} \tag{2.10}$$

and an associated null vector D are given locally in a space–time. The most general null tetrad system containing D and preserving the orthonormality conditions (2.2) is $\{D, \Delta, \delta, \bar{\delta}\}$, where

$$D = \partial/\partial r, \tag{2.11a}$$

$$\Delta = \frac{\partial}{\partial u} + \frac{U\partial}{\partial r} + \frac{X^m\partial}{\partial x^m}, \tag{2.11b}$$

$$\delta = \frac{\omega\partial}{\partial r} + \frac{\xi^m\partial}{\partial x^m}. \tag{2.11c}$$

The null coordinate system (2.10) and associated null tetrad system (2.11) are not unique. (2.10) could be replaced by any coordinate system in which the coordinate conditions

$$g^{0\mu} = \delta_1^\mu \tag{2.12}$$

hold. Also (2.11) could be replaced by any null tetrad system related to it by null rotations about D,

$$\begin{aligned}
&\tilde{D} = D,\\
&\tilde{\Delta} = \Delta + a\bar{\delta} + \bar{a}\delta + a\bar{a}D,\\
&\tilde{\delta} = \delta + aD,
\end{aligned} \tag{2.13}$$

where a is complex, and/or spatial rotations,

$$\tilde{D} = D, \qquad \tilde{\Delta} = \Delta, \qquad \text{and } \tilde{\delta} = e^{iC}\delta, \tag{2.14}$$

where C is real, since the conditions (2.2) and (2.8) are preserved by these transformations.

Much of the ambiguity in the null tetrad system can be removed by choosing (2.11c) in a particular way. Under (2.13) with $a = -\omega$, δ becomes

$$\tilde{\delta} = \delta + aD = \xi^m \frac{\partial}{\partial x^m}.$$

Therefore, the ambiguity between (2.11) and any null tetrad system related to it by null rotations about D can be eliminated by choosing

$$\omega = 0 \tag{2.15}$$

in (2.11c). Furthermore, under (2.14), $(\epsilon - \bar{\epsilon})$ becomes

$$(\epsilon \tilde{-} \bar{\epsilon}) = (\epsilon - \bar{\epsilon}) - iDC.$$

Therefore, the ambiguity between (2.11) and any null tetrad system related to it by spatial rotations can be reduced to those rotations with $C = C(u, x^m)$ by choosing ϵ real. This choice and (2.9) imply that

$$\epsilon = 0. \tag{2.16}$$

With (2.15) and (2.16) adopted, it has been established that:

In a spacetime there exists a class of null coordinate systems such that any one of these coordinate systems

$$\{u, r, x^m\} \tag{2.17}$$

satisfies the coordinate conditions (2.12) and has associated with it a particular null tetrad system $\{D, \Delta, \delta, \bar{\delta}\}$ with

$$D = \frac{\partial}{\partial r}, \tag{2.18a}$$

$$\Delta = \frac{\partial}{\partial u} + U\frac{\partial}{\partial r} + X^m\frac{\partial}{\partial x^m}, \tag{2.18b}$$

$$\delta = \xi^m \frac{\partial}{\partial x^m}, \tag{2.18c}$$

which satisfies the orthonormality conditions (2.2), is unique up to spatial rotations (2.14) with $C = C(u, x^m)$, and has spin coefficients satisfying (2.9) and (2.16).

From (2.1), (2.3), and (2.18) the components $g^{\mu\nu}$ of the contravariant metric are

$$\begin{aligned}
&g^{0\mu} = g_1^\mu, \qquad g^{11} = 2U, \qquad g^{1m} = X^m,\\
&g^{mn} = -(\xi^m\bar{\xi}^n + \bar{\xi}^m\xi^n).
\end{aligned} \tag{2.19}$$

With the null tetrad system (2.18) chosen, the NP equations will now be exhibited[10] in three classes: the commutator equations applied to the coordinates, the spin coefficient equations, and the spin-coefficient form of the Bianchi identities. The commutator equations applied to the coordinates imply that the spin coefficients τ, π, and μ satisfy

$$\tau = \bar{\pi}, \quad \mu = \bar{\mu} \tag{2.20}$$

and the metric variables U, X^m, and ξ^m satisfy

$$D\xi^m = \rho\xi^m + \sigma\bar{\xi}^m, \tag{2.21a}$$

$$DX^m = 2(\bar{\tau}\xi^m + \tau\bar{\xi}^m), \tag{2.21b}$$

$$DU = -(\gamma + \bar{\gamma}), \tag{2.21c}$$

$$\delta\bar{\xi}^m - \bar{\delta}\xi^m = (\bar{\alpha} - \beta)\bar{\xi}^m + (\bar{\beta} - \alpha)\xi^m, \tag{2.21d}$$

$$\Delta\xi^m - \delta X^m = -(\mu + \bar{\gamma} - \gamma)\xi^m - \bar{\lambda}\bar{\xi}^m, \tag{2.21e}$$

$$\nu = -\bar{\delta}U. \tag{2.21f}$$

With (2.9), (2.16), and (2.20) satisfied, the spin-coefficient equations are

$$D\rho = \rho^2 + \sigma\bar{\sigma} + \Phi_{00}, \tag{2.22a}$$

$$D\sigma = 2\rho\sigma + \Psi_0, \tag{2.22b}$$

$$D\tau = 2\rho\tau + 2\sigma\bar{\tau} + \Psi_1 + \Phi_{01}, \tag{2.22c}$$

$$D\alpha = (\alpha + \bar{\tau})\rho + \beta\bar{\sigma} + \Phi_{10}, \tag{2.22d}$$

$$D\beta = \rho\beta + (\alpha + \bar{\tau})\sigma + \Psi_1, \tag{2.22e}$$

$$D\gamma = 2\tau\alpha + 2\bar{\tau}\beta + \tau\bar{\tau} + \Psi_2 - \tfrac{1}{24}R + \Phi_{11}, \tag{2.22f}$$

$$D\lambda - \bar{\delta}\bar{\tau} = \rho\lambda + \bar{\sigma}\mu + \bar{\tau}^2 + (\alpha - \bar{\beta})\bar{\tau} + \Phi_{20}, \tag{2.22g}$$

$$D\mu - \delta\bar{\tau} = \rho\mu + \sigma\lambda + \tau\bar{\tau} - (\bar{\alpha} - \beta)\bar{\tau} + \Psi_2 + \tfrac{1}{12}R, \tag{2.22h}$$

$$\delta\rho - \bar{\delta}\sigma = \rho\tau - (3\alpha - \bar{\beta})\sigma - \Psi_1 + \Phi_{01}, \tag{2.22i}$$

$$\delta\alpha - \bar{\delta}\beta = \rho\mu - \sigma\lambda + \alpha\bar{\alpha} + \beta\bar{\beta} - 2\alpha\beta - \Psi_2 + \tfrac{1}{24}R + \Phi_{11}, \tag{2.22j}$$

$$\delta\lambda - \bar{\delta}\mu = \mu\bar{\tau} + (\bar{\alpha} - 3\beta)\lambda - \Psi_3 + \Phi_{21}, \tag{2.22k}$$

$$\Delta\tau - D\bar{\nu} = -2\mu\tau - 2\overline{\tau\lambda} + (\gamma - \bar{\gamma})\tau - \bar{\Psi}_3 - \Phi_{12}, \tag{2.22l}$$

$$\Delta\lambda - \bar{\delta}\nu = (\bar{\gamma} - 3\gamma - 2\mu)\lambda + (3\alpha + \bar{\beta})\nu - \Psi_4, \tag{2.22m}$$

$$\Delta\mu - \delta\nu = -\mu^2 - \lambda\bar{\lambda} - (\gamma + \bar{\gamma})\mu + 2\beta\nu + \overline{\nu\tau} - \Phi_{22}, \tag{2.22n}$$

$$\Delta\beta - \delta\gamma = -\mu\tau + \sigma\nu + (\gamma - \bar{\gamma} - \mu)\beta - \alpha\bar{\lambda} - \Phi_{12}, \tag{2.22o}$$

$$\Delta\sigma - \delta\tau = -\mu\sigma - \rho\bar{\lambda} - 2\beta\tau + (3\gamma - \bar{\gamma})\sigma - \Phi_{02}, \tag{2.22p}$$

$$\Delta\rho - \bar{\delta}\tau = -\mu\rho - \sigma\lambda - 2\alpha\tau + (\gamma - \bar{\gamma})\rho - \Psi_2 - \tfrac{1}{12}R, \tag{2.22q}$$

$$\Delta\alpha - \bar{\delta}\gamma = \rho\nu - (\tau + \beta)\lambda + (\bar{\gamma} - \gamma - \mu)\alpha - \Psi_3; \tag{2.22r}$$

and the spin-coefficient form of the Bianchi identities are

$$D\Psi_1 - \bar{\delta}\Psi_0 = 4\rho\Psi_1 - (4\alpha - \bar{\tau})\Psi_0 + D\Phi_{01} - \delta\Phi_{00} - 2\rho\Phi_{01} + \tau\Phi_{00} - 2\sigma\Phi_{10}, \tag{2.23a}$$

$$\Delta\Psi_0 - \delta\Psi_1 = (4\gamma - \mu)\Psi_0 - 2(2\tau + \beta)\Psi_1 + 3\sigma\Psi_2 - D\Phi_{02} + \delta\Phi_{01} + \rho\Phi_{02} + 2\bar{\alpha}\Phi_{01} - \bar{\lambda}\Phi_{00} + 2\sigma\Phi_{11}, \tag{2.23b}$$

$$\begin{aligned} D\Psi_2 - \bar{\delta}\Psi_1 = {} & 3\rho\Psi_2 + 2\bar{\beta}\Psi_1 - \lambda\Psi_0 + \tfrac{2}{3}D\Phi_{11} \\ & - \tfrac{2}{3}\delta\Phi_{10} + \tfrac{1}{3}\bar{\delta}\Phi_{01} - \tfrac{1}{3}\Delta\Phi_{00} - \tfrac{2}{3}\rho\Phi_{11} - \tfrac{2}{3}(2\bar{\tau} + \alpha)\Phi_{01} \\ & - \tfrac{4}{3}\beta\Phi_{10} - \tfrac{2}{3}\sigma\Phi_{20} + \tfrac{1}{3}\bar{\sigma}\Phi_{02} + \tfrac{1}{3}(\mu + 2\gamma + 2\bar{\gamma})\Phi_{00}, \end{aligned} \tag{2.23c}$$

$$\begin{aligned} \Delta\Psi_1 - \delta\Psi_2 = {} & \nu\Psi_0 + 2(\gamma - \mu)\Psi_1 - 3\tau\Psi_2 + 2\sigma\Psi_3 \\ & - \tfrac{2}{3}D\Phi_{12} + \tfrac{2}{3}\delta\Phi_{11} - \tfrac{1}{3}\bar{\delta}\Phi_{02} + \tfrac{1}{3}\Delta\Phi_{01} - \tfrac{1}{3}\bar{\nu}\Phi_{00} \\ & - \tfrac{2}{3}\gamma\Phi_{01} - \tfrac{2}{3}\bar{\lambda}\Phi_{10} + 2\tau\Phi_{11} + \tfrac{1}{3}(4\alpha + \bar{\tau})\Phi_{02} \\ & + \tfrac{2}{3}\sigma\Phi_{21}, \end{aligned} \tag{2.23d}$$

$$\begin{aligned} D\Psi_3 - \bar{\delta}\Psi_2 = {} & 2\rho\Psi_3 + 3\bar{\tau}\Psi_2 - 2\lambda\Psi_1 + \tfrac{1}{3}D\Phi_{21} \\ & - \tfrac{1}{3}\delta\Phi_{20} + \tfrac{2}{3}\bar{\delta}\Phi_{11} - \tfrac{2}{3}\Delta\Phi_{10} + \tfrac{2}{3}\nu\Phi_{00} - \tfrac{2}{3}\lambda\Phi_{01} \\ & - \tfrac{4}{3}\bar{\gamma}\Phi_{10} - 2\bar{\tau}\Phi_{11} - \tfrac{1}{3}(4\beta + \tau)\Phi_{20} + \tfrac{2}{3}\bar{\sigma}\Phi_{12}, \end{aligned} \tag{2.23e}$$

$$\begin{aligned} \Delta\Psi_2 - \delta\Psi_3 = {} & 2\nu\Psi_1 - 3\mu\Psi_2 - 2\bar{\alpha}\Psi_3 + \sigma\Psi_4 \\ & - \tfrac{1}{3}D\Phi_{22} + \tfrac{1}{3}\delta\Phi_{21} - \tfrac{2}{3}\bar{\delta}\Phi_{12} + \tfrac{2}{3}\Delta\Phi_{11} - \tfrac{2}{3}\nu\Phi_{01} \\ & - \tfrac{2}{3}\bar{\nu}\Phi_{10} + \tfrac{2}{3}\mu\Phi_{11} + \tfrac{2}{3}\lambda\Phi_{02} - \tfrac{1}{3}\bar{\lambda}\Phi_{20} + \tfrac{4}{3}\alpha\Phi_{12} \\ & + \tfrac{2}{3}(\beta + 2\tau)\Phi_{21} - \tfrac{1}{3}\rho\Phi_{22}, \end{aligned} \tag{2.23f}$$

$$\begin{aligned} D\Psi_4 - \bar{\delta}\Psi_3 = {} & \rho\Psi_4 + 2(\alpha + 2\bar{\tau})\Psi_3 - 3\lambda\Psi_2 + \bar{\delta}\Phi_{21} \\ & - \Delta\Phi_{20} + 2\nu\Phi_{10} - 2\lambda\Phi_{11} - (2\gamma - 2\bar{\gamma} + \mu)\Phi_{20} \\ & - 2(\bar{\tau} - \alpha)\Phi_{21} + \bar{\sigma}\Phi_{22}, \end{aligned} \tag{2.23g}$$

$$\begin{aligned} \Delta\Psi_3 - \delta\Psi_4 = {} & 3\nu\Psi_2 - 2(\gamma + 2\mu)\Psi_3 + (4\beta - \tau)\Psi_4 \\ & - \bar{\delta}\Phi_{22} + \Delta\Phi_{21} - 2\nu\Phi_{11} - \bar{\nu}\Phi_{20} + 2\lambda\Phi_{12} \\ & + 2(\gamma + \mu)\Phi_{21} - \bar{\tau}\Phi_{22}, \end{aligned} \tag{2.23h}$$

$$\begin{aligned} & D\Phi_{11} - \delta\Phi_{10} - \bar{\delta}\Phi_{01} + \Delta\Phi_{00} + \tfrac{1}{8}DR \\ & = 2(\gamma + \bar{\gamma} - \mu)\Phi_{00} - (2\alpha + \bar{\tau})\Phi_{01} - (2\bar{\alpha} + \tau)\Phi_{10} \\ & \quad + 4\rho\Phi_{11} + \bar{\sigma}\Phi_{02} + \sigma\Phi_{20}, \end{aligned} \tag{2.23i}$$

$$\begin{aligned} & D\Phi_{12} - \delta\Phi_{11} - \bar{\delta}\Phi_{02} + \Delta\Phi_{01} + \tfrac{1}{8}\delta R \\ & = (2\gamma - 3\mu)\Phi_{01} + \bar{\nu}\Phi_{00} - \bar{\lambda}\Phi_{10} \\ & \quad + 2(\bar{\beta} - \alpha)\Phi_{02} + 3\rho\Phi_{12} + \sigma\Phi_{21}, \end{aligned} \tag{2.23j}$$

$$\begin{aligned} & D\Phi_{22} - \delta\Phi_{21} - \bar{\delta}\Phi_{12} + \Delta\Phi_{11} + \tfrac{1}{8}\Delta R \\ & = \nu\Phi_{01} + \bar{\nu}\Phi_{10} - 4\mu\Phi_{11} - \lambda\Phi_{02} \\ & \quad - \bar{\lambda}\Phi_{20} + (\bar{\tau} + 2\bar{\beta})\Phi_{12} + (\tau + 2\beta)\Phi_{21} + 2\rho\Phi_{22}. \end{aligned} \tag{2.23k}$$

The class of space-times investigated were the Einstein-scalar space-times. An Einstein-scalar space-time is a space-time whose Einstein tensor $G_{\mu\nu}$ satisfies the Einstein field equations[11]

$$G_{\mu\nu} = -T_{\mu\nu} \tag{2.24}$$

with

$$T_{\mu\nu} = \phi,_\mu\phi,_\nu - \tfrac{1}{2}(g^{\rho\sigma}\phi,_\rho\phi,_\sigma)g_{\mu\nu}, \tag{2.25}$$

where ϕ is a massless scalar field satisfying the equation

$$(g^{\mu\nu}\phi,_{\nu});_{\mu} = 0. \tag{2.26}$$

From (2.24) and (2.25) it is easily seen that the Ricci tensor for an Einstein-scalar space-time is

$$R_{\mu\nu} = -\phi,_{\mu}\phi,_{\nu}. \tag{2.27}$$

The scalar field equation (2.26) can be expressed in spin-coefficient form as

$$\eta^{ab}(\phi,_{ab} - \gamma^{c}_{ab}\phi,_{c}) = 0, \tag{2.28}$$

where $\phi,_a$ and $\phi,_{ab}$ are the physical components of $\phi,_\mu$ and $(\phi,_a);_\mu$, respectively,

$$(\eta_{ab}) = (\eta^{ab}) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \tag{2.29}$$

is the null form of the Minkowski metric, and the $\gamma_{abc} = z_{a\mu;\nu}z^{\mu}_{b}z^{\nu}_{c}$ with $(z^{\mu}_{a}) = (\ell^{\mu}, n^{\mu}, m^{\mu}, \overline{m}^{\mu})$ are the Ricci rotation coefficients. With (2.1), (2.6), (2.9), and (2.16), Eq. (2.28) becomes

$$(D\Delta + \mu D - \rho\Delta - \bar{\delta}\delta - 2\bar{\beta}\delta - \tau\bar{\delta})\phi = 0. \tag{2.30}$$

Not only was the spin-coefficient form of the scalar field equation used in this investigation, but also the spin-coefficient form of the optical scalars.[12] The spin-coefficient form of the divergence, rotation, and shear of a vector $k = k^{\mu}\dfrac{\partial}{\partial x^{\mu}}$ can be shown to be[13]

$$d(k) = \tfrac{1}{2}\eta^{ab}(k_{a,b} - \gamma_{acb}k^{c}), \tag{2.31a}$$

$$r(k) = [\tfrac{1}{2}(k_{[a,b]} - \gamma_{[a|c|b]}k^{c})(k^{a,b} - \gamma^{adb}k_{d})]^{1/2}, \tag{2.31b}$$

$$s(k) = [\tfrac{1}{2}(k_{(a,b)} - \gamma_{(a|c|b)}k^{c})(k^{a,b} - \gamma^{adb}k_{d}) - d^{2}]^{1/2}, \tag{2.31c}$$

respectively.

The formalism presented here provides not only for the determination of the metric variables, spin coefficients, physical Weyl tensor components, and scalar field of an Einstein-scalar spacetime, but also for a convenient characterization of trapped surfaces. To discover this characterization, consider the spacelike two-surface

$$S_{(u,r)} = \{(u, r, x^{m}): u \text{ and } r \text{ are constant}\}.$$

In order for $S_{(u,r)}$ to be a trapped surface, it must be compact and have the property that all null geodesics meeting it orthogonally coverge locally to the future. A vector tangent to one of these null geodesics must coincide with either D or Δ on $S_{(u,r)}$. That which coincides with D must be D, since a geodesic is specified uniquely by a point and a direction in the tangent space at that point. Therefore, from (2.31a) its divergence must be $-\rho$. However, the divergence of the null vector that coincides with Δ on $S_{(u,r)}$ is not so easily determined since

$$n^{\mu};_{\nu}n^{\nu} = -(\gamma + \bar{\gamma})n^{\mu} + \bar{\nu}m^{\mu} + \nu\overline{m}^{\mu}$$

implies that Δ is not everywhere tangent to a null geodesic unless $\nu = 0$. For the general, case where $\nu \neq 0$, the divergence of this vector can be calculated directly from (2.31a). Let k be this vector parameterized with an affine parameter. Then the components of k, k^{μ}, can be expressed as

$$k_{\mu} = S,_{\mu}$$

where S is a solution to $g^{\mu\nu}S,_{\mu}S,_{\nu} = 0$.

From this it follows that

$$k = (\Delta S)D + (DS)\Delta + (\delta S)\bar{\delta} + (\bar{\delta}S)\delta.$$

$k = \Delta$ on $S_{(u,r)}$ implies that $DS = 1$, $\delta S = 0 = \bar{\delta}S$ on $S_{(u,r)}$. Therefore, from (2.31a) the divergence of k on $S_{(u,r)}$, θ_0, is

$$\theta_0 = D\Delta S + \tfrac{1}{2}(\delta\bar{\delta}S + \bar{\delta}\delta S) + \tfrac{1}{2}(\mu + \bar{\mu}),$$

where the derivatives are evaluated on $S_{(u,r)}$. Now

$$D\Delta S = D(k_{\mu}n^{\mu}) = k_{\mu;\nu}n^{\mu}\ell^{\nu} + n_{\mu;\nu}k^{\mu}\ell^{\nu}.$$

Since $n^{\mu}{}_{;\nu}\ell^{\nu} = -\tau\overline{m}^{\mu} - \bar{\tau}m^{\mu}$, it follows that $n_{\mu;\nu}n^{\mu}\ell^{\nu} = 0$. Also since $k_{\mu} = S_{,\mu}$, it follows that $k_{\mu;\nu}n^{\mu}\ell^{\nu} = k_{\mu;\nu}\ell^{\mu}n^{\nu}$. This vanishes on $S_{(u,r)}$ since $k = \Delta$ there and k is tangent to a null geodesic parameterized with an affine parameter. Therefore, $D\Delta S = 0$ on $S_{(u,r)}$. Also $\delta\bar{\delta}S = 0 = \bar{\delta}\delta S$ on $S_{(u,r)}$ since δ and $\bar{\delta}$ are differentiations in $S_{(u,r)}$ over which $\bar{\delta}S$ and δS are constant. $D\Delta S = 0 = \delta\bar{\delta}S = \bar{\delta}\delta S$ on $S_{(u,r)}$ and $\mu = \bar{\mu}$ imply that $\theta_0 = \mu$. With the determination of θ_0, it has been established that:

The spacelike two-surface

$$S_{(u,r)} = \{(u, r, x^{m}): u \text{ and } r \text{ are constant}\}$$

is a trapped surface if and only if it is compact and everywhere on it the spin coefficients ρ and μ satisfy

$$\rho > 0 \quad \text{and} \quad \mu < 0. \tag{2.32}$$

3. SPACE-TIMES CONTAINING A NONDIVERGING NULL HYPERSURFACE

The possibility that there exist nonspherically symmetric Einstein-scalar spacetimes containing both a nondiverging null hypersurface and trapped surfaces was investigated.[14] This investigation began by obtaining the metric variables, spin coefficients, physical Weyl tensor components, and scalar field of an Einstein-scalar space-time containing a nondiverging null hypersurface. Then the characteristic data for this space-time were determined and examined for restrictions placed on them for which trapped surfaces develop.

The problem of determining all Einstein-scalar space-times containing a nondiverging null hypersurface was solved using the formalism presented in Sec. 2. For any one of these space-times with a particular null coordinate system (2.17) and associated null tetrad system (2.18) introduced in it, the main conditions that were adopted are that the nondiverging null hypersurface is given by $u = 0$ and the metric variables, spin coefficients, physical Weyl tensor components, and scalar field are analytic functions of (u, r, x^{m}) in the region $\{(u, r, x^{m})\}$. Subject to these conditions, the NP equations involving D and the scalar field equation (2.30) yielded these quantities in terms of a set of functions of (x^{m}) given on the spacelike two-surface

$$S_0 = \{(u, r, x^{m}): u = 0 = r\}.$$

Some of these functions had conditions placed on them in order that the coordinate and tetrad systems chosen initially be specified up to scale transformations.

$$\tilde{u} = Au, \quad \tilde{r} = A^{-1}r, \quad \tilde{x}^m = x^m, \qquad\qquad (3.1)$$
$$\tilde{D} = A^{-1}D, \quad \tilde{\Delta} = A\Delta, \quad \tilde{\delta} = \delta;$$

whereas others were determined by the remaining NP equations. The functions that remained constitute the characteristic data for this space-time.

The condition that the nondiverging null hypersurface is given by $u = 0$ implies that

$$\rho(0, r, x^m) = 0. \qquad (3.2)$$

This and Eq. (2.22a) imply that

$$\sigma(0, r, x^m) = 0 \quad \text{and} \quad \phi(0, r, x^m) = \phi_0(x^m),$$

where $\phi_0(x^m)$ is an arbitrary function of (x^m). These and Eq. (2.22b) imply that

$$\Psi_0(0, r, x^m) = 0.$$

$\rho = 0 = \sigma$ on $u = 0$ and Eq. (2.21a) imply that

$$\xi^m(0, r, x^n) = \xi_0^m(x^n),$$

where the ξ_0^m are arbitrary functions of (x^n).

The presentation of subsequent results will be simplified considerably by further specification of (ξ_0^m). Consider the contravariant metric induced on S_0,

$$-(\xi_0^m \bar{\xi}_0^n + \bar{\xi}_0^m \xi_0^n) \frac{\partial}{\partial x^m} \otimes \frac{\partial}{\partial x^n}.$$

Since this is a two-metric, it can be made conformally flat by some transformation[15]

$$\tilde{u} = u, \quad \tilde{r} = r, \quad \tilde{x}^m = \tilde{x}^m(x^n).$$

Therefore, it may be taken to be

$$-P\bar{P} \frac{\partial}{\partial z} \otimes \frac{\partial}{\partial \bar{z}}, \qquad (3.3)$$

where $z = (x^2 - ix^3)/\sqrt{2}$ and P is an arbitrary function of (x^m). With this choice (ξ_0^m) becomes

$$(\xi_0^m) = (P, iP)/\sqrt{2}.$$

Now consider the spatial rotations (2.14) with

$$C(u, x^m) = C_0(x^m) + C_1(x^m)u + \cdots.$$

Under this transformation δ becomes

$$\tilde{\delta} = [\exp(iC)]\delta = \frac{1}{\sqrt{2}} [\exp(iC_0)] P \frac{\partial}{\partial z} + \cdots.$$

Therefore, if P is required to be real, then δ is specified up to spatial rotations with $C_0 = 0$. Once the contravariant metric induced on S_0 is chosen to be (3.3) with P real, it is convenient to introduce the differential operators $\eth$ and $\bar{\eth}$[16] as

$$\eth\eta = P^{1-s} \frac{\partial}{\partial z} (P^s \eta) \quad \text{and} \quad \bar{\eth}\eta = P^{1+s} \frac{\partial}{\partial \bar{z}} (P^{-s}\eta),$$

where η is a quantity of spin weight s, that is, a quantity that transforms as

$$\tilde{\eta} = [\exp(is\psi)]\eta$$

under the transformation

$$\tilde{\delta}_0 = [\exp(i\psi)]\delta_0 \quad \text{where } \delta_0 = \frac{1}{\sqrt{2}} P \frac{\partial}{\partial z}.$$

The metric variables, spin coefficients, physical Weyl tensor components, and scalar field obtained thus far on $u = 0$ were calculated using (3.2) and the NP equations expressed in a particular null coordinate system (2.17) and associated null tetrad system (2.18). Similarly, one can obtain on $u = 0$: Ψ_1 from (2.22i), τ from (2.22c), α and β from (2.22d), (2.22e), (2.21d), and (2.9), X^m from (2.21b), Ψ_2 from (2.22j), γ from (2.22f), U from (2.21c), ν from (2.20), λ from (22.22g), and μ from (2.22h). However, (2.17) and (2.18), each being but one of the class given in Sec. 2, are not uniquely specified. By considering the coordinate transformation between (2.17) and any coordinate system $\{u, r, x^m\}$ satisfying the coordinate conditions (2.12), where

$$\tilde{u} = A_0(r, x^m) + A_1(r, x^m)u + \cdots,$$
$$\tilde{r} = B_0(r, x^m) + B_1(r, x^m)u + \cdots,$$
$$\tilde{x} = Y_0^m(r, x^n) + Y_1^m(r, x^n)u + \cdots,$$

and the tetrad transformation between (2.18) and any tetrad system $\{\tilde{D}, \tilde{\Delta}, \tilde{\delta}, \bar{\tilde{\delta}}\}$ related to (2.18) by spatial rotations (2.14), where

$$C(u, x^m) = C_1(x^m)u + C_2(x^m)u^2 + \cdots,$$

and then considering the effects of these coordinate and tetrad transformations on certain metric variables and spin coefficients, it can be established that[14]:

The null coordinate system (2.17) and associated null tetrad system (2.18) can be specified up to scale transformations (3.1) by imposing the conditions

$$[\xi^m(0, 0, x^n)] = (P, iP)/\sqrt{2}, \qquad (3.4a)$$

where P is real,

$$\mu(0, 0, x^m) = 0, \qquad (3.4b)$$

$$X^m(u, 0, x^n) = 0, \qquad (3.4c)$$

$$\tau_0(x^m) = \tau(0, 0, x^m) = i\eth g, \qquad (3.4d)$$

where g is real and an analytic function of $(z, \bar{z})$ with spin weight zero,

$$U(u, 0, x^m) = 0, \qquad (3.4e)$$

and

$$\gamma(u, 0, x^m) = 0. \qquad (3.4f)$$

After Ψ_1 through μ are obtained on $u = 0$ as previously indicated, only Ψ_3 and Ψ_4 remain to be determined there. They can be calculated from Eqs. (2.22k) and (2.23g), respectively, once $\Delta\phi(0, r, x^m)$ is calculated from Eq. (2.30). After this is done, the metric variables, spin coefficients, physical Weyl tensor components, and scalar field will be known on $u = 0$. Once they are known on $u = 0$, they can be determined off $u = 0$ in a straightforward but tedious manner from the NP equations and the scalar field equation. By doing this, subject to the conditions (3.4), it can be established that:

An Einstein-scalar space-time containing a nondiverging null hypersurface has over the region $\{(u, r, x^m)\}$ metric variables

$$U = (\tfrac{1}{2}K - 3\tau_0\bar{\tau}_0 - \tfrac{1}{4}\eth\phi_0\bar{\eth}\phi_0)r^2 + (\dot{U}_2 r^2 + \dot{U}_3 r^3)u + \cdots, \tag{3.5a}$$

$$X^m = 2(\bar{\tau}_0\xi_0^m + \tau_0\bar{\xi}_0^m)r + (\dot{X}_1^m r + \dot{X}_2^m r^2)u + \cdots, \tag{3.5b}$$

$$\xi^m = \xi_0^m + \{-\bar{\lambda}_0\bar{\xi}_0^m + [(\tfrac{1}{2}K - \tau_0\bar{\tau}_0 - \tfrac{1}{4}\eth\phi_0\bar{\eth}\phi_0)\xi_0^m + (\eth\tau_0/\sqrt{2} - \tau_0^2 - \tfrac{1}{4}(\eth\phi_0)^2)\bar{\xi}_0^m]r\}u + \cdots; \tag{3.5c}$$

nonzero spin coefficients

$$\rho = (\tfrac{1}{2}K - \tau_0\bar{\tau}_0 - \tfrac{1}{4}\eth\phi_0\bar{\eth}\phi_0)u + \cdots, \tag{3.6a}$$

$$\sigma = (\eth\tau_0/\sqrt{2} - \tau_0^2 - \tfrac{1}{4}(\eth\phi_0)^2)u + \cdots, \tag{3.6b}$$

$$\mu = -(\tfrac{1}{2}K - \tau_0\bar{\tau}_0 - \tfrac{1}{4}\eth\phi_0\bar{\eth}\phi_0)r + \{(-\lambda_0\bar{\lambda}_0 - \tfrac{1}{2}\dot{\phi}_0^2) + \dot{\mu}_1 r + \dot{\mu}_2 r^2\}u + \cdots, \tag{3.6c}$$

$$\lambda = \{\lambda_0 + (\bar{\tau}_0^2 + \bar{\eth}\bar{\tau}_0/\sqrt{2} + \tfrac{1}{4}(\bar{\eth}\phi_0)^2)r\} + (\dot{\lambda}_0 + \dot{\lambda}_1 r + \dot{\lambda}_2 r^2)u + \cdots, \tag{3.6d}$$

$$\tau = \tau_0 + (\dot{\tau}_0 + \dot{\tau}_1 r)u + \cdots, \tag{3.6e}$$

$$\nu = (1/\sqrt{2})\bar{\eth}(-\tfrac{1}{2}K + 3\tau_0\bar{\tau}_0 + \tfrac{1}{4}\eth\phi_0\bar{\eth}\phi_0)r^2 + (\dot{\nu}_2 r^2 + \dot{\nu}_3 r^3)u + \cdots, \tag{3.6f}$$

$$\alpha = (\alpha_0 + \tfrac{1}{2}\bar{\tau}_0) + (\dot{\alpha}_0 + \dot{\alpha}_1 r)u + \cdots, \tag{3.6g}$$

$$\beta = (-\bar{\alpha}_0 + \tfrac{1}{2}\tau_0) + (\dot{\beta}_0 + \dot{\beta}_1 r)u + \cdots, \tag{3.6h}$$

$$\gamma = (-\tfrac{1}{2}K + 3\tau_0\bar{\tau}_0 + \bar{\eth}\tau_0/\sqrt{2} + \tfrac{1}{4}\eth\phi_0\bar{\eth}\phi_0 + 2\alpha_0\tau_0 - 2\bar{\alpha}_0\bar{\tau}_0)r + (\dot{\gamma}_1 r + \dot{\gamma}_2 r^2)u + \cdots; \tag{3.6i}$$

physical Weyl tensor components

$$\Psi_0 = \ddot{\Psi}_0 u^2 + \cdots, \tag{3.7a}$$

$$\Psi_1 = (-\eth K/2\sqrt{2} + \tfrac{3}{2}\tau_0 K + \tfrac{1}{2}\eth\bar{\eth}\tau_0 - 3\tau_0\bar{\eth}\tau_0/\sqrt{2} + \bar{\eth}\phi_0\eth^2\phi_0/4\sqrt{2} - \tfrac{1}{2}\tau_0\eth\phi_0\bar{\eth}\phi_0)u + \cdots, \tag{3.7b}$$

$$\Psi_2 = (-\tfrac{1}{2}K + \bar{\eth}\tau_0/\sqrt{2} + \tfrac{1}{6}\eth\phi_0\bar{\eth}\phi_0) + (\dot{\Psi}_2^0 + \dot{\Psi}_2^1 r)u + \cdots, \tag{3.7c}$$

$$\Psi_3 = \{(-\eth\lambda_0/\sqrt{2} - \tau_0\lambda_0 + \dot{\phi}_0\bar{\eth}\phi_0/2\sqrt{2}) + (-\bar{\eth}K/2\sqrt{2} - \tfrac{3}{2}\bar{\tau}_0K + \tfrac{1}{2}\bar{\eth}^2\tau_0 + 3\bar{\tau}_0\bar{\eth}\tau_0/\sqrt{2} + \eth\phi_0\bar{\eth}^2\phi_0/4\sqrt{2} + \tfrac{1}{2}\bar{\tau}_0\eth\phi_0\bar{\eth}\phi_0)r\} + (\dot{\Psi}_3^0 + \dot{\Psi}_3^1 r + \dot{\Psi}_3^2 r^2)u + \cdots, \tag{3.7d}$$

$$\Psi_4 = (\Psi_4^0 + \Psi_4^1 r + \Psi_4^2 r^2) + (\dot{\Psi}_4^0 + \dot{\Psi}_4^1 r + \dot{\Psi}_4^2 r^2 + \dot{\Psi}_4^3 r^3)u + \cdots; \tag{3.7e}$$

and scalar field

$$\phi = \phi_0 + \{\dot{\phi}_0 + (\tfrac{1}{2}\bar{\eth}\eth\phi_0 - \bar{\tau}_0\eth\phi_0/\sqrt{2} - \tau_0\bar{\eth}\phi_0/\sqrt{2})r\}u + \cdots; \tag{3.8}$$

where $K = \bar{\eth}\eth \ln P$, $\alpha_0 = (1/2\sqrt{2})\eth \ln P$, and

$\dot{U}_2, \dot{U}_3, \dot{X}_1^m, \dot{X}_2^m, \dot{\mu}_1, \dot{\mu}_2, \dot{\lambda}_0, \dot{\lambda}_1, \dot{\lambda}_2, \dot{\tau}_0, \dot{\tau}_1, \dot{\nu}_2, \dot{\nu}_3, \dot{\alpha}_0, \dot{\alpha}_1, \dot{\beta}_1, \dot{\beta}_2, \dot{\gamma}_1, \dot{\gamma}_2, \ddot{\Psi}_0, \dot{\Psi}_2^0, \dot{\Psi}_2^1, \dot{\Psi}_3^0, \dot{\Psi}_3^1, \dot{\Psi}_3^2, \dot{\Psi}_4^1, \dot{\Psi}_4^2, \dot{\Psi}_4^1, \dot{\Psi}_4^2$ and $\dot{\Psi}_4^3$

are functions of (x^m) that can be given explicitly[14] in terms of the arbitrary functions of (x^m),

$$P, \tau_0, \lambda_0, \Psi_4^0, \dot{\Psi}_4^0, \phi_0, \text{ and } \dot{\phi}_0. \tag{3.9}$$

The quantities given by (3.5) through (3.8) are determined up to terms linear in u by specifying (3.9). Terms involving higher powers of u can similarly be obtained using Eqs. (2.21), (2.22), and (2.23), and (2.30), but they will depend on additional arbitrary functions of (x^m). To find these functions, consider Eqs. (2.23g) and (2.30). Equation (2.23g) determines $\Psi_4(u, r, x^m)$ but only up to an arbitrary function $\Psi_4(u, 0, x^m)$. Moreover, this function cannot be determined from the remaining NP equations. The scalar field equation (2.30) determines $\Delta\phi(u, r, x^m)$ but only up to an arbitrary function which from (2.18b) is $\dot{\phi}(u, 0, x^m)$. Moreover, this function cannot be determined from any of the NP equations. Since $\Psi_4(u, 0, x^m)$ and $\dot{\phi}(w, 0, x^m)$ are the only quantities in addition to (3.9) that are not determined by the NP equations or the scalar field equation, it has been established that:

The metric variables, spin coefficients, physical Weyl tensor components, and scalar field of an Einstein-scalar space-time containing a nondiverging null hypersurface are completely determined in the region $\{(u, r, x^m)\}$ from Eqs. (2.21), (2.22), (2.23), and (2.30) by specifying the arbitrary functions

$$P(x^m), \quad \tau_0(x^m) = \tau(0, 0, x^m), \quad \lambda_0(x^m) = \lambda(0, 0, x^m), \quad \Psi_4(u, 0, x^m), \quad \text{and} \quad \phi(u, 0, x^m). \tag{3.10}$$

The arbitrary functions (3.10) constitute the characteristic data for an Einstein-scalar spacetime containing a nondiverging null hypersurface. Before these data are discussed, the restrictions placed on them for which trapped surfaces develop will be determined. From (2.32) the spacelike two-surface $S_{(u,r)}$ is a trapped surface if and only if it is compact and everywhere on it $\rho > 0$ and $\mu < 0$. Since for fixed u and r the mapping

$$f_{(u,r)}: S_0 \to S_{(u,r)} \quad \text{where } f_{(u,r)}(0, 0, x^m) = (u, r, x^m)$$

is a homeomorphism of S_0 onto $S_{(u,r)}$, $S_{(u,r)}$ is compact if and only if S_0 is compact. From (3.6a) and (3.6c) it is seen that if everywhere on S_0, the functions K, τ_0, and ϕ_0 satisfy

$$\tfrac{1}{2}K - \tau_0\bar{\tau}_0 - \tfrac{1}{4}\eth\phi_0\bar{\eth}\phi_0 > 0, \tag{3.11}$$

then for any positive value of r, r_0, there exists a sufficiently small value of u, u_0, such that $\rho(u, r, x^m) > 0$ and $\mu(u, r, x^m) < 0$ for $0 < u \le u_0$ and $0 < r \le r_0$. Furthermore, since on $u = 0$ (3.5) through (3.8) are polynomials in r whose coefficients are analytic functions of (x^m), r_0 can be taken arbitrarily large on $u = 0$. With this it has been established that:

In the region $\{(u, r, x^m)\}$ of an Einstein-scalar space-time containing a nondiverging null hypersurface $u = 0$, trapped surfaces develop to the future of the $r > 0$ branch of this hypersurface if S_0 is compact and everywhere on it K, τ_0, and ϕ_0 satisfy

$$\tfrac{1}{2}K - \tau_0\bar{\tau}_0 - \tfrac{1}{4}\eth\phi_0\bar{\eth}\phi_0 > 0. \tag{3.12}$$

This result establishes the existence of nonspherically symmetric Einstein-scalar space–times that contain both a nondiverging null hypersurface and trapped surfaces.

In order to better understand (3.12), the characteristic data (3.10) will now be discussed. The function P is the most important of (3.10) for the development of trapped surfaces, since unless the spacelike two-surface with induced covariant metric,

$$-P^{-2}dz \otimes d\bar{z},$$

has strictly positive Gaussian curvature, $K = \bar{\eth}\eth \ln P$, there is no possibility of satisfying (3.11). In the case of a spherically symmetric space–time it is known that S_0 is a two-sphere with $K > 0$.[14] Although in the case of an arbitrary space–time, S_0 may be chosen to be a two-sphere, there do exist other compact two-surfaces with strictly positive Gaussian curvature and hence this choice is not imperative.

The function τ_0 is also very important for the development of trapped surfaces, since even if $K > 0$ and even in vacuum, the magnitude of τ_0 could be sufficiently large that (3.11) is violated. This possibility suggests that there may exist a relationship between τ_0 and angular momentum. Such a relationship can be obtained in the case of the linearized Kerr space–time, whose metric depends on two parameters m and a, where m is the mass and a is the angular momentum per unit mass. The components $(g^{\mu\nu})$ of the contravariant metric of this space–time are obtained relative to $\{u, r, \theta, \Phi\}$ from the components of the contravariant metric of the Kerr space–time,[17]

$$\frac{1}{(r^2 + a^2\cos^2\theta)} \times \begin{bmatrix} -a^2\sin^2\theta & (r^2+a^2) & 0 & -a \\ (r^2+a^2) & -(r^2 - 2mr + a^2) & 0 & a \\ 0 & 0 & -1 & 0 \\ -a & a & 0 & -\csc^2\theta \end{bmatrix},$$

by neglecting all terms that are quadratic in a. Therefore,

$$(g^{\mu\nu}) = \frac{1}{r^2}\begin{bmatrix} 0 & r^2 & 0 & -a \\ r^2 & -(r^2 - 2mr) & 0 & a \\ 0 & 0 & -1 & 0 \\ -a & a & 0 & -\csc^2\theta \end{bmatrix}.$$

Since $(g^{\mu\nu})$ does not satisfy the coordinate conditions (2.12), it is necessary to transform to a coordinate system $\{\tilde{u}, \tilde{r}, \tilde{\theta}, \tilde{\Phi}\}$ in which (2.12) is satisfied by $(\tilde{g}^{\mu\nu})$. This is accomplished by the coordinate transformation

$$\tilde{u} = -\exp(-u/4m), \quad \tilde{r} = 4m(r - 2m)\exp(u/4m),$$
$$\tilde{\theta} = \theta, \quad \tilde{\Phi} = \Phi - \frac{a}{r} - (a/4m^2)u,$$

whose inverse transformation is

$$u = -4m\ln(-\tilde{u}), \quad r = (8m^2 - \tilde{u}\tilde{r})/4m,$$
$$\theta = \tilde{\theta}, \quad \Phi = \tilde{\Phi} + \frac{4am}{(8m^2 - \tilde{u}\tilde{r})} - \frac{a}{m}\ln(-\tilde{u}).$$

Under this transformation $(g^{\mu\nu})$ becomes

$$(\tilde{g}^{\mu\nu}) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 2\tilde{r}^2(8m^2 - \tilde{u}\tilde{r})^{-1} & 0 & \tilde{g}^{13} \\ 0 & 0 & -r^{-2} & 0 \\ 0 & \tilde{g}^{13} & 0 & -r^{-2}\csc^2\tilde{\theta} \end{bmatrix},$$

where

$$\tilde{g}^{13} = -a\left(\frac{1}{4m^2 r} + \frac{1}{2mr^2} + \frac{1}{r^3}\right)\tilde{r}.$$

The null tetrad system that yields $(\tilde{g}^{\mu\nu})$ is $\{\tilde{D}, \tilde{\Delta}, \tilde{\delta}, \bar{\tilde{\delta}}\}$, where

$$\tilde{D} = \frac{\partial}{\partial\tilde{r}},$$
$$\tilde{\Delta} = \frac{\partial}{\partial\tilde{u}} + \frac{\tilde{r}^2}{(8m^2 - \tilde{u}\tilde{r})}\frac{\partial}{\partial\tilde{r}} - a\left(\frac{1}{4m^2 r} + \frac{1}{2mr^2} + \frac{1}{r^3}\right)\tilde{r}\frac{\partial}{\partial\tilde{\Phi}},$$
$$\tilde{\delta} = \frac{1}{\sqrt{2}r}\left(\frac{\partial}{\partial\tilde{\theta}} + i\csc\tilde{\theta}\,\frac{\partial}{\partial\tilde{\Phi}}\right).$$

The Lie brackets of this null tetrad system yield after a straightforward calculation the nonzero spin coefficients

$$\rho = \frac{\tilde{u}}{(8m^2 - \tilde{u}\tilde{r})}, \quad \tilde{\mu} = -\frac{8m^2\tilde{r}}{(8m^2 - \tilde{u}\tilde{r})^2},$$
$$\tilde{\tau} = -\frac{ia\sin\tilde{\theta}}{2\sqrt{2}}\left[\left(\frac{1}{4m^2} + \frac{1}{2mr} + \frac{1}{r^2}\right) + \left(\frac{1}{4m^2 r} + \frac{1}{mr^2} + \frac{3}{r^3}\right)\tilde{u}\tilde{r}\right],$$
$$\tilde{\alpha} = -\cot\tilde{\theta}/2\sqrt{2}r + \tfrac{1}{2}\bar{\tilde{\tau}}, \quad \tilde{\beta} = \cot\tilde{\theta}/2\sqrt{2}r + \tfrac{1}{2}\tilde{\tau},$$
$$\tilde{\gamma} = -\frac{(8m^2 - \tilde{u}\tilde{r}/2)\tilde{r}}{(8m^2 - \tilde{u}\tilde{r})^2}.$$

The null tetrad system, these spin coefficients, and Eqs. (2.22b) and (2.22m) imply that the only nonzero physical Weyl tensor components are Ψ_1, Ψ_2, and Ψ_3. Furthermore, the metric variables and spin coefficients imply that the nonzero characteristic data for linearized Kerr space–time are

$$(\tilde{\xi}^n_0) = (1, i\csc\tilde{\theta})/2\sqrt{2}m \quad \text{and} \quad \tilde{\tau}_0 = -3ia\sin\tilde{\theta}/8\sqrt{2}m^2.$$

Additional evidence that τ_0 is related to angular momentum can be given by considering the propagation of the null tetrad system (2.18) along the generators of $u = 0$. A tetrad system is normally said to be propagated without rotation along a timelike curve if and only if it is Fermi propagated[18] along this curve, which in the case of a timelike geodesic is equivalent to being parallelly propagated. If this notion is extended to null geodesics, then it can be said that the null tetrad system (2.18) is propagated without rotation along the generators of $u = 0$ if and only if it is parallelly propagated along them. From (2.6) it can be shown that

$$n^{\mu}{}_{;\nu}\ell^{\nu} = \bar{\tau}m^{\mu} + \tau\bar{m}^{\mu} \quad \text{and} \quad m^{\mu}{}_{;\nu}\ell^{\nu} = \tau\ell^{\mu}.$$

Therefore Δ, δ, and $\bar{\delta}$ are parallelly propagated along the generators of $u = 0$ if and only if $\tau_0 = 0$.

Like τ_0, ϕ_0 is important for the development of trapped surfaces, since even if $K > 0$ and $\tau_0 = 0$, the magnitude of $\eth\phi_0$ could be sufficiently large that (3.11) is violated. While the metric and scalar field are determined on $u = 0$ by specifying P, τ_0 and ϕ_0, they are determined in a neighborhood of $u = 0$ only by specifying $\lambda(0,0,x^m)$, $\Psi_4(u,0,x^m)$, and $\dot\phi(u,0,x^m)$. Additional significance of $\lambda(0,0,x^m)$, $\Psi_4(u,0,x^m)$, and $\dot\phi(u,0,x^m)$ can be shown by considering the $r = 0$ hypersurface. The general $r =$ constant hypersurface has a normal k where

$$k = k^\mu \frac{\partial}{\partial x^\mu} = g^{\mu\nu} r_{,\nu} \frac{\partial}{\partial x^\mu} = \delta_0^\mu \frac{\partial}{\partial x^\mu} = \frac{\partial}{\partial u}.$$

From this the $r =$ const hypersurface is spacelike, null, or timelike according to the sign of

$$g_{\mu\nu} k^\mu k^\nu = -2U$$

being positive, zero, or negative respectively on $r =$ const. The condition (3.4e) implies that the $r = 0$ hypersurface is null. From (3.4c) and (3.4e)

$$\Delta = \partial/\partial u$$

on $r = 0$. Therefore, the $r = 0$ hypersurface is a null hypersurface generated by null geodesics each with Δ as its tangent vector and each parameterized with an affine parameter u. From (2.31) the vectors tangent to these generators have divergence $\mu(u,0,x^m)$, zero rotation, and shear $\lambda(u,0,x^m)$. Equations (2.22m) and (2.22n), conditions (3.4), and (2.20) imply that

$$\begin{aligned}
&\dot\mu(u,0,x^m) = -\mu^2(u,0,x^m) - \lambda(u,0,x^m)\bar\lambda(u,0,x^m) \\
&\qquad - \tfrac{1}{2}\dot\phi^2(u,0,x^m), \\
&\mu(0,0,x^m) = 0, \qquad (3.13)\\
&\dot\lambda(u,0,x^m) = -2\mu(u,0,x^m)\lambda(u,0,x^m) - \Psi_4(u,0,x^m), \\
&\lambda(0,0,x^m) = \lambda_0.
\end{aligned}$$

Therefore through these $\lambda_0, \Psi_4(u,0,x^m)$, and $\dot\phi(u,0,x^m)$ determine $\mu(u,0,x^m)$ and $\lambda(u,0,x^m)$ and hence the optical properties of the generators of $r = 0$.

4. GENERAL SPACE-TIMES CONTAINING TRAPPED SURFACES

The possibility that there exist Einstein-scalar space-times more general than those containing a nondiverging null hypersurface that also contain trapped surfaces was investigated.[14] This investigation began by obtaining the metric variables, spin coefficients, physical Weyl tensor components, and scalar field of an Einstein-scalar space-time. Then the characteristic data for this space-time were determined and examined for restrictions placed on them for which trapped surfaces develop.

The metric variables, spin coefficients, physical Weyl tensor components, and scalar field of an Einstein-scalar space-time were obtained using the formalism presented in Sec. 2. To accomplish this, the main condition adopted was that, in a particular null coordinate system (2.17) and associated null tetrad system (2.18), these quantities are analytic functions of (u,r,x^m) in the region $\{(u,r,x^m)\}$. Since the procedure involved in obtaining these quantities is exactly that employed in Sec. 3, only the results will be stated here. By using the techniques of Sec. 3 subject to the conditions (3.4), it can be established that:

An Einstein-scalar space-time has over the region $\{(u,r,x^m)\}$

metric variables

$$\begin{aligned}
U = \{[\tfrac{1}{2}K - 3\tau_0\bar\tau_0 + (\sigma_0\lambda_0 + \bar\sigma_0\bar\lambda_0) - \tfrac{1}{2}\phi_1\dot\phi_0 \\
- \tfrac{1}{4}\eth\phi_0\bar\eth\phi_0]r^2 + \cdots\} + (\cdots)u + \cdots, \qquad (4.1a)
\end{aligned}$$

$$\begin{aligned}
X^m = \{2(\bar\tau_0\xi_0{}^m + \tau_0\bar\xi_0{}^m)r + \cdots\} + \{2(\eth\lambda_0/\sqrt{2} - 2\tau_0\lambda_0 \\
- \dot\phi_0\bar\eth\phi_0/\sqrt{2})\xi_0^m + \bar\eth\bar\lambda_0/\sqrt{2} - 2\bar\tau_0\bar\lambda_0 \\
- \dot\phi_0\eth\phi_0)\bar\xi_0^m]r + \cdots\}u + \cdots, \qquad (4.1b)
\end{aligned}$$

$$\begin{aligned}
\xi^m = \{\xi_0^m + (\rho_0\xi_0^m + \sigma_0\bar\xi_0^m)r + \cdots\} + (-\bar\lambda_0\bar\xi_0^m \\
+ \cdots)u + \cdots; \qquad (4.1c)
\end{aligned}$$

nonzero spin coefficients

$$\begin{aligned}
\rho = \{\rho_0 + (\rho_0^2 + \sigma_0\bar\sigma_0 + \tfrac{1}{2}(\phi_1)^2)r + \cdots\} + \{(\tfrac{1}{2}K - \tau_0\bar\tau_0 \\
- \tfrac{1}{4}\eth\phi_0\bar\eth\phi_0) + \cdots\}u + \cdots, \qquad (4.2a)
\end{aligned}$$

$$\begin{aligned}
\sigma = \{\sigma_0 + (2\rho_0\sigma_0 + \Psi_0^0)r + \cdots\} + \{(-\rho_0\bar\lambda_0 + \eth\tau_0/\sqrt{2} - \tau_0^2 \\
- \tfrac{1}{4}(\eth\phi_0)^2) + \cdots\}u + \cdots, \qquad (4.2b)
\end{aligned}$$

$$\begin{aligned}
\mu = \{-(\tfrac{1}{2}K - \tau_0\bar\tau_0 - \tfrac{1}{4}\eth\phi_0\bar\eth\phi_0)r + \cdots\} \\
+ \{(-\lambda_0\bar\lambda_0 - \tfrac{1}{2}(\dot\phi_0)^2)r + \cdots\} + \cdots, \qquad (4.2c)
\end{aligned}$$

$$\begin{aligned}
\lambda = \{\lambda_0 + (\rho_0\lambda_0 + \bar\eth\bar\tau_0/\sqrt{2} + \bar\tau_0^2 + \tfrac{1}{4}(\bar\eth\phi_0)^2)r + \cdots\} \\
+ (-\Psi_4^0 + \cdots)u + \cdots, \qquad (4.2d)
\end{aligned}$$

$$\begin{aligned}
\tau = \{\tau_0 + (3\rho_0\tau_0 + \sigma_0\bar\tau_0 - \eth\rho_0/\sqrt{2} + \bar\eth\sigma_0/\sqrt{2} \\
+ \phi_1\eth\phi_0)r + \cdots\} \\
+ \{(\bar\eth\bar\lambda_0/\sqrt{2} - \bar\tau_0\bar\lambda_0 - \dot\phi_0\eth\phi_0/\sqrt{2}) + \cdots\}u + \cdots, \qquad (4.2e)
\end{aligned}$$

$$\nu = -\eth U, \qquad (4.2f)$$

$$\begin{aligned}
\alpha = \{(\alpha_0 + \tfrac{1}{2}\bar\tau_0) + (\tfrac{3}{2}\rho_0\bar\tau_0 + \tfrac{1}{2}\bar\sigma_0\tau_0 + \phi_1\bar\eth\phi_0/2\sqrt{2} \\
+ \rho_0\alpha_0 - \bar\sigma_0\bar\alpha_0)r + \cdots\} + (\cdots)u + \cdots, \qquad (4.2g)
\end{aligned}$$

$$\begin{aligned}
\beta = \{(-\bar\alpha_0 + \tfrac{1}{2}\tau_0) + (\tfrac{3}{2}\rho_0\tau_0 + \tfrac{1}{2}\sigma_0\bar\tau_0 - \eth\rho_0/\sqrt{2} \\
+ \bar\eth\sigma_0/\sqrt{2} + \phi_1\eth\phi_0/2\sqrt{2} + \sigma_0\alpha_0 - \rho_0\bar\alpha_0)r + \cdots\} \\
+ (\cdots)u + \cdots, \qquad (4.2h)
\end{aligned}$$

$$\begin{aligned}
\gamma = \{(-\tfrac{1}{2}K + 3\tau_0\bar\tau_0 + \bar\eth\tau_0/\sqrt{2} - \sigma_0\lambda_0 - \bar\sigma_0\bar\lambda_0 \\
+ \tfrac{1}{2}\phi_1\dot\phi_0 + \tfrac{1}{4}\eth\phi_0\bar\eth\phi_0 + 2\alpha_0\tau_0 - 2\bar\alpha_0\bar\tau_0)r \\
+ \cdots\} + (\cdots)u + \cdots; \qquad (4.2i)
\end{aligned}$$

physical Weyl tensor components

$$\Psi_0 = (\Psi_0^0 + \Psi_0^1 r + \cdots) + (\cdots)u + \cdots, \qquad (4.3a)$$

$$\Psi_1 = (\cdots) + (\cdots)u + \cdots, \qquad (4.3b)$$

$$\begin{aligned}
\Psi_2 = \{(-\tfrac{1}{2}K + \bar\eth\tau_0/\sqrt{2} - \sigma_0\lambda_0 + \tfrac{1}{6}\eth\phi_0\bar\eth\phi_0 \\
+ \tfrac{1}{6}\phi_1\dot\phi_0) + \cdots\} + (\cdots)u + \cdots, \qquad (4.3c)
\end{aligned}$$

$$\begin{aligned}
\Psi_3 = \{(-\eth\lambda_0/\sqrt{2} - \tau_0\lambda_0 + \dot\phi_0\bar\eth\phi_0/2\sqrt{2}) + \cdots\} \\
+ \{\cdots\}u + \cdots, \qquad (4.3d)
\end{aligned}$$

$$\Psi_4 = (\Psi_4^0 + \cdots) + (\dot\Psi_4^0 + \cdots)u + \cdots; \qquad (4.3e)$$

and scalar field

$$\begin{aligned}
\phi = (\phi_0 + \phi_1 r + \cdots) + \{\dot\phi_0 + (\rho_0\dot\phi_0 + \tfrac{1}{2}\bar\eth\eth\phi_0 \\
- \tau_0\bar\eth\phi_0/\sqrt{2} - \bar\tau_0\eth\phi_0/\sqrt{2})r + \cdots\}u + \cdots; \qquad (4.4)
\end{aligned}$$

where $K = \bar{\eth}\eth \ln P$, $\alpha_0 = (1/2\sqrt{2})\eth \ln P$, and

$$P, \rho_0, \sigma_0, \tau_0, \lambda_0, \Psi_0^0, \Psi_0^1, \Psi_4^0, \dot{\Psi}_4^0, \phi_0, \phi_1, \text{ and } \dot{\phi}_0 \qquad (4.5)$$

are arbitrary functions of (x^m).

The terms displayed in (4.1) through (4.4) depend on the arbitrary functions (4.5). Terms involving higher powers of u and r can be obtained using Eqs. (2.21), (2.22), (2.23), and (2.30), but they will depend on additional arbitrary functions of (x^m). By reasoning similar to that used to obtain (3.10), it can be shown that:

The metric variables, spin coefficients, physical Weyl tensor components, and scalar field of an Einstein-scalar space–time are completely determined from Eqs. (2.21), (2.22), (2.23), and (2.30) in the region $\{(u, r, x^m)\}$ by specifying the arbitrary functions

$$P(x^m), \rho_0(x^m), \sigma_0(x^m), \tau_0(x^m), \lambda_0(x^m), \Psi_0(0, r, x^m), \Psi_4(u, 0, x^m), \phi(0, r, x^m), \phi(u, 0, x^m). \qquad (4.6)$$

The arbitrary functions (4.6) constitute the characteristic data for an Einstein-scalar space–time. The functions $P(x^m), \tau_0(x^m), \lambda_0(x^m), \Psi_4(u, 0, x^m), \phi(u, 0, x^m)$ were already discussed in Sec. 3. The remaining functions $\rho_0(x^m), \sigma_0(x^m), \Psi_0(0, r, x^m)$, and $\phi(0, r, x^m)$ through Eqs. (2.22a) and (2.22b) determine the divergence of D on $u = 0$, $\rho(0, r, x^m)$, and the shear of D on $u = 0, \sigma(0, r, x^m)$. Hence these additional functions determine the optical properties of the generators of $u = 0$.

Of (4.6) it is P, ρ_0, τ_0, and ϕ_0 that are important for the development of trapped surfaces; since from (4.2a) and (4.2c), if everywhere on S_0

$$\rho_0 \geq 0 \quad \text{and} \quad \tfrac{1}{2}K - \tau_0\bar{\tau}_0 - \tfrac{1}{4}\eth\phi_0\bar{\eth}\phi_0 > 0,$$

then there exist sufficiently small positive real numbers u_0 and r_0 such that $\rho > 0$ and $\mu < 0$ for $0 < u \leq u_0$ and $0 < r \leq r_0$. Therefore by (2.32) the two-surface $S_{(u,r)}$ for $0 < u \leq u_0$ and $0 < r \leq r_0$ are trapped surfaces if and only if they are compact, which was shown in Sec. 3 to be equivalent to S_0 being compact. With this it has been established that:

The region $\{(u, r, x^m)\}$ of an Einstein-scalar space–time contains trapped surfaces $S_{(u,r)}$ for some range of u and r, $0 < u \leq u_0$ and $0 < r \leq r_0$, if S_0 is compact and everywhere on it ρ_0, K, τ_0 and ϕ_0 satisfy

$$\rho_0 \geq 0 \quad \text{and} \quad \tfrac{1}{2}K - \tau_0\bar{\tau}_0 - \tfrac{1}{4}\eth\phi_0\bar{\eth}\phi_0 > 0. \qquad (4.7)$$

This result establishes the existence of Einstein-scalar space–times more general than those containing a nondiverging null hypersurface that also contain trapped surfaces.

5. SUMMARY AND CONCLUSIONS

An investigation of the characteristic development of trapped surfaces in Einstein-scalar space–times, of which the empty space–times are a special case, was discussed in this paper. After presenting the formalism used in this investigation in Sec. 2, the characteristic development of trapped surfaces in Einstein-scalar space–times containing a nondiverging null hypersurface was considered in Sec. 3. The main result of this section is (3.12), which states that in a region $\{(u, r, x^m)\}$ of an Einstein-scalar space–time containing a nondiverging null hypersurface $u = 0$, trapped surfaces develop to the future of the $r = 0$ branch of this hypersurface if the spacelike two-surface $S_0 = \{(u, r, x^m) : u = 0 = r\}$ is compact and everywhere on it the characteristic data $P(x^m)$, $\tau_0(x^m)$, and $\phi_0(x^m)$ satisfy (3.11), $\tfrac{1}{2}K - \tau_0\bar{\tau}_0 - \tfrac{1}{4}\eth\phi_0\bar{\eth}\phi_0 > 0$, where $K = \bar{\eth}\eth \ln P$ is the Gaussian curvature of S_0. If (3.11) is to be satisfied at all, then K must be strictly positive. This suggests that the collapse of an object maintaining a cylindrical or toroidal shape during collapse cannot result in the formation of trapped surfaces, since a cylinder has zero Gaussian curvature[19] and a torus has a region of negative Gaussian curvature.[19] Indeed, it has been shown that at least one cylindrical collapse model results in complete collapse without the formation of trapped surfaces.[20] The possibility that the magnitude of τ_0 could be sufficiently large that (3.11) is not satisfied even if $K > 0$, and even in vacuum, suggests that τ_0 is related to angular momentum. The role of τ_0 in the criteria for the existence of trapped surfaces, even in vacuum, and the evidence presented in support of its interpretation in terms of angular momentum are considered to be important results of this investigation. With the presence of angular momentum and asymmetries in the scalar field indicated by τ_0 and $\eth\phi_0$ respectively, (3.12) emphasizes the importance of these quantities in determining whether or not trapped surfaces develop in Einstein-scalar space–times containing a nondiverging null hypersurface. The metric variables, spin coefficients, physical Weyl tensor components, and scalar field for these space–times are determined in some neighborhood of $u = 0$ by the characteristic data (3.10) through (3.5), (3.6), (3.7), and (3.8), respectively. The data $\lambda_0(x^m), \psi_4(u, 0, x^m)$, and $\dot{\phi}(u, 0, x^m)$ have the additional significance of determining the optical properties of the $r = 0$ null hypersurface through (3.13).

In Sec. 4 the existence of Einstein-scalar space–times more general than those containing a nondiverging null hypersurface that also contain trapped surfaces was established. The main result of this section is that in a region $\{(u, r, x^m)\}$ of an Einstein-scalar space–time, trapped surfaces develop to the future in some neighborhood of S_0 if S_0 is compact and everywhere on it the characteristic data $\rho_0(x^m), P(x^m), \tau_0(x^m)$, and $\phi_0(x^m)$ satisfy (3.11) and $\rho_0 \geq 0$. The metric variables, spin coefficients, physical Weyl tensor components, and scalar field are determined in some neighborhood of S_0 by the characteristic data (4.6) through (4.1), (4.2), (4.3), and (4.4), respectively. The additional characteristic data for these space–times, $\rho_0(x^m), \sigma_0(x^m), \Psi_0(0, r, x^m)$, and $\phi(0, r, x^m)$, were shown to determine the optical properties of the generators of the $u = 0$ null hypersurface through (2.22a) and (2.22b).

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. E. T. Newman and Dr. D. W. Pajerski for many helpful comments.

*This work was supported in part by the National Science Foundation. It incorporates some of the results of the first author's doctoral dissertation at the University of Pittsburgh.

[1] R. Penrose, Phys. Rev. Lett. 14, 57 (1965).
[2] See, for example, S. W. Hawking and R. Penrose, Proc. Roy. Soc. (London) A 314, 529 (1970).
[3] D. W. Pajerski and E. T. Newman, J. Math. Phys. 12, 1929 (1971).
[4] E. T. Newman and R. Penrose, J. Math. Phys. 3, 566 (1962).

[5] Greek indices and the lower-case Latin indices a through d range and sum over the values (0, 1, 2, 3) and the remaining lower case Latin indices range and sum over the values (2, 3).

[6] The parentheses denote symmetrization: $A_{(\mu\nu)} = \frac{1}{2}(A_{\mu\nu} + A_{\nu\mu})$; and the brackets denote antisymmetrization: $A_{[\mu\nu]} = \frac{1}{2}(A_{\mu\nu} - A_{\nu\mu})$.

[7] The physical components $T_{a_1 \ldots a_s}{}^{b_1 \ldots b_t}$ relative to a tetrad system $\{z^\mu_a \partial/\partial x^\mu\}$ of a tensor $T_{\mu_1 \ldots \mu_s}{}^{\nu_1 \ldots \nu_t}$ are given by $T_{a_1 \ldots a_s}{}^{b_1 \ldots b_t} = z_{a_1}{}^{\mu_1} \ldots z_{a_s}{}^{\mu_s} z_{\nu_1}{}^{b_1} \ldots z_{\nu_t}{}^{b_t} T_{\mu_1 \ldots \mu_s}{}^{\nu_1 \ldots \nu_t}$.

[8] A comma denotes partial differentiation and a semicolon denotes covariant differentiation.

[9] $\{x^\mu\}$ is the coordinate system; $\{(x^\mu)\}$ is the region consisting of all points of the manifold covered by $\{x^\mu\}$, and (x^μ) is a point in the region $\{(x^\mu)\}$.

[10] These equations can also be found in S. W. Hawking, J. Math. Phys. **9**, 598 (1968). Misprints in Hawking's Eqs. (3.32) and (3.41) are corrected in Eqs. (2.23b) and (2.23k), respectively.

[11] Units are employed in which the gravitational constant and the speed of light are of unit magnitude.

[12] R. Sachs, Proc. Roy. Soc. (London) A **264**, 309 (1961).

[13] The subscript notations $(a|c|b)$ and $[a|c|b]$ denote symmetrization and antisymmetrization, respectively, with respect to a and b.

[14] P. Demmie, Ph.D. thesis (University of Pittsburgh, 1971).

[15] L. P. Eisenhart, *An Introduction to Differential Geometry* (Princeton U.P., Princeton, New Jersey, 1947), pp. 261–65.

[16] E. T. Newman and R. Penrose, J. Math. Phys. **7**, 863 (1966).

[17] This metric follows from that given in E. Newman *et al.*, J. Math. Phys. **6**, 918 (1965), by setting $e = 0$. It should be emphasized that by linearized it is meant linearized in a.

[18] See, for example, J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960), Chap. III, Sec. 8.

[19] See, for example, B. O'Neill, *Elementary Differential Geometry* (Academic, New York, 1966), pp. 231–36.

[20] K. S. Thorne, "Nonspherical Gravitational Collapse—A Short Review," (preprint, 1971).

Erratum: Absence of long-range order in thin films [J. Math. Phys.13, 1735 (1972)]

John C. Garrison

Lawrence Livermore Laboratory, University of California, P.O. Box 808, Livermore, California 94550

(Received 17 January 1973)

Several unfortunate misprints occurred in the above paper. Corrections are as follows:

(1) In the first sentence of the second paragraph in Sec. 2, the expression $\pi : \mathfrak{H} \to \mathcal{L}(\mathfrak{H})$ should be replaced by $\pi : \mathfrak{A} \to \mathcal{L}(\mathfrak{H})$.

(2) In the last sentence of the same paragraph the expression "ω-*affiliated* to $\mathfrak{U}$ if" should be replaced by "ω-*affiliated* to $\mathfrak{A}$ if".

(3) The second sentence of the last paragraph in Sec. 3 should read: "Since a state ω exhibiting long-range order cannot be T-ergodic, it has a nontrivial decomposition into T-ergodic states; and ω can, for all physical purposes, be replaced by any one of the states in the decomposition."

(4) In the last sentence in the first paragraph of Sec. 5 the term "spin dependent" should be replaced by "spin independent".