JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 13, NUMBER 2 FEBRUARY 1972

Mapping onto Solutions of the Gravitational Initial Value Problem*

James W. York, Jr.
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540
(Received 28 June 1971)

Two approaches to the Einstein initial value problem for vacuum gravitational fields are considered. In the first, the metric of a spacelike slice is prescribed arbitrarily and it is shown that momenta satisfying the constraints can be constructed by exploiting the well-known relation of three of the four constraints to the three-dimensional coordinate transformation group. Specifically, it is shown that there exists a coordinate mapping of a certain specified set of functions onto momenta satisfying the constraints for a specified 3-metric. A further interpretation of this procedure is discussed. In the second approach the 3-geometry of a spacelike slice is specified up to a conformal factor. It is shown that, using a coordinate transformation method similar to the above, transverse traceless momenta can be constructed and that this construction depends essentially only on the conformal geometry of the spacelike hypersurface. The remaining constraint is satisfied by a choice of the conformal factor. As a result, it follows that the initial value equations can be satisfied by mapping certain specified sets of functions onto solutions by using coordinate transformations and a group of scale transformations which include conformal transformation of the metric. This is significant because the unconstrained initial data (gravitational degrees of freedom) are represented by a pair of scale-invariant transverse, traceless tensors of weight $\frac{5}{3}$. These objects, in turn, give irreducible representations of the coordinate and scaling groups which are used to effect solutions of the initial value equations.

1. INTRODUCTION

The purpose of this paper is to show that the initial data satisfying the Einstein gravitational constraint equations can be constructed by mapping certain specified sets of arbitrary functions to solutions in such a way that the unconstrained data (gravitational degrees of freedom) are left essentially invariant. The mappings involve three-dimensional coordinate transformation of the three-dimensional metric. The unconstrained data are represented by a pair of scale invariant transverse traceless tensors of weight $\frac{5}{3}$. These objects give irreducible representations of the groups which transform the sets of arbitrary functions to solutions.

The initial value problem[1] arises because four of Einstein's ten field equations for gravitation involve only initial data. In the Hamiltonian form of general relativity,[2,3] the data which must be given compatibily on an initial spacelike hypersurface are the metric γ_{ab} on the slice (canonical field coordinate) and a tensor density π^{ab} measuring the bending of the slice relative to the surrounding space-time by means of the relation $\pi^{ab} = \gamma^{1/2}(K\gamma^{ab} - K^{ab})$, where K^{ab} is the second fundamental tensor of the surface, K is its trace, and γ is the determinant of γ_{ab}. Therefore, π^{ab} measures "extrinsic curvature" and plays the role of canonical field momentum. Thus, the problem of initial conditions is to give γ_{ab} and π^{ab} in such a way that the four constraint equations are satisfied at the initial instant. If the initial data are consistent on the initial surface, they continue to satisfy the constraints for a finite time into the future as a consequence of Einstein's dynamical equations and the contracted Bianchi identities satisfied by the latter. Satisfactory understanding of the initial value problem is essential for full appreciation of the dynamics of gravitational fields, for the constraints serve as (vanishing) Hamiltonians in the canonical representation of dynamics. Therefore, they contain implicitly the complete description of gravitational dynamics in Einstein's theory.

The key issue in the initial value problem is to determine which variables may be specified freely on the initial hypersurface and which are to be thought of as constrained. Among the former, one expects some information depending merely on the choice of space-time coordinates and some corresponding to true gravitational degrees of freedom. A number of schemes have been proposed in order to settle this issue in a physically satisfying manner. Some of these approaches do not work directly with the canonical momentum π^{ab}, notably the "thin sandwich" approach[4] which introduces in place of π^{ab} other less physically meaningful variables. The sandwich problem is not dealt with in the present paper. All approaches, however, do deal more or less directly with the metric γ_{ab}. This is natural because, if the initial surface is designated t = const, the only coefficients of the space-time metric ${}^4g_{\mu\nu}$ whose second time derivatives appear in Einstein's equations are those corresponding to the spacelike 3-metric on the surface t = const. From this fact, plus the fact that

the three-dimensional coordinate system contains no physical information and may itself be changing arbitrarily in time, one concludes that the dynamical object in general relativity is three-dimensional spatial geometry. This 3-geometry ${}^3\mathcal{G}$ is described by an equivalence class of metrics with respect to diffeomorphisms D_3. In other words, all metrics γ_{ab} which transform into each other by coordinate mappings represent one and the same 3-geometry. In the following, by "metric" we refer to component functions $\gamma_{ab}(x)$ representing the metric tensor field as displayed in a definite coordinate system; the term "3-geometry" refers to the entire D_3-equivalence class of $\gamma_{ab}(x)$. Therefore, specification of the metric (i.e., ${}^3\mathcal{G}$ plus a coordinate system) requires more information than specification of a three-geometry alone. Superspace,[5] the collection of all 3-geometries, is a configuration space for Einstein's geometrodynamics. It is therefore natural in treating the initial value problem to assume that the three-geometry can be specified freely while the momentum can be constructed in accordance with the constraints. In the early sections of the following work, it is convenient to adopt this point of view or the related one in which the 3-metric is arbitrarily specified.

It has long been known that three of the four constraints are intimately related to the group of all three-dimensional coordinate mappings D_3.[6] Because the constraints are covariant, it would not at first appear to be possible to use the coordinate group to effect a solution to these equations. Nevertheless, it turns out that there is procedure available for doing just this. In this paper, it is shown explicitly how a set of arbitrary functions can be mapped by a certain coordinate transformation onto local solutions of the initial value equations, with the metric completely fixed in advance. Further, an alternative interpretation of this process is given which indicates that it can also be viewed as an explicit method of choosing coordinate conditions, for a fixed 3-geometry (not a fixed metric), such that a given set of arbitrary functions will satisfy the constraints. On the other hand, in the last three sections we do not fully specify either the metric or the 3-geometry. Rather, we use the method of specifying the metric only up to a conformal factor. This useful mathematical technique is known from the work of Lichnerowicz[7] and others.[1,8] However, the compelling physical arguments for this approach were only recently demonstrated.[9] It was shown that conformal 3-geometry is characterized by a covariantly transverse and traceless tensor and that it therefore represents "pure spin-two" geometrodynamics. In addition, it was shown that the existence of this tensor is closely related to the fact that there is a group of scale transformations $\bar{C}$, which include conformal transformations of the metric, that leave the unconstrained initial data invariant. As a result, there are two groups, $\bar{C}$ and D_3, which are fundamental to the initial value problem. In this paper, it is shown that by the action of these "gaugelike" groups, certain specified sets of arbitrary functions can be mapped onto solutions of initial value equations.

2. INITIAL VALUE PROBLEM IN CANONICAL VARIABLES

In terms of the standard canonical variables, the constraint equations for empty space-time have the form[3]

$$\nabla_b \pi^{ab} = 0, \tag{1}$$

$$\gamma^{-1/2}(\pi_{ab}\pi^{ab} - 1/2\pi^2) - \gamma^{1/2}R = 0. \tag{2}$$

Here, γ_{ab} is the metric on a spacelike hypersurface, R is its scalar curvature, γ its determinant, and ∇_b denotes the associated covariant derivative. Geometrically, the initial value equations (1) and (2) are simply the Gauss–Codazzi equations[10] for Ricci-flat space-times[11] ${}^4R_{\mu\nu} = 0$, that is, necessary and sufficient conditions that a slice with metric γ_{ab} and extrinsic curvature π^{ab} can be embedded in some space-time satisfying the Einstein field equations ${}^4R_{\mu\nu} = 0$.

One can regard the transversality conditions (1) and the Hamiltonian constraint (2) as four equations on the twelve functions γ_{ab}, π^{ab}. However, four of these variables can be fixed in advance by choosing four space-time coordinate conditions, so the general solution should contain four essential arbitrary functions of three spatial coordinates. This familiar conclusion[12] is just the expected one for a field with two degrees of freedom per space point, in accord with familiar result that there are two independent states of polarization for gravitational waves.

Which variables should be regarded as unconstrained, and, therefore freely specifiable on the spacelike slice? Let us initially choose the metric γ_{ab} in advance and construct the appropriate momenta π^{ab}. One recognizes that three of the six γ_{ab} may be regarded as containing only information about the three-dimensional coordinate system on the slice. The other three fix the intrinsic geometry ${}^3\mathcal{G}$, which specifies a point in superspace. In this approach one argues that there are four constraints on the six π^{ab}, leaving two of them as essential arbitrary functions. There are also the three arbitrary functions specifying ${}^3\mathcal{G}$, giving a total of five, which would appear to be one too many. The dynamically superfluous variable owes its presence to the circumstance that no condition has been chosen regarding choice of "time", that is, the manner of slicing up space-time. One often regards this extra variable corresponding to "time" as residing in the metric. For example, in Misner's Mixmaster Universe,[13] the determinant of γ_{ab} (i.e., "volume" of the slice) plays the role of "time". However, there has been no universally accepted method of extracting time directly from the 3-geometry. In fact, Kuchař[14] has successfully applied an "extrinsic time" representation of gravitational dynamics, in which time is not regarded as residing in the intrinsic 3-geometry, but is a momentumlike variable instead. Another inconvenience of the above approach is that the Hamiltonian constraint (2) is quadratic algebraic in the momenta. As a consequence, it is not obvious how to single out momentum components or functions of these components to regard as dependent in the simultaneous solution of all four constraints. Nevertheless, it is easy to show that with γ_{ab} fixed, there exist solutions of the initial value equations involving two arbitrary functions of three variables and three arbitrary functions of two variables. The latter correspond, for example, to values of three of the π^{ab} (for example, π^{13}, π^{23}, π^{33}) on a 2-surface (for example, $x^3 = \text{const}$). These functions of only two variables are unimportant for the purposes of the present paper and will not be considered further.

3. MAPPING ONTO TRANSVERSE MOMENTA

If we consider the metric $\gamma_{ab}(x)$ to be fixed, then the momentum constraints are a system of three linear first-order partial differential equations for the π^{ab}:

$$\nabla_b \pi^{ab} = \partial_b \pi^{ab} + \pi^{bc}\{_{b\ c}^{\ a}\} = 0, \tag{3}$$

where the Christoffel symbols are now treated as known functions of the coordinates. It is well known that there is an intimate connection between the momentum constraints and the group of three-dimensional coordinate transformations with respect to which the standard canonical formalism is convariant.[6] For example, the expression[15]

$$\int (-2\xi_a \nabla_b \pi^{ab}) d^3x$$

generates spatial Lie derivatives along ξ^a of any function of the canonical variables with which its Poisson bracket is taken, provided the function is not explicitly coordinate dependent.

A related approach,[15] also well known, employs the Hamilton–Jacobi version of dynamics. In anology to the relation $p_i = \partial S/\partial q^i$ of classical mechanics, we put

$$\pi^{ab} = \delta S/\delta \gamma_{ab}$$

where S is the Jacobi action, or Hamilton's principal functional, and $\delta/\delta\gamma_{ab}$ denotes functional differentiation with respect to the metric. If S is any functional of γ_{ab} that does not depend explicitly on the choice of coordinates on the 3-space, then the conditions (1) are automatically satisfied. This conclusion follows in a manner exactly analogous to the way that the contracted Bianchi identities ${}^4G^{\mu\nu}{}_{;\nu} = 0$ follow from the four-dimensional coordinate invariance of the full Einstein action principle of general relativity. In these well-known results, one recognizes a deep connection between the vanishing covariant divergence of a symmetric tensor and the coordinate invariance of an appropriate functional. This connection has another manifestation first exploited by Pereira,[16] who was concerned not with the present problem but with the space–time equations $[(-g)^{1/2} T^{\mu\nu}]_{;\nu} = 0$ of Einstein's theory. Reasoning similar in part is applied below. However, the present application and the general interpretation of the process presented here are those of the present author.

We are concerned with the construction of transverse momenta, specifically, with how they may be generated by three-dimensional coordinate mappings. Let there be given a metric $\gamma_{ab}(x)$ and an arbitrary set of functions $\bar{\pi}^{ab}(\bar{x})$. The metric is defined in one-coordinate system (x); whereas, the momentum $\bar{\pi}^{ab}(\bar{x})$ is defined in another set of coordinates $(\bar{x})$. We are concerned with a mapping $\Theta : \bar{x} \to x$ defined by some functions $x^a = \theta^a(\bar{x})$. The objective is to find the mapping Θ that transforms $\bar{\pi}^{ab}(\bar{x})$ onto solutions $\pi^{ab}(x)$ of Eqs. (3). In the process, the transversality conditions are converted into a system of three second-order equations in three unknown functions $\theta^a(\bar{x})$. The metric $\gamma_{ab}(x)$ is specified in a given coordinate system (x). Define $\bar{h}^a(\bar{x}) \equiv \partial \bar{\pi}^{ab}/\partial \bar{x}^b$ which are known functions. Then the transversality constraint in barred variables is equivalent to

$$\{\overline{_{b\ c}^{\ a}}\}\,\bar{\pi}^{bc} + \bar{h}^a(\bar{x}) = 0, \tag{4}$$

where, according to the familiar rule of transformation,

$$\{\overline{_{b\ c}^{\ a}}\} = \{_{e\ f}^{\ d}\}\frac{\partial \bar{x}^a}{\partial x^d}\frac{\partial x^e}{\partial \bar{x}^b}\frac{\partial x^f}{\partial \bar{x}^c} + \frac{\partial \bar{x}^a}{\partial x^e}\frac{\partial^2 x^e}{\partial \bar{x}^b \partial \bar{x}^c}.$$

Making this substitution, multiplying by $\partial x^r/\partial \bar{x}^a$, and summing on ^{a} lead to[17]

$$\bar{\pi}^{bc}\frac{\partial^2 \theta^r}{\partial \bar{x}^b \partial \bar{x}^c} + \bar{\pi}^{bc}\{_{m\ n}^{\ r}\}\frac{\partial \theta^m}{\partial \bar{x}^b}\frac{\partial \theta^n}{\partial \bar{x}^c} + \bar{h}^b \frac{\partial \theta^r}{\partial \bar{x}^b} = 0 \tag{5}$$

as the equations governing $\theta^a(\bar{x})$. Were this only one equation for one function θ, it would be classified[18] as "semilinear" because the second derivatives appear linearly with coefficients $\bar{\pi}^{bc}$ depending only on the $\bar{x}$. Even in this case, however, the equation would not necessarily be of a definite type. In the case of space, as opposed to space–time, one would not expect there to be any preferred direction and, therefore, one might suspect the equation to be of elliptic type. Ellipticity would obtain provided $\bar{\pi}^{ab}\lambda_a\lambda_b > 0$ for all real $\lambda_a \neq 0$. On the contrary, however, there is no reason to suppose the $\bar{\pi}^{ab}$ to form coefficients of a positive definite quadratic form. This is easily seen from the fact that traceless momenta exist and play a fundamental role in the initial value problem (See sec. 6). The momentum tensor has no fixed signature in general. Of course, Eqs, (5) form a nonlinear system of three equations for three unknown functions of a kind concerning which apparently no useful general results are know.[18] Therefore, although in space we expect no preferred direction, we nevertheless single one out only in order to be able to apply the Cauchy–Kowalewski existence theorem.[19] The equations can be partially summed and rearranged so as to have the form

$$\frac{\partial^2 \theta^r}{(\partial \bar{x}^3)^2} = (\bar{\pi}^{33})^{-1} M^r\left(\bar{x}; \theta^a; \frac{\partial \theta^a}{\partial x^b}\right), \tag{6}$$

where the $\bar{x}^3$ direction is singled out only to show that the equations assume the "normal" form provided $\bar{\pi}^{33} \neq 0$. Assuming analyticity and invoking the Cauchy–Kowalewski theorem,[19] we see that in a sufficiently small domain including the surface $\bar{x}^3 =$ const, the θ^r exist and are unique up to functions of two variables ($\bar{x}^1$ and $\bar{x}^2$ in the present case), that is, up to the values of θ^r and $\partial\theta^r/\partial\bar{x}^3$ on $\bar{x}^3 =$ const, which values can be chosen arbitrarily.

Supposing that the functions $x^a = \theta^a(\bar{x})$ are known, we use them to transform the functions $\bar{\pi}^{ab}(\bar{x})$ as tensor densities of weight unity by the usual rule

$$\pi^{ab} = \left[\det\left(\frac{\partial \theta^c}{\partial x^d}\right)\right]\frac{\partial \theta^a}{\partial \bar{x}^m}\frac{\partial \theta^b}{\partial \bar{x}^n}\bar{\pi}^{mn}.$$

The functions $\pi^{ab}(x)$ were constructed in such a way that they automatically satisfy the transversality requirement.[20]

This result deserves a brief discussion because, at first sight, there might seem to be too much arbitrariness in $\pi^{ab}(x)$, which is seen to depend on six arbitrary functions of three variables $\bar{\pi}^{mn}(\bar{x})$. However, three of the six may be regarded as carrying information about a coordinate system, in this case, the coordinate system $(\bar{x})$. Although in this procedure one regards the coordinates (x) as fixed, the auxiliary coordinates $(\bar{x})$ are completely arbitrary and may therefore be transformed before the construction of a solution to (6) by means of *any* mapping $\Lambda : \bar{x} \to \bar{\bar{x}}$. This means we start from $\bar{\bar{\pi}}^{mn}(\bar{\bar{x}})$ instead of $\bar{\pi}^{mn}(\bar{x})$ in determining the solution to (6). However, this new

solution $\bar{\Theta}:\bar{\bar{x}} \to x$ is simply given by $\bar{\Theta} = \Theta \circ \Lambda^{-1}$, where $\circ$ denotes the composition of mappings. The details of the proof will not be given as it follows readily from the tensor character of the transversality conditions (1). It follows that all diffeomorphically equivalent tensor densities $\bar{\pi}^{mn}$ lead to the same $\pi^{ab}(x)$. Hence the $\pi^{ab}(x)$ depend on only three *essential* arbitrary functions of three variables, as expected. It was suggested earlier (Sec. 1) that the transversality condition can be satisfied by choosing three-dimensional coordinate conditions. In the procedure presented here, however, no coordinate conditions are imposed. The 3-geometry and the coordinates (x) in which the metric is displayed are picked freely before transverse momenta are constructed. There are no *a priori* restrictions on the coordinates $(\bar{x})$. Nevertheless, we now describe another interpretation of the coordinate mapping procedure which does show clearly the relation between the transversality conditions and coordinate conditions.

In the mapping procedure, one can equally well consider the inverse mapping $\Theta^{-1}: x \to \bar{x}$ and allow it to act on the metric while the momentum functions $\bar{\pi}^{ab}(\bar{x})$ are in the privileged position of being held fixed. Let us adopt this second viewpoint for the present. The mapping, taking $\gamma_{ab}(x)$ to $\bar{\gamma}_{ab}(\bar{x})$, can in no way influence the intrinsic 3-geometry itself. Therefore, given (i) any set of functions $\bar{\pi}^{ab}(\bar{x})$ and (ii) any 3-geometry (point in superspace), then there exists a metric $\bar{\gamma}_{ab}(\bar{x})$ describing the given 3-geometry such that $\bar{\nabla}_b \bar{\pi}^{ab} = 0$. Again, the reason that there are not actually six essential arbitrary functions in the transverse momentum, is that all diffeomorphically equivalent sets of momentum functions are physically and geometrically equivalent, as discussed earlier.

Despite the seeming reciprocity in the treatment of γ_{ab} and π^{ab} afforded by the mapping method, there is still an inherent difference for the following reason. Given a metric in any coordinate system, the 3-geometry is determined. But given a tensor π^{ab} in any coordinates, its relevant properties (e.g., divergence, trace) depend on knowing the metric in the same coordinate system. Unlike the metric, the momentum is not sufficient unto itself to define its own invariant properties.

We see that the method of constructing coordinate mappings as a way of treating the momentum constraint has two interpretations, of which one lends itself naturally to the view that the momentum constraints can be satisfied, for a given 3-geometry, by imposing coordinate conditions. It has been shown that the appropriate coordinate systems are generated by the mapping Θ that satisfies (6). The existence of Θ and its two interpretations give a further reflection of well-known and deep connections which exist between the momentum constraints and the three-dimensional coordinate group.

4. HAMILTONIAN CONSTRAINT

Thus far, the Hamiltonian constraint $\mathcal{H} = 0$ has not been taken into account. If we regard the metric as given, then $\mathcal{H}$ is quadratic algebraic in the momentum components, which prevents any neat method of solution because it is coupled to the momentum problem. What one might desire, particularly on physical grounds, is to decouple the momentum and Hamiltonian constraints, but this cannot be achieved if one regards the initial intrinsic 3-geometry as completely fixed in advance. However, the decoupling can be achieved in connection with a different choice of variables arbitrarily specifiable on the initial surface. This treatment is dealt with in Sec. 6. For the rest of the present section, we continue to regard either the full metric or the intrinsic 3-geometry as fixed.

To solve Eq. (2) simultaneously with constructing the map Θ, such that Θ will map the $\bar{\pi}^{ab}$ onto solutions of the complete initial value equations, it is necessary to choose another function to be unknown. In accord with our desire to keep the metric fixed, we choose one of the $\bar{\pi}^{ab}$, say $\bar{\pi}^{11}$, to be dependent. First we write $\mathcal{H} = 0$ in terms of the functions $\gamma_{ab}(x)$ and $\bar{\pi}^{ab}(\bar{x})$, which are known except for $\bar{\pi}^{11}$, and the unknown transformation coefficients $\partial\theta^r/\partial\bar{x}^a$. Then we solve algebraically for $\bar{\pi}^{11}$ as a function of the quantities listed above. The algebra is unenlightening and a little messy, so we omit it here. The resulting expression can then be substituted into Eq. (6). Observe that $\bar{\pi}^{11}$ enters this expression only by means of $\bar{\pi}^{11}$ and $\partial\bar{\pi}^{11}/\partial\bar{x}^1$. Therefore, no terms of the type $\partial^2\theta^r/(\partial\bar{x}^3)^2$ are introduced in eliminating $\bar{\pi}^{11}$. Since the system retains the Cauchy normal form,[19] solutions exist under the same conditions as in the previous section, except that $\bar{\pi}^{11}$ is no longer arbitrary. Hence when $\pi^{ab}(x)$ is constructed by means of Θ, it will be a solution to all the constraints with $\gamma_{ab}(x)$ fixed and will contain only *two* essential arbitrary functions of three variables.

As in the previous section, we may also consider that the momentum $\bar{\pi}^{ab}(\bar{x})$ is in the privileged position of being fixed in advance while the mapping acts on the metric $\gamma_{ab}(x)$. Now, however, only five of the six $\bar{\pi}^{ab}$ are arbitrary, the other being the unknown function permitting an algebraic elimination of the Hamiltonian constraint. We conclude that given the momentum $\bar{\pi}^{ab}$, of which functions five are arbitrary, and given any 3-geometry, there exists a metric representation $\bar{\gamma}_{ab}(\bar{x})$ of the 3-geometry such that $\bar{\nabla}_b \bar{\pi}^{ab} = 0$ and $\bar{\mathcal{H}} = 0$.

Although the problem of interest here is the initial value problem for vacuum space-times, there is a simple extension of the above result for the case when sources are present. There, one may specify in addition to $\gamma_{ab}(x)$ and five of the six $\bar{\pi}^{ab}$ also the current density $\mathcal{S}^a(x)$ of the sources and their energy density $\epsilon(x)$ as functions of the coordinates. The gravity initial value equations for this case,

$$\nabla_b \pi^{ab} = -\tfrac{1}{2}\mathcal{S}^a, \tag{7}$$

$$\gamma^{-1/2}(\pi^{ab}\pi_{ab} - 1/2\,\pi^2) - \gamma^{1/2}R = \epsilon, \tag{8}$$

again can be converted and solved by the coordinate mapping method just as above.[21]

5. CONFORMAL MAPPINGS AND $\tilde{C}$-MAPPINGS

The Hamiltonian constraint was accounted for by choosing one of the $\bar{\pi}^{ab}$ to be dependent. This is a rather inelegant method, but we saw no other choice because of the requirement that the metric or the 3-geometry be completely fixed in advance. However, we shall see below that a much better method is available if the transverse momenta constructed by the map Θ are also traceless. In this case, we shall show that the Hamiltonian constraint can be decoupled from the momentum constraints and can be accounted for by a method closely related to conform-

al mapping of the metric. Although we have to abandon the idea that the 3-geometry is completely fixed in advance, in return, we gain a more satisfying view of the initial value problem.

A "TT" tensor is one that is traceless and has vanishing covariant divergence (transverse). Suppose that the momentum tensor density is transverse and traceless relative to a given metric γ_{ab}, that is $\pi^{ab} = \pi^{a}_{\mathrm{T}}{}^{b}_{\mathrm{T}}$. We wish to show that there is a certain scale transformation denoted by $\bar{C}$, which preserves the TT character of the momentum. Define the mapping[9]

$$\begin{aligned} \bar{C} &: \gamma_{ab} \to \bar{\gamma}_{ab} = \phi^4 \gamma_{ab}, \\ &\pi^{ab} \to \bar{\pi}^{ab} = \phi^{-4} \pi^{ab}, \end{aligned} \tag{9}$$

where $\phi(x)$ is an arbitrary function. The first half of the $\bar{C}$ transformation is merely a conformal transformation of the metric. This leads to a transformation of the Christoffel symbols given by

$$\{b\ a\ c\} \to \{\overline{b\ a\ c}\} = \{b\ a\ c\} + 2\phi^{-1}(\delta^a_b \partial_c \phi + \delta^a_c \partial_b \phi - \gamma_{bc}\gamma^{ad}\partial_d \phi).$$

Substituting this expression and the transformation rule for π^{ab} into $\nabla_b \pi^{ab}$ yields

$$\nabla_b \pi^{ab} = \phi^4 \bar{\nabla}_b \bar{\pi}^{ab},$$

if π^{ab} is traceless. Therefore, the $\bar{C}$-map has the fundamental property that if π^{ab} is TT with respect to γ_{ab}, then $\bar{\pi}^{ab}$ will be TT with respect to $\bar{\gamma}_{ab}$. Further, it follows that the action of $\bar{C}$ on the weight $\frac{5}{3}$ momentum[22] $\tilde{\pi}^{ab} = \gamma^{1/3}\pi^{ab}$ is the identity: $\bar{C} : \tilde{\pi}^{ab} \to \tilde{\pi}^{ab}$. Of course, the multiplicative factor $\gamma^{1/3}$ does not affect the TT property. Therefore, once $\pi^{a}_{\mathrm{T}}{}^{b}_{\mathrm{T}}$ has been constructed by any means whatsoever for a given metric γ_{ab}, there is an infinite family of them generated by $\bar{C}$ corresponding to the conformally transformed metrics. This may be stated in another way. The construction of TT tensors depends only on the conformal structure of the underlying space. The stated invariance of $\tilde{\pi}^{a}_{\mathrm{T}}{}^{b}_{\mathrm{T}}$ with respect to $\bar{C}$ decouples the momentum and Hamiltonian constraints, as we now show.

For given γ_{ab} and $\pi^{a}_{\mathrm{T}}{}^{b}_{\mathrm{T}}$, the Hamiltonian constraint will not in general be satisfied. However we can now use a $\bar{C}$ transformation to insure that the Hamiltonian constraint will hold in barred variables, while the TT property is still maintained, as we showed above. By using Eq. (9) and the well-known conformal transformation of scalar curvature,

$$\bar{R} = \phi^{-4}R - 8\phi^{-5}\nabla^2\phi,$$

we find that ϕ must satisfy the Lichnerowicz[7] equation

$$\nabla^2\phi + \tfrac{1}{8}M\phi^{-7} - \tfrac{1}{8}R\phi = 0. \tag{10}$$

Here $\nabla^2 \equiv \gamma^{ab}\nabla_a\nabla_b$ is the Laplacian associated with the given metric γ_{ab} and

$$M \equiv \gamma^{-7/6}\gamma_{ac}\gamma_{bd}\tilde{\pi}^{a}_{\mathrm{T}}{}^{b}_{\mathrm{T}}\tilde{\pi}^{c}_{\mathrm{T}}{}^{d}_{\mathrm{T}}.$$

All the coefficients in this equation for ϕ are known, inasmuch as γ_{ab} is specified arbitrarily and $\tilde{\pi}^{a}_{\mathrm{T}}{}^{b}_{\mathrm{T}}$ is assumed to be given. This equation is elliptic and in different but equivalent form was studied by Lichnerowicz.[7] It will not in itself be subject to further inquiry here. The result we wish to stress at this point is that $\tilde{\pi}^{a}_{\mathrm{T}}{}^{b}_{\mathrm{T}}$ may be regarded as completely unconstrained (independent of all four constraints) because it is insensitive to the solution ϕ of (10). Thus, we regard the Hamiltonian constraint as determining the conformal factor of the metric, with $\tilde{\pi}^{a}_{\mathrm{T}}{}^{b}_{\mathrm{T}}$ unaffected by this determination. The Hamiltonian constraint, as well as the momentum constraints, can thus be treated by mapping arbitrary functions onto solutions. The physical naturalness of this interpretation is discussed in the final section.

6. TRANSVERSE TRACELESS MOMENTA

The additional requirement that the momentum be traceless is often regarded as a condition on the choice of a timelike coordinate–that is, space–time is sliced in such a way that the spacelike hypersurface with metric γ_{ab} is embedded "maximally". This means that the volume of any finite portion of the 3-space is unchanged by infinitesimal deformations in the orthogonal (timelike) direction. To maintain the criterion of tracelessness for a finite time cannot always be achieved as a time-coordinate condition on all compact 3-manifolds, as pointed out by DeWitt.[15] On the other hand, Deser[23] and others,[24] have shown how to decompose symmetric tensors covariantly with respect to a given metric in such a way that their TT parts are obtained without the necessity of imposing slicing conditions. The method of covariant decomposition is probably the most natural and powerful tool in the treatment of the momentum constraints; however, this particular matter must be deferred at present. The objective of the present section is to show that the traceless requirement can be incorporated into the problem of finding the mapping Θ discussed earlier.

We must now construct Θ subject to constraint $\pi = 0$, or

$$\pi = \gamma_{ab}\pi^{ab} = \left[\det\left(\frac{\partial\theta^c}{\partial\bar{x}^d}\right)\right]\frac{\partial\theta^a}{\partial\bar{x}^m}\frac{\partial\theta^b}{\partial\bar{x}^n}\bar{\pi}^{mn} = 0, \tag{11}$$

where the Jacobian of the mapping can be divided out since we assume it does not vanish. We may consider that $\pi = 0$ merely makes one of the six $\bar{\pi}^{ab}$ dependent so that five of them remain arbitrary. In turn, this means that there will be only two essential arbitrary functions of three variables in the solution for Θ. Solving (11) for $\bar{\pi}^{11}$, for example, and substituting into the basic equation (5), we see that no terms of the type $\partial^2\theta^r/(\partial\bar{x}^3)^2$ are introduced. Therefore the character of the equations is unaffected and the Cauchy–Kowalewski theorem[19] applies as before, subject to the same requirements. Also as before, it is helpful to think of this new coordinate mapping Θ_{TT} as acting on the metric with the $\bar{\pi}^{ab}$ remaining untransformed. Hence, for a given 3-geometry and a set of functions $\bar{\pi}^{ab}(\bar{x})$, of which all but one are arbitrary, there exists a metric representation $\bar{\gamma}_{ab}(\bar{x})$ of the given 3-geometry such that $\bar{\gamma}_{ab}\bar{\pi}^{ab} = 0$ and $\bar{\nabla}_b\bar{\pi}^{ab} = 0$.

7. THE INITIAL VALUE EQUATIONS AND GAUGE-LIKE TRANSFORMATIONS

Once one has constructed $\pi^{a}_{\mathrm{T}}{}^{b}_{\mathrm{T}}$ by determining the mapping Θ_{TT} or by any other method, then $\tilde{\pi}^{a}_{\mathrm{T}}{}^{b}_{\mathrm{T}}$ is immediately known. This object is $\bar{C}$ invariant and is therefore independent of the Hamiltonian constraint. The latter determines ϕ and, hence, the $\bar{C}$ transformation which is needed to complete the solution of the problem. Thus, the four unknowns determined by the four constraint equations in this approach are three functions θ^a, which define a coordinate

transformation and one function ϕ, which defines a scaling transformation $\bar{C}$.

We see that in this approach the full 3-geometry is not arbitrarily specifiable in advance; only the *conformally invariant* part of 3-geometry has the status of being completely unconstrained. That this result is fundamental in the theory of gravitation has been shown recently[9] to follow from conformally invariant representations of three-dimensional geometry. What we may call the three-dimensional conformal curvature tensor[25] is defined by[26]

$$\tilde{\beta}^{ab} \equiv \gamma^{1/3}\epsilon^{ef(a}\gamma^{b)m}\nabla_e R_{mf}, \tag{12}$$

where ϵ^{efa} is the completely antisymmetric unit tensor with $\epsilon^{123} = +1$ and R_{mf} is the Ricci tensor. This object has quite remarkable properties which can be summarized in a theorem.

Theorem: For every spacelike 3-geometry, there exists a tensor density $\tilde{\beta}^{ab}$ of weight 5/3 defined by (12), which is (i) symmetric, (ii) trace-free, (iii) conformally invariant, (iv) covariantly transverse, and (v) vanishes if and only if the 3-space is conformally flat.

That $\tilde{\beta}^{ab}$ is purely intrinsic to 3-geometry and is conformally invariant, and therefore $\bar{C}$-invariant, show that it is unconstrained by the initial value equations. Its further properties, notably that it is identically transverse and traceless, show that it gives the general "pure spin-two" representation of 3-geometry. These properties lead one to identify the three-dimensional conformal curvature tensor $\tilde{\beta}^{ab}$ with gravitational radiation.[27] Characterizing gravitational radiation by means of $\tilde{\beta}^{ab}$ and $\tilde{\pi}^{a}_{\mathrm{T}}{}^{b}_{\mathrm{T}}$ enables the recognition that the $\bar{C}$-transformations for *arbitrary* ϕ are "gaugelike" in that they leave the radiation field invariant. Unlike the case in electrodynamics, however, this gauge function is not arbitrary in terms of the *complete* initial value problem, which does not deal solely with purely transverse fields. On the contrary, this factor is determined by the Hamiltonian constraint, which thus may be thought of as playing the role of a built-in "gauge" condition.

In conclusion, we see that general unconstrained initial data, i.e., the gravitational degrees of freedom, are represented by a pair of transverse, traceless tensors $\tilde{\beta}^{ab}$ and $\tilde{\pi}^{a}_{\mathrm{T}}{}^{b}_{\mathrm{T}}$. Each of these objects provides an irreducible representation of both the coordinate group $\mathfrak{D}_3$ and the group of scale transformations defined by $\bar{C}$. It has been shown that elements of these groups can be used in mapping from sets of arbitrary functions to solutions of the initial value equations.

ACKNOWLEDGEMENT

I wish to thank K. Kuchař for many helpful discussions of the initial value problem and related issues and for his remarks on this paper.

---

* Work assisted in part by National Science Foundation Grant GP7669 and USAF Office of Scientific Research Grant AF49(638)-1545 to Princeton University.

1 The initial value problem is reviewed by Y. Choquet-Bruhat, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).

2 P. A. M. Dirac, Proc. Roy. Soc. (London) **A246**, 333 (1958).

3 R. Arnowitt, S. Deser, and C. W. Misner, Ref. 1, Chap. 7.

4 R. F. Baierlein, D. H. Sharp, and J. A. Wheeler, Phys. Rev. **126**, 1864 (1962); E. P. Belasco and H. C. Ohanian, J. Math. Phys. **10**, 1503 (1969).

5 See, for example, "Superspace," by J. A. Wheeler, in *Analytic Methods in Mathematical Physics*, edited by R. P. Gilbert and R. G. Newton (Gordon and Breach, New York, 1970), and "The Theory of Superspace" by A. E. Fischer, in *Relativity*, edited by M. Carmeli, S. I. Fickler, and L. Witten (Plenum, New York, 1970).

6 P. W. Higgs, Phys. Rev. Letters **1**, 373 (1958); **3**, 66 (1959).

7 A. Lichnerowicz, J. Math. Pure Appl. **23**, 37 (1944).

8 D. Brill, Ann. Phys. (N.Y.) **7**, 466 (1959).

9 J. W. York, Phys. Rev. Letters **26**, 1656 (1971).

10 See, for example, L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, N.J., 1926), p. 146.

11 Four-dimensional objects are denoted explicitly by a superscript[4]; other objects refer to 3-space.

12 See, for example, R. K. Sachs, "Gravitational Radiation," in *Relativity, Groups, and Topology*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964).

13 C. W. Misner, Phys. Rev. Letters **22**, 1071 (1969); "Classical and Quantum Dynamics of a Closed Universe," see Ref. 5, Carmeli *et al.*

14 K. Kuchař, Phys. Rev. D **4**, 955 (1971).

15 See, for example, B. S. DeWitt, Phys. Rev. **160**, 1113 (1967).

16 C. M. Pereira, Ph. D. thesis (Case Western Reserve University, 1967); also "Non-Metrical Specification of Space-Time Sources," (preprint, 1971).

17 This equation, with $\bar{\pi}^{bc}$ replaced by $(-g)^{1/2}\bar{T}^{\mu\nu}$ and all indices and functions referring to four dimensions, was given by Pereira, Ref. 16. The present application and the remainder of this discussion are those of the present author.

18 C. Miranda, *Partial Differential Equations of Elliptic Type*, transl. by Z. C. Motteler (Springer-Verlag, Berlin, 1970), p. 180.

19 See, for example, R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. II.

20 This method of constructing transverse momenta might seem rather "local" and perhaps restricted because it works directly with the coordinates and because the Eqs. (5) are of a type about which we can say very little beyond what is afforded by the Cauchy-Kowalewski kind of analysis. However, any limitations are probably not as serious as they might appear. This can be inferred from the fact that powerful global results are available in the closely related problem of decomposing symmetric tensors. For example, on any compact C^∞ Riemannian manifold, the transverse part of an arbitrary symmetric tensor can be found by a method of decomposition similar to the well-known decomposition of vectors into their transverse and longitudinal parts. See Ref. 23 and a paper by M. Berger and D. Ebin, J. Diff. Geom. **3**, 379 (1960). Thanks are due to S. Deser for the latter reference.

21 Essentially, all the results up to this point were summarized in the author's report "Initial Value Problem of Geometrodynamics," abstract published in Bull. Am. Phys. Soc. **15**, 660 (1970).

22 The weight 5/3 momentum was used by P. A. M. Dirac, Phys. Rev. **114**, 924 (1959).

23 S. Deser, Ann. Inst. Henri Poincaré **7**, 149 (1967).

24 D. Brill and S. Deser, Ann. Phys. (N.Y.) **50**, 584 (1968); D. Brill, "Isolated Solutions in General Relativity," to be published in the Petrov jubilee volume *Gravitation: Problems and Prospects*.

25 It is well known that the Weyl conformal curvature tensor vanishes identically for 3-space. An object playing an analogous role for 3-space, however, can be found in L. P. Eisenhart, *Riemannian Geometry* (Princeton U.P., Princeton, N.J., 1926), p. 91. This object, called R_{abc}, is related to (12) through the relation $\tilde{\beta}^{ab} = -\frac{1}{2}\gamma^{1/3}\epsilon^{ef(a}\gamma^{b)m}R_{mef}$.

26 The definition (12) leads us to call $\tilde{\beta}^{ab}$ the "symmetrized curl of the Ricci tensor" and is completely equivalent to the definition given in Ref. 9. In (12), to the Ricci tensor may be added any term of the form $\mu\gamma_{mf}$, where μ is an arbitrary function.

27 See Ref. 3, p. 250 the tensor R_{abc} of reference 25 in connection with gravitational radiation was mentioned, but its full role was not understood (S. Deser, personal communication).

Operator Treatment of the Gel'fand-Naimark Basis for $SL(2, C)$*

Itzhak Bars† and Feza Gürsey‡
Physics Department, Yale University, New Haven, Connecticut 06520
(Received 6 July 1971)

An operator form is developed to treat the Gel'fand-Naimark z basis for the homogeneous Lorentz group. It is shown that the operator Z with eigenvalues z is a definite operator-valued function of the generators of $SL(2,C)$. A unified formulation of the unitary representations of the Lorentz group is obtained in a Hilbert space endowed with an affine metric operator G whose functional dependence on the generators is derived explicitly. The Dirac bra-ket formalism is extended by making a distinction between covariant and contravariant state vectors. The matrix elements of G are shown to coincide with the intertwining operator of Gel'fand and co-workers. The principal series, the supplementary series, and the two kinds of integer point representations are unified by means of a single scalar product involving the metric operator.

1. INTRODUCTION

The unitary representations of the homogeneous Lorentz group have been obtained and classified by Bargmann[1] and Gel'fand and Naimark[2] who used different methods. The discussion of representations at integer points is due to Gel'fand and Vilenkin and is treated in the book on Generalized Functions of Gel'fand, Graev, and Vilenkin.[3] For extensive reviews of the representations of the Lorentz group, the reader is referred to Ref. 3 and the books by Gel'fand, Minlos, and Shapiro,[4] Naimark,[5] and Rühl.[6]

Physicists who, until recently, were content to use the finite dimensional nonunitary representations of $SL(2,C)$ in connection with covariant fields, are now motivated to familiarize themselves with the general representation theory in the light of several new developments in particle physics. One is the harmonic analysis of scattering amplitudes with respect to $SL(2,C)$ started by Toller.[7] Another approach, initiated by Nambu,[8] is the use of infinite component fields to represent an infinite family of particles suggested by the Regge classification of hadrons. A third reason is the formal similarity between $SL(2,C)$ and the internal invariance group for dual amplitudes, as first noticed by Domokos *et al.*[9] The Gel'fand-Naimark basis is particularly useful for the study of dual amplitudes in the Koba-Nielsen form[9-11] and the integer point representations of $SL(2,C)$ play an important role.

In view of such considerations, it might be useful to develop a compact operator formulation, in accordance with the quantum mechanical treatment of groups of physical relevance to reformulate the Gel'fand-Naimark theory of the representations of $SL(2,C)$ which was originally formulated with the method of homogeneous functions.

Our method can be outlined as follows. We start with the Lie algebra of the abstract operators that represent the infinitesimal generators of the Lorentz group. We then construct two functions of the generators which, together with the Casimir operators, form a maximal set of commuting operators. The simultaneous eigenstates of this set define a certain vector space. The z basis is labeled by the Hermitian and anti-Hermitian parts of an operator valued function

$$Z = Z(J_{\mu\nu}), \tag{1.1}$$

which becomes a rational function of the generators within an irreducible representation. Thus it belongs to an operator space which constitutes a generalization of the enveloping algebra of the elements $J_{\mu\nu}$ of the Lie algebra. The eigenstates, labeled by the eigenvalue z of Z, form basis vectors in this space. A general vector $|f)$ is regarded as a wave packet with components $f(z)$ in this vector space. Now the operator Z has been chosen in such a way that

$$U^{-1}(\Lambda)ZU(\Lambda) = (aZ + b)/(cZ + d), \tag{1.2}$$

where

$$\Lambda = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,C). \tag{1.3}$$

We also introduce an operator $\bar{Z}$ such that its eigenvalue z^* transforms with the complex conjugate $SL(2,C)$ matrix Λ^*. We show that

$$\bar{Z} = GZ^\dagger G^{-1} \neq Z^\dagger, \tag{1.4}$$

hence the simultaneous eigenstates of Z and $\bar{Z}$ are different from the simultaneous eigenstates of $\bar{Z}^\dagger$ and $Z^\dagger$. Thus we are led to introduce two kinds of basis vectors which we call covariant and contravariant. They are related by the operator G which is also a definite operator-valued function of the generators. Then G plays the role of the metric in an infinite dimensional affine vector space. With its matrix elements which coincide with Gel'fand's intertwining operator,[3] we can transform the covariant components $f(z)$ of a vector into its contravariant components $\tilde{f}(z)$.

This is a new situation in physics. The metric in Hilbert space is usually taken to be Euclidean. Although spaces with indefinite metric have been considered in quantum field theory, the representation theory of $SL(2,C)$ yields naturally a Hilbert space endowed with a metric G. We are then led to generalize Dirac's bra and ket formalism in quantum mechanics and distinguish covariant and contravariant bras and kets. Just as in ordinary affine space, the scalar product of two vectors can be written in terms of their covariant components and the elements of the metric tensor. Hence the scalar product of two vectors $|f)$ and $(g|$ involves, besides their covariant components $f(z)$ and $g(z')$, also the matrix elements of the metric operator G. The form of these matrix elements depends on the particular class of unitary representations considered. Thus, from the general form of G associated with the representation (j_1, j_2) which we show to be

$$G = \Pi^{-\rho}(\Pi^{-\rho})^\dagger \tag{1.5}$$

with

$$\rho = j_1 + j_2^* + 1 \tag{1.6}$$

and

$$\Pi = J_{23} + J_{14} + i(J_{31} + J_{24}), \tag{1.7}$$

we derive the various forms of the scalar product for the principal series, the supplementary series, and the two kinds of integer point representations from a single formula. Although partial unifications

of the different scalar products without the benefit of the operator formalism can be found in Refs. 3–10, we believe this is the first unified treatment of all the unitary representations of $SL(2,C)$.

The affine Hilbert spaces and the operator method introduced in this paper may find applications in the representation theory of other noncompact groups relevant to physics such as the de Sitter groups or the conformal group.

2. THE Z BASIS OF $SL(2,C)$

The generators of the infinitesimal transformations of $SL(2,C)$ are $J_{\mu\nu} = -J_{\nu\mu}$ with the commutation relations

$$[J_{\kappa\lambda}, J_{\mu\nu}] = i(\delta_{\kappa\mu}J_{\lambda\nu} + \delta_{\lambda\nu}J_{\kappa\mu} - \delta_{\kappa\nu}J_{\lambda\mu} - \delta_{\lambda\mu}J_{\kappa\nu}), \quad (2.1)$$

where the greek indices run from 1 to 4.

Define the rotation operators J_k and the boost operators K_n through the equations

$$J_k = \tfrac{1}{2}\epsilon_{klm}J_{lm}, \quad K_n = iJ_{4n}, \quad (2.2)$$

where the latin indices take the values 1, 2, 3. The commutation relations take the familiar form

$$[J_k, J_l] = i\epsilon_{klm}J_m, \quad (2.3a)$$

$$[K_k, K_l] = -i\epsilon_{klm}J_m, \quad (2.3b)$$

$$[J_k, K_l] = i\epsilon_{klm}K_m. \quad (2.3c)$$

We shall also use the self-dual and anti-self-dual combinations

$$M_{\mu\nu} = \tfrac{1}{2}(J_{\mu\nu} + \tilde{J}_{\mu\nu}), \quad N_{\mu\nu} = \tfrac{1}{2}(J_{\mu\nu} - \tilde{J}_{\mu\nu}) \quad (2.4)$$

with

$$\tilde{J}_{\mu\nu} = \tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}J_{\alpha\beta} \quad (2.5)$$

and define the left- and right-handed vector operators

$$X^L_k = M_{k4} = \tfrac{1}{2}\epsilon_{klm}M_{lm}, \quad X^R_k = -N_{k4} = \tfrac{1}{2}\epsilon_{klm}N_{lm}. \quad (2.6)$$

In vector notation, we have

$$\mathbf{X}_L = \tfrac{1}{2}(\mathbf{J} + i\mathbf{K}), \quad \mathbf{X}_R = \tfrac{1}{2}(\mathbf{J} - i\mathbf{K}) \quad (2.7)$$

with commutators

$$[X^L_m, X^R_n] = 0, \quad (2.8a)$$

$$[X^L_k, X^L_l] = i\epsilon_{klm}X^L_m, \quad (2.8b)$$

$$[X^R_k, X^R_l] = i\epsilon_{klm}X^R_m. \quad (2.8c)$$

The two Casimir operators that commute with $J_{\mu\nu}$ are

$$F_1 = \tfrac{1}{4}J_{\mu\nu}J_{\mu\nu} = \tfrac{1}{2}(\mathbf{J}\cdot\mathbf{J} - \mathbf{K}\cdot\mathbf{K}) = \mathbf{X}_L\cdot\mathbf{X}_L + \mathbf{X}_R\cdot\mathbf{X}_R, \quad (2.9a)$$

$$F_2 = \tfrac{1}{4}J_{\mu\nu}\tilde{J}_{\mu\nu} = i\mathbf{J}\cdot\mathbf{K} = \mathbf{X}_L\cdot\mathbf{X}_L - \mathbf{X}_R\cdot\mathbf{X}_R. \quad (2.9b)$$

Under a parity transformation

$$I_S : \mathbf{J} \to \mathbf{J}, \quad \mathbf{K} \to -\mathbf{K}, \quad (2.10a)$$

we have

$$F_1 \to F_1, \quad F_2 \to -F_2, \quad (2.10b)$$

so that F_1 is a Lorentz scalar and F_2 a Lorentz pseudoscalar. Instead of these parity eigenstates we shall often use the alternative Casimir operators

$$C_1 = \tfrac{1}{2}(F_1 + F_2) = \mathbf{X}_L\cdot\mathbf{X}_L,$$
$$C_2 = \tfrac{1}{2}(F_1 - F_2) = \mathbf{X}_R\cdot\mathbf{X}_R, \quad (2.11)$$

which are transformed into each other under parity.

The irreducible representations are labeled by the eigenvalues of C_1 and C_2 which define the numbers (in general complex) j_1 and j_2 through the equations

$$C_1 = j_1(j_1 + 1), \quad C_2 = j_2(j_2 + 1). \quad (2.12)$$

The bases usually employed are

(1) The covariant (m_1, m_2) basis defined by the simultaneous eigenstates of the four commuting operators C_1, C_2, X^L_3, and X^R_3, with m_1 and m_2 being the eigenvalues of the last two operators;

(2) The canonical (j, m) basis defined by the simultaneous eigenstates of the four commuting operators C_1, C_2 (or $F_1, -iF_2$), $\mathbf{J}\cdot\mathbf{J}$ with eigenvalues $j(j+1)$ and J_3 with eigenvalues m.

We now introduce the Gel'fand–Naimark z basis as the simultaneous eigenstates of C_1, C_2, and Z, where $Z = Z(C_1, C_2, X^L_k)$ is a certain non-Hermitian operator that is a rational function of $J_{\mu\nu}$. This operator should satisfy the following conditions.

(i) In order to have four commuting operators C_1, C_2, and the Hermitian and anti-Hermitian parts of Z, we must have

$$[Z + Z^\dagger, Z - Z^\dagger] = 0$$
or
$$[Z, Z^\dagger] = 0. \quad (2.13)$$

(ii) Let z be the eigenvalue of Z. In order for z to transform as

$$z' = \Lambda z = \frac{az + b}{cz + d} \quad (2.14)$$

under an $SL(2,C)$ transformation with elements

$$\Lambda = e^{i(\sigma\cdot\alpha)/2} = e^{i[\sigma\cdot(\omega - i\nu)]/2} = \begin{pmatrix} ab \\ cd \end{pmatrix}, \quad ad - bc = 1, \quad (2.15)$$

we must have Z transforming as the ratio of two operators Ψ_1 and Ψ_2 that, under $SL(2,C)$, transform like the components of a spinor.

We may define Ψ_1 and Ψ_2 as functions of $J_{\mu\nu}$ if Ψ with components Ψ_1 and Ψ_2 satisfies the equation

$$(\sigma\cdot\mathbf{X}_L)\Psi = \Omega\Psi, \quad \Psi = \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix}, \quad (2.16)$$

where σ_i are the 2×2 Pauli matrices. In other words, we define Ψ as one of the eigenspinors of the 2×2 matrix of generators and Ω as the eigenvalue operator associated with Ψ. The invariant Ω is a function of C_1 [Eq. (2.19)]. Using

$$\mathbf{X}_L \times \mathbf{X}_L = i\mathbf{X}_L, \quad (2.17)$$

which is equivalent to (18.b), we obtain

$$(\sigma\cdot\mathbf{X}_L)^2\Psi = (\mathbf{X}_L\cdot\mathbf{X}_L - \sigma\cdot\mathbf{X}_L)\Psi,$$

or, with the help of the definition (2.11),

$$C_1\Psi = \sigma\cdot\mathbf{X}_L(\sigma\cdot\mathbf{X}_L + 1)\Psi = \Omega(\Omega + 1)\Psi. \tag{2.18}$$

Hence we find for Ω two operator solutions

$$\Omega = \Omega_1 = -\tfrac{1}{2} + (C_1 + \tfrac{1}{4})^{1/2} \tag{2.19a}$$

or

$$\Omega = \Omega_1' = -\tfrac{1}{2} - (C_1 + \tfrac{1}{4})^{1/2}, \tag{2.19b}$$

which satisfy

$$\Omega_1 + \Omega_1' = -1 \tag{2.20a}$$

and

$$-\Omega_1\Omega_1' = C_1 = \Omega_1(\Omega_1 + 1) = \Omega_1'(\Omega_1' + 1). \tag{2.20b}$$

In the following we shall call Ψ the spinor operator associated with Ω_1. Thus we have the homogeneous operator-matrix equation

$$\begin{pmatrix} X_3^L - \Omega_1 & X_1^L - iX_2^L \\ X_1^L + iX_2^L & -X_3^L - \Omega_1 \end{pmatrix}\begin{pmatrix}\Psi_1 \\ \Psi_2\end{pmatrix} = 0. \tag{2.21}$$

The compatibility of the two linear equations is assured by the relations (2.11) and (2.20b). We now define the operator Z by

$$\begin{aligned} Z = \Psi_1\Psi_2^{-1} &= (\Omega_1 - X_3^L)^{-1}(X_1^L - iX_2^L) \\ &= (X_1^L + iX_2^L)^{-1}(\Omega_1 + X_3^L). \end{aligned} \tag{2.22}$$

Now, consider a finite Lorentz transformation with rotation parameters ω and boost parameters ν. Let

$$U = U(\Lambda) = e^{i\mathbf{J}\cdot\omega + i\mathbf{K}\cdot\nu} = e^{i\boldsymbol{\alpha}\cdot\mathbf{X}_L + i\boldsymbol{\alpha}^*\cdot\mathbf{X}_R}, \tag{2.23a}$$

where

$$\boldsymbol{\alpha} = \omega - i\nu, \qquad \boldsymbol{\alpha}^* = \omega + i\nu. \tag{2.23b}$$

We find

$$U^{-1}\sigma\cdot\mathbf{X}_L U = e^{(i\alpha/2)\cdot\alpha}\sigma\cdot\mathbf{X}_L e^{(-i/2)\sigma\cdot\alpha} = \Lambda\sigma\cdot\mathbf{X}_L\Lambda^{-1}, \tag{2.24}$$

where Λ is defined by (2.15). This relation can easily be verified for infinitesimal $\boldsymbol{\alpha}$. Similarly, we find

$$U^{-1}\sigma\cdot\mathbf{X}_R U = e^{(i/2)\sigma\cdot\alpha^*}\sigma\cdot\mathbf{X}_R e^{(-i/2)\sigma\cdot\alpha^*} = \hat{\Lambda}\sigma\cdot\mathbf{X}_R\hat{\Lambda}^{-1}, \tag{2.25}$$

where

$$\hat{\Lambda} = \sigma_2\Lambda^*\sigma_2 = \begin{pmatrix} d^* & -c^* \\ -b^* & a^* \end{pmatrix}. \tag{2.26}$$

The 2×2 matrix operator that transforms with Λ^* is then $-\sigma^*\cdot\mathbf{X}_R$, since we have

$$\begin{aligned} U^{-1}(-\sigma^*\cdot\mathbf{X}_R)U &= U^{-1}\sigma_2\sigma\cdot\mathbf{X}_R\sigma_2 U = \sigma_2\hat{\Lambda}\sigma\cdot\mathbf{X}_R\hat{\Lambda}^{-1}\sigma_2 \\ &= \Lambda^*(-\sigma^*\cdot\mathbf{X}_R)\Lambda^{*-1} \end{aligned} \tag{2.27a}$$

with

$$\Lambda^* = \begin{pmatrix} a^* & b^* \\ c^* & d^* \end{pmatrix}. \tag{2.27b}$$

Let

$$-\sigma^*\cdot\mathbf{X}_R\Phi = \Omega_2\Phi. \tag{2.28}$$

As before we have

$$C_2 = \mathbf{X}_R\cdot\mathbf{X}_R = \Omega_2(\Omega_2 + 1) = \Omega_2'(\Omega_2' + 1), \tag{2.29}$$

where Ω_2 and Ω_2' are defined as the roots

$$\begin{aligned} \Omega_2 &= -\tfrac{1}{2} + (C_2 + \tfrac{1}{4})^{1/2}, \\ \Omega_2' &= -\Omega_2 - 1 = -\tfrac{1}{2} - (C_2 + \tfrac{1}{4})^{1/2}. \end{aligned} \tag{2.30}$$

Φ satisfies the homogeneous equation

$$\begin{pmatrix} -X_3^R - \Omega_2 & -X_1^R - iX_2^R \\ -X_1^R + iX_2^R & X_3^R - \Omega_2 \end{pmatrix}\begin{pmatrix}\Phi_1 \\ \Phi_2\end{pmatrix} = 0 \tag{2.31}$$

which is the right-handed analog of (2.21). We now define a non-Hermitian operator $\overline{Z}$ by

$$\begin{aligned} \overline{Z} = \Phi_1\Phi_2^{-1} &= (-\Omega_2 - X_3^R)^{-1}(X_1^R + iX_2^R) \\ &= (X_1^R - iX_2^R)^{-1}(-\Omega_2 + X_3^R). \end{aligned} \tag{2.32}$$

To find the transformation law for Z and $\overline{Z}$, we must first obtain the transformation law for Ψ and Φ. To do so we apply U^{-1} and U to both sides of Eqs. (2.16) and (2.28):

$$\begin{aligned} U^{-1}(\Lambda)\sigma\cdot\mathbf{X}_L U(\Lambda)U^{-1}(\Lambda)\Psi U(\Lambda) &= \Omega_1 U^{-1}(\Lambda)\Psi U(\Lambda), \\ U^{-1}(\Lambda)\sigma^*\cdot\mathbf{X}_R U(\Lambda)U^{-1}(\Lambda)\Phi U(\Lambda) &= -\Omega_2 U^{-1}(\Lambda)\Phi U(\Lambda), \end{aligned}$$

where we have commuted $U(\Lambda)$ with $\Omega_{1,2}$ since they are functions of the Casimir operators.

Using Eqs. (2.24) and (2.27a), we can write

$$\begin{aligned} \sigma\cdot\mathbf{X}_L[\Lambda^{-1}U^{-1}(\Lambda)\Psi U(\Lambda)] &= \Omega_1[\Lambda^{-1}U^{-1}(\Lambda)\Psi U(\Lambda)], \\ \sigma^*\cdot\mathbf{X}_R[\Lambda^{-1*}U^{-1}(\Lambda)\Phi U(\Lambda)] &= -\Omega_2[\Lambda^{-1*}U^{-1}(\Lambda)\Phi U(\Lambda)]. \end{aligned}$$

Therefore, comparing to (2.16) and (2.28), we must have

$$\begin{aligned} \Lambda^{-1}U^{-1}(\Lambda)\Psi U(\Lambda) &= \lambda_1\Psi, \\ \Lambda^{-1*}U^{-1}(\Lambda)\Phi U(\Lambda) &= \lambda_2\Phi, \end{aligned}$$

where λ_1 and λ_2 are c-numbers. Thus, we obtain

$$\begin{aligned} U^{-1}(\Lambda)\Psi U(\Lambda) &= \lambda_1\Lambda\Psi, \\ U^{-1}(\Lambda)\Phi U(\Lambda) &= \lambda_2\Lambda^*\Phi. \end{aligned} \tag{2.33}$$

From the expressions (2.15) and (2.27b) for Λ and Λ^*, respectively, and the definitions (2.22) and (2.32), it follows that Z and $\overline{Z}$ obey the transformation laws

$$T(\Lambda)Z \equiv U^{-1}ZU = (aZ + b)/(cZ + d), \tag{2.34a}$$

$$T(\Lambda)\overline{Z} \equiv U^{-1}\overline{Z}U = (a^*\overline{Z} + b^*)/(c^*\overline{Z} + d^*). \tag{2.34b}$$

Furthermore, the commutation relations (2.8a) imply

$$[Z, \overline{Z}] = 0. \tag{2.35}$$

We note that Z and $\overline{Z}$ are not independent if we use the parity operators I_S of Eq. (2.10a) such that

$$I_S\mathbf{X}_L I_S^{-1} = \mathbf{X}_R, \qquad I_S\Omega_1 I_S^{-1} = \Omega_2. \tag{2.36}$$

From (2.22) and (2.32) we find the relation

$$-(\overline{Z})^{-1} = I_S Z I_S^{-1}. \tag{2.37}$$

The z basis is now defined by the simultaneous eigenstates of C_1, C_2, Z, and $\bar{Z}$. Let z represent the eigenvalue of Z on these states. Then the state $|z\rangle$ will transform to

$$|z\rangle \to U|z\rangle = \lambda(\Lambda, z)|(az + b)/(cz + d)\rangle, \tag{2.38}$$

where $\lambda(\Lambda, z)$ is a certain multiplier to be determined later. In order for the eigenvalue ζ of $\bar{Z}$ to transform as

$$\zeta \to (a^*\zeta + b^*)/(c^*\zeta + d^*),$$

we must have $\zeta = z^*$. Thus, we define the states $|j_1, j_2; z, z^*\rangle$ by

$$\Omega_1|j_1, j_2; z, z^*\rangle = j_1|j_1, j_2; z, z^*\rangle, \tag{2.39a}$$

$$\Omega_2|j_1, j_2; z, z^*\rangle = j_2|j_1, j_2; z, z^*\rangle, \tag{2.39b}$$

$$Z|j_1, j_2; z, z^*\rangle = z|j_1, j_2; z, z^*\rangle, \tag{2.39c}$$

$$\bar{Z}|j_1, j_2; z, z^*\rangle = z^*|j_1, j_2; z, z^*\rangle. \tag{2.39d}$$

Within a given representation (j_1, j_2) using the commutation relations, we obtain the expressions

$$\begin{aligned} Z &= (j_1 - X_3^L)^{-1}(X_1^L - iX_2^L) \\ &= (X_1^L - iX_2^L)(j_1 + 1 - X_3^L)^{-1} \\ &= (X_1^L + iX_2^L)^{-1}(j_1 + X_3^L) \\ &= (j_1 + 1 + X_3^L)(X_1^L + iX_2^L)^{-1}, \end{aligned} \tag{2.40a}$$

$$\begin{aligned} \bar{Z} &= (-j_2 - X_3^R)^{-1}(X_1^R + iX_2^R) \\ &= (X_1^R + iX_2^R)(-j_2 - 1 - X_3^R)^{-1} \\ &= (X_1^R - iX_2^R)^{-1}(-j_2 + X_3^R) \\ &= (-j_2 - 1 + X_3^R)(X_1^R - iX_2^R)^{-1}. \end{aligned} \tag{2.40b}$$

3. UNITARY REPRESENTATIONS

In this paper we shall be concerned with unitary representations so that $U(\Lambda)$ of Eq. (2.23a) is unitary. Thus $\mathbf{J}$ and $\mathbf{K}$ are Hermitian

$$\mathbf{J} = \mathbf{J}^\dagger, \quad \mathbf{K} = \mathbf{K}^\dagger \tag{3.1}$$

and

$$(\mathbf{X}_L)^\dagger = \mathbf{X}_R, \quad (C_1)^\dagger = C_2. \tag{3.2}$$

Thus, for the eigenvalues of C_1 and C_2 we have the relation

$$j_1^*(j_1^* + 1) = j_2(j_2 + 1) \tag{3.3a}$$

or

$$j_2 = -\tfrac{1}{2} \pm (j_1^* + \tfrac{1}{2}). \tag{3.3b}$$

We define $\mathbf{X}$ as

$$\mathbf{X}_L = \mathbf{X}, \quad \mathbf{X}_R = \mathbf{X}^\dagger. \tag{3.4}$$

and introduce j_0, k by

$$2j_1 + 1 = j_0 + k, \quad 2j_2 + 1 = -j_0 + k, \tag{3.5}$$

so that the Casimir operators F_1 and F_2 of Eqs. (2.9) have the eigenvalues

$$F_1 = \tfrac{1}{2}(\mathbf{J}\cdot\mathbf{J} - \mathbf{K}\cdot\mathbf{K}) = \tfrac{1}{2}(j_0^2 + k^2 - 1), \tag{3.6a}$$

$$F_2 = i\mathbf{J}\cdot\mathbf{K} = (j_1 + j_2 + 1)(j_1 - j_2) = j_0 k. \tag{3.6b}$$

Since F_1 and $-iF_2$ are Hermitian for unitary representations and, as we shall see below, j_0 is a real integer or half-integer, we have the well-known two cases for $k = \rho + i\sigma$:

(1) The principal series for which

$$\rho = j_1 + j_2^* + 1 = j_1^* + j_2 + 1 = 0, \quad k = -k^*; \tag{3.7}$$

(2) The supplementary series and integer points for which

$$j_0 + i\sigma = j_1 - j_2^* = 0, \quad k = k^*. \tag{3.8a}$$

Equation (3.8a) together with j_0 = real imply

$$j_1 = j_1^* = j_2 = j_2^*. \tag{3.8b}$$

These two cases come from

$$\rho(j_0 + i\sigma) = 0,$$

which is equivalent to (3.3a).

From the expressions (2.40) we obtain

$$Z = (j_1 - X_3)^{-1}X_- = (j_1 + 1 + X_3)X_+^{-1}, \tag{3.9}$$

$$\bar{Z} = (-j_2 - X_3^\dagger)^{-1}(X_-)^\dagger = (-j_2 - 1 + X_3^\dagger)(X_+^{-1})^\dagger, \tag{3.10}$$

where

$$X_\pm = X_1 \pm iX_2, \quad (X_\pm)^\dagger = X_1^\dagger \mp iX_2^\dagger. \tag{3.11}$$

We shall now derive a relation between $\bar{Z}$ and $Z^\dagger$. We have, from Eq. (2.40a),

$$Z^\dagger = (j_1^* + 1 - X_3^\dagger)^{-1}(X_-)^\dagger = (j_1^* + X_3^\dagger)(X_+^{-1})^\dagger. \tag{3.12}$$

$Z^\dagger$ satisfies the relation

$$[Z, Z^\dagger] = 0, \tag{3.13}$$

which allows its Hermitian and anti-Hermitian parts to be diagonalized simultaneously. We note the important commutation relation

$$[Z, \Pi] = 1, \quad \Pi = X_+ \tag{3.14a}$$

$$[Z^\dagger, \Pi^\dagger] = -1. \tag{3.14b}$$

The easily proved lemma,

$$[(\Pi^n)^\dagger, Z^\dagger] = [(X_1^\dagger - iX_2^\dagger)^n, Z^\dagger] = n(X_1^\dagger - iX_2^\dagger)^{n-1}, \tag{3.15}$$

leads to

$$f(\Pi^\dagger)Z^\dagger = Z^\dagger f(\Pi^\dagger) + f'(\Pi^\dagger), \tag{3.16}$$

where f' is the derivative of the function f with respect to its argument. Taking

$$f(\Pi^\dagger) = (\Pi^\dagger)^{-\rho} = (X_1^\dagger - iX_2^\dagger)^{-j_1^* - j_2 - 1}, \tag{3.17}$$

where ρ is defined by (3.7), we obtain from (3.16)

the relation

$$\overline{Z} = (\Pi^\dagger)^{-\rho} Z^\dagger (\Pi^\dagger)^\rho. \tag{3.18}$$

On the other hand, by virtue of

$$[X_l, X_m^\dagger] = 0, \tag{3.19}$$

we also have

$$[\Pi^\rho, Z^\dagger] = 0, \tag{3.20}$$

so that we can write

$$\overline{Z} = GZ^\dagger G^{-1} = Z^\dagger - \rho(X_1^\dagger - iX_2^\dagger)^{-1}, \tag{3.21}$$

where

$$G = G^\dagger = (\Pi^\dagger)^{-\rho}\Pi^{-\rho} = (\Pi^\dagger\Pi)^{-\rho} = (X_1^\dagger - iX_2^\dagger)^{-\rho}(X_1 + iX_2)^{-\rho}. \tag{3.22}$$

From (3.21) it follows that $\overline{Z}$ satisfies the commutation relation

$$[\overline{Z}, \Pi^\dagger] = -1. \tag{3.23}$$

Finally we note the relations

$$\overline{Z}^\dagger = G^{-1}ZG = Z - \rho(X_1 + iX_2)^{-1}, \tag{3.24}$$

$$[\overline{Z}^\dagger, \Pi] = 1, \tag{3.25}$$

$$[\overline{Z}^\dagger, Z^\dagger] = 0, [\Pi, \Pi^\dagger] = 0, \tag{3.26}$$

which show that both Z and $\overline{Z}^\dagger$ are canonical conjugate to Π, while $\overline{Z}$ and $Z^\dagger$ are canonical conjugate to $\Pi^\dagger$.

The operator G collapses to the unit operator in the case of the principal series for which we get

$$\overline{Z} = Z^\dagger \text{ when } \rho = 0. \tag{3.27}$$

As we shall see later, in general, G plays the role of a metric operator in Hilbert space.

The generators $\mathbf{X}$ and $\mathbf{X}^\dagger$ can now be expressed by means of $\Pi, \Pi^\dagger$ and their canonical conjugates. The result is

$$X_1 + iX_2 = \Pi, \tag{3.28a}$$

$$X_3 = -j_1 + \Pi Z, \tag{3.28b}$$

$$X_1 - iX_2 = 2j_1 Z - \Pi Z^2. \tag{3.28c}$$

The first equation is the definition of Π, while the second follows from

$$Z = (X_1 + iX_2)^{-1}(j_1 + X_3), \tag{3.29}$$

which is one of the forms of Z in (2.40a). Equation (3.28c) follows from (3.9) after replacing X_3 by its value given by Eq. (3, 28b).

Similarly, from the expressions of $\overline{Z}$ we find

$$X_1^\dagger - iX_2^\dagger = \Pi^\dagger, \tag{3.30a}$$

$$X_3^\dagger = j_2 + \Pi^\dagger \overline{Z}, \tag{3.30b}$$

$$X_1^\dagger + iX_2^\dagger = -2j_2\overline{Z} - \Pi^\dagger \overline{Z}^2. \tag{3.30c}$$

The generators $\mathbf{X}$ and $\mathbf{X}^\dagger$ can also be expressed by means of $\overline{Z}^\dagger$ and $Z^\dagger$. Equation (3.21) yields

$$\overline{Z} = Z^\dagger - \rho\Pi^{\dagger -1} \tag{3.31a}$$

and

$$Z = \overline{Z}^\dagger + \rho\Pi^{-1}, \tag{3.31b}$$

while from Eq. (3.25) we get

$$\Pi\overline{Z}^\dagger\Pi^{-1} = \overline{Z}^\dagger - \Pi^{-1}. \tag{3.32}$$

Using these results and (3.3a), we find

$$X_3 = j_2^* + 1 + \Pi\overline{Z}^\dagger, \tag{3.33a}$$

$$X_1 - iX_2 = -2(j_2^* + 1)\overline{Z}^\dagger - \Pi(\overline{Z}^\dagger)^2, \tag{3.33b}$$

$$X_3^\dagger = -(j_1^* + 1) + \Pi^\dagger Z^\dagger, \tag{3.34a}$$

$$X_1^\dagger + iX_2^\dagger = 2(j_1^* + 1)Z^\dagger - \Pi^\dagger Z^{\dagger 2}. \tag{3.34b}$$

Using the commutation relations (3.14), (3.23), and (3.24) we can now derive the Lie algebra (2.8) and the relations (2.12) from either set (3.28), (3.30) or (3.33), (3.34) together with (3.28a) and (3.30a). Thus we have obtained a unitary representation of the Lorentz group in terms of (Π, Z) and $(\Pi^\dagger, \overline{Z})$ or $(\Pi, \overline{Z}^\dagger)$ and $(\Pi^\dagger, Z^\dagger)$.

Note that from (3.28b) and (3.30b) we have

$$J_3 = X_3 + X_3^\dagger = -(j_1 - j_2) + \Pi Z + \Pi^\dagger \overline{Z}. \tag{3.35}$$

Introduce the operators

$$A = (1/\sqrt{2})(Z + \Pi^\dagger), \quad \overline{A} = (1/\sqrt{2})(\overline{Z} + \Pi), \tag{3.36a}$$

$$B = (1/\sqrt{2})(\overline{Z} - \Pi), \quad \overline{B} = (1/\sqrt{2})(Z - \Pi^\dagger), \tag{3.36b}$$

which obey the commutation relations

$$[A, \overline{B}] = 0, \quad [A, \overline{B}] = 0, \quad [\overline{A}, B] = 0, \quad [\overline{A}, \overline{B}] = 0, \tag{3.37a}$$

$$[A, \overline{A}] = 1, \quad [B, \overline{B}] = 1, \tag{3.37b}$$

typical of harmonic oscillator creation and annihilation operators. We can then recast (3.35) in the form

$$J_3 = -(j_1 - j_2) + \overline{A}A - \overline{B}B. \tag{3.38}$$

Since $\overline{A}A$ and $\overline{B}B$ must have integer values and J_3 has integer or half-integer eigenvalues m, it follows that $j_0 = j_1 - j_2$ is real and allowed to take integer or half-integer values.

4. DEFINITION AND TRANSFORMATION PROPERTIES OF COVARIANT AND CONTRAVARIANT KET VECTORS

The states $|j_1, j_2; z, z^*\rangle$ which are common eigenstates of C_1, C_2, Z, and $\overline{Z}$ will be called *covariant kets* and will be denoted by the more concise notation

$$|j_1, j_2; z, z^*\rangle = \left|\begin{matrix} \\ z \end{matrix}\right), \tag{4.1}$$

where we can omit (j_1, j_2) if we always work within

an irreducible representation of $SL(2,C)$. This notation is in analogy to the vector notation in a finite dimensional complex space where the basis vectors are chosen to be e_n, with the correspondence

$$e_n \leftrightarrow \left|\, z\right\rangle. \tag{4.2}$$

In the finite case, n is discrete and varies from 1 to N, while in our case the label z varies continuously in the complex plane. We have

$$\Omega_1 \left|\, z\right\rangle = \{-\tfrac{1}{2} + (C_1 + \tfrac{1}{4})^{1/2}\} \left|\, z\right\rangle = j_1 \left|\, z\right\rangle, \tag{4.3a}$$

$$\Omega_2 \left|\, z\right\rangle = \{-\tfrac{1}{2} + (C_2 + \tfrac{1}{4})^{1/2}\} \left|\, z\right\rangle = j_2 \left|\, z\right\rangle, \tag{4.3b}$$

$$Z \left|\, z\right\rangle = z \left|\, z\right\rangle, \quad \overline{Z} \left|\, z\right\rangle = z^* \left|\, z\right\rangle. \tag{4.3c}$$

We now introduce the *contravariant* basis ket vectors as common eigenstates of $C_1, C_2, \overline{Z}^\dagger$, and $Z^\dagger$ and denote them by an upper z label:

$$\Omega_1^\dagger \left|^z\right\rangle = j_1^* \left|^z\right\rangle, \quad \Omega_2^\dagger \left|^z\right\rangle = j_2^* \left|^z\right\rangle, \tag{4.4a}$$

$$\overline{Z}^\dagger \left|^z\right\rangle = z \left|^z\right\rangle, \quad Z^\dagger \left|^z\right\rangle = z^* \left|^z\right\rangle. \tag{4.4b}$$

Because of the relations (3.21) and (3.24), Eq. (4.3c) can be rewritten as

$$\overline{Z}^\dagger G^{-1} \left|\, z\right\rangle = z G^{-1} \left|\, z\right\rangle, \quad Z^\dagger G^{-1} \left|\, z\right\rangle = z^* G^{-1} \left|\, z\right\rangle. \tag{4.5}$$

This shows that with a suitable normalization we can write

$$\left|^z\right\rangle = G^{-1} \left|\, z\right\rangle, \quad G \left|^z\right\rangle = \left|\, z\right\rangle. \tag{4.6}$$

In case the basis vectors e_n are not orthogonal, we must distinguish between the contravariant basis vectors e^n and the covariant basis vectors e_n in the finite dimensional case. They are related by

$$e^m = e_n g^{nm}, \quad e^m g_{mn} = e_n \tag{4.7}$$

with

$$g^{ml} g_{ln} = \delta_n^m \tag{4.8}$$

in complete analogy with (4.6) if g_{mn} is a Hermitian metric tensor in N-dimensional complex space.

Let us now apply the unitary operator

$$U(\Lambda) = e^{i\mathbf{J}\cdot\boldsymbol{\omega} + i\mathbf{K}\cdot\boldsymbol{\nu}} = e^{i\boldsymbol{\alpha}\cdot\mathbf{X} + i\boldsymbol{\alpha}^*\cdot\mathbf{X}^\dagger} \tag{4.9}$$

on the covariant or contravariant ket vectors. [The complex parameters $\boldsymbol{\alpha}$ are given by (2.23b)]. The infinitesimal transformation has the form

$$U(\Lambda) \approx 1 + i\boldsymbol{\alpha}\cdot\mathbf{X} + i\boldsymbol{\alpha}^*\cdot\mathbf{X}^\dagger \tag{4.10}$$

with $\mathbf{X}$ and $\mathbf{X}^\dagger$ given by (3.28), (3.30) or (3.33), and (3.34). From the commutation relations (3.14), (3.23), and (3.25) and from the properties of kets as spelled out by (4.3) and (4.4), we find

$$\begin{aligned} \left|\, z\right\rangle &= Z\Pi \left|\, z\right\rangle - \Pi Z \left|\, z\right\rangle = Z\Pi \left|\, z\right\rangle - \Pi z \left|\, z\right\rangle \\ &= Z\Pi \left|\, z\right\rangle - z\Pi \left|\, z\right\rangle, \end{aligned} \tag{4.11}$$

where we have commuted the C-number z with the operator Π. Since Π is the canonical conjugate of Z, it acts on the states as

$$\Pi \left|\, z\right\rangle = \frac{\partial}{\partial z} \left|\, z\right\rangle \tag{4.12}$$

just as the momentum in quantum mechanics.

Indeed, since $\partial/\partial z$ is a C-number, it commutes with the operator Z and we can write

$$\begin{aligned} Z\Pi \left|\, z\right\rangle &= Z \frac{\partial}{\partial z} \left|\, z\right\rangle = \frac{\partial}{\partial z} Z \left|\, z\right\rangle = \frac{\partial}{\partial z} \{z \left|\, z\right\rangle\} \\ &= \left|\, z\right\rangle + z \frac{\partial}{\partial z} \left|\, z\right\rangle \end{aligned} \tag{4.13}$$

so that (4.11) is satisfied. In the same way

$$[\overline{Z}, \Pi^\dagger] \left|\, z\right\rangle = - \left|\, z\right\rangle$$

gives

$$\Pi^\dagger \left|\, z\right\rangle = -\frac{\partial}{\partial z^*} \left|\, z\right\rangle. \tag{4.14}$$

Finally, from

$$[\overline{Z}^\dagger, \Pi] \left|^z\right\rangle = \left|^z\right\rangle, \quad [Z^\dagger, \Pi^\dagger] \left|^z\right\rangle = - \left|^z\right\rangle, \tag{4.15}$$

we obtain

$$\Pi \left|^z\right\rangle = \frac{\partial}{\partial z} \left|^z\right\rangle, \quad \Pi^\dagger \left|^z\right\rangle = -\frac{\partial}{\partial z^*} \left|^z\right\rangle. \tag{4.16}$$

Therefore the rule is to replace Π by $\partial/\partial z$ and $\Pi^\dagger$ by $-\partial/\partial z^*$ on any ket vector, while $(Z, \overline{Z})$ give (z, z^*) on a covariant ket and $(\overline{Z}^\dagger, Z^\dagger)$ give (z, z^*) on a contravariant ket.

Applying the rules, we can replace $\mathbf{X}$ and $\mathbf{X}^\dagger$ by differential operators when they are applied on kets. The resulting expressions are found to be

$$(X_1 + iX_2) \left|\, z\right\rangle = \frac{\partial}{\partial z} \left|\, z\right\rangle, \tag{4.17a}$$

$$X_3 \left|\, z\right\rangle = \left(-j_1 + z \frac{\partial}{\partial z}\right) \left|\, z\right\rangle, \tag{4.17b}$$

$$(X_1 - iX_2) \left|\, z\right\rangle = \left(2j_1 z - z^2 \frac{\partial}{\partial z}\right) \left|\, z\right\rangle, \tag{4.17c}$$

$$(X_1^\dagger - iX_2^\dagger) \left|\, z\right\rangle = -\frac{\partial}{\partial z^*} \left|\, z\right\rangle, \tag{4.17d}$$

$$X_3^\dagger \left|\, z\right\rangle = \left(j_2 - z^* \frac{\partial}{\partial z^*}\right) \left|\, z\right\rangle, \tag{4.17e}$$

$$(X_1^\dagger + iX_2^\dagger) \left|\, z\right\rangle = \left(-2j_2 z^* + z^{*2} \frac{\partial}{\partial z^*}\right) \left|\, z\right\rangle, \tag{4.17f}$$

and

$$(X_1 + iX_2) \left|^z\right\rangle = \frac{\partial}{\partial z} \left|^z\right\rangle, \tag{4.18a}$$

$$X_3 \left|^z\right\rangle = \left(j_2^* + 1 + z \frac{\partial}{\partial z}\right) \left|^z\right\rangle, \tag{4.18b}$$

$$(X_1 - iX_2) \left|^z\right\rangle = \left\{-2(j_2^* + 1)z - z^2 \frac{\partial}{\partial z}\right\} \left|^z\right\rangle, \tag{4.18c}$$

$$(X_1^\dagger - iX_2^\dagger) \left|^z\right\rangle = -\frac{\partial}{\partial z^*} \left|^z\right\rangle, \tag{4.18d}$$

$$X_3^\dagger \left|^z\right\rangle = \left(-j_1^* - 1 - z^* \frac{\partial}{\partial z^*}\right) \left|^z\right\rangle, \tag{4.18e}$$

$$(X_1^\dagger + iX_2^\dagger)\left|{}^{z}\right\rangle = \left\{2(j_1^* + 1)z^* + z^{*2}\frac{\partial}{\partial z^*}\right\}\left|{}^{z}\right\rangle. \tag{4.18f}$$

From the formulas we derive the transformation law of kets for infinitesimal Lorentz transformations in the form

$$(1 + i\boldsymbol{\alpha}\cdot\mathbf{X} + i\boldsymbol{\alpha}^*\mathbf{X}^\dagger)\left|{}_{z,z^*}\right\rangle = \lambda(\boldsymbol{\alpha}, z, z^*)\left|{}_{z',z'^*}\right\rangle, \tag{4.19}$$

where

$$\lambda(\boldsymbol{\alpha}, z, z^*) = 1 + i\alpha_3 j_1 + (\alpha_2 - i\alpha_1)j_1 z - i\alpha_3^* j_2 + (\alpha_2^* + i\alpha_1^*)j_2 z^*, \tag{4.20}$$

$$z' = z - i\alpha_3 z - \tfrac{1}{2}(\alpha_2 + i\alpha_1) - \tfrac{1}{2}(\alpha_2 - i\alpha_1)z^2, \tag{4.21}$$

$$z'^* = z^* + i\alpha_3^* z^* - \tfrac{1}{2}(\alpha_2^* - i\alpha_1^*) - \tfrac{1}{2}(\alpha_2^* + i\alpha_1^*)z^{*2}. \tag{4.22}$$

Because $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$ are infinitesimal, we can rewrite these expressions as

$$\lambda(\boldsymbol{\alpha}, z, z^*) = (cz + d)^{2j_1}(c^*z^* + d^*)^{2j_2}, \tag{4.23}$$

$$z' = \Lambda z = \frac{az + b}{cz + d}, \qquad z'^* = \Lambda^* z^* = \frac{a^*z^* + b^*}{c^*z^* + d^*}, \tag{4.24}$$

with

$$\Lambda = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = 1 - \tfrac{1}{2}\boldsymbol{\sigma}\cdot\boldsymbol{\alpha} = \begin{pmatrix} 1 - \frac{1}{2}i\alpha & -\frac{1}{2}i\alpha_1 - \frac{1}{2}\alpha_2 \\ -\frac{1}{2}i\alpha_1 + \frac{1}{2}\alpha_2 & 1 + \frac{1}{2}i\alpha_3 \end{pmatrix}. \tag{4.25}$$

Hence, the infinitesimal transformation (4.19) can be integrated to give

$$U(\Lambda)\left|{}_{z}\right\rangle = (cz + d)^{2j_1}(c^*z^* + d^*)^{2j_2}\left|{}_{\Lambda z}\right\rangle, \tag{4.26}$$

where $U(\Lambda)$ is given by (4.9) and Λz by

$$\Lambda z = \frac{az + b}{cz + d} \quad \text{with} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = e^{-i(\boldsymbol{\sigma}\cdot\boldsymbol{\alpha})/2} = e^{-i[\boldsymbol{\sigma}\cdot(\boldsymbol{\omega} - i\boldsymbol{\nu})]/2}. \tag{4.27}$$

Thus we have derived an analog of the Gel'fand-Naimark transformation law on covariant basis ket vectors. The transformation law has the form anticipated in Eq. (2.38) with the multiplier

$$\lambda(\Lambda, z) = (cz + d)^{2j_1}(c^*z^* + d^*)^{2j_2}. \tag{4.28}$$

In a similar way, using the set (4.18) we find

$$U(\Lambda)\left|{}^{z}\right\rangle = (cz + d)^{-2j_2^* - 2}(c^*z^* + d^*)^{-2j_1^* - 2}\left|{}^{\Lambda z}\right\rangle. \tag{4.29}$$

We note that going from covariant to contravariant kets only results in a change for the multiplier in the transformation law. The necessary substitution

$$j_1 \to -j_2^* - 1, \qquad j_2 \to -j_1^* - 1 \tag{4.30}$$

is consistent with the unitary relation (3.3a). Therefore the metric operator G which relates the two transformation laws can also be interpreted as an "intertwining operator" in Gel'fand's language.[3]

5. THE BASIS BRA VECTORS AND SCALAR PRODUCTS OF GENERAL VECTORS

In the finite complex space, the contravariant basis vectors e^m are said to be orthogonal to the covariant basis vectors e_n if we have

$$(e^m, e_n) = (e^m)^*\cdot e_n = \delta_n^m \tag{5.1}$$

because of the Hermitian form of the scalar product. Using the metric tensor and (4.7) we can write

$$(e_m, e_n) = (e^l g_{lm})^*\cdot e_n = g_{lm}^*\,\delta_n^l = g_{nm}^*.$$

Because of the hermiticity of the metric tensor

$$g_{nm}^* = g_{mn}, \tag{5.2}$$

we find

$$(e_m, e_n) = g_{mn}. \tag{5.3}$$

Similarly

$$(e^m, e^n) = (e^m)^*\cdot e_l g^{ln} = g^{mn} \tag{5.4}$$

and

$$(e_m, e^n) = (e^k g_{km})^*\cdot e_l g^{ln} = \delta_l^k g_{km}^* g^{ln} = g_{km}^* g^{kn} = g_{mk}g^{kn} = \delta_m^n, \tag{5.5}$$

where we have used (5.2) and (4.8).

In a generalized Dirac notation, we can write

$$e_n = \left|{}_{n}\right\rangle, \quad e^m = \left|{}^{m}\right\rangle, \quad (e^m)^* = \left\langle{}^{m}\right|, \quad e_n^* = \left\langle{}_{n}\right|. \tag{5.6}$$

Denoting scalar products by the notation

$$e_n^*\cdot e_m = \left\langle{}_{n}\middle|{}_{m}\right\rangle, \tag{5.7}$$

we find

$$\left\langle{}^{m}\middle|{}_{n}\right\rangle = \left\langle{}_{m}\middle|{}^{n}\right\rangle = \delta_n^m \tag{5.8}$$

and

$$\left\langle{}_{m}\middle|{}_{n}\right\rangle = g_{mn}, \qquad \left\langle{}^{m}\middle|{}^{n}\right\rangle = g^{mn}. \tag{5.9}$$

Following Dirac, we also represent dyadics by kets followed by bras, so that, for example,

$$e_n e_m^* = \left|{}_{n}\right\rangle\left\langle{}_{m}\right|. \tag{5.10}$$

If we only apply the dyadic $e_m e^{*m}$ on e_n, we have

$$e_m e^{*m} e_n = e_m(e^{*m}\cdot e_n) = e_n, \tag{5.11}$$

so that the dyadic $e_m e^{*m}$ is equivalent to the unit matrix. In modified Dirac notation, we write

$$\left|{}_{m}\right\rangle\left\langle{}^{m}\right| \equiv \sum_m \left|{}_{m}\right\rangle\left\langle{}^{m}\right| = I, \tag{5.12}$$

I denoting the identity operation. This is the completeness relation.

In the same spirit we introduce *contravariant bra* basis vectors in the vector space in which $U(\Lambda)$ operates, through the orthogonality relation

$$\left\langle{}^{z}\middle|{}_{z'}\right\rangle = \delta^{(2)}(z - z'). \tag{5.13}$$

If

$$z = x + iy, \quad z' = x' + iy',$$

the Dirac delta function in (5.13) is defined by

$$\delta^{(2)}(z - z') = \delta(x - x')\delta(y - y'). \tag{5.14}$$

The relation (5.13) is consistent with the contravariant bra vectors being simultaneous eigenstates of Z and $\bar{Z}$ with respective eigenvalues z and z^*. Indeed we have

$$\left(z\left|Z\right|z'\right) = \left(z\left|\{Z\}\right|z'\right) = z'\left(z\middle|z'\right),$$

but also

$$\left(z\left|Z\right|z'\right) = \left\{\left(z\left|Z\right\}\right|z'\right) = z\left(z\middle|z'\right),$$

so that we find

$$(z - z')\left(z\middle|z'\right) = 0. \tag{5.15}$$

Hence the scalar product (5.13) must vanish for $z \neq z'$, and a delta function normalization is possible.

The generalization of the completeness relation reads

$$\int d^2z \left|z\right)\left(z\right| = \mathrm{I}. \tag{5.16}$$

The *covariant bra* vectors are now defined by

$$\left(z\right| = \left(z\right|G \tag{5.17}$$

and they are simultaneous eigenstates of $\bar{Z}^\dagger$ and $Z^\dagger$:

$$\left(z\middle|\bar{Z}^\dagger = \left(z\right|z, \quad \left(z\middle|Z^\dagger = \left(\bar{z}\right|z^*, \tag{5.18}$$

with the properties

$$\left(z\middle|z'\right) = \delta^{(2)}(z - z'), \quad \int d^2z \left|z\right)\left(z\right| = \mathrm{I}. \tag{5.19}$$

In analogy to (5.9) we also have

$$\left(z\middle|z'\right) = \left(z\left|G\right|z'\right) = G(z, z'), \tag{5.20}$$

where the function $G(z, z')$ represents the covariant components of the metric operator, which will be given explicitly in Sec. 7. Its contravariant components are represented by the function

$$\tilde{G}(z, z') = \left(z\middle|z'\right) = \left(z\left|G^{-1}\right|z'\right), \tag{5.21}$$

corresponding to the matrix elements of the inverse metric tensor.

A general ket is represented by the vector

$$|g) = \int d^2\zeta g^*(\zeta)\left|\zeta\right) = \int d^2\zeta \tilde{g}^*(\zeta)\left|\zeta\right), \tag{5.22}$$

in analogy with

$$|v) = v_n^* e^n = v^{*n} e_n = v_n^*\left|n\right) = v^{*n}\left|n\right), \tag{5.23}$$

in the N-dimensional complex space. Then we may regard the functions $g^*(z)$ and $\tilde{g}^*(z)$ as, respectively, the covariant and contravariant "components" of the vector $|g)$.

A general bra will have the expansion

$$(f| = \int d^2z \left(z\middle|f(z) = \int d^2z \left(z\right|\tilde{f}(z), \tag{5.24}$$

corresponding to the finite form

$$(u| = |u)^\dagger = e^{*n} u_n = e_n^* u^n. \tag{5.25}$$

We can also write, inserting a complete set of states through Eq. (5.16) and using (5.21),

$$\begin{aligned}(f| &= \iint d^2z d^2\zeta \left(z\middle|\zeta\right)\left(\zeta\right|f(z) \\ &= \iint d^2z d^2\zeta \left(\zeta\right|f(z)\tilde{G}(z, \zeta)\end{aligned} \tag{5.26}$$

so that the contravariant components of $|f)$ are related to its covariant components by the relation

$$\tilde{f}(\zeta) = \int d^2z \tilde{G}(z, \zeta) f(z), \tag{5.27}$$

through the inverse metric tensor.

We define the product $(f|g)$ which will be shown to be a Hermitian scalar product in Sec. 6:

$$(f|g) = \iint d^2z d^2\zeta \left(z\right|\tilde{f}(z)g^*(\zeta)\left|\zeta\right) = \int d^2z \tilde{f}(z) g^*(z), \tag{5.28}$$

where we have used (5.19). In terms of covariant components only, we have

$$(f|g) = \iint d^2z d^2\zeta \tilde{G}(z, \zeta) f(z) g^*(\zeta) \tag{5.29}$$

corresponding to the finite expression

$$(u|v) = g^{mn} u_m v_n^*. \tag{5.30}$$

Working only with the components of vectors has the meaning of the realization of the group transformations in a Hilbert space of functions. The norm of $|f)$ is defined by

$$\|f\|^2 = (f|f) = \iint d^2z d^2\zeta \tilde{G}(z, \zeta) f(z) f^*(\zeta). \tag{5.31}$$

It is positive definite because the inverse metric operator G^{-1} has the form

$$G^{-1} = \Pi^\rho (\Pi^\rho)^\dagger = (\Pi\Pi^\dagger)^\rho, \tag{5.32}$$

so that it is both Hermitian and positive definite. Thus we have a genuine Hilbert space, provided the function space is chosen so that the integrals that define the scalar products are convergent. Then the normed vectors $|v)$, unlike the basis vectors $\left|z\right)$, span the Hilbert space.

6. TRANSFORMATION PROPERTIES OF THE BRA VECTORS AND THE INVARIANCE OF THE SCALAR PRODUCT

To derive the transformation law for the bra basis vectors we first calculate the matrix elements of $U(\Lambda)$. From the transformation law (4.26) and the orthogonality relation (5.13), we find

$$\left(z'\middle| U(\Lambda)\middle|_{z}\right) = (cz+d)^{2j_1}(c^*z^*+d^*)^{2j_2}\delta^{(2)}(z'-\Lambda z), \tag{6.1}$$

while (4.29) gives

$$\left(_{z'}\middle| U(\Lambda)\middle|^{z}\right) = (cz+d)^{-2j_2^*-2} \times (c^*z^*+d^*)^{-2j_1^*-2}\delta^{(2)}(z'-\Lambda z). \tag{6.2}$$

Now, using the completeness relation, we can write

$$\left(z'\middle| U(\Lambda) = \int d^2z \left(z'\middle| U(\Lambda)\middle|_{z}\right)\left(z\right|$$
$$= \int d^2z \left(z\middle| (cz+d)^{2j_1} \times (c^*z^*+d^*)^{2j_2}\delta^{(2)}(z-\Lambda z).$$

Let

$$\zeta = \Lambda z = (az+b)/(cz+d). \tag{6.3}$$

Then,

$$d^2\zeta = (cz+d)^{-2}(c^*z^*+d^*)^{-2}d^2z, \tag{6.4}$$

so that,

$$\left(z'\middle| U(\Lambda) = \int d^2\zeta \left(\Lambda^{-1}\zeta\middle| (cz+d)^{2j_1+2} \times (c^*z^*+d^*)^{2j_2+2}\delta^{(2)}(z'-\zeta)\right. \tag{6.5}$$

Equation (6.3) implies

$$cz+d = (-c\zeta+a)^{-1}.$$

Inserting in (6.5), we find

$$\left(z'\middle| U(\Lambda) = \left(\Lambda^{-1}z'\middle| (-cz'+a)^{-2j_1-2}(-c^*z'^*+a^*)^{-2j_2-2}.\right.\right. \tag{6.6}$$

Another form is

$$\left(z'\middle| U^\dagger(\Lambda) = \left(z'\middle| U(\Lambda^{-1}) = \left(\Lambda z'\middle| (cz'+d)^{-2j_1-2}(c^*z'^*+d^*)^{-2j_2-2}.\right.\right.\right. \tag{6.7}$$

This last formula is consistent with (4.26) since we have

$$\left(z'\middle| U^\dagger U\middle|_{z}\right) = (cz'+d)^{-2}(c^*z'^*+d^*)^{-2}\left(\Lambda z'\middle|_{\Lambda z}\right). \tag{6.8}$$

Now

$$\left(\Lambda z'\middle|_{\Lambda z}\right) = \delta^{(2)}(\Lambda z'-\Lambda z) = (cz'+d)^2(c^*z'^*+d^*)^2\delta^{(2)}(z'-z), \tag{6.9}$$

so that

$$\left(z'\middle| U^\dagger U\middle|_{z}\right) = \delta^{(2)}(z'-z) = \left(z'\middle|_{z}\right), \tag{6.10}$$

which is the unitarity condition for U.

In the same way we find

$$\left(_{z'}\middle| U^\dagger(\Lambda) = \left(_{z'}\middle| U(\Lambda^{-1}) = \left(\Lambda z'\middle| (cz'+d)^{2j_2^*}(c^*z'^*+d^*)^{2j_1^*},\right.\right.\right. \tag{6.11}$$

for the transformation law of covariant basis bra vectors.

By applying the commutation relations (3.14), (3.23), and (3.25) on bra vectors, proceeding exactly as in Sec. 4, we find the effect of Π and $\Pi^\dagger$ on these states in the form

$$\left(z\middle|\Pi = -\frac{\partial}{\partial z}\left(z\right|, \quad \left(z\middle|\Pi^\dagger = \frac{\partial}{\partial z^*}\left(z\right|, \tag{6.12}$$

$$\left(_{z}\middle|\Pi = -\frac{\partial}{\partial z}\left(^{z}\right|, \quad \left(_{z}\middle|\Pi^\dagger = \frac{\partial}{\partial z^*}\left(^{z}\right|. \tag{6.13}$$

Note the change of sign from Eqs. (4.12), (4.14), and (4.16). The rule is to replace Π by $-\partial/\partial z$ and $\Pi^\dagger$ by $\partial/\partial z^*$ when applied on bra vectors. A differential form for $\mathbf{X}$ and $\mathbf{X}^\dagger$ can be found from the operator expressions (3.28), (3.30), (3.33), and (3.34) when these generators act on bra vectors. Integrating the infinitesimal transformation law we recover (6.6) and (6.11). The transformation law for the covariant vector components $f(z)$ and $g^*(\zeta)$ can now be derived from the invariance of the scalar product. We must have

$$(f|g) = (f|UU^\dagger|g) = (f'|g'), \tag{6.14}$$

where we define the transformed vector as

$$(f'| = (f|U(\Lambda) = \int d^2z \left(z\middle| (T(\Lambda)f(z))\right.$$
$$= \int d^2z' \left(z'\middle| Uf(z').\right. \tag{6.15}$$

Using (6.6) we find

$$(f'| = \int d^2z' \left(\Lambda^{-1}z'\middle| (-cz'+a)^{-2j_1-2} \times (-c^*z'^*+a^*)^{-2j_2-2}f(z'). \right. \tag{6.16}$$

As in (6.3) and (6.4), making the change of variable,

$$z' = \Lambda z, \quad d^2z = (-cz'+a)^{-2}(-c^*z'^*+a^*)^{-2}d^2z', \quad -cz'+a = (cz+d)^{-1},$$

we find

$$(f'| = \int d^2z \left(z\middle| (cz+d)^{2j_1}(c^*z^*+d^*)^{2j_2}f(\Lambda z). \right. \tag{6.17}$$

Comparing with (6.15) we obtain

$$\left(f|U(\Lambda)\middle|_{z}\right) = T(\Lambda)f(z) = (cz+d)^{2j_1}(c^*z^*+d^*)^{2j_2}f(\Lambda z), \tag{6.18}$$

as the transformation law for the covariant components of a vector.

Thus we recover the Gel'fand–Naimark law without using the method of homogeneous functions.

The components $f(z)$ are projected out of the vector $(f|$ by means of the formula

$$f(z) = \left(f\middle|_{z}\right) = \int d^2z' \left(z'\middle|_{z}\right) f(z'). \tag{6.19}$$

Then, the functions

$$F(z_1,z_2) = z_2^{2j_1}z_2^{*2j_2}\left(f\middle|_{z_1/z_2}\right) = z_2^{2j_1}z_2^{*2j_2}f\left(\frac{z_1}{z_2}\right) \tag{6.20}$$

are homogeneous of degree $2j_1$ in (z_1, z_2) and $2j_2$ in (z_1^*, z_2^*). The transformation law is then linear for z_1 and z_2 so that

$$\psi = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \to \Lambda\psi, \tag{6.21}$$

transforms like a spinor. This establishes the connection with the representation theory on spaces of homogeneous functions.

7. THE Π REPRESENTATION AND CALCULATION OF $\tilde{G}(z,z')$. UNIFIED SCALAR PRODUCT FOR VARIOUS CLASSES OF REPRESENTATIONS.

In the Π representation,[12] Π and $\Pi^\dagger$ are taken as diagonal, so that we consider simultaneous eigenstates of the commuting set C_1, C_2, Π, and $\Pi^\dagger$. Let us write

$$\Omega_1 |j_1,j_2;p\rangle = j_1 |j_1,j_2;p\rangle, \tag{7.1a}$$

$$\Omega_2 |j_1,j_2;p\rangle = j_2 |j_1,j_2;p\rangle, \tag{7.1b}$$

$$\Pi |j_1,j_2;p\rangle = -\tfrac{1}{2}ip\,|j_1,j_2;p\rangle, \tag{7.1c}$$

$$\Pi^\dagger |j_1,j_2;p\rangle = \tfrac{1}{2}ip^*|j_1,j_2;p\rangle. \tag{7.1d}$$

There is no essential distinction between covariant and contravariant states in p space except for a p-dependent normalization factor. Hence, dropping the j_1, j_2 labels, the bra vectors can be introduced by

$$\langle p|\Pi = -\tfrac{1}{2}ip\langle p|, \quad \langle p|\Pi^\dagger = \tfrac{1}{2}ip^*\langle p|. \tag{7.2}$$

Since

$$\langle p|\Pi|p'\rangle = -\tfrac{1}{2}ip\langle p|p'\rangle = -\tfrac{1}{2}ip'\langle p|p'\rangle,$$

we require the normalization

$$\langle p|p'\rangle = \delta^{(2)}(p-p'). \tag{7.3}$$

Because $i\Pi$ and Z are conjugate, the p and z eigenstates must be Fourier transforms of each other. Indeed, writing

$$\left|\ \right\rangle_z = \frac{1}{2\pi}\int d^2p e^{-i\mathrm{Re}(pz)}|p\rangle \tag{7.4a}$$

and

$$\left\langle\ \right|^{z'} = \frac{1}{2\pi}\int d^2p' e^{i\mathrm{Re}(p'z')}\langle p'|, \tag{7.4b}$$

we have

$$\left\langle\ \right|^{z'}\left|\ \right\rangle_z = \frac{1}{4\pi^2}\iint d^2p'd^2p e^{i[\mathrm{Re}(p'z')-\mathrm{Re}(pz)]}\delta^{(2)}(p-p')$$

$$= \frac{1}{4\pi^2}\int d^2p e^{i\mathrm{Re}[p(z'-z)]} = \delta^{(2)}(z'-z), \tag{7.5}$$

which shows that the normalizations in z and p spaces are consistent.

We have

$$\Pi\left|\ \right\rangle_z = \frac{\partial}{\partial z}\left|\ \right\rangle_z = \frac{1}{2\pi}\int d^2p\,\frac{\partial}{\partial z}\{e^{-i(pz+p^*z^*)/2}\}|p\rangle$$

$$= \frac{1}{2\pi}\int d^2p e^{-i\mathrm{Re}(pz)}(-\tfrac{1}{2}ip)\,|p\rangle$$

$$= \frac{1}{2\pi}\int d^2p e^{-i\mathrm{Re}(pz)}\Pi\,|p\rangle. \tag{7.6}$$

Hence, from (7.4a) we have derived (7.1c). We are therefore led to the relations

$$\Pi^\rho|p\rangle = (-\tfrac{1}{2}ip)^\rho|p\rangle, \quad (\Pi^\rho)^\dagger|p\rangle = (\tfrac{1}{2}ip^*)^\rho|p\rangle, \tag{7.7}$$

which allow us to calculate the matrix elements of the inverse metric tensor. We have

$$\tilde{G}(z',z) = \left\langle\ \right|^{z'}\left|G^{-1}\right|_z\left.\ \right\rangle = \left\langle\ \right|^{z'}\left|\Pi^\rho(\Pi^\rho)^\dagger\right|_z\left.\ \right\rangle$$

$$= (2\pi)^{-2}\int d^2pd^2p' e^{i\mathrm{Re}(zp-z'p')}$$

$$\times(-\tfrac{1}{2}ip)^\rho(\tfrac{1}{2}ip^*)^\rho\delta^{(2)}(p'-p)$$

$$= (2\pi)^{-2}\int d^2p e^{i\mathrm{Re}[(z-z')p]}(\tfrac{1}{2})^{2\rho}|p|^{2\rho}.$$

Using the Fourier transform of $|p|^{2\rho}$ as given in Gel'fand *et al.*,[13] we find

$$\left\langle\ \right|^{z'}\left|G^{-1}\right|_z\left.\ \right\rangle = \frac{\Gamma(1+\rho)}{\pi\Gamma(-\rho)}|z-z'|^{-2\rho-2},$$

$$\rho = \rho^* = j_1 + j_2^* + 1, \tag{7.8}$$

which is the general form of the "intertwining operator" to within a factor.[3]

This expression now allows us to discuss the different explicit forms of the scalar product defined in Eq. (5.29):

(1) *The principal series:* As we have seen before, in this case we have $\rho = 0$, so that $G = 1$ and we get

$$\lim_{\rho\to 0}\tilde{G}(z,z') = \lim_{\rho\to 0}\left\langle\ \right|^{z'}\left|G^{-1}\right|_z\left.\ \right\rangle = \left\langle\ \right|^{z'}\left|\ \right\rangle_z$$

$$= \delta^{(2)}(z'-z) = \lim_{\rho\to 0}\frac{\Gamma(1+\rho)}{\pi\Gamma(-\rho)}|z-z'|^{-2\rho-2}, \tag{7.9}$$

so that the scalar product takes the form

$$(f|g) = \int d^2z f(z)g^*(z). \tag{7.10}$$

Thus the functions $f(z)$ must be $L^{(2)}$ functions. For the principal series, the distinction between covariant and contravariant states disappears as we have $\bar{Z} = Z^\dagger$.

(2) *The supplementary series:* In this case $\rho \neq 0$ and

$$(f|g) = \frac{\Gamma(1+\rho)}{\pi\Gamma(-\rho)}\int d^2zd^2z'|z-z'|^{-2\rho-2}f(z)g^*(z'). \tag{7.11}$$

Apart from the over-all normalization factor, this is the familiar form for the supplementary series scalar product.[3] In the same Hilbert space as in the preceding case, the integral exists for

$$0 < |\rho| < 1. \tag{7.12}$$

(3) *The integer point case with $\rho = n$:* n being a positive integer or zero.

In this case we have the representation

$$(j_1,j_2) = \left(\frac{n-1}{2}, \frac{n-1}{2}\right). \tag{7.13}$$

The inverse metric operator takes the form

$$G^{-1} = (\Pi\Pi^\dagger)^n \tag{7.14}$$

so that, from (4.12) and (4.14), we have

$$\Pi\Pi^\dagger \Big|_z\Big) = -\frac{\partial^2}{\partial z \partial z^*}\Big|_z\Big), \qquad (\Pi\Pi^\dagger)^n \Big|_z\Big) = (-1)^n \frac{\partial^{2n}}{\partial z^n \partial z^{*n}}\Big|_z\Big)$$

leading to

$$\begin{aligned}\tilde{G}(z,z') &= (-1)^n \frac{\partial^{2n}}{\partial z^n \partial z^{*n}} \delta^{(2)}(z-z') \\ &= (-1)^n \delta^{[n,n]}(z-z'). \qquad (7.15)\\ &= \lim_{\rho \to 0} \frac{\Gamma(1+\rho)}{\pi\Gamma(-\rho)} |z-z'|^{-2\rho-2}.\end{aligned}$$

The scalar product takes the form

$$(f|g) = (-1)^n \int d^2z f^{[n,m]}(z) g^*(z), \qquad (7.16)$$

where

$$f^{[n,m]}(z) = \frac{\partial^{2n}}{\partial z^n \partial z^{*n}} f(z). \qquad (7.17)$$

This representation is formally the same as finite nonunitary representations. In fact, the polynomials $P_n(z,z^*)$ which correspond to the latter are annihilated by the metric operator, so that

$$P_n^{[n,n]}(z,z^*) = 0, \qquad (7.18)$$

and therefore form a subspace of zero norm. Hence, as discussed by Gel'fand *et al.*[3] the functional space must be restricted to functions such that

$$f^{[n,n]}(z) \neq 0$$

(4) *The integer point case with* $\rho = -n$: n being a positive integer.

In this case the representation is

$$(j_1, j_2) = \left(-\frac{n+1}{2}, -\frac{n+1}{2}\right) \qquad (7.19)$$

To find $\tilde{G}(z,z')$, we write $\rho = -n + \epsilon$ and let $\epsilon \to 0$:

$$\begin{aligned}\tilde{G}(z,z') &= \lim_{\epsilon \to 0} \frac{\Gamma(1-n+\epsilon)}{\pi\Gamma(n-\epsilon)} |z-z'|^{2n-2-2\epsilon} \qquad (7.20)\\ &= \lim_{\epsilon \to 0} \left\{ \frac{(-1)^{n-1}|z-z'|^{2n-2}}{\pi\epsilon[\Gamma(n)]^2} \right.\\ &\quad \left. + \frac{2(-1)^n}{\pi[\Gamma(n)]^2} |z-z'|^{2n-2} \log|z-z'| + 0(\epsilon) \right\}.\end{aligned}$$

To eliminate the first term which blows up like $1/\epsilon$, we must restrict ourselves to the subset of functions which satisfy

$$\int z^k z^{*l} f(z) d^2z = 0 \qquad (7.21)$$

for $k, l \leq n-1$. Hence, the scalar product reads

$$(f,g) = \{2(-1)^n/\pi[\Gamma(n)]^2\} \iint d^2z d^2z' |z-z'|^{2n-2} \times \log|z-z'| f(z) g^*(z'). \qquad (7.22)$$

Thus, we have shown that the form (5.29) with the expression (7.8) for the matrix elements of the inverse metric operator encompasses and unifies all cases in which a unitary representation of $SL(2,C)$ exists.

8. REMARKS ON THE INTEGER POINT CASE WITH $\rho = n$

In this section we shall study further the case $(j_1, j_2) = (\frac{1}{2}(n-1), \frac{1}{2}(n-1))$, where n is a positive integer and show its intimate connection with the same nonunitary (j_1, j_2) representation. For the corresponding nonunitary representation, the function $P_n(z)$ is a polynomial of order $n-1$ in z and order $n-1$ in z^* so that its norm given by the formula

$$||P_n(z)||^2 = (-1)^n \int d^2z P_n^{[n,n]}(z) P_n^*(z), \qquad (8.1)$$

which follows from (7.16) becomes zero.

As explained by Gel'fand *et al.*,[3] in order to avoid a degenerate scalar product, we must restrict ourselves to a subspace of functions $f(z)$ such that for $\rho = n$ their derivatives of order n, $f^{[n,n]}(z)$ do not vanish. Then all the polynomials

$$P_n(z) = \sum_{k,l=n}^{n-1} a_{kl} z^k z^{*l} \qquad (8.2)$$

should be excluded from our space. Note that for $\rho = n$, the set of such polynomials (E_n) forms an invariant subspace under Lorentz transformations, connected with the nonunitary representations mentioned above. The subspace of functions which form an infinite dimensional unitary representation for $\rho = n = 1, 2, \ldots$ is the set of functions obtained from homogeneous functions through (6.20) and defined up to a polynomial in E_n. Following Gel'fand *et al.*[3] we denote this set by F_n. Thus if $f(z) \in F_n$ and $P_n(z) \in E_n$, then

$$\{f(z) + P_n(z)\} \in F_n.$$

Let us denote the part of $f(z)$ which contributes to the scalar product (7.16) by $f^{(u)}(z)$. Then we can write

$$f^{(u)}(z) = \sum_{k,l=n}^{\infty} a_{kl} z^k z^{*l} \qquad (8.3)$$

and

$$f(z) = f^{(u)}(z) + P_n(z), \qquad (8.4)$$

where $P_n(z)$ is *any* polynomial in E_n. From the way it is defined, the only part of $f(z)$ which is "relevant" to unitary representations is obviously $f^{(u)}(z)$. Under a Lorentz transformation, as shown in (6.18), when

$$T(\Lambda)f(z) = |cz+d|^{2\rho-2} f\left(\frac{az+b}{cz+d}\right), \qquad (8.5)$$

$P_n(z)$ goes into another polynomial, but $f^{(u)}(z)$ does not transform into a function of the form (8.3). Therefore, to extract the relevant part of $T(\Lambda)f(z)$, we must subtract a polynomial. In other words, the set of functions of the form $f^{(u)}(z)$ as in (8.3) is not an invariant set, and an additional "gauge transformation" (subtracting a polynomial) is needed to make it invariant. This is analogous to the use of the transverse electromagnetic potential A_μ^T, which is the only dynamical variable but is only covariant with an additional gauge transformation. Thus, in the case of the integer point representations, it is possible to choose a gauge just as we do in electrodynamics. If the gauge is fixed in such a way that we extract only $f^{(u)}(z)$ out of $f(z)$, then, as described above, we shall need an additional gauge transformation to obtain

a covariant $f(z)$. Instead, we choose the "gauge" as follows.

Each term in Eq. (8.2) is a linearly independent polynomial which is annihilated by the metric G^{-1} at integer points $\rho = 1, 2, \ldots$:

$$\int P_n(z')\delta^{[n,n]}(z - z')d^2z' = P_n^{[n,n]}(z) = 0. \tag{8.6}$$

Thus, for $2j_1 + 1 = 2j_2 + 1 = \rho = n$, there are

$$n^2 = (2j_1 + 1)(2j_2 + 1)$$

linearly independent polynomials which satisfy Eq. (8.6). Using these polynomials, we can choose the gauge so that we can write $f(z) \in F_n$ as

$$f(z) = \sum_{k,l=0}^{n-1} z^k z^{*l} \phi_{kl}(z), \tag{8.7}$$

where each ϕ_{kl} can be written in the form

$$\phi_{kl}(z) = \sum_{m,n=0}^{\infty} (b^{kl})_{mn} z^m z^{*n}. \tag{8.8}$$

The advantage of this new form is that under a Lorentz transformation the structure (8.7) remains invariant, so that (8.7), unlike (8.4) is a covariant form, that is, the functions ϕ_{kl} transform into each other and acquire no extra factors of $(cz + d)$ and $(c^*z^* + d^*)$.

To give a definite example, let us consider the case $j_1 = j_2 = \frac{1}{2}$, or $\rho = 2$ representation. This is the representation used in Ref. 11. We can write

$$f(z) = \left(f \,\middle|\, \rho = 2; \underset{z}{\ }\right)$$
$$= zz^*\phi_{11}(z) + z^*\phi_{10}(z) + z\phi_{01}(z) + \phi_{00}(z), \tag{8.9}$$

or

$$f(z) = (z^* \; 1) \begin{pmatrix} \phi_{11}(z) & \phi_{10}(z) \\ \phi_{01}(z) & \phi_{00}(z) \end{pmatrix} \begin{pmatrix} z \\ 1 \end{pmatrix}. \tag{8.10}$$

Applying a Lorentz transformation we get

$$T(\Lambda)f(z) = (cz + d)(c^*z^* + d^*)f(z') \tag{8.11}$$

with

$$z' = \frac{az + b}{cz + d}. \tag{8.12}$$

Using Eq. (8.10) we can rewrite Eq. (8.11) as

$$T(\Lambda)f(z) = (z^* \; 1)\Lambda^\dagger \begin{pmatrix} \phi_{11}(z') & \phi_{10}(z') \\ \phi_{01}(z') & \phi_{00}(z') \end{pmatrix} \Lambda \begin{pmatrix} z \\ 1 \end{pmatrix}. \tag{8.13}$$

This is because we have

$$\Lambda \begin{pmatrix} z \\ 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} z \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{az + b}{cz + d} \\ 1 \end{pmatrix} (cz + d).$$

Therefore, to describe the $(\frac{1}{2}, \frac{1}{2})$ representation, we could just as well use the functions $\phi_{kl}(z)$ by specifying their Lorentz transformation property as

$$T(\Lambda)\phi_{ij}(z) = (\Lambda^\dagger)_{ik} \phi_{kl}(z')(\Lambda)_{lj}. \tag{8.14}$$

Notice that we can write

$$\begin{pmatrix} \phi_{11}(z) & \phi_{10}(z) \\ \phi_{01}(z) & \phi_{00}(z) \end{pmatrix} = \phi_0(z) + \sigma \cdot \phi(z), \tag{8.15}$$

where σ_i are the usual 2×2 Pauli matrices. Then, from Eq. (8.14) we obtain the transformation properties of

$$\phi_\mu(z) = (\phi_0(z), (\tfrac{1}{2}, \tfrac{1}{2})) \tag{8.16}$$

in the form

$$T(\Lambda)\phi_\mu(z) = \Lambda_\mu^\nu \phi_\nu \left(\frac{az + b}{cz + d}\right), \tag{8.17}$$

where Λ_μ^ν is the 4×4 nonunitary representation of the Lorentz transformation. Thus, the functions $\phi_\mu(z)$ describe the unitary integer point representation $(\frac{1}{2}, \frac{1}{2})$ and transform like the direct product of two representations, namely the finite dimensional $(\frac{1}{2}, \frac{1}{2})$ representation, and the infinite dimensional unitary representation $j_1 = j_2 = 0$, or $\rho = 1$.

We can regard the functions $\phi_\mu(z)$ as limiting functions of the supplementary series as ρ tends to one and write

$$\phi_\mu(z) = \lim_{\rho \to 1} \left(\phi_\mu \,\middle|\, \rho; \underset{z}{\ }\right). \tag{8.18}$$

For the representation $(\frac{1}{2}(n-1), \frac{1}{2}(n-1))$, we write $f(z)$ in the form of a polynomial of order $n - 1$ in z and $n - 1$ in z^* with coefficients that are functions of the form (8.8) transforming with $\rho = 1$ and getting mixed like the components of the nonunitary representation $(\frac{1}{2}(n-1), \frac{1}{2}(n-1))$.

ACKNOWLEDGMENTS

The present paper has evolved from ideas and methods presented in a course on Noncompact Groups given at Yale University by one of the authors (F.G.). He feels it is his duty and pleasure to acknowledge the active participation of his class in the clarification and sharpening of the material and thank, particularly, M. Günaydin who wrote the lecture notes,[14] S. Orfanidis, S. Krinsky, M. Serdaroglu, and H. Georgi for their enthusiasm and constructive criticism. The authors would also like to express their gratitude to S. MacDowell and R. Roskies for stimulating discussions and keen interest during the course of this work.

---

* Research (Report Yale 2726-608) supported by the U.S. Atomic Energy Commission under contract AEC (30-1) 2726.

† New Address: Department of Physics, University of California, Berkeley, California 94720.

‡ On leave of absence from the Middle East Technical University, Ankara, Turkey.

1 V. Bargmann, Ann. Math. 48, 568 (1947).

2 I. M. Gel'fand and M. A. Naimark, Izv. Akad. Nauk SSSR Ser. Mat. 11, 411 (1947).

3 I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5, Chap. 3.

4 I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups* (Pergamon, New York, 1963).

5 M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964).

6 W. Rühl, *The Lorentz Group and Harmonic Analysis* (Benjamin, New York, 1970).

7 M. Toller, Nuovo Cimento 53, 671 (1968).

8 Y. Nambu, Progr. Theoret. Phys. Suppl. No. 37-38, 368 (1966); Y. Nambu, Phys. Rev. 160, 1171 (1967).

9 G. Domokos, S. Kovesi-Domokos, and E. Schonberg, Phys. Rev. D 2, 1026 (1970).

[10] I. Montvay, Phys. Rev. D 3, 2532 (1971).
[11] I. Bars and F. Gürsey, "Duality and the Lorentz Group," Phys. Rev. (to be published).
[12] The Π representation was first used and its properties studied by S. J. Chang and L. O'Raifertaigh, J. Math. Phys. 10, 21 (1969). We thank Dr. S. J. Chang for bringing this reference to our attention after our work was completed.
[13] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 1, Appendix B, Sec. 1.7.
[14] F. Gürsey, Representations of Some Non-Compact Groups Related to the Poincaré Group, Yale Lecture Notes compiled by Murat Günaydin, Yale Preprint (1971).

Deformable Magnetically Saturated Media. I. Field Equations *

Gérard A. Maugin and A. Cemal Eringen
Department of Aerospace and Mechanical Sciences, Princeton University, Princeton, New Jersey
(Received 23 July 1971; Revised Manuscript Received 30 August 1971)

In this article we propose a variational approach to the study of nonlinear elastic solids in which magnetization is constant in magnitude. The emphasis is placed upon the application of the different invariances used in modern continuum mechanics: Euclidean invariance, objectivity, and material symmetry. In Part I, a variational treatment is given in the spirit of "oriented media theory." A comparison is made with the results of a direct treatment starting with the postulation of balance laws. Part II is devoted to the development of constitutive equations for a variety of material classes.

1. INTRODUCTION

In the literature there exist extensive works on the theory of ferromagnetic materials. Earlier works due to Weiss, Heisenberg, Dirac, Bloch, Van Vleck, and Néel have been reviewed[1-4]. Most of them deal with a quantum mechanical approach. However, we are here interested in the theory of *micromagnetics* where ferromagnetic bodies are described by means of a vector field, the magnetization, whose magnitude is constant and whose direction varies continuously with position.[5-9] Discrete details are omitted and the theory is phenomenological, that is, in the spirit of the theory developed decades ago by Landau and Lifshitz.[4,10] This description is particularly accepted at temperatures significantly lower than the Curie temperature as emphasized by Minnaja[8] who recently gave a theory at high temperature with no restriction on the magnetization amplitude. The formulation given below does not consider the case of high temperatures and, in fact, neglects the dependence on this physical factor for most of the development.

The main purpose of the present paper is to formulate a continuum theory of deformable magnetically saturated media. The approach being phenomenological, ultimately it must be verified by experiments. We are, however, guided by the quantum mechanical results for the concept of spin. For instance, following Brown,[11] we will make use of certain hypotheses of the H.D.V.V. (for Heisenberg–Dirac–Van Vleck) model. In this regard the following remarks are in order.

(i) The ferromagnetic materials possess high tendency to orientation, i.e., without any important applied magnetic field, the neighboring electron spins tend to align parallel. Heisenberg[12] explains this fact in taking account of *exchange forces* in the electronic spin continuum through a potential representing the *spin–spin coupling*. This effect leads to the notions of *magnetic anisotropy energy* and *spin–wave* motions (Bloch[13]). Brown, in a series of books and articles[11] dealing with *micromagnetism*, shows that this leads to the introduction of the material gradients of the magnetization in the free energy.[14] We do not repeat these arguments but take this result for granted.

(ii) The *spin–orbit interactions* and the *quadrupoles* tend to orient the aligned spins along particular crystalline axes; hence the concept of *magnetocrystalline anisotropy*.

(iii) The *temperature agitation* disturbs the alignment and causes the magnetization to decrease with increasing temperature.

(iv) The *applied magnetic field* tends to align the magnetization vector along the field. In contrast to the effect (ii), opposite directions are not equivalent.

(v) The *internal magnetic interactions: Dipole–dipole* long range interactions tend to rotate the magnetization vector toward a direction of smallest magnetostatic self-energy (in the case of uniform magnetization).

The effects (ii) will be taken into account in the nonlinear constitutive relations giving rise to a local magnetic field due to anisotropy. Effects (iv) and (v) manifest themselves in the form of the angular momentum equation. We disregard (iii) since we shall be dealing with nondissipative systems (with the exception of Sec. 8) for which we ignore the influence of the temperature.

The theory is developed in the frame of *quasi-magnetostatics*, electric fields being ignored. Thus, we assume that the velocity of dynamical phenomena is small in comparison with the propagation velocity of electromagnetic perturbations; hence, the jump conditions are written for stationary discontinuity surfaces. We present a finite deformation theory.

The present work may be considered in the same spirit as those of Brown[11] and Tiersten.[15] Some improvements of the latter works are due to Amari[16] and Alblas[17] of whom the former developed a theory with linear constitutive equations and the latter author follows Brown closely. The basic results of the present article, Eqs. (5.7), (5.8), (5.4), (5.10) and (5.12), can be brought in general agreement with those of Tiersten and Brown. However we have reasons to believe that our presentation is more suited to generalizations (presence of current, dynamical case, inclusion of extra degrees of freedom of structural origin):

(i) We tried to emphasize the physical starting point and the analogy of the magnetization vector with a *rigid director*. By applying the Euclidean invariance requirement, we have established the formal analogy of the present theory of deformable magnetically saturated media with the *indeterminate* theory of couple stresses[18] (cf. Eqs. (7.26), (7.27), (8.19)-(8.21). The formalism used is likely to encompass the introduction of "mechanical" directors.

[10] I. Montvay, Phys. Rev. D 3, 2532 (1971).
[11] I. Bars and F. Gürsey, "Duality and the Lorentz Group," Phys. Rev. (to be published).
[12] The Π representation was first used and its properties studied by S. J. Chang and L. O'Raifertaigh, J. Math. Phys. 10, 21 (1969). We thank Dr. S. J. Chang for bringing this reference to our attention after our work was completed.
[13] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 1, Appendix B, Sec. 1.7.
[14] F. Gürsey, Representations of Some Non-Compact Groups Related to the Poincaré Group, Yale Lecture Notes compiled by Murat Günaydin, Yale Preprint (1971).

Deformable Magnetically Saturated Media. I. Field Equations *

Gérard A. Maugin and A. Cemal Eringen
Department of Aerospace and Mechanical Sciences, Princeton University, Princeton, New Jersey
(Received 23 July 1971; Revised Manuscript Received 30 August 1971)

In this article we propose a variational approach to the study of nonlinear elastic solids in which magnetization is constant in magnitude. The emphasis is placed upon the application of the different invariances used in modern continuum mechanics: Euclidean invariance, objectivity, and material symmetry. In Part I, a variational treatment is given in the spirit of "oriented media theory." A comparison is made with the results of a direct treatment starting with the postulation of balance laws. Part II is devoted to the development of constitutive equations for a variety of material classes.

1. INTRODUCTION

In the literature there exist extensive works on the theory of ferromagnetic materials. Earlier works due to Weiss, Heisenberg, Dirac, Bloch, Van Vleck, and Néel have been reviewed[1-4]. Most of them deal with a quantum mechanical approach. However, we are here interested in the theory of *micromagnetics* where ferromagnetic bodies are described by means of a vector field, the magnetization, whose magnitude is constant and whose direction varies continuously with position.[5-9] Discrete details are omitted and the theory is phenomenological, that is, in the spirit of the theory developed decades ago by Landau and Lifshitz.[4,10] This description is particularly accepted at temperatures significantly lower than the Curie temperature as emphasized by Minnaja[8] who recently gave a theory at high temperature with no restriction on the magnetization amplitude. The formulation given below does not consider the case of high temperatures and, in fact, neglects the dependence on this physical factor for most of the development.

The main purpose of the present paper is to formulate a continuum theory of deformable magnetically saturated media. The approach being phenomenological, ultimately it must be verified by experiments. We are, however, guided by the quantum mechanical results for the concept of spin. For instance, following Brown,[11] we will make use of certain hypotheses of the H.D.V.V. (for Heisenberg–Dirac–Van Vleck) model. In this regard the following remarks are in order.

(i) The ferromagnetic materials possess high tendency to orientation, i.e., without any important applied magnetic field, the neighboring electron spins tend to align parallel. Heisenberg[12] explains this fact in taking account of *exchange forces* in the electronic spin continuum through a potential representing the *spin–spin coupling*. This effect leads to the notions of *magnetic anisotropy energy* and *spin–wave* motions (Bloch[13]). Brown, in a series of books and articles[11] dealing with *micromagnetism*, shows that this leads to the introduction of the material gradients of the magnetization in the free energy.[14] We do not repeat these arguments but take this result for granted.

(ii) The *spin–orbit interactions* and the *quadrupoles* tend to orient the aligned spins along particular crystalline axes; hence the concept of *magnetocrystalline anisotropy*.

(iii) The *temperature agitation* disturbs the alignment and causes the magnetization to decrease with increasing temperature.

(iv) The *applied magnetic field* tends to align the magnetization vector along the field. In contrast to the effect (ii), opposite directions are not equivalent.

(v) The *internal magnetic interactions: Dipole–dipole* long range interactions tend to rotate the magnetization vector toward a direction of smallest magnetostatic self-energy (in the case of uniform magnetization).

The effects (ii) will be taken into account in the nonlinear constitutive relations giving rise to a local magnetic field due to anisotropy. Effects (iv) and (v) manifest themselves in the form of the angular momentum equation. We disregard (iii) since we shall be dealing with nondissipative systems (with the exception of Sec. 8) for which we ignore the influence of the temperature.

The theory is developed in the frame of *quasi-magnetostatics*, electric fields being ignored. Thus, we assume that the velocity of dynamical phenomena is small in comparison with the propagation velocity of electromagnetic perturbations; hence, the jump conditions are written for stationary discontinuity surfaces. We present a finite deformation theory.

The present work may be considered in the same spirit as those of Brown[11] and Tiersten.[15] Some improvements of the latter works are due to Amari[16] and Alblas[17] of whom the former developed a theory with linear constitutive equations and the latter author follows Brown closely. The basic results of the present article, Eqs. (5.7), (5.8), (5.4), (5.10) and (5.12), can be brought in general agreement with those of Tiersten and Brown. However we have reasons to believe that our presentation is more suited to generalizations (presence of current, dynamical case, inclusion of extra degrees of freedom of structural origin):

(i) We tried to emphasize the physical starting point and the analogy of the magnetization vector with a *rigid director*. By applying the Euclidean invariance requirement, we have established the formal analogy of the present theory of deformable magnetically saturated media with the *indeterminate* theory of couple stresses[18] (cf. Eqs. (7.26), (7.27), (8.19)-(8.21). The formalism used is likely to encompass the introduction of "mechanical" directors.

(ii) The variational principle given here in the frame of quasimagnetostatics may be considered as first step toward a fully dynamical theory. It is easily shown that the introduction of our magnetic force is equivalent to adding a magnetic energy $\frac{1}{2}\mathbf{B}^2$—$\mathbf{B}\cdot\mathbf{M}$ to the strain energy $\rho\mathfrak{F}$. This magnetic energy is to be varied under the constraint $\nabla\cdot\mathbf{B} = 0$; thus, a *vector* potential $\mathbf{A}$ defined by $\mathbf{B} = \nabla\times\mathbf{A}$ should be introduced to carry out the variation. The extension to the dynamical case including currents, polarization, and time derivative terms in the Maxwell's equations is easy since the constraint $\nabla\cdot\mathbf{B} = 0$ is unchanged for such a generalization. Whereas Tiersten chose to vary a magnetic energy $\frac{1}{2}\mathbf{H}^2 + \mathbf{H}\cdot\mathbf{M}$ under the constraint $\nabla\times\mathbf{H} = 0$. He thus introduced a *scalar* magnetic potential φ through the relation $\mathbf{H} = -\nabla\varphi$. This cannot be easily extended to the dynamical case (see our generalization in a forthcoming article treated in special relativity).
(iii) We performed the variation first in the undeformed configuration. The equations obtained, Eqs. (5.1) and (5.2) can be of interest for certain applications. For instance, if we go to the dynamical case, it is after these equations that we shall study wave front propagation (e.g., second-order discontinuities) for the case of finite deformations.
(iv) For the case of dissipative media in Sec. 8, we restricted the form of the constitutive equations by use of the thermodynamic admissibility (Clausius–Duhem inequality).

We may consider that the electric counterpart of the theory presented has been approached by different authors.[19-25] A synthesized theory of both fields in the dynamical case remains to be done. We deal with this problem in a forthcoming article.

The notations used hereafter are similar to those used in the treatises of Truesdell and Toupin[26] and/or Eringen,[27] the first chapter of the latter being worth reading for the reader not familiar with the formalism of modern continuum mechanics. The most frequent notations appearing in the subsequent developments are the following ones. Capital kernel letters and indices refer to material coordinates, i.e., Lagrangian coordinates. Lower case kernel letters and indices refer to spatial coordinates. A capital index following a comma, or a symbol ∇_R indicates partial differentiation with respect to the material coordinates $X^K, K = 1, 2, 3$. A lower case index following a semicolon or a symbol ∇ indicates covariant partial differentiation with respect to the spatial coordinates $x^k, k = 1, 2, 3$. Colons are used to denote covariant total differentiation of two-point tensor fields. A superposed dot or the usual symbol d/dt indicates material derivative. Parentheses around a set of indices denote alteration. ϵ^{ijk} is the permutation symbol. For instance, we write

$$\nabla\cdot\mathbf{a} = a^i{}_{;i}, \qquad (\nabla\mathbf{a})_{ij} = a_{i;j}$$
$$\nabla_R\cdot\mathbf{A} = A^K{}_{;K}, \qquad (\nabla_R\mathbf{a})_{iK} = a_{i,K}$$
$$(\nabla\mathbf{a})\cdot\mathbf{b}|_i = a_{k;i}b^k, \qquad (\mathbf{a}\cdot\nabla)\mathbf{b}|_i = a^k b_{i;k}.$$

2. THE SPIN ANGULAR MOMENTUM

In classical physics, it is well known that the electrons and other particles possess magnetic moments. Thus, in an electronic spin continuum, one can define a magnetic moment per unit mass. For a deformable medium, a magnetic moment per unit of undeformed mass, $\boldsymbol{\mu}$, is posited to exist at each point of the body. The magnetic moment of a specimen $(B_R) \subset E^3$ is

$$\mathfrak{M} = \int_{(B_R)} \rho_R\,\boldsymbol{\mu}\,dv_R = \int_{(B_R)} \boldsymbol{\mu}\,dm_R \tag{2.1}$$

where ρ_R is the material density in the undeformed body (B_R) and m_R a mass measure in (B_R) embedded in Euclidean space E^3.

We assign an intrinsic angular momentum $\mathbf{s}$, called *spin*, to particles such as electrons, muons, protons.[28,29] The *spin angular momentum* is related to the magnetic moment $\boldsymbol{\mu}$ by[30]

$$\boldsymbol{\mu} = \Gamma\mathbf{s}, \tag{2.2}$$

where the quantity

$$\Gamma = -ge/2mc, \qquad e > 0, \tag{2.3}$$

is called the *gyromagnetic ratio* with:

$e \equiv$ electric charge of the particle,

$m \equiv$ mass of the particle,

$c \equiv$ velocity of light,

$g \equiv$ coupling constant (≈ 2 for electrons).

From here on we consider the case of electrons. The medium is thus described by an ensemble of particles to each of which there is attached a vector $\boldsymbol{\mu}$. We propose in Sec. 3 a Lagrangian formulation for the equations governing the behavior of a deformable magnetized medium. The first question that must be answered is what is the form of the spin angular momentum term in the variational principle?

It is clear that this effect is of gyroscopic nature. This is shown by considering the analogy of a spinning magnetized element of material with an axisymmetric top (cf. Art. 5.7 of Ref. 31) of moments of inertia $(A \to 0, A \to 0, C > 0)$.[32] It then appears that the rotation angle Ψ of the top about its own axis is an *ignorable* variable. It follows that the corresponding generalized momentum p_Ψ is conserved. Since the magnetized medium is considered to be saturated, i.e.,

$$\boldsymbol{\mu}\cdot\boldsymbol{\mu} = \mu_s^2 = \text{const}, \tag{2.4}$$

the analogy is carried on by taking

$$p_\Psi = \rho_R\,\mu_s/\Gamma = \text{const}. \tag{2.5}$$

It can then be shown that, in cartesian coordinates, the Routhian[31] or reduced Lagrangian density R of this special top is an expression *linear* in the components of $\dot{\boldsymbol{\mu}}$ (cf., Ref. 7, p. 28, also Ref. 17). This asserts the gyroscopic character of R. Thus, introducing the spin angular momentum per unit volume of undeformed material $\mathbf{G} = -\rho_R\,\boldsymbol{\mu}/\Gamma$, we can consider the quantity $-d\mathbf{G}/dt$ as *d'Alembertian inertia couple*.[15] In an actual motion, its rate of work vanishes as can be easily verified,

$$-\frac{d\mathbf{G}}{dt}\cdot\boldsymbol{\omega} = 0 \tag{2.6}$$

where $\boldsymbol{\omega}$ is the angular velocity of the magnetization

$\boldsymbol{\mu}$. Indeed, differentiating Eq. (2.4) with respect to time, we have

$$\boldsymbol{\mu}\cdot\dot{\boldsymbol{\mu}} = 0 \tag{2.7}$$

which implies

$$\dot{\boldsymbol{\mu}} = \boldsymbol{\omega}\times\boldsymbol{\mu}, \tag{2.8}$$

where

$$\boldsymbol{\omega} = \mu_s^{-2}\boldsymbol{\mu}\times\dot{\boldsymbol{\mu}} + \mu_s^{-2}(\boldsymbol{\mu}\cdot\boldsymbol{\omega})\boldsymbol{\mu}. \tag{2.9}$$

In a virtual motion, the expression (2.6) is different from zero. Note that the angular velocity $\boldsymbol{\omega}$ of $\boldsymbol{\mu}$ is a *kinematical* or *nonholonomic* vector[33,15] i.e., it is not the time derivative of an actual vector function. Thus there exists a vector $\Delta\boldsymbol{\theta}$ such that $\Delta\boldsymbol{\theta} = \boldsymbol{\omega}\Delta t$, which is merely an infinitesimal change of angle and not the differential of a vector function. Therefore, in a virtual motion, we write, instead of Eqs. (2.7) and (2.8),

$$\boldsymbol{\mu}\cdot\delta\boldsymbol{\mu} = 0 \tag{2.10}$$

$$\delta\boldsymbol{\mu} = \delta\boldsymbol{\theta}\times\boldsymbol{\mu}, \tag{2.11}$$

with

$$\delta\boldsymbol{\theta} = \mu_s^{-2}\boldsymbol{\mu}\times\delta\boldsymbol{\mu} + \mu_s^{-2}(\boldsymbol{\mu}\cdot\delta\boldsymbol{\theta})\boldsymbol{\mu}, \tag{2.12}$$

and the virtual work δw done by the d'Alembertian inertia couple in an arbitrary variation is

$$\delta w = \rho_R\Gamma^{-1}\dot{\boldsymbol{\mu}}\cdot\delta\boldsymbol{\theta}. \tag{2.13}$$

Hence, we take account of the spin angular momentum in the variational principle through an already-varied term δW that is not included in the Lagrangian density. Integrating Eq. (2.13) over the whole body (B_R) and over time we obtain

$$\delta W = \int_t dt \int_{(B_R)} \rho_R\Gamma^{-1}\dot{\boldsymbol{\mu}}\cdot\delta\boldsymbol{\theta}\,dv_R. \tag{2.14}$$

An alternate derivation of Eq. (2.14) based on thermodynamical considerations is given in Appendix A.

3. THE VARIATIONAL PRINCIPLE

To every material point $\mathbf{X}$ of a deformable body we assign a vector $\boldsymbol{\mu}$ called magnetization. This vector considered as a director provides a *non-constrained Cosserat continuum.*[18] The origin of $\boldsymbol{\mu}$ cannot translate with respect to the material point, but $\boldsymbol{\mu}$ can rotate independently of the material motion. The complete motion of the continuum is therefore described by two sets of equations:

$$\mathbf{x} = \chi(\mathbf{X}, t), \tag{3.1}$$

$$\boldsymbol{\mu} = \chi_{(\mu)}(\mathbf{X}, t). \tag{3.2}$$

The material gradients defined by

$$F_x = \nabla_R\chi, \qquad F_{(\mu)} = \nabla_R\chi_{(\mu)} \tag{3.3}$$

are essential in the following development.

The *action* associated with the body (B) enclosed within a surface (∂B) and considered in the interval of time $[t_1, t_2]$ is given by

$$A = \int_{t_1}^{t_2} dt \int_{(B_R)} \mathfrak{L}(\mathbf{x}, \boldsymbol{\mu}, \dot{\mathbf{x}}, \dot{\boldsymbol{\mu}}, \mathbf{F}_x, \mathbf{F}_{(\mu)}, \mathbf{X}, t)dv_R, \tag{3.4}$$

where the subscript R indicates the reference configuration. The presence of $\mathbf{X}$ in the Lagrangian density indicates the possible inhomogeneity and those of the gradients $\mathbf{F}_x$ and $\mathbf{F}_{(\mu)}$, the mechanical and magnetic stresses. The absence of higher order spatial derivatives of the motion other than the first shows that we are satisfied with the study of a hyperelastic medium,[34] i.e., a nonlinear elastic medium with constitutive equations derivable from a potential (this is to be compared with the classical theories of magnetostriction and ferromagnetic bodies). The body is assumed to be magnetically saturated. Thus,

$$\boldsymbol{\mu}\cdot\boldsymbol{\mu} = \mu_s^2 = \text{const} \quad \text{and} \quad \boldsymbol{\mu}\cdot\nabla_R\boldsymbol{\mu} = 0 \quad \text{in } (B). \tag{3.5}$$

These constraints may be introduced into the field equations by means of Lagrange multipliers λ and L^K, $K = 1, 2, 3$. Following the tradition set by Lagrange and Piola in formulating the principle of action, we introduce indeterminate multipliers for each term that can arise in varying $\mathbf{x}$ and $\boldsymbol{\mu}$ independently in (B) and on (∂B). We define

$$\begin{aligned}\delta W^* = &\int_t dt \int_{(B_R)} \rho_R(\mathbf{b}\cdot\delta\mathbf{x} + \mathbf{b}^{(\mu)}\cdot\delta\boldsymbol{\mu})dv_R \\ &+ \int_t dt \int_{(\partial B_R)} (\mathbf{t}_R\cdot\delta\mathbf{x} + \mathbf{t}_R^{(\mu)}\cdot\delta\boldsymbol{\mu})ds_R \\ &- \int_{(B_R)} \rho_R(\mathbf{m}\cdot\delta\mathbf{x} + \mathbf{m}^{(\mu)}\cdot\delta\boldsymbol{\mu})dv_R\Big]_{t_1}^{t_2},\end{aligned} \tag{3.6}$$

$$\begin{aligned}\delta W = -\delta\int_t dt \int_{(B_R)} &[\tfrac{1}{2}\lambda(\boldsymbol{\mu}\cdot\boldsymbol{\mu} - \mu_s^2) + \mathbf{L}\cdot(\boldsymbol{\mu}\cdot\nabla_R\boldsymbol{\mu})]dv_R \\ &+ \int_t dt \int_{(B_R)} (\rho_R/\Gamma)\dot{\boldsymbol{\mu}}\cdot\delta\boldsymbol{\theta}\,dv_R.\end{aligned} \tag{3.7}$$

The variational principle that follows earlier formulations of Hamilton's principle[35,36,46] can now be expressed as:

$$\delta A + \delta W^* + \delta W = 0 \tag{3.8}$$

In (3.6) and (3.7) we introduced the following quantities associated with the reference configuration:

ρ_R = mass density,
$\mathbf{b}$ = body force per unit mass,
$\mathbf{b}^{(\mu)}$ = magnetic field,
$\mathbf{t}_R$ = stress vector,
$\mathbf{t}_R^{(\mu)}$ = magnetic stress vector,
$\mathbf{m}$ = momentum per unit mass,
$\mathbf{m}^{(\mu)}$ = magnetic momentum per unit mass.

The expression (3.6) represents the virtual work of body and surface loads acting in (B_R), on (∂B_R) and at the time limits in (B_R). The first term of (3.7) takes account of the holonomic constraints (3.5), while the second term is an already varied term (so called Hertz' nonholonomic constraint[33]).

4. VARIATION OF THE ACTION

Since (B_R) is the reference configuration of the body, we have

$$\delta A = \delta\int_t dt \int_{(B_R)} \mathfrak{L}dv_R = \int_t dt \int_{(B_R)} \delta\mathfrak{L}\,dv_R, \tag{4.1}$$

where

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\mathbf{x}}\cdot\delta\mathbf{x} + \frac{\partial\mathcal{L}}{\partial\dot{\mathbf{x}}}\cdot\delta\dot{\mathbf{x}} + \frac{\partial\mathcal{L}}{\partial\boldsymbol{\mu}}\cdot\delta\boldsymbol{\mu} + \frac{\partial\mathcal{L}}{\partial\dot{\boldsymbol{\mu}}}\cdot\delta\dot{\boldsymbol{\mu}} + \frac{\partial\mathcal{L}}{\partial\mathbf{F}_x}\cdot\delta\mathbf{F}_x + \frac{\partial\mathcal{L}}{\partial\mathbf{F}_{(\mu)}}\cdot\delta\mathbf{F}_{(\mu)} \quad (4.2)$$

and according to Sec. 2

$$\delta\boldsymbol{\mu} = -\boldsymbol{\mu}\times\delta\boldsymbol{\theta}. \quad (4.3)$$

We make use of the following identities:

$$\int_t \frac{\partial\mathcal{L}}{\partial\dot{\mathbf{x}}}\cdot\delta\dot{\mathbf{x}}\,dt = \frac{\partial\mathcal{L}}{\partial\dot{\mathbf{x}}}\cdot\delta\mathbf{x}\Big]_{t_1}^{t_2} - \int_t \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{\mathbf{x}}}\right)\cdot\delta\mathbf{x}\,dt, \quad (4.4)$$

$$\int_{(B_R)} \frac{\partial\mathcal{L}}{\partial\mathbf{F}_x}\cdot\delta\mathbf{F}_x\,dv_R = \int_{(B_R)} \frac{\partial\mathcal{L}}{\partial\mathbf{F}_x}\cdot\nabla_R(\delta\mathbf{x})\,dv_R = \int_{(\partial B_R)} \frac{\partial\mathcal{L}}{\partial\mathbf{F}_x}\cdot\mathbf{N}_R\,\delta\mathbf{x}\,ds_R - \int_{(B_R)} \nabla_R\cdot\left(\frac{\partial\mathcal{L}}{\partial\mathbf{F}_x}\right)\delta\mathbf{x}\,dv_R, \quad (4.5)$$

where $\mathbf{N}_R$ is the unit exterior normal to (∂B_R). In (4.4) we interchanged the δ variation with the material derivatives, and in (4.5) we applied the Green–Gauss theorem. Similar expressions are valid for terms involving $\partial\mathcal{L}/\partial\boldsymbol{\mu}$ and $\partial\mathcal{L}/\partial\mathbf{F}_{(\mu)}$.

Equation (3.8) now is

$$\begin{aligned}\int_t dt \int_{(B_R)} \Big\{ &\left[\frac{\partial\mathcal{L}}{\partial\mathbf{x}} - \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{\mathbf{x}}}\right) - \nabla_R\cdot\left(\frac{\partial\mathcal{L}}{\partial\mathbf{F}_x}\right) + \rho_R\mathbf{b}\right]\cdot\delta\mathbf{x} \\ &+ \left[\frac{\partial\mathcal{L}}{\partial\boldsymbol{\mu}} - \frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{\boldsymbol{\mu}}}\right) - \nabla_R\cdot\left(\frac{\partial\mathcal{L}}{\partial\mathbf{F}_{(\mu)}}\right) - (\lambda - \nabla_R\cdot\mathbf{L})\boldsymbol{\mu} + \rho_R\mathbf{b}^{(\mu)}\right]\cdot\delta\boldsymbol{\mu}\Big\} dv_R \\ &+ \int_t dt \int_{(\partial B_R)} \Big\{\left[\frac{\partial\mathcal{L}}{\partial\mathbf{F}_x}\cdot\mathbf{N}_R + \mathbf{t}_R\right]\cdot\delta\mathbf{x} + \left[\frac{\partial\mathcal{L}}{\partial\mathbf{F}_{(\mu)}}\cdot\mathbf{N}_R + \mathbf{t}_R^{(\mu)} - (\mathbf{L}\cdot\mathbf{N}_R)\boldsymbol{\mu}\right]\cdot\delta\boldsymbol{\mu}\Big\} ds_R + \int_t dt \int_{(B_R)} (\rho_R/\Gamma)\dot{\boldsymbol{\mu}}\cdot\delta\boldsymbol{\theta}\,dv_R \\ &- \int_{(B_R)} \Big[\left(\rho_R\mathbf{m} - \frac{\partial\mathcal{L}}{\partial\dot{\mathbf{x}}}\right)\cdot\delta\mathbf{x} + \left(\rho_R\mathbf{m}^{(\mu)} - \frac{\partial\mathcal{L}}{\partial\dot{\boldsymbol{\mu}}}\right)\cdot\delta\boldsymbol{\mu}\Big] dv_R\Big]_{t_1}^{t_2} = 0 \end{aligned} \quad (4.6)$$

and holds for independent variation of $\mathbf{x}$ and $\boldsymbol{\mu}$ (equivalently $\delta\boldsymbol{\theta}$) in (B_R) and on (∂B_R). Hence

$$\rho_R\mathbf{m} = \frac{\partial\mathcal{L}}{\partial\dot{\mathbf{x}}}, \quad \rho_R\mathbf{m}^{(\mu)} = \frac{\partial\mathcal{L}}{\partial\dot{\boldsymbol{\mu}}} \text{ at } t = t_1, \quad t = t_2, \quad (4.7)$$

$$\nabla_R\cdot\mathbf{T}_R + \rho_R\mathbf{b} = \rho_R\dot{\mathbf{m}} - \frac{\partial\mathcal{L}}{\partial\mathbf{x}} \text{ in } (B_R), \quad (4.8)$$

$$[\nabla_R\cdot\mathbf{T}_R^{(\mu)} + \rho_R(\mathbf{b}^{(\mu)} + \mathbf{b}^{*(\mu)}) - (\lambda - \nabla_R\cdot\mathbf{L})\boldsymbol{\mu}]\times\boldsymbol{\mu} = \rho_R\dot{\mathbf{m}}^{(\mu)}\times\boldsymbol{\mu} + (\rho_R/\Gamma)\dot{\boldsymbol{\mu}} \text{ in } (B_R), \quad (4.9)$$

and

$$\mathbf{T}_R\cdot\mathbf{N}_R = \mathbf{t}_R \text{ on } (\partial B_R), \quad (4.10)$$

$$[\mathbf{T}_R^{(\mu)}\cdot\mathbf{N}_R - \mathbf{t}_R^{(\mu)} + (\mathbf{L}\cdot\mathbf{N}_R)\boldsymbol{\mu}]\times\boldsymbol{\mu} = 0 \text{ on } (\partial B_R), \quad (4.11)$$

where we defined the following suggestive symbols:

$$\mathbf{T}_R \equiv -\frac{\partial\mathcal{L}}{\partial\mathbf{F}_x}, \quad \mathbf{T}_R^{(\mu)} \equiv -\frac{\partial\mathcal{L}}{\partial\mathbf{F}_{(\mu)}}, \quad \rho_R\mathbf{b}^{*(\mu)} \equiv \frac{\partial\mathcal{L}}{\partial\boldsymbol{\mu}}. \quad (4.12)$$

In order to associate various mathematical notions obtained with physical concepts, we now consider a particular Lagrangian:

$$\mathcal{L} = \rho_R[\tfrac{1}{2}\dot{\mathbf{x}}^2 - \mathcal{F}(\mathbf{F}_x, \boldsymbol{\mu}, \mathbf{F}_{(\mu)}, \mathbf{X})]; \quad (4.13)$$

hence, we have

$$\rho_R\mathbf{m} = \rho_R\dot{\mathbf{x}} \equiv \rho_R\mathbf{v}, \quad \rho_R\mathbf{m}^{(\mu)} \equiv 0, \quad \mathbf{b}^{*(\mu)} = -\frac{\partial\mathcal{F}}{\partial\boldsymbol{\mu}}, \quad (4.14)$$

$$\mathbf{T}_R = \rho_R\frac{\partial\mathcal{F}}{\partial\mathbf{F}_x}, \quad \mathbf{T}_R^{(\mu)} = \rho_R\frac{\partial\mathcal{F}}{\partial\mathbf{F}_{(\mu)}}, \quad \frac{\partial\mathcal{L}}{\partial\mathbf{x}} \equiv 0, \quad \frac{\partial\mathcal{L}}{\partial t} \equiv 0. \quad (4.15)$$

In the expression of the body force $\mathbf{b}$, we distinguish the purely mechanical part and the magnetic contribution, i.e., we write

$$\rho_R\mathbf{b} = \rho_R\mathbf{b}_{(\text{mech})} + \rho_R\mathbf{b}_{(\text{magn})}, \quad (4.16)$$

Without any damage to the calculated values of observable physical forces, one is at liberty to choose a formula for the ponderomotive force with a large degree of arbitrariness. This was emphasized by Sedov and others.[37,38] Here we take

$$\mathbf{b}_{(\text{magn})} = (\nabla\mathbf{B})\cdot\boldsymbol{\mu}, \text{ i.e., } \mathbf{b}_{(\text{magn})k} = B_{i;k}\mu^i \quad (4.17)$$

and

$$\mathbf{b}_{(\text{mech})} = \mathbf{f}, \quad \mathbf{b}^{(\mu)} \equiv \mathbf{B}, \quad (4.18)$$

where $\mathbf{f}$ stands for the mechanical body force per unit mass (e.g., gravity) and $\mathbf{B}$ is the *Maxwellian* magnetic field subject to Maxwell's equations given by:

$$\begin{aligned} &\nabla\times\mathbf{B} = \nabla\times\rho\boldsymbol{\mu}, \quad \nabla\cdot\mathbf{B} = 0 \text{ in } (B), \\ &\nabla\times\mathbf{B} = 0, \quad \nabla\cdot\mathbf{B} = 0 \text{ outside } (B), \\ &\mathbf{n}\times[\mathbf{B} - \rho\boldsymbol{\mu}] = 0, \quad \mathbf{n}\cdot[\mathbf{B}] = 0 \text{ across } (\partial B). \end{aligned} \quad (4.19)$$

Here ρ is the mass density in the deformed configuration, $\mathbf{n}$ is the unit exterior normal to (∂B) and $[A] = A^+ - A^-$.

The expression (4.17) is that resulting from the analysis of Dixon and Eringen[22] for the case of *magnetostatics*, when no discontinuity surface exists in (B_R) and the *currents are neglected*. It also results in the same conditions from the relativistic treatment of Grot and Eringen.[39] For the analysis of surface tractions on (∂B_R) and for the case for which a discontinuity surface exists in (B_R), the surface terms corresponding to (4.17) are derived in Appendix B.

5. FIELD EQUATIONS IN THE DEFORMED CONFIGURATION

We call $\mathbf{T}_R$ the "macro" Piola stress tensor and, by analogy with the couple theory, $\mathbf{T}_R^{(\mu)}$ the "micro" Piola stress tensor.[40] Eqs. (4.8) and (4.9) with (4.14), (4.15), (4.17) and (4.18) take the form:

$$\nabla_R\cdot\left(\rho_R\frac{\partial\mathcal{F}}{\partial\mathbf{F}_x}\right) + \rho_R\mathbf{f} + \rho_R(\nabla\mathbf{B})\cdot\boldsymbol{\mu} = \rho_R\dot{\mathbf{v}}, \text{ in } (B_R), \quad (5.1)$$

$$\left[\nabla_R \cdot \left(\rho_R \frac{\partial \mathfrak{F}}{\partial \mathbf{F}_{(\mu)}}\right) + \rho_R\left(\mathbf{B} - \frac{\partial \mathfrak{F}}{\partial \boldsymbol{\mu}}\right)\right] \times \boldsymbol{\mu} = \frac{\rho_R}{\Gamma}\dot{\boldsymbol{\mu}}, \text{ in } (B_R). \quad (5.2)$$

In component notations, we have

$$(\mathbf{T}_R)_k^{\ K} = \rho_R \frac{\partial \mathfrak{F}}{\partial x^k_{,K}} \quad (5.3)$$

Introducing the Cauchy stress t_k^l by

$$T_k^{\ K} = JX^K_{\ ,l} t_k^{\ l}, \qquad t_k^{\ l} = \rho \frac{\partial \mathfrak{F}}{\partial x^k_{,K}} x^l_{1K}, \quad (5.4)$$

where

$$J = \rho_R/\rho, \quad (5.5)$$

and, via the identity[41]

$$(JX^K_{\ ,l})_{:K} = 0 \quad (5.6)$$

Eq. (5.1), in the deformed configuration, takes the form

$$t_k^{\ l}{}_{;l} + \rho\mu_l B^l_{\ ;k} + \rho f_k = \rho \dot{v}_k \text{ in } (B) \quad (5.7)$$

If the same approach is considered for the "magnetic stress" (not to be mistaken with the Maxwell stress tensor)

$$(\mathbf{T}^{(\mu)})_k^{\ K} = JX^K_{\ ,l} t^{(\mu)}_k{}^l, \qquad t^{(\mu)}_k{}^l = \rho \frac{\partial \mathfrak{F}}{\partial \mu^k_{,K}} x^l_{,K}, \quad (5.8)$$

then (5.2) takes the form

$$\epsilon^{ikm} t^{(\mu)}_k{}^l{}_{;l}\mu_m + \epsilon^{ikm}(B_k + {}_L B_k)\mu_m = (\rho/\Gamma)\dot{\mu}^i, \quad (5.9)$$

where we have set

$${}_L\mathbf{B} \equiv -\frac{\partial \mathfrak{F}}{\partial \boldsymbol{\mu}} = \mathbf{b}^{*(\mu)}. \quad (5.10)$$

Eq. (5.9) is a torque equation which may be written concisely

$$\rho\dot{\boldsymbol{\mu}} = \Gamma \mathbf{B}_{(\text{eff})} \times \mathbf{M}, \qquad \mathbf{M} \equiv \rho\boldsymbol{\mu}, \quad (5.11)$$

with

$$\mathbf{B}_{(\text{eff})} \equiv \mathbf{B} + {}_L\mathbf{B} + \rho^{-1}\left(\rho \frac{\partial \mathfrak{F}}{\partial \boldsymbol{\mu}_{,K}} x^l_{\ ,K}\right)_{;l}. \quad (5.12)$$

Equation (5.11) is an equation for the balance of magnetic angular momentum where the angular velocity is given by an effective magnetic field which includes the anisotropy effect through ${}_L\mathbf{B}$ and the exchange forces due to the presence of material gradients of the magnetization. These latter reflect the neighboring spin interactions according to Brown.[14]

Introducing the dual $\mathbf{S}$ of $\boldsymbol{\mu}$:

$$\mathbf{S} = \text{dual } \boldsymbol{\mu}, \text{ i.e., } S_{ln} = \epsilon_{lni}\mu^i, \quad (5.13)$$

with

$$\mu_s^2 = \mu_k\mu^k = \tfrac{1}{2}S_{kl}S^{lk}.$$

Eq. (5.11) may be written in the following form

$$\rho\dot{S}^{kl} = 2\Gamma B^{[k}_{(\text{eff})} M^{l]}, \quad (5.14)$$

which provides a useful suggestion for the equation satisfied by the spin tensor in four-dimensional analysis.

6. BOUNDARY CONDITIONS

Using the relations (4.10), (4.11), (5.4), and (5.8) we can give a set of boundary conditions on (∂B) in the deformed configuration. In the absence of any mechanical couple stress vector on (∂B) and the magnetization being zero outside (B), it is reasonable to take $\mathbf{t}_R^{(\mu)} = 0$ on (∂B). From (4.10) and (4.11), it then follows that:

$$\rho \frac{\partial \mathfrak{F}}{\partial x^k_{,K}} x^l_{\ ,K} n_l = t_k = \tau_k - (\overline{\mathbf{M}} \cdot [\mathbf{B}])n_k \text{ on } (\partial B) \quad (6.1)$$

since, similarly to the decomposition (4.16), we set on the frontier (∂B)

$$t_k = \tau_k + t_{(em)k}, \qquad t_{(em)k} = -(\overline{\mathbf{M}} \cdot [\mathbf{B}])n_k,$$

where τ_k is a mechanical stress vector prescribed on (∂B) and the second term is the magnetic contribution to the surface traction (cf. Appendix B).

With the foregoing assumptions, Eq. (4.11) yields

$$\rho\mu_{[m} \frac{\partial \mathfrak{F}}{\partial \mu^{k]}_{,K}} x^l_{\ ,K} n_l = 0 \text{ on } (\partial B). \quad (6.2)$$

It is clear that the Lagrange multipliers became irrelevant in the formulation. Nevertheless one can determine them (or at least the combination $\lambda - \nabla_R \cdot \mathbf{L}$) since we should take in all rigor from (4.6)

$${}_L\mathbf{B} = -\frac{\partial \mathfrak{F}}{\partial \boldsymbol{\mu}} - (\lambda - \nabla_R \cdot \mathbf{L})\boldsymbol{\mu}, \quad (6.3)$$

but clearly from Eq. (5.9), only the part of ${}_L\mathbf{B}$ orthogonal to $\boldsymbol{\mu}$ is of importance, and, without loss of generality, we may take

$${}_L\mathbf{B} \cdot \boldsymbol{\mu} = 0, \quad (6.4)$$

hence the value of $(\lambda - \nabla_R \cdot \mathbf{L})$

$$\lambda - \nabla_R \cdot \mathbf{L} = -\frac{1}{\mu_s^2} \frac{\partial \mathfrak{F}}{\partial \boldsymbol{\mu}} \cdot \boldsymbol{\mu}. \quad (6.5)$$

Remarks

(i) The equation of energy can be arrived at by manipulating Eqs. (5.7) and (5.9) but we provide a more direct approach in Sec. 7.

(ii) Steady discontinuity surfaces will be dealt with in the next section. Their introduction requires an appropriate extension of the Green–Gauss theorem, (see Generalized Green–Gauss theorem in Eringen, Appendix of Ref. 34.

(iii) The equation (5.11) obtained from the Lagrangian density (4.13) describes a rotation of $\boldsymbol{\mu}$ in the plane formed by the magnetization $\boldsymbol{\mu}$ (at $t = t_1$) and the effective field $\mathbf{B}_{(\text{eff})}$ of the equilibrium condition. It is however observed[43] that the vector $\boldsymbol{\mu}$ spirals into parallelism with $\mathbf{B}_{(\text{eff})}$, an effect that can be represented by adding a damping term to Eq. (5.11). Though we restrain from introducing a too special form of dissipation (this would be inconsistent with the nonlinearity of the nondissipative terms derived from the potential $\mathfrak{F}$), this could indeed be taken into account in the present formulation. All we need to do is to add a term

$$\int dt \int \frac{\partial \mathfrak{R}}{\partial \dot{\boldsymbol{\mu}}} \cdot \delta\boldsymbol{\mu} dv_R$$

to the expression (3.8). With α a damping coefficient, the Rayleigh dissipation density $\mathcal{R}$ is defined as

$$\text{(a)}\quad \mathcal{R} = \tfrac{1}{2}\rho_R \alpha \mu_s^{-1}\dot{\boldsymbol{\mu}}^2.$$

Equation (4.9) when transformed into the deformed configuration would then yield

$$\text{(b)}\quad \Gamma^{-1}\dot{\boldsymbol{\mu}} = \mathbf{B}_{(\text{eff})} \times \boldsymbol{\mu} - (\alpha/\mu_s)\dot{\boldsymbol{\mu}} \times \boldsymbol{\mu}.$$

The last term of this equation which describes the approach of $\boldsymbol{\mu}$ to $\mathbf{B}_{(\text{eff})}$ in a more realistic manner than Eq. (5.11), has already been introduced by Gilbert and Kelley[44] and is known to give a good description of the loss mechanisms in a number of applications in ferromagnetic resonance; in fact a better description[43] than the supplementary term $\lambda(\mathbf{B}_{(\text{eff})} \times \boldsymbol{\mu}) \times \boldsymbol{\mu}$ introduced earlier by Landau and Lifshitz.[10] For *small damping*, we can replace $\dot{\boldsymbol{\mu}}$ in the last term of Eq. (b) by its value given by Eq. (5.11) and, setting $\lambda = -\alpha\Gamma/\mu_s$, we obtain an expression similar to that of Landau and Lifshitz:

$$\Gamma^{-1}\dot{\boldsymbol{\mu}} = \mathbf{B}_{(\text{eff})} \times \boldsymbol{\mu} + \lambda\,(\mathbf{B}_{(\text{eff})} \times \boldsymbol{\mu}) \times \boldsymbol{\mu},$$

where $\mathbf{B}_{(\text{eff})}$ in general contains nonlinear expressions.

7. EUCLIDEAN INVARIANCE REQUIREMENT

We now apply the requirement of Euclidean invariance (i.e., Galilean invariance) first used by the Cosserats[45] and taken over by Toupin[18] (see also Maugin[46]). It allows us to obtain the balance equations including the equation of energy balance by application of Noether's theorem of invariance for the group of transformations

$$\begin{aligned} x^{*k} &= x^{*k}(\mathbf{X}, t^*) = Q^k{}_l x^l(\mathbf{X}, t) + c^k, \\ \mu^{*k} &= \mu^{*k}(\mathbf{X}, t^*) = Q^k{}_l \mu^l(\mathbf{X}, t), \\ t^* &= t - a, \end{aligned} \tag{7.1}$$

where $\mathbf{c}$ is an arbitrary constant vector, $\mathbf{Q}$ is an arbitrary constant proper orthogonal tensor and a is an arbitrary constant. Invariance under (7.1) represents the restrictions to be placed on the field equations deduced from the variational formulation under the group of rigid body motions and time shifts.

We consider infinitesimal transformations generated by (7.1) specifically, three sets of variations corresponding to these infinitesimal transformations:

(i) shift of coordinates:

$$\delta x^k = d^k, \qquad \delta\mu^k = 0, \tag{7.2}$$

where $\mathbf{d}$ is an infinitesimally small constant vector.

(ii) rotation:

$$\delta x^k = eQ^k{}_l x^l, \qquad \delta\mu^k = eQ^k{}_l \mu^l, \tag{7.3}$$

where e is infinitesimally small and $\mathbf{Q}$ is an arbitrary constant skew symmetric tensor.

(iii) shift of time:

$$\delta x^k \to \dot{x}^k, \qquad \delta\mu^k \to \dot{\mu}^k = \epsilon^{klm}\omega_l \mu_m, \tag{7.4}$$

where $\boldsymbol{\omega}$ is the angular velocity of $\boldsymbol{\mu}$. We shall also use $\boldsymbol{\Omega}$, the dual of $\boldsymbol{\omega}$ defined by

$$\Omega^{ij} \equiv \epsilon^{ijk}\omega_k. \tag{7.5}$$

We set

$$A_{kl} \equiv x_{[k}m_{l]} + \mu_{[k}m^{(\mu)}_{l]}, \qquad A_{kl} = -A_{lk}, \tag{7.6}$$

$$\epsilon \equiv m_k \dot{x}^k + m^{(\mu)}_k \dot{\mu}^k - (\mathfrak{L}/\rho_R), \tag{7.7}$$

$$\rho_R K_{kl} \equiv x_k \frac{\partial \mathfrak{L}}{\partial x^l} + \mu_k \frac{\partial \mathfrak{L}}{\partial \mu^l} + \dot{x}_k \frac{\partial \mathfrak{L}}{\partial \dot{x}^l} + \dot{\mu}_k \frac{\partial \mathfrak{L}}{\partial \dot{\mu}^l} + x_{k,K} \frac{\partial \mathfrak{L}}{\partial x^l{}_{,K}} + \mu_{k,K} \frac{\partial \mathfrak{L}}{\partial \mu^l{}_{,K}}. \tag{7.8}$$

ϵ is called the energy. A form more interesting can be given for δW. First the constraints (3.5) can be discarded as has been shown above. Second, since $\delta\boldsymbol{\theta}$ is arbitrary and dual $\mathbf{Q}$ represents a rotation of the same nature, we can take

$$\delta\boldsymbol{\theta} = e \text{ dual } Q, \text{ i.e.,} \qquad \delta\theta^i = \tfrac{1}{2}e\epsilon^{ijk}Q_{jk}. \tag{7.9}$$

Thus, using definition (5.13), we have

$$\dot{\boldsymbol{\mu}} \cdot \delta\boldsymbol{\theta} = \tfrac{1}{2}\, e\dot{S}_{kl}Q^{lk}. \tag{7.10}$$

Necessary and sufficient conditions that (3.8) holds for the special variations (7.2)–(7.4) are:

$$\int_{(B_R-\Gamma_R)} \rho_R m_i dv_R \Big]_{t_1}^{t_2} - \int_t dt \int_{(B_R)} \rho_R b_i dv_R - \int_t dt \int_{(\partial B_R-\Gamma_R)} (\mathbf{t}_R)_i ds_R = \int_t dt \int_{(B_R-\Gamma_R)} \frac{\partial \mathfrak{L}}{\partial x^i} dv_R, \tag{7.11}$$

$$\begin{aligned} &\int_{(B_R-\Gamma_R)} \rho_R A_{kl} dv_R \Big]_{t_1}^{t_2} - \int_t dt \int_{(B_R)} \rho_R (x_{[k}b_{l]} + \mu_{[k}b^{(\mu)}_{l]}) dv_R \\ &\quad - \int_t dt \int_{(B_R-\Gamma_R)} \{x_{[k}(\mathbf{t}_R)_{l]} + \mu_{[k}(\mathbf{t}^{(\mu)}_R)_{l]}\} ds_R \\ &\quad + \int dt \int_{(B_R-\Gamma_R)} (\rho_R/2\Gamma)\dot{S}_{kl}\, dv_R = \int_t dt \int_{(B_R-\Gamma)} \rho_R K_{[kl]} dv_R, \end{aligned} \tag{7.12}$$

$$\begin{aligned} &\int_{(B_R-\Gamma_R)} \rho_R \epsilon dv_R \Big]_{t_1}^{t_2} - \int_t dt \int_{(B_R)} \rho_R (b_k \dot{x}^k + \mu_k b^{(\mu)}_l \Omega^{lk}) dv_R \\ &\quad + \int_t dt \int_{(B_R-\Gamma_R)} (\rho_R/2\Gamma)\dot{S}_{kl}\Omega^{lk} dv_R \\ &\quad - \int_t dt \int_{(\partial B_R-\Gamma_R)} \{(\mathbf{t}_R)_k \dot{x}^k + \mu_k (\mathbf{t}^{(\mu)}_R)_l \Omega^{lk}\} ds_R \\ &\quad = -\int_t dt \int_{(B_R-\Gamma_R)} \frac{\partial \mathfrak{L}}{\partial t} dv_R. \end{aligned} \tag{7.13}$$

Here we have considered a steady discontinuity surface (Γ_R) in (B_R).

We say that the variational principle satisfies the Euclidean invariance requirement if and only if:

$$\frac{\partial \mathfrak{L}}{\partial \mathbf{x}} = 0, \qquad K_{[kl]} = 0, \qquad \frac{\partial \mathfrak{L}}{\partial t} = 0. \tag{7.14}$$

If these conditions are fulfilled, then (7.11)–(7.14) are the global balance laws in the reference configuration, i.e., the conservation of momentum, moment of momentum, and the conservation of energy for the body (B_R) as a whole for nondissipative processes. To them must be adjoined the integral form of the con-

servation of mass:

$$\int_{(B_R-\Gamma_R)} \rho_R dv_R = \text{const.} \tag{7.15}$$

Note that, according to Appendix B, the total body force (for the case of magnetostatics, with the currents neglected) reads:

$$\int_{(B_R)} \rho_R \mathbf{b} dv_R = \int_{(B_R-\Gamma_R)} \rho_R \{\mathbf{f} + (\nabla \mathbf{B})\cdot\boldsymbol{\mu}\} dv_R + \int_{(\Gamma_R)} J\mathbf{F}_x^{-1}\cdot\mathbf{N}_R(\overline{\mathbf{M}}\cdot[\mathbf{B}])\, ds_{\Gamma_R}. \tag{7.16}$$

In order to establish the second term in the reference configuration, we have made use of the identity

$$[J\mathbf{F}_x^{-1}]\cdot\mathbf{N}_R = 0 \tag{7.17}$$

which follows from (5.6).

Upon using a procedure similar to that used by Toupin[18] and Maugin[46], the special form (4.13) and formulas (4.12) and (4.15), it is not difficult to show that (7.13) can be written in the form

$$\int_t dt \int_{(B_R-\Gamma_R)} \rho_R \dot{\mathfrak{F}} dv_R - \int_t dt \int_{(B_R-\Gamma_R)} \rho_R \mu_k b_l^{(\mu)} \Omega^{lk} dv_R - \int_t dt \int_{(\Gamma_R)} [T^{kK}\dot{x}_k + \mu_k T^{(\mu)K}_l \Omega^{lk}] N_K ds_{\Gamma_R} + \int_t dt \int_{(B_R-\Gamma_R)} (T^{kK}\dot{x}_{k,K} + \mu_k T^{(\mu)K}_l \Omega^{lk}{}_{,K}) dv_R = 0. \tag{7.18}$$

Equations (7.11), (7.12), (7.15), and (7.18), subject to (7.14), are posited to be valid for every part of the body and the discontinuity surface (Γ_R). The local balance laws in the reference configuration (and after some transformations, in the deformed configuration) follow.

(a) Equation of momentum balance:

$$\frac{\partial}{\partial t}(\rho v^k) = (t^{kp} + t^{kp}_{(em)} - \rho v^k v^p)_{;p} + \rho f \quad \text{in } (B-\Gamma), \tag{7.19}$$

$$[t^{kp} + t^{kp}_{(em)} - \rho v^k v^p] n_p = 0 \text{ on } (\Gamma), \tag{7.20}$$

where we have used (4.13). Here t^{kp} is given by the second of Eqs. (5.4) and we have set

$$t^{kp}_{(em);p} = \rho B^{p;k}\mu_p, \tag{7.21a}$$

$$[t^{kp}_{(em)}] n_p = \overline{M}_p [B^p] n^k. \tag{7.21b}$$

(b) Equation of spin angular momentum:

Introducing the notations:

$$L_{kl} = -\mu_k b_l^{(\mu)}, \quad D_{kl} = \mu_k M_l^{(\mu)}, \quad \mathcal{S}_{kl} = D_{kl} + (1/2\Gamma) S_{kl}, \tag{7.22}$$

$$M_{pqr} = -\mu_{[p} t^{(\mu)}_{q]r} \tag{7.23}$$

where, for the lack of better terminology, we call $\mathbf{L}$ the *body couple*, $\mathbf{D}$ the *extrinsic spin* (in opposition with the *intrinsic spin* $S/2\Gamma$), $\mathcal{S}$ the *total spin* and $\mathbf{M}$ the *magnetic couple stress* (a third order tensor). We obtain from (7.12):

$$\rho(\mathcal{S}_{[kl]} + \dot{x}_{[k} m_{l]}) = M_{kl}{}^{p}{}_{;p} + \rho L_{[kl]} + t_{[lk]} \text{ in } (B-\Gamma), \tag{7.24}$$

$$[M_{klp} + x^{[k} t^{l]p} + x^{[k} t^{l]p}_{(em)}] n_p = 0 \text{ on } (\Gamma). \tag{7.25}$$

Upon use of (4.13), (4.19), and (7.21b), these formulas reduce to

$$(\rho/2\Gamma)\dot{S}^{pq} = M^{pqr}{}_{;r} + t^{[qp]} + \rho B^{[p}\mu^{q]} \text{ in } (B-\Gamma), \tag{7.26}$$

$$[x^{[p} t^{q]r} + M^{pqr}] n_r + (\overline{\mathbf{M}}\cdot[\mathbf{B}]) x^{[p} n^{q]} = 0 \text{ on } (\Gamma). \tag{7.27}$$

Eq. (7.26) looks like a *couple stress equation*.[18] We emphasize here the analogy existing between the manifestation of extra degrees of freedom due to external causes (i.e., magnetic field) and those due to pure mechanical (or structural) causes (e.g., micropolar theory[47]). This analogy persists throughout the treatment.

(c) Equation of energy:

From (7.18), on account of (4.13), (7.22), and (7.23), we obtain

$$\rho\dot{\mathfrak{F}} = \rho L^{kl}\Omega_{kl} + t^{kl}v_{k;l} + M^{klm}\Omega_{kl;m} \quad \text{in } (B-\Gamma) \tag{7.28}$$

$$[t^{lk}v_l + t^{lk}_{(em)}v_l + M^{pqk}\Omega_{pq}] n_k = 0 \quad \text{on } (\Gamma) \tag{7.29}$$

(d) Equation of continuity:

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0 \quad \text{in } (B-\Gamma), \tag{7.30a}$$

$$[\rho\mathbf{v}]\cdot\mathbf{n} = 0 \quad \text{on } (\Gamma). \tag{7.30b}$$

(e) Boundary conditions:

They are

$$t^k = t^{kl} n_l \quad \text{on } (\partial B - \Gamma), \tag{7.31}$$

where t_k is given by the right-hand side of Eq. (6.1), and

$$M^{pqr}_{(\text{in})} n_r = 0 \quad \text{on } (\partial B - \Gamma). \tag{7.32}$$

The latter is obtained by letting the surface of discontinuity (Γ) coincide with the frontier (∂B) in formula (7.27), the magnetization vanishing outside (B). $M^{pqr}_{(\text{in})}$ is the interior value of M^{pqr} on the surface (∂B).

Finally we recall the constitutive equations. They are given by the *definitions*

$$t_k{}^l = \rho\frac{\partial\mathfrak{F}}{\partial x^k{}_{,K}} x^l{}_{,K}, \qquad M^{pqr} = -\rho\mu^{[p}\frac{\partial\mathfrak{F}}{\partial\mu_{q],K}} x^{r}{}_{,K}, \tag{7.33}$$

Equations (7.19), (7.20), *and* (7.26)–(7.33) *constitute the fundamental set of field equations, boundary conditions and constitutive equations for the theory of magnetically saturated elastic solids (for the case of quasimagnetostatics with the currents neglected).* They are supplemented with the Maxwell's equations (4.19) to which must be adjoined the jump relations:

$$\mathbf{n}\cdot[\mathbf{B}] = 0, \qquad \mathbf{n}\times[\mathbf{B}] = \mathbf{n}\times[\rho\mu] \text{ on } (\Gamma). \tag{7.34}$$

The field equations are identical to those found in Sec. 5, except for the spin angular momentum equation.

In fact, even Eqs. (7.26) and (5.11) are identical. This is shown as follows:

From the definition (7.8) and the expression (4.13), we find that the second of Eqs. (7.14) gives:

$$\frac{\partial \mathfrak{F}}{\partial x^{[k}{}_{,K}}\, x_{l],K} + \frac{\partial \mathfrak{F}}{\partial \mu^{[k}}\, \mu_{l]} + \frac{\partial \mathfrak{F}}{\partial \mu^{[k}{}_{,K}}\, \mu_{l],K} = 0. \quad (7.35)$$

Upon use of the first of Eqs. (7.33), Eq. (7.35) (which is referred to as the "Euclidean invariance requirement") yields a constitutive equation for the antisymmetric part of the stress tensor:

$$t_{[kl]} = -\rho\left(\frac{\partial \mathfrak{F}}{\partial \mu^{[k}}\, \mu_{l]} + \frac{\partial \mathfrak{F}}{\partial \mu^{[k}{}_{,K}}\, \mu_{l],K}\right). \quad (7.36)$$

This is identical to Eq. (7.53) of Brown.[11] It will be shown in Part II that, as a consequence of the objectivity requirement, the last term of (7.36) vanishes. Therefore if we set, as in Sec. 5:

$$_{L}B_{k} = -\frac{\partial \mathfrak{F}}{\partial \mu^{k}} \quad (7.37)$$

we get

$$t_{[kl]} = \rho\, {}_{L}B_{[k}\mu_{l]}. \quad (7.38)$$

Finally, upon carrying (7.36) into (7.26) and using the second of Eqs. (7.33), we obtain Eq. (5.11).

8. DIRECT APPROACH THROUGH BALANCE LAWS

A. Nondissipative Case

It is of interest to compare the foregoing variational treatment with the direct approach based on the statement of the global balance laws. The latter follows a thorough analysis of the forces and couples acting upon the continuum. In a paper,[15] Tiersten has constructed such a model. In his description, the electronic spin continuum built up of magnetization vectors attached to each material point is assumed to possess properties such as those enunciated in Sec. 3. The balance laws are stated and are similar to the set (7.11)–(7.13), and (7.15) but for the moment of momentum equation which is split into two parts due to the coexistence of two continua that can rotate independently. The moment of momentum equation of the continuum built up of the material points gives in our notations

$$t_{[kl]} = \rho\, {}_{L}B_{[k}\mu_{l]} \text{ in } (B - \Gamma), \quad (8.1)$$

and the spin angular momentum equation of the electronic spin continuum reads

$$(\rho/\Gamma)\dot{\mu}^{i} = \rho\epsilon^{ijk}(B_{j} + {}_{L}B_{j} + \rho^{-1}t_{j}^{(\mu)\,m}{}_{;m})\mu^{k} + \epsilon^{ijk}t_{j}^{(\mu)\,m}\mu_{k;m} \quad \text{in } (B - \Gamma), \quad (8.2)$$

$$\epsilon^{ij}\,[\mu_{j}t^{(\mu)}{}_{k}^{;m}]n_{m} = 0 \quad \text{on } (\Gamma). \quad (8.3)$$

It will be shown in Part II, while studying the objectivity, that the last term of (8.2) is zero. Therefore (8.2) is similar to (5.9). We may say that the structure of Eq. (8.2) resulted from two postulates:

Postulate A: Motivated by the classical formula of Larmor precession, it is posited that the couples acting on the electronic spin continuum are of the form $\mathfrak{G} \times \mathbf{M}$ where $\mathbf{M}$ is the magnetization per unit volume and $\mathfrak{G}$ is a quantity proportional to a magnetic field.

This is of course a kind of *petitio principii* since the postulate gives at once the final form of the equation satisfied by the magnetization field. No such postulate, except for the action of the Maxwellian field, has been used in the variational treatment.

Postulate B: The short-range vector field describing the interaction between neighboring spins satisfies a principle similar to the Cauchy principle for stresses, i.e.,

$$\mathbf{t}_{k}^{(\mu)} = t_{kl}^{(\mu)}n^{l}, \quad (8.4)$$

where $\mathbf{n}$ the unit positive normal to a two-dimensional surface element.

According to Postulate A, the neighboring spin interaction produces on the magnetization field a surface couple of the form:

$$\mathbf{t}^{(\mu)} \times \mathbf{M}. \quad (8.5)$$

Hence, the following form of the global electronic spin momentum balance:

$$\frac{d}{dt}\int_{(B-\Gamma)} \frac{\rho\mu}{\Gamma}\, dv = \int_{(B-\Gamma)} (\mathbf{B} + {}_{L}\mathbf{B}) \times \mathbf{M}\, dv + \int_{(\partial B-\Gamma)} \mathbf{t}^{(\mu)} \times \mathbf{M}\, ds \quad (8.6)$$

which yields (8.2).

Note that Tiersten's approach permits us to grant a physical significance to ${}_{L}\mathbf{B}$. It represents the local interaction between the material continuum and the magnetization field. This interaction verifies the third law of Newton. Thus, ${}_{L}\mathbf{B}$ appears in both equations (8.1) and (8.2).

Furthermore, we note that Tiersten uses a form for the magnetic forces which differs from ours. It can be shown, by a redefinition of the stresses and of the potential $\mathfrak{F}$, that the two forms yield equivalent field equations. The statement of balance laws as a starting point, in contrast to the variational treatment given above,[48] allows us to deal with dissipative processes. In that case, the balance of energy equation takes all its importance and we need consider the second law of thermodynamics in order to obtain restrictions on the constitutive equations.

B. Dissipative Processes

According to Tiersten,[15] the global balance of energy equation is

$$\frac{d}{dt}\int_{(B-\Gamma)} (\tfrac{1}{2}\rho\mathbf{v}^{2} + \rho\epsilon)\, dv = \int_{(\partial B-\Gamma)} \{\mathbf{t}\cdot\mathbf{v} + (\mathbf{t}^{(\mu)} \times \mathbf{M})\cdot\boldsymbol{\omega} + \mathbf{q}\cdot\mathbf{n}\} ds + \int_{(B-\Gamma)} \{(\mathbf{B} \times \mathbf{M})\cdot\boldsymbol{\omega} + \mathbf{f}_{(\mathrm{em})}\cdot\mathbf{v} + \rho\mathbf{f}\cdot\mathbf{v} + \rho\mathbf{h}\} dv, \quad (8.7)$$

to which we add the second law of thermodynamics postulated as usual as[49]

$$\frac{d}{dt}\int_{(B-\Gamma)} \rho\eta\, dv - \int_{(B-\Gamma)} \rho\frac{h}{\theta}\, dv - \int_{(\partial B-\Gamma)} \theta^{-1}\mathbf{q}\cdot\mathbf{n}\, ds \geq 0. \quad (8.8)$$

In these equations, we have defined the following quantities:

$\epsilon \equiv$ internal energy per unit mass,

η = entropy per unit mass,

θ = thermodynamical temperature,

h = heat supply per unit mass,

$\mathbf{q}$ = heat flux vector,

$\boldsymbol{\omega}$ = angular velocity of the magnetization,

$f^{l}_{(\mathrm{em})} = t^{lk}_{(\mathrm{em});k} = \rho B^{k;l}\mu_k$.

Upon using the standard procedure, we deduce from (8.7) and (8.8) the local equations

$$\rho\dot{\epsilon} = t^{kl}v_{k;l} + t^{(\mu)qr}\dot{\mu}_{q;r} - \rho_L B^q\dot{\mu}_q + q^k{}_{;k} + \rho h \quad \text{in } (B-\Gamma), \tag{8.9}$$

$$\rho\dot{\eta} - (q^k/\theta)_{;k} - \theta^{-1}\rho h \geq 0 \quad \text{in } (B-\Gamma), \tag{8.10}$$

$$[(\tfrac{1}{2}\rho\mathbf{v}^2 + \rho\epsilon)v^k + t^{lk}v_l + t^{(\mu)lk}\dot{\mu}_l - q^k]n_k = 0 \text{ on } (\Gamma), \tag{8.11}$$

$$[q^k/\theta + \rho\eta v^k]n_k \geq 0 \text{ on } (\Gamma). \tag{8.12}$$

We decompose $v_{l;k}$ into its symmetric and skew symmetric parts:

$$d_{lk} = v_{(l;k)}, \qquad \varpi_{lk} = v_{[l;k]}. \tag{8.13}$$

On account of (8.1) and (8.13), Eq. (8.9) can be written in the following form which proves to be adequate for the discussion of the Clausius–Duhem inequality[50]:

$$\rho\dot{\epsilon} = t^{(kl)}d_{kl} + t^{(\mu)qr}\dot{\mu}_{q;r} - \rho_L B^q\dot{\mu}_q - \rho\mu^{[k}{}_L B^{l]}\varpi_{kl} + q^k{}_{;k} + \rho h. \tag{8.14}$$

Upon substitution of h from (8.14) into (8.10) and introduction of the free energy per unit mass Ψ by the relation

$$\Psi \equiv \epsilon - \eta\theta, \tag{8.15}$$

we obtain the Clausius–Duhem inequality:

$$-(\rho/\theta)(\dot{\Psi} + \dot{\theta}\eta) + \theta^{-1}t^{(kl)}d_{kl} - \frac{\rho}{\theta}{}_L B^k(\dot{\mu}_k - \varpi_{lk}\mu^l) + \theta^{-1}t^{(\mu)qr}\dot{\mu}_{q;r} + \theta^{-2}q^k\theta_{,k} \geq 0 \quad \text{in } (B-\Gamma), \tag{8.16}$$

Note that

$$\Omega_{pq} \equiv \mu_s^{-2}\mu_{[p}\dot{\mu}_{q]}. \tag{8.17}$$

Thus,

$$t^{(\mu)qr}\dot{\mu}_{q;r} \equiv M^{pqr}\Omega_{pq;r}, \qquad \rho_L B^q\mu_q \equiv \rho_L B^{[p}\mu^{q]}\Omega_{pq}, \tag{8.18}$$

and Eqs. (8.9), (8.11), and (8.16) can be equivalently written in the "couple stress" form

$$\rho\dot{\epsilon} = t^{kl}v_{k;l} + M^{pqr}\Omega_{pq;r} + \rho\mu^q{}_L B^p\Omega_{pq} + q^k{}_{;k} + \rho^k \text{ in } (B-\Gamma), \tag{8.19}$$

$$[(\tfrac{1}{2}\rho\mathbf{v}^2 + \rho\epsilon)v^k + t^{lk}v_l + M^{pqk}\Omega_{pq} - q^k]n^k = 0 \text{ on } (\Gamma), \tag{8.20}$$

$$-(\rho/\theta)(\dot{\Psi} + \dot{\theta}\eta) + \theta^{-1}t^{kl}v^k{}_{;l} + \theta^{-1}M^{pqr}\Omega_{pq;r} + (\rho/\theta)\mu^q{}_L B^p\Omega_{pq} + \theta^{-2}q^k\theta_{,k} \geq 0 \text{ in } (B-\Gamma). \tag{8.21}$$

If the medium considered is *nondissipative*, then $q^k = 0, h = 0$, and ϵ can be replaced by the strain energy function $\mathfrak{F}$ in (8.19). Hence (8.19) becomes identical to (7.28) while (8.21) takes the equality form.

It remains, in this approach, to determine the constitutive equations. In this regard, the following remark is in order: In Sec. 4, the first two of Eqs. (4.12) are pure definitions. In this subsection t^{kl} and $t^{(\mu)kl}$ have been granted a physical significance since the postulation of balance laws but their form as function of the constitutive arguments is still unknown. It is the purpose of the next paragraph to arrive at their form in order to compare to preceding results (5.4b), (5.8b) and (5.10).

C. The Positive Entropy Production Requirement

Following Eringen,[51] we require (8.16) or (8.21) to be verified for any independent dynamical processes. Constitutive relations for the nondissipative parts of the constitutive variables will follow. We assume that the constitutive variables t^{kl}, $t^{(\mu)kl}$, and ${}_L B^k$ present recoverable and dissipative parts. Thus we set

$$t^{kl} = {}^R t^{kl} + {}^D t^{kl}, \qquad {}_L B^k = {}^R_L B^k + {}^D_L B^k, \qquad t^{(\mu)kl} = {}^R t^{(\mu)kl} + {}^D t^{(\mu)kl}. \tag{8.22}$$

In the sequel, the recoverable parts are shown to be derivable from a potential, the free energy Ψ. In agreement with the form (4.13), we consider a nonlinear elastic solid of grade one of which the free energy may be written as

$$\Psi = \Psi(\mathbf{F}_x, \boldsymbol{\mu}, \mathbf{F}_{(\mu)}, \theta, \mathbf{X}). \tag{8.23}$$

In taking the time derivative of Ψ, we take account of the constraints (3.5) by introducing four Lagrange multipliers λ and L^K, $K = 1, 2, 3$. Thus

$$\dot{\Psi} = \frac{\partial\Psi}{\partial x^i_{,K}}v^i_{;j}x^j_{,K} + \frac{\partial\Psi}{\partial\mu^i}\dot{\mu}^i + \frac{\partial\Psi}{\partial\mu^i_{,K}}(\dot{\mu}^i)_{;j}x^j_{,K} + \lambda\mu_i\dot{\mu}^i + L^K(\dot{\mu}^i\mu_{i,K} + \mu_i\dot{\mu}^i_{;j}x^j_{,K}), \tag{8.24}$$

in which the following identities have been used:

$$\dot{\overline{x^i_{,K}}} \equiv v^i_{;j}x^j_{,K}, \qquad \dot{\overline{\mu^i_{,K}}} \equiv (\dot{\mu}^i)_{;j}x^j_{,K}. \tag{8.25}$$

Upon carrying the expression (8.24) into (8.16), we obtain

$$-\frac{\rho}{\theta}\left(\frac{\partial\Psi}{\partial x^{(i}_{,K}}x_{j),K} - \rho^{-1}{}^R t_{ij}\right)d^{ij} - \frac{\rho}{\theta}\left(\frac{\partial\Psi}{\partial\theta} + \eta\right)\dot{\theta} - \frac{\rho}{\theta}\left(\frac{\partial\Psi}{\partial x_{[i,K}}x^{j]}_{,K} + {}^R_L B^{[j}\mu^{i]}\right)\varpi_{ij} - \frac{\rho}{\theta}\left(\frac{\partial\Psi}{\partial\mu^i} + \lambda\mu_i + L^K\mu_{i,K} + {}^R_L B_i\right)\dot{\mu}^i - \frac{\rho}{\theta}\left(\frac{\partial\Psi}{\partial\mu^i_{,K}}x^j_{,K} + L^K\mu_i x^j_{,K} - \rho^{-1}t^{(\mu)\ j}_{\ i}\right)\dot{\mu}^i_{;j}$$

$$+ \theta^{-1}\, {}^{D}t^{ji}\, v_{j;i} - (\rho/\theta)\, {}^{D}_{L}B^{k}(\dot{\mu}_k - \varpi_{lk}\mu^l)$$
$$+ (\rho/\theta)\, {}^{D}t^{(\mu)kl}\, \dot{\mu}_{k;l} + \theta^{-2} q^k \theta_{,k} \geq 0. \tag{8.26}$$

If this is to hold for any independent dynamical processes (i.e., variations $d_{ij}, \varpi_{ij}, \dot{\mu}^i, \dot{\mu}^i_{;j}, \dot{\theta}$), the following equations must be satisfied:

$$\begin{aligned} {}^{R}t^{(ij)} &= \rho \frac{\partial \Psi}{\partial x_{(i,K}}\, x^{j)}_{,K}\,, \\ {}^{R}t^{[ij]} &= \rho \frac{\partial \Psi}{\partial x_{[i,K}}\, x^{j]}_{,K} = \rho_L B^{[i}\mu^{j]}, \end{aligned} \tag{8.27}$$

$${}^{R}_{L}B^{i} = -\left(\frac{\partial \Psi}{\partial \mu_i} + \lambda \mu^i + L^K \mu^i_{,K}\right), \tag{8.28}$$

$${}^{R}t^{(\mu)ij} = \rho \frac{\partial \Psi}{\partial \mu_{i,K}}\, x^j_{,K} + \rho L^K \mu^i x^j_{,K}, \tag{8.29}$$

$$\eta = -\frac{\partial \Psi}{\partial \theta}. \tag{8.30}$$

Equivalent to (8.29), we have after (7.23)

$${}^{R}M^{pqr} = \rho \frac{\partial \Psi}{\partial \mu_{[p,K}}\, \mu^{q]} x^r_{,K}, \tag{8.31}$$

with

$$M^{pqr} = {}^{R}M^{pqr} + {}^{D}M^{pqr}.$$

The nonvanishing part that remains in the lhs of (8.26) is a constraint imposed upon the entropy production:

$$\rho\theta\dot{\eta} = {}^{D}t^{kl}\, v_{k;l} - \rho\, {}^{D}_{L}B_k \dot{\mu}^k + {}^{D}t^{(\mu)kl}\dot{\mu}_{k;l} + q^k{}_{;k} + \rho h. \tag{8.32}$$

The Lagrange multipliers are found by noting that only the components of ${}_{L}\mathbf{B}$ and $\mathbf{t}^{(\mu)}$ orthogonal to $\boldsymbol{\mu}$ are of importance. We find

$$\begin{aligned} L^M &= -\frac{\partial \Psi}{\partial \mu^k{}_{,M}}\, \frac{\mu^k}{\mu_s^2}\,, \\ \lambda &= -\frac{\partial \Psi}{\partial \mu^k}\, \frac{\mu^k}{\mu_s^2}\,. \end{aligned} \tag{8.33}$$

We set

$${}^{D}t^{kl} = {}^{D}_{L}B^k = {}^{D}t^{(\mu)kl} = {}^{D}M^{pqr} = 0. \tag{8.34}$$

Then, upon use of (8.28) and (8.33), (8.1) becomes

$$t^{[kl]} = -\rho \frac{\partial \Psi}{\partial \mu_{[k}}\, \mu^{l]} + \rho \frac{\partial \Psi}{\partial \mu^i{}_{,M}}\, \frac{\mu^i}{\mu_s^2}\, \mu^{[k}{}_{,M}\, \mu^{l]}. \tag{8.35}$$

The last term will be shown to vanish when we study the objectivity in Part II. Furthermore, upon carrying (8.28) and (8.29) into (8.2), we obtain explicitly:

$$\begin{aligned} (\rho/\Gamma)\dot{\mu}^i = \rho \epsilon^{ijk} \Big[& B_j - \frac{\partial \Psi}{\partial \mu^j} - L^K \mu_{j,K} - (\lambda - \nabla_R \cdot \mathbf{L})\mu_j \\ & + \rho^{-1} \left(\rho \frac{\partial \Psi}{\partial \mu^j{}_{,K}}\, x^m_{,K}\right)_{;m} \Big] \mu_k \\ & + \epsilon^{ijk} \left(\rho \frac{\partial \Psi}{\partial \mu^j_{,K}}\, x^m_{,K}\right) \mu_{k;m}. \end{aligned} \tag{8.36}$$

Given the degree of arbitrariness in the definition of the constitutive variables when one constructs a model, we can adopt the new definitions

$${}_{L}B_j \equiv \frac{\partial \Psi}{\partial \mu^j} - L^K \mu_{j,K}, \tag{8.37a}$$

$$t^{(\mu)}{}_j{}^m \equiv \rho \frac{\partial \Psi}{\partial \mu^j_{,K}}\, x^m_{,K}. \tag{8.37b}$$

Note that the combination $(\lambda - \nabla_R \cdot \mathbf{L})$ already encountered in Sec. 6 is of no importance and the corresponding term can be dropped in Eq. (8.36).

The set of results is in full agreement with those of the variational treatment except for the constitutive equation (8.37a). The complete similarity is arrived at after the study of the objectivity requirement[51] in Part II.

9. THE EINSTEIN–DE HAAS EFFECT FOR A BODY AS A WHOLE

It is well known that a freely suspended body, on being magnetized, begins to rotate. This is the Einstein–de Haas effect.[52-54] The inverse effect exists: A uniform rotation causes a magnetization linearly dependent on the angular velocity (this without any external magnetic field). It is called the Barnett effect.[54-56] Let us show that these effects are contained in the foregoing theory.

We add Eq. (7.26) to Eq. (7.19) to gather all contributions, that is, mechanical and magnetic, in the *total* local law of conservation of moment of momentum. After integration over the whole body (B) and transformation of some terms, we use the Green-Gauss theorem (discarding discontinuity surfaces which add nothing to the present result) for the divergence terms. On account of the fact that the conservation of mass (7.30a) can be written in local form

$$\frac{d}{dt}(\rho dv) = 0,$$

we obtain the *global* conservation of moment of momentum in the deformed configuration

$$\begin{aligned} \frac{d}{dt} \int_{(B)} \rho x_{[k} \dot{x}_{l]} dv = & -\frac{d}{dt} \int_{(B)} \frac{\rho}{2\Gamma}\, S_{kl}\, dv \\ & + \mathfrak{L}_{kl\,(\mathrm{mech})} + \mathfrak{L}_{kl\,(\mathrm{magn})}, \end{aligned} \tag{9.1}$$

where we have used (7.16) and set

$$\mathfrak{L}_{kl\,(\mathrm{mech})} \equiv \int_{(B)} \rho x_{[k} f_{l]}\, dv + \int_{(\partial B)} x_{[k} T_{l]} ds, \tag{9.2}$$

$$\begin{aligned} \mathfrak{L}_{kl\,(\mathrm{magn})} \equiv & \int_{(B)} \rho \left\{\mu_{[k} B_{l]} + \mu^m x_{[k} B_{m;l]}\right\} dv \\ & + \int_{(\partial B)} \left\{M_{kl}{}^{m}{}_{n_m} - (\overline{\rho\boldsymbol{\mu}} \cdot [\mathbf{B}]) x_{[k} n_{l]}\right\} ds. \end{aligned} \tag{9.3}$$

Equations (9.2) and (9.3) represent the total mechanical couple and the total couple due to the presence of magnetization in the body, respectively. In absence of these couples and, if the initial angular momentum of the body (B) is nil, Eq. (9.1) yields, after integration with respect to time,

$$\mathbf{G} = -\Gamma^{-1} \int_{(B)} \rho \boldsymbol{\mu}\, dv, \tag{9.4}$$

where the *angular momentum* of the body (B) has been defined as

$$G^m = \int_{(B)} \rho \epsilon^{mkl} x_k \dot{x}_l\, dv. \tag{9.5}$$

Eq. (9.4) represents the conventional Einstein–de Haas effect for the whole body (B), while Eq. (9.1) is

a more general equation, given in differential form, which involves effects of other sources.

ACKNOWLEDGMENT

The authors are indebted to the referee for a careful critique of this work.

APPENDIX A: THERMODYNAMICS OF ROTATING SYSTEMS

Based on the Thermodynamics of rotating systems, one can give another argument leading to formula (2.14). Let ϵ and ϵ' respectively be the density of energy in a fixed coordinate system and in a coordinate system moving with an element of matter having the angular velocity ϖ about a fixed axis. Classical mechanics[57] gives the relations

$$\epsilon' = \epsilon - \varpi \cdot \mathbf{l}, \tag{A1}$$

with

$$\mathbf{l} = -\frac{\partial \epsilon'}{\partial \varpi}$$

where $\mathbf{l}$ is the angular momentum. In a rotating system, we have

$$\delta\epsilon' = \theta\delta\eta - \mathbf{l}\cdot\delta\varpi\,, \tag{A2}$$

where the scalar θ is the thermodynamic temperature and η is the entropy density. If we select

$$\begin{aligned} &\mathbf{l} \equiv -\boldsymbol{\mu}/\Gamma\,, \\ &\varpi \equiv (\boldsymbol{\mu}\times\dot{\boldsymbol{\mu}}/\mu_s^2) + (\boldsymbol{\mu}\cdot\varpi/\mu_s^2)\boldsymbol{\mu}, \end{aligned} \tag{A3}$$

it follows that

$$\delta\epsilon = \theta\delta\eta - (\boldsymbol{\mu}\times\dot{\boldsymbol{\mu}}/\Gamma\mu_s^2)\cdot\delta\boldsymbol{\mu}. \tag{A4}$$

Note that the last term differs from zero since $\delta\mu$ is a virtual variation. Eq. (A4) can be written as

$$\delta\epsilon = \theta\delta\eta - (\delta\boldsymbol{\mu}\times\boldsymbol{\mu}/\Gamma\mu_s^2)\cdot\boldsymbol{\mu}. \tag{A5}$$

We call $\delta\theta$ the anholonomic quantity $\boldsymbol{\mu}\times\delta\boldsymbol{\mu}/\mu_s^2$, hence the relation (A5) reads

$$\delta\epsilon = \theta\delta\eta + \Gamma^{-1}\,\delta\boldsymbol{\theta}\cdot\boldsymbol{\mu}\,. \tag{A6}$$

($\delta\boldsymbol{\theta}$ is a vector not to be confused with the temperature θ.) Integrating over the undeformed volume and over time, we get

$$\begin{aligned} \delta W &= \delta \int_t dt \int_{(B_R)} \rho_R\,\epsilon\, dv_R \\ &= \int_t dt \int_{(B_R)} \rho_R\,\theta\delta\eta\, dv_R + \int_t dt \int_{(B_R)} \rho_R\,\Gamma^{-1}\delta\boldsymbol{\theta}\cdot\boldsymbol{\mu}\, dv_R. \end{aligned} \tag{A7}$$

One must take account of the first term of the latter relation if one introduces η in $\mathfrak{F}$.

APPENDIX B: THE PONDEROMOTIVE FORCE AND COUPLE IN A POLARIZED AND MAGNETIZED MEDIUM

(a) It was pointed out in Sec. 4 that we are at liberty to choose a formula for the volume ponderomotive force. However, once this is chosen, the corresponding surface term on the boundary of the body and the jump term across a discontinuity surface are fixed. In order to determine these terms, we examine, in this appendix, the *total ponderomotive electromagnetic* force and couple acting upon a body (B) when a steady discontinuity surface (Γ) exists in (B), the currents not being neglected. We give a derivation for magnetostatics; that for electrostatics follows the same line.

To start with, consider the magnetic stress tensor $t_{(em)}^{kp}$ given by the relation

$$t_{(em)}^{kp} = H^k B^p - (\tfrac{1}{2}\,\mathbf{B}^2 - \mathbf{B}\cdot\mathbf{M})\,g^{kp} \tag{B1}$$

and the Maxwell's equations for magnetostatics in matter

$$\nabla\times\mathbf{H} = c^{-1}\mathbf{J}_c, \quad \nabla\cdot\mathbf{B} = 0, \quad \nabla\cdot\mathbf{J} = 0, \quad \text{in } (B-\Gamma), \tag{B2}$$

$$\mathbf{n}\times[\mathbf{H}] = c^{-1}\mathbf{K}, \quad [\mathbf{B}]\cdot\mathbf{n} = 0, \quad \mathbf{n}\cdot[\mathbf{J}] = 0, \quad \text{on } (\partial B - \Gamma) \text{ and on } (\Gamma), \tag{B3}$$

In these equations, $\mathbf{B}, \mathbf{H}, \mathbf{M}, \mathbf{J}, \mathbf{J}_c$ and $\mathbf{K}$ are the magnetic intensity, the magnetic field, the magnetization per unit volume, the total volume current, the conduction current, and the surface current, respectively. $\mathbf{n}$ is the positive oriented normal to $(\partial B - \Gamma)$ or to (Γ) and the familiar symbolism $[\cdots]$ represents the jump, i.e., $[A] = A_{(+)} - A_{(-)}$. g^{kp} is the metric tensor.

Upon using (B2) and the relations

$$\mathbf{H} = \mathbf{B} - \mathbf{M}, \quad \mathbf{M} = \rho\boldsymbol{\mu}, \tag{B4}$$

where ρ is the density of matter and $\boldsymbol{\mu}$ is the magnetization per unit mass, we obtain the ponderomotive force $\mathbf{f}_{(em)}$ per unit volume as

$$f_{(em)}^k = t_{(em);p}^{kp} = c^{-1}(\mathbf{J}\times\mathbf{B})^k + \rho B^{p;k}\mu_p \tag{B5}$$

or, equivalently,

$$\mathbf{f}_{(em)} = c^{-1}\mathbf{J}\times\mathbf{H} + (\mathbf{M}\cdot\nabla)\mathbf{B} + \tfrac{1}{2}\nabla\mathbf{M}^2. \tag{B6}$$

The corresponding ponderomotive couple per unit volume is:

$$\mathbf{c}_{(em)} = \mathbf{p}\times\mathbf{f}_P + \mathbf{M}\times\mathbf{B}, \tag{B7}$$

where $\mathbf{p}$ is the position vector and $\mathbf{f}_P$ is given by

$$\mathbf{f}_P = \mathbf{f}_{(em)} - \mathbf{f}_L, \tag{B8}$$

with

$$\mathbf{f}_L \equiv c^{-1}\mathbf{J}\times\mathbf{B}.$$

(b) In order to compute the total ponderomotive force, we consider a control surface (S) enclosing the material body (B) and very close to the boundary (∂B) of (B). Thus, we consider (∂B) to be a discontinuity surface in the total volume enclosed within (S). We write

$$F_{(em)}^i = \int_{(S)} t_{(em)}^{ik}\, n_k\, ds. \tag{B9}$$

Upon using the generalized Green–Gauss theorem[58] when a discontinuity surface (Γ) exists in (B), we have

$$F_{(em)}^i = \int_{S-(\Gamma+\partial B)} t_{(em);k}^{ik}\, dv + \int_{(\Gamma)} [t_{(em)}^{ik}]\, n_k\, ds_\Gamma + \int_{(\partial B)} [t_{(em)}^{ik}]\, n_k\, ds. \tag{B10}$$

From (B1) it follows that the first term of (B10) is nothing but the volume integral of (B5). We analyze the second or third term in detail. The following definitions for the jump and the mean value on a surface are used:

$$[A] = A_{(+)} - A_{(-)}, \qquad \overline{A} = \tfrac{1}{2}(A_{(+)} + A_{(-)}). \tag{B11}$$

The following identities hold:

$$[AB] = \overline{A}[B] + [A]\overline{B}, \tag{B12}$$

$$[A^2] = 2\overline{A}[A]. \tag{B13}$$

For a quantity B such that $[\mathbf{B}]\cdot\mathbf{n} = 0$, we have

$$\mathbf{B}\cdot\mathbf{n} \equiv \overline{\mathbf{B}}\cdot\mathbf{n}, \tag{B14}$$

From (B1), we see that

$$d^{i}_{(\mathrm{magn})} = [t^{ik}_{(em)}]\, n_k = [H^i B^k - \tfrac{1}{2}\mathbf{B}^2 g^{ik} + \mathbf{M}\cdot\mathbf{B} g^{ik}]\, n_k. \tag{B15}$$

We set

$$\mathbf{B} = \mathbf{B}_{(n)} + \mathbf{B}_{(t)}, \qquad \mathbf{B}_{(n)} = (\mathbf{B}\cdot\mathbf{n})\mathbf{n}, \qquad \mathbf{B}_{(t)} = \mathbf{n}\times(\mathbf{B}\times\mathbf{n}). \tag{B16}$$

Similar formulas are written for **M**. From the first of Eqs. (B2), we have

$$[\mathbf{B}_{(n)}] = 0, \qquad [\mathbf{B}] = [\mathbf{B}_{(t)}]. \tag{B17}$$

Upon use of (B17), (B14) and (B13), Eq. (B15) yields

$$\mathbf{d}_{(\mathrm{magn})} = (\overline{\mathbf{B}}\cdot\mathbf{n})[\mathbf{H}] - (\overline{\mathbf{B}}\cdot[\mathbf{B}])\mathbf{n} + [\mathbf{M}\cdot\mathbf{B}]\mathbf{n}. \tag{B18}$$

Transforming the second term by use of (B4) and using (B12) for the last term, we obtain

$$\mathbf{d}_{(\mathrm{magn})} = (\overline{\mathbf{B}}\cdot\mathbf{n})[\mathbf{H}] - (\overline{\mathbf{B}}\cdot[\mathbf{H}])\mathbf{n} + (\overline{\mathbf{M}}\cdot[\mathbf{B}])\mathbf{n}$$

Note that the two first terms of the last equation combine to give $(\mathbf{n}\times[\mathbf{H}])\times\overline{\mathbf{B}}$ after the vectorial identity

$$(\mathbf{A}\times\mathbf{B})\times\mathbf{C} \equiv (\mathbf{A}\cdot\mathbf{C})\mathbf{B} - (\mathbf{B}\cdot\mathbf{C})\mathbf{A}.$$

Thus, with the first of Eqs. (B3), we finally have

$$\mathbf{d}_{(\mathrm{magn})} = c^{-1}\mathbf{K}\times\overline{\mathbf{B}} + (\overline{\mathbf{M}}\cdot[\mathbf{B}])\mathbf{n}, \tag{B19}$$

Upon carrying this result into Eq. (B10), we get the total ponderomotive force acting upon a magnetized body (B) for the case of statics when a discontinuity surface (Γ) exists in (B). It reads

$$\mathbf{F}_{(em)} = \int_{(B-\Gamma)} \{c^{-1}\mathbf{J}\times\mathbf{B} + \rho(\nabla\mathbf{B})\cdot\boldsymbol{\mu}\}\,dv + \int_{(\Gamma)} \{c^{-1}\mathbf{K}\times\overline{\mathbf{B}} + (\overline{\rho\boldsymbol{\mu}}\cdot[\mathbf{B}])\mathbf{n}\}\,ds_\Gamma + \int_{(\partial B-\Gamma)} \{c^{-1}\mathbf{K}\times\overline{\mathbf{B}} + (\overline{\rho\boldsymbol{\mu}}\cdot[\mathbf{B}])\mathbf{n}\}\,ds. \tag{B20}$$

The total ponderomotive torque acting upon the body (B) under the same condition is simply given by the integral of (B7). That is,

$$\mathbf{C}_{(em)} = \int_{(B)} (\mathbf{p}\times\mathbf{f}_P + \rho\boldsymbol{\mu}\times\mathbf{B})\,dv. \tag{B21}$$

For the case of electromagnetostatics, it is shown that, starting from the electromagnetic stress tensor

$$t^{ik}_{(em)} = E^i D^k + H^i B^k - \{\tfrac{1}{2}(\mathbf{B}^2 + \mathbf{E}^2) - \mathbf{M}\cdot\mathbf{B}\}g^{ik}$$

and using the Maxwell's equations

$$\nabla\times\mathbf{E} = 0, \qquad \nabla\cdot\mathbf{D} = q \text{ in } (B-\Gamma),$$
$$\mathbf{n}\times[\mathbf{E}] = 0, \qquad \mathbf{n}\cdot[\mathbf{D}] = \sigma \text{ on } (\partial B - \Gamma)\ (\Gamma),$$

where **E** is the electric field, **D** is the electric displacement, q is the volume charge density and σ is the surface charge density, we would be led to adding terms of the form $\int\{q\mathbf{E} + (\nabla\mathbf{E})\cdot\mathbf{P}\}\,dv$ and $\int\{\sigma\overline{E} + (\overline{\mathbf{P}}\cdot[\mathbf{E}])\mathbf{n}\}\,ds$ to the first and to the second and third term of Eq. (B20), respectively, and a term $\int\mathbf{P}\times\mathbf{E}\,dv$ to Eq. (B21) with

$$\mathbf{P} = \mathbf{D} - \mathbf{E}.$$

---

* This work was supported by the National Science Foundation

1 C. Kittel, Rev. Mod. Phys. **21**, 541 (1949).
2 J. H. Van Vleck, Ann. Inst. Henri Poincaré **10**, 150 (1947).
3 J. C. Anderson, *Magnetism and Magnetic Materials* (Chapman and Hall, London, 1968), Chaps. 5–7.
4 L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, Oxford, 1960), Chap. 5.
5 G. T. Rado and H. Suhl, Eds., *Magnetism* (Academic, New York, 1963), Vol. III, pp. 395, 415.
6 W. Döring, in *Handbuch der Physik*, Vol. 18/2, edited by S. Flügge (Springer-Verlag, Berlin, 1966), p. 341.
7 W. F. Brown, *Micromagnetics* (Wiley, New York, 1963) and references cited therein.
8 N. Minnaja, Phys. Rev. B 1, 1151 (1970).
9 A. Aharoni, to be published in *CRC Critical Reviews in Solid State Sciences*.
10 L. D. Landau and E. M. Lifshitz, Phys. Z. Sowjet **8**, 153 (1935).
11 W. F. Brown, *Magnetoelastic Interactions* (Springer-Verlag, Berlin, 1966).
12 W. Heisenberg, Z. Physik **49**, 619 (1928).
13 F. Bloch, Z. Physik **74**, 295 (1932).
14 Cf. Ref. 11, pp. 90–92.
15 H. Tiersten, J. Math. Phys **5**, 1298 (1964); **6**, 779 (1965).
16 S. I. Amari, in *RAAG Memoirs*, edited by K. Kondo (Tokyo, 1962), Vol. 3, p. 257.
17 J. B. Alblas, in *Mechanics of Generalized Continua*, IUTAM Symposium 1967, edited by E. Kröner (Springer-Verlag, Berlin, 1968), p. 350.
18 R. A. Toupin, Arch. Ratl. Mech. Anal. **17**, 85 (1964).
19 R. A. Toupin J. Ratl. Mech. Anal. **5**, 849 (1956).
20 R. A. Toupin, Intern. J. Eng. Sci. **1**, 101 (1963).
21 A. C. Eringen, *Nonlinear Theory of Continuous Media* (McGraw-Hill, New York, 1962), Chap. II.
22 R. C. Dixon and A. C. Eringen, Intern. J. Eng. Sci. **3**, 359 (1965).
23 J. Grindlay, *An Introduction to the Phenomenological Theory of Ferroelectricity* (Pergamon, Oxford, 1970).
24 R. D. Mindlin, Intern. J. Solids Structures **4**, 637 (1968).
25 E. S. Suhubi, Intern. J. Eng. Sci. **7**, 993 (1969).
26 C. Truesdell and R. A. Toupin, in *Handbuch der Physik*, Vol. 3/1, edited by S. Flügge (Springer-Verlag, Berlin, 1960), p. 226.
27 Cf. Chap. I of Ref. 21.
28 J. Jackson, *Classical Electrodynamics* (Wiley, New York, 1967), p. 148.
29 J. Anderson, *Principles of Relativity Physics* (Academic, New York, 1967), p. 248.
30 In doing so, we disregard any anisotropy in the gyromagnetic effect (for anisotropy, see Ref. 4, p. 145). For anisotropy, the relation (2. 2) would read $\mu_i = \Gamma\lambda_{ik}s^k$, where λ_{ik} is a constant second-order tensor, in general nonsymmetric. It would be trivial to use this relation throughout.
31 H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1950), p. 219.
32 While, for a real body, two principal moments of inertia cannot vanish simultaneously; for the concept of spin, this analogy is a familiar one from quantum mechanics.

[33] C. Lanczos, *The Variational Principles of Mechanics* (Toronto U. P., Toronto, 1966), p. 146.
[34] A. C. Eringen, *Mechanics of Continua* (Wiley, New York, 1967), p. 119.
[35] See Ref. 18.
[36] G. A. Maugin, thesis (Princeton University, 1971), Chap. I.
[37] P. Penfield and H. Haus, *Electrodynamics of Moving Media* (MIT Press, Cambridge, Mass., 1967), Chap. 7 and pp. 236–37.
[38] L. Sedov, Prikl. Mat. Mekh. **29**, 4 (1965).
[39] R. A. Grot and A. C. Eringen, Intern. J. Eng. Sci. **4**, 638 (1966).
[40] These names are given by analogy with the couple-stress theory, cf. Ref. 18.
[41] Cf. Ref. 26, p. 247.
[42] Cf. Ref. 34, Chap. I.
[43] Cf. Ref. 3, p. 180.
[44] T. L. Gilbert and J. M. Kelley, *Proceedings of the Conference on Magnetism and Magnetic Materials* (A.I.E.E. Publ., New York, 1955).
[45] E. and F. Cosserat, *Théorie des corps déformables* (Hermann, Paris, 1909).
[46] G. A. Maugin, Compt. Rend. **271A**, 807 (1970).
[47] A. C. Eringen, "Micropolar Elasticity", in *Fracture*, Vol. 2, edited by H. Liebowitz (Academic, New York, 1968), p. 621
[48] It is, of course, possible to introduce very special forms of dissipation in a variational formulation, for instance, through a Rayleigh function of dissipation as it is done in Remark (iii) of Sec. 6. However, to the knowledge of the authors, there exist no means to introduce dissipation in terms as general as the potential giving the recoverable parts of the nonlinear constitutive equations.
[49] Cf. Ref. 34, p. 128
[50] Ref. 34, p. 129.
[51] Ref. 34, Chap. 5.
[52] A. Einstein and de Haas, Verhandl.-Physik Ges. **17**, 152 (1915).
[53] E. Beck, Ann. Physik **60**, 109 (1919).
[54] S. J. Bennet, Phys. Rev. **6**, 239 (1915).
[55] S. J. Bennet, Proc. Am. Acad. **66**, 273 (1931).
[56] R. Becker, *Theorie der Electrizität* (Teubner Verlag, Stuttgart, 1933), Vol. 2, p. 100.
[57] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Addison-Wesley, Reading, Mass., 1969), 2nd ed., Sec. 26.
[58] Cf. Ref. 34, p. 471.

Variational Methods in the Wave Operator Formalism: Applications in Variation-Perturbation Theory and the Theory of Energy Bounds*

Erkki J. Brändas and David A. Micha†
Departments of Chemistry and Physics, Nuclear Sciences Building, University of Florida, Gainesville, Florida 32601

(Received 1 July 1971)

Variational principles of the Lippmann–Schwinger type are used to develop approximations to eigenenergies and eigenfunctions within the wave-operator formalism. The present approach starts with exactly soluble "primary" eigenvalue equations to give explicit results valid beyond the limits of conventional perturbation theory. The variational functionals are expressed in terms of resolvents of the primary Hamiltonian, and bounds to the functionals are constructed also for cases where the resolvents are only partly known. Approximations to eigenenergies and eigenfunctions are obtained in terms of quantities in the Brillouin–Wigner perturbation theory. Connections with methods for upper and lower energy bounds are discussed, and the convergence properties of the nonlinear Padé summation is recovered in this way. Closed formulas within the double perturbation theory framework are presented as a logical extension.

1. INTRODUCTION

The wave operator formalism of Löwdin is a convenient tool for discussing several aspects of the eigenvalue problem in quantum mechanics.[1] It has been used to develop compact expressions for the Brillouin–Wigner and the Rayleigh-Schrödinger perturbation theory and more recently to obtain Fredholm type expansions for bound states.[2] It has also proved helpful in the derivation of upper and lower bounds to eigenvalues and second-order properties.[3,4] Parallel developments have occurred in scattering theory, giving expansions and bounds for phase shifts.[5,6]

In this contribution we want to emphasize the use of variational principles of the Lippmann–Schwinger type[7] in the wave operator formalism. Although the following procedure looks very similar to that for scattering states, there are several new aspects worth studying. Since this principle does not require normalizable trial functions, it is particularly useful in connection with intermediate normalization, and may be applied (unlike the Ritz variational principle) to both bound and quasibound states. Most of the following developments refer to bound states, while the connection between these and quasibound states has been considered elsewhere.[8]

In Sec. 2 we briefly review the wave operator formalism and variational principles of the Lippmann–Schwinger type. These principles are expressed in terms of reduced resolvents after performing splitting of the Hamiltonian into a primary part H_1 and H_2.

Upper and lower bounds for the variation functional, in the case when the primary reduced resolvent is only partly known, are given in Sec. 3. Section 4 finally contains results for specific trial functions. It deals with expansions in given basis sets and in particular with the energies and states of the Brillouin–Wigner perturbation theory. This leads in a natural way to the concepts of inner projections and Padé approximants for perturbation sums. Connections with the theory of upper and lower bounds are pointed out where appropriate.

Finally a formulation within the double perturbation theory framework follows as a logical extension of the previous analysis.

The nomenclature is for the most part that of previous work on the wave operator formalism.[3(b)]

2. THE WAVE-OPERATOR FORMALISM AND LIPPMANN–SCHWINGER-TYPE VARIATION PRINCIPLES

In the wave-operator formalism the solutions to the Schrödinger equation

$$(H - E)\psi = 0 \tag{2.1}$$

are obtained from

$$(e - H)\psi(e) = [e - f(e)]\varphi \tag{2.2}$$

by finding the roots of

$$e = f(e). \tag{2.3}$$

[33] C. Lanczos, *The Variational Principles of Mechanics* (Toronto U. P., Toronto, 1966), p. 146.
[34] A. C. Eringen, *Mechanics of Continua* (Wiley, New York, 1967), p. 119.
[35] See Ref. 18.
[36] G. A. Maugin, thesis (Princeton University, 1971), Chap. I.
[37] P. Penfield and H. Haus, *Electrodynamics of Moving Media* (MIT Press, Cambridge, Mass., 1967), Chap. 7 and pp. 236–37.
[38] L. Sedov, Prikl. Mat. Mekh. **29**, 4 (1965).
[39] R. A. Grot and A. C. Eringen, Intern. J. Eng. Sci. **4**, 638 (1966).
[40] These names are given by analogy with the couple-stress theory, cf. Ref. 18.
[41] Cf. Ref. 26, p. 247.
[42] Cf. Ref. 34, Chap. I.
[43] Cf. Ref. 3, p. 180.
[44] T. L. Gilbert and J. M. Kelley, *Proceedings of the Conference on Magnetism and Magnetic Materials* (A.I.E.E. Publ., New York, 1955).
[45] E. and F. Cosserat, *Théorie des corps déformables* (Hermann, Paris, 1909).
[46] G. A. Maugin, Compt. Rend. **271A**, 807 (1970).
[47] A. C. Eringen, "Micropolar Elasticity", in *Fracture*, Vol. 2, edited by H. Liebowitz (Academic, New York, 1968), p. 621
[48] It is, of course, possible to introduce very special forms of dissipation in a variational formulation, for instance, through a Rayleigh function of dissipation as it is done in Remark (iii) of Sec. 6. However, to the knowledge of the authors, there exist no means to introduce dissipation in terms as general as the potential giving the recoverable parts of the nonlinear constitutive equations.
[49] Cf. Ref. 34, p. 128
[50] Ref. 34, p. 129.
[51] Ref. 34, Chap. 5.
[52] A. Einstein and de Haas, Verhandl.-Physik Ges. **17**, 152 (1915).
[53] E. Beck, Ann. Physik **60**, 109 (1919).
[54] S. J. Bennet, Phys. Rev. **6**, 239 (1915).
[55] S. J. Bennet, Proc. Am. Acad. **66**, 273 (1931).
[56] R. Becker, *Theorie der Electrizität* (Teubner Verlag, Stuttgart, 1933), Vol. 2, p. 100.
[57] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Addison-Wesley, Reading, Mass., 1969), 2nd ed., Sec. 26.
[58] Cf. Ref. 34, p. 471.

Variational Methods in the Wave Operator Formalism: Applications in Variation-Perturbation Theory and the Theory of Energy Bounds*

Erkki J. Brändas and David A. Micha†
Departments of Chemistry and Physics, Nuclear Sciences Building, University of Florida, Gainesville, Florida 32601

(Received 1 July 1971)

Variational principles of the Lippmann–Schwinger type are used to develop approximations to eigenenergies and eigenfunctions within the wave-operator formalism. The present approach starts with exactly soluble "primary" eigenvalue equations to give explicit results valid beyond the limits of conventional perturbation theory. The variational functionals are expressed in terms of resolvents of the primary Hamiltonian, and bounds to the functionals are constructed also for cases where the resolvents are only partly known. Approximations to eigenenergies and eigenfunctions are obtained in terms of quantities in the Brillouin–Wigner perturbation theory. Connections with methods for upper and lower energy bounds are discussed, and the convergence properties of the nonlinear Padé summation is recovered in this way. Closed formulas within the double perturbation theory framework are presented as a logical extension.

1. INTRODUCTION

The wave operator formalism of Löwdin is a convenient tool for discussing several aspects of the eigenvalue problem in quantum mechanics.[1] It has been used to develop compact expressions for the Brillouin–Wigner and the Rayleigh-Schrödinger perturbation theory and more recently to obtain Fredholm type expansions for bound states.[2] It has also proved helpful in the derivation of upper and lower bounds to eigenvalues and second-order properties.[3,4] Parallel developments have occurred in scattering theory, giving expansions and bounds for phase shifts.[5,6]

In this contribution we want to emphasize the use of variational principles of the Lippmann–Schwinger type[7] in the wave operator formalism. Although the following procedure looks very similar to that for scattering states, there are several new aspects worth studying. Since this principle does not require normalizable trial functions, it is particularly useful in connection with intermediate normalization, and may be applied (unlike the Ritz variational principle) to both bound and quasibound states. Most of the following developments refer to bound states, while the connection between these and quasibound states has been considered elsewhere.[8]

In Sec. 2 we briefly review the wave operator formalism and variational principles of the Lippmann–Schwinger type. These principles are expressed in terms of reduced resolvents after performing splitting of the Hamiltonian into a primary part H_1 and H_2.

Upper and lower bounds for the variation functional, in the case when the primary reduced resolvent is only partly known, are given in Sec. 3. Section 4 finally contains results for specific trial functions. It deals with expansions in given basis sets and in particular with the energies and states of the Brillouin–Wigner perturbation theory. This leads in a natural way to the concepts of inner projections and Padé approximants for perturbation sums. Connections with the theory of upper and lower bounds are pointed out where appropriate.

Finally a formulation within the double perturbation theory framework follows as a logical extension of the previous analysis.

The nomenclature is for the most part that of previous work on the wave operator formalism.[3(b)]

2. THE WAVE-OPERATOR FORMALISM AND LIPPMANN–SCHWINGER-TYPE VARIATION PRINCIPLES

In the wave-operator formalism the solutions to the Schrödinger equation

$$(H - E)\psi = 0 \tag{2.1}$$

are obtained from

$$(e - H)\psi(e) = [e - f(e)]\varphi \tag{2.2}$$

by finding the roots of

$$e = f(e). \tag{2.3}$$

In Eqs. (2.2) and (2.3), φ is an arbitrary normalized reference function, e is a complex energy, $f(e)$ is given by

$$\langle\varphi|H|\psi(e)\rangle = \langle\varphi|HW(e)|\varphi\rangle, \tag{2.4}$$

where the wave operator is defined as

$$W(e) = 1 + R(e)H, \tag{2.5}$$

$$R(e) = (e - PHP)^{-1}P, \tag{2.6}$$

and $R(e)$ is the reduced resolvent at energy e. P is the projection operator for the orthogonal complement to φ, which means that

$$\psi(e) = W(e)\varphi \tag{2.7}$$

satisfies intermediate normalization, i.e.,

$$\langle\varphi|\psi(e)\rangle = 1. \tag{2.8}$$

Consider a general splitting of the Hamiltonian H into

$$H = H_1 + H_2, \tag{2.9}$$

where H_1 will be called the *primary Hamiltonian*. Introducing the auxiliary quantities

$$R_1(e) = (e - PH_1P)^{-1}P \tag{2.10}$$

$$W_1(e) = 1 + R_1(e)H_1, \tag{2.11}$$

and

$$\psi_1(e) = W_1(e)\varphi \tag{2.12}$$

we obtain

$$\psi(e) = (1 + RH_2)\psi_1(e) = \psi_1(e) + R_1(e)H_2\psi(e), \tag{2.13}$$

$$f(e) = f_1(e) + g(e), \tag{2.14}$$

$$f_1(e) = \langle\varphi|H_1W_1(e)|\varphi\rangle, \tag{2.15}$$

$$g(e) = \langle\psi_1(e^*)|[1 + H_2R(e)]H_2|\psi_1(e)\rangle. \tag{2.16}$$

In this way the formalism allows us to independently choose a convenient reference function φ as well as the primary Hamiltonian H_1.[8]

In order to calculate $g(e)$ we construct a functional $\mathcal{J}[\psi', e]$, which is stationary with respect to variations of the trial function $\psi'(e)$ around the function $\psi(e)$.[8] This is easily done by defining the functionals

$$A[\psi'] = \langle\psi_1(e^*)|H_2|\psi'(e)\rangle, \tag{2.17}$$

$$B[\psi'] = \langle\psi'(e^*)|H_2|\psi_1(e)\rangle, \tag{2.18}$$

and

$$C[\psi', e] = \langle\psi'(e^*)|H_2 - H_2R_1(e)H_2|\psi'(e)\rangle \tag{2.19}$$

It then follows that

$$\mathcal{J}[\psi', e] = A[\psi'] + B[\psi'] - C[\psi', e] \tag{2.20}$$

has the desired properties, i.e., if $\psi'(e) = \psi(e) + \delta\psi(e)$, then

$$\mathcal{J}[\psi', e] = g(e) - \langle\delta\psi(e^*)|H_2 - H_2R_1(e)H_2|\delta\psi(e)\rangle. \tag{2.21}$$

Thus $\psi'(e)$ is determined by varying $\mathcal{J}$, and using the stationary value of $\mathcal{J}$ around ψ' in the space of trial functions S' it follows that

$$e = f'(e) + \mathcal{O}(\delta^2), \tag{2.22}$$

$$f'(e) = f_1(e) + g'(e), \tag{2.23}$$

$$g'(e) = \mathcal{J}[\psi', e], \tag{2.24}$$

where

$$\mathcal{O}(\delta^2) = \langle\delta\psi(e^*)|H_2 - H_2R_1H_2|\delta\psi(e)\rangle, \tag{2.25}$$

i.e., a quantity of second order in $\delta\psi$.

Equation (2.23) has an added significance for real energies in the discrete spectrum. For opposite signs of H_2 and R_1, $\mathcal{O}(\delta^2)$ is either positive or negative; hence a solution to

$$e = f'(e) = E' \tag{2.26}$$

gives, respectively, a lower bound to the energy E_S if $\mathcal{O}(\delta^2) > 0$ and an upper bound if $\mathcal{O}(\delta^2) < 0$.

3. EXPRESSIONS FOR THE FUNCTIONAL

In order to use the variational principle in practical applications, we must deal with two features in $\mathcal{J}$: calculation of the reduced resolvent $R_1(e)$ and the dependence of $\mathcal{J}$ on e.

In Sec. 2 we developed equations where H_1 could be conveniently chosen, so as to simplify the calculation of $R_1(e)$. Here we shall obtain expressions for $\mathcal{J}$ that apply even when R_1 is not completely known. To prove the following relations, we make use of the eigenvalues and eigenoperators of $\overline{H}_1 = PH_1P$ that satisfy

$$\overline{H}_1\overline{O}_k^{(1)} = \overline{E}_k^{(1)}\overline{O}_k^{(1)}, \qquad k = 1, 2, \cdots, \tag{3.1}$$

and

$$P = \sum_{k\geq 1}\overline{O}_k^{(1)}. \tag{3.2}$$

The expression for R_1 is immediately obtained from

$$R_1(e) = \sum_{k\geq 1} O_k^{(1)}/(e - E_k^{(1)}), \tag{3.3}$$

and $\mathcal{J}$ follows replacing Eq. (3.3) in Eq. (2.20). Further results may be obtained in terms of upper and lower bounds to $\mathcal{J}$. Assuming that only the lowest eigenvalues and eigenoperators of $\overline{H}_1$ are known, it is possible to give upper and lower bounds for $\mathcal{J}$, following Ref. 5.

Given a positive operator $A > 0$, one can write

$$A \geq |\mathbf{h}\rangle\langle\mathbf{h}|A^{-1}|\mathbf{h}\rangle^{-1}\langle\mathbf{h}|, \tag{3.4}$$

where $|\mathbf{h}\rangle$ is a column vector whose elements are the functions $h_1, h_2, \ldots, h_M$ belonging to the space of A, and

$$(\langle\mathbf{h}|A^{-1}|\mathbf{h}\rangle)_{ij} = \langle h_i|A^{-1}|h_j\rangle. \tag{3.5}$$

It follows that for $B < 0$ it is

$$B^{-1} \leq |\mathbf{h}\rangle\langle\mathbf{h}|B|\mathbf{h}\rangle^{-1}\langle\mathbf{h}|. \tag{3.6}$$

These bounds may be applied to the operator $R_1(e)$, when e is real and $e \leq E_1^{(1)}$. Then $B = (e - \overline{H}_1)^{-1} < 0$ and, from Eq. (6),

$$R_1(e) \leq |\mathbf{h}\rangle\langle\mathbf{h}|e - \bar{H}_1|\mathbf{h}\rangle^{-1}\langle\mathbf{h}| = R_1^{(u)}(e) \tag{3.7}$$

with $P|\mathbf{h}\rangle = |\mathbf{h}\rangle$. To obtain a lower bound, one writes

$$R_1(e) - P(e - \bar{E}_1^{(1)})^{-1} = -P(e - \bar{E}_1^{(1)})^{-1}(\bar{E}_1^{(1)} - \bar{H}_1)(e - \bar{H}_1)^{-1} \tag{3.8}$$

and, noticing that

$$A = P(E_1^{(1)} - \bar{H}_1)(e - \bar{H}_1)^{-1} \geq 0, \tag{3.9}$$

one gets, applying Eq. (4),

$$R_1 \geq P(e - \bar{E}_1^{(1)})^{-1} - (e - \bar{E}_1^{(1)})^{-1}(\bar{E}_1^{(1)} - \bar{H}_1) \times A'(\bar{E}_1^{(1)} - \bar{H}_1) = R_1^{(l)}(e),$$
$$A' = |\mathbf{k}\rangle\langle\mathbf{k}|(\bar{E}_1^{(1)} - \bar{H}_1)(e - \bar{H}_1)|\mathbf{k}\rangle^{-1}\langle\mathbf{k}|, \tag{3.10}$$

using $|\mathbf{h}\rangle = (\bar{H}_1 - \bar{E}_1^{(1)})|\mathbf{k}\rangle$, with $P|\mathbf{k}\rangle = |\mathbf{k}\rangle$ and $\bar{O}_1^{(1)}|\mathbf{k}\rangle \neq 0$.

Replacing $R_1(e)$ in Eq. (2.20) by $R_1^{(u)}$ and $R_1^{(l)}$, we obtain the bounds

$$\mathcal{J}^{(u)}[\psi', e] \geq \mathcal{J}[\psi', e] \geq \mathcal{J}^{(l)}[\psi', e], \tag{3.11}$$

where

$$\mathcal{J}^{(b)} = A + B - C^{(b)},\ C^{(b)} = \langle\psi'(e^*)|H_2 - H_2R_1^{(b)}(e)H_2|\psi'(e)\rangle,\ \text{for } b = u, l. \tag{3.12}$$

Choosing the function ψ' to approximate ψ, we get, from Eqs. (2.22)–(2.25)

$$f_{1u}(e) + g_u'(e) + \mathcal{O}(\delta^2) \geq f(e) \geq f_{1l}(e) + g_l'(e) + \mathcal{O}(\delta^2) \tag{3.13}$$

where the subindices indicate that R_1 has been replaced by the corresponding bounds. Equation (3.13) is in a form convenient for iteration. The iteration limit now gives upper and lower bounds to E, within errors of second order in $\delta\psi = \psi' - \psi$. Furthermore, the right side of Eq. (3.13) gives a rigorous lower bound to $f(e)$ if $\mathcal{O}(\delta^2) \geq 0$, and the left-hand side a rigorous upper bound if $\mathcal{O}(\delta^2) \leq 0$.

Although the previous results were obtained for real $e \leq \bar{E}_1^{(1)}$, it is clear that the previous development could be applied, with few changes, to the case where $e \leq \bar{E}_n^{(1)}$, for any n.[5]

4. APPLICATIONS WITH SPECIFIC TRIAL FUNCTIONS

Explicit expressions may be obtained for the functional $\mathcal{J}$ and for f' only after specific choices of the trial function ψ' are made. These choices are dictated by the nature of the problem being investigated and would take the form of parametrized functions or expansions in given basis sets.

We will consider here expansions of the type

$$\psi'(e) = |\boldsymbol{\xi}\rangle\mathbf{c}(e) \tag{4.1}$$

which makes use of the basis set $\xi_1, \xi_2, \cdots, \xi_N$. Replacing Eq. (4.1) in Eq. (2.20)

$$\mathcal{J}(\mathbf{c}, e) = \langle\psi_1(e^*)|H_2|\boldsymbol{\xi}\rangle\mathbf{c}(e) + \mathbf{c}^\dagger(e^*)\langle\boldsymbol{\xi}|H_2|\psi_1(e)\rangle - \mathbf{c}^\dagger(e^*)\langle\boldsymbol{\xi}|H_2 - H_2R_1(e)H_2|\boldsymbol{\xi}\rangle\mathbf{c}(e). \tag{4.2}$$

Differentiating with respect to $\mathbf{c}$ and $\mathbf{c}^\dagger$, we get

$$\mathbf{c}(e) = \langle\boldsymbol{\xi}|H_2 - H_2R_1(e)H_2|\boldsymbol{\xi}\rangle^{-1}\langle\boldsymbol{\xi}|H_2|\psi_1(e)\rangle \tag{4.3}$$

and, replacing this in Eq. (2),

$$g'(e) = \mathbf{a}(e)[\mathbf{C}(e)]^{-1}\mathbf{b}(e) = \operatorname{stat}_{\mathbf{c}}\mathcal{J}(\mathbf{c}, e) \tag{4.4}$$

with

$$\mathbf{a}(e) = \langle\psi_1(e^*)|H_2|\boldsymbol{\xi}\rangle, \tag{4.5}$$

$$\mathbf{b}(e) = \langle\boldsymbol{\xi}|H_2|\psi_1(e)\rangle = \mathbf{a}^\dagger(e^*) \tag{4.6}$$

and

$$\mathbf{C}(e) = \langle\boldsymbol{\xi}|H_2 - H_2R_1(e)H_2|\boldsymbol{\xi}\rangle = \mathbf{C}^\dagger(e^*), \tag{4.7}$$

provided $\mathbf{C}^{-1}$ exists for values of e around E', which we can achieve by a proper choice of $\boldsymbol{\xi}$ and H_1.

Due to the construction of the difference form for $\mathcal{J}$, we obtain, if $|\boldsymbol{\xi}\rangle$ is complete,

$$1 = \langle\varphi|\psi(e)\rangle = \langle\varphi|\boldsymbol{\xi}\rangle[\mathbf{C}(e)]^{-1}\mathbf{b}(e),$$
$$1 = \langle\psi(e)|\varphi\rangle = \mathbf{a}(e)[\mathbf{C}(e)]^{-1}\langle\boldsymbol{\xi}|\varphi\rangle. \tag{4.8}$$

We shall see later that Eq. (8) does not automatically hold for an approximate trial function ψ' determined by Eqs. (4.1)–(4.3), but can be invoked as a constraint.

As an application of the previous equations we shall develop approximants to $\psi(e)$ and $f(e)$ in terms of quantities similar to those in the Brillouin–Wigner perturbation theory, but quite general insofar H_1 will be unspecified. We introduce the Brillouin–Wigner type quantities

$$g_n(e) = \langle\psi_1(e^*)|H_2[R_1(e)H_2]^{n-1}|\psi_1(e)\rangle, \quad n \neq 0 \tag{4.9}$$

$$\Phi_n(e) = [R_1(e)H_2]^n\psi_1(e), \tag{4.10}$$

with

$$R(e) = R_1(e) + R_1(e)H_2R(e), \tag{4.11}$$

$$R_1(e) = P(e - PH_1P)^{-1}P, \quad P = |-|\varphi\rangle\langle\varphi|, \tag{4.12}$$

and e real.

Choosing $\xi_i = \Phi_{i-1}(e)$, $i = 1, 2, \ldots, N$, it follows from Eqs. (4.1)–(4.7) that

$$\operatorname{stat}_{\mathbf{c}}\mathcal{J}(\mathbf{c}, e) = g'(e) = \mathbf{g}_1^\dagger(\mathbf{G}_1 - \mathbf{G}_2)^{-1}\mathbf{g}_1 \tag{4.13}$$

and introducing $|\Phi\rangle = |\Phi_0, \Phi_1, \ldots, \Phi_{N-1}\rangle$

$$\psi'(e) = |\Phi\rangle(\mathbf{G}_1 - \mathbf{G}_2)^{-1}\mathbf{g}_1 \tag{4.14}$$

Here we have used the row matrices

$$\mathbf{g}_i^\dagger = (g_i, g_{i+1}, \ldots, g_{N+i-1}), \quad i = 1, 2, \ldots, N+1 \tag{4.15}$$

and the square matrices

$$\mathbf{G}_i = (\mathbf{g}_i, \mathbf{g}_{i+1}, \ldots, \mathbf{g}_{N+i-1}), \quad i = 1, 2, \cdots. \tag{4.16}$$

We observe that, rewriting $f_1(e) = g_0(e)$,

$$f'(e) = g_0(e) + \mathbf{g}_1^\dagger(\mathbf{G}_1 - \mathbf{G}_2)^{-1}\mathbf{g}_1 \tag{4.17a}$$

is the $[N, N]$ Padé approximant[9,10,2] to the energy series

$$f(e) = g_0(e) + g_1(e) + g_2(e) + \cdots \tag{4.17b}$$

based now upon expansion of Eq. (2.16) in a general

basis set given by Eq. (4.10). Secondly, we obtain for $H_2 \geq 0$ and $R_1(e) \leq 0$ that the error term is

$$\Theta(\delta^2) = \langle \delta\psi | H_2 - H_2 R_1 H_2 | \delta\psi \rangle \geq 0 \tag{4.18}$$

so that

$$f(e) \geq [N, N](e). \tag{4.19}$$

A simple geometric approximation is obtained for $N = 1$, using $|\xi\rangle = \Phi_0 = \psi_1(e)$ which gives[2]

$$f'(e) = g_0(e) + [g_1(e)]^2/[g_1(e) - g_2(e)], \tag{4.20}$$

$$\psi'(e) = \{g_1(e)/[g_1(e) - g_2(e)]\}\psi_1(e). \tag{4.21}$$

We see here that $\langle \varphi | \psi'(e) \rangle \neq 1$, for $g_2(e) \not\equiv 0$. In order to incorporate intermediate normalization, we write

$$\psi = \psi_1(e) + \chi, \tag{4.22a}$$

$$\langle \chi | \varphi \rangle = 0 \tag{4.22b}$$

and then proceed to construct a variational principle for χ. It follows immediately that the functional

$$\mathcal{K}[\chi, e] = \langle \psi_1(e) | H_2 R_1 H_2 | \chi \rangle + \langle \chi | H_2 R_1 H_2 | \psi_1(e) \rangle - \langle \chi | H_2 - H_2 R_1 H_2 | \chi \rangle \tag{4.23a}$$

is of the Lippman–Schwinger type and has the required stationary properties discussed in Sec. 2.

It is also easily seen that

$$f(e) = g_0(e) + g_1(e) + g_2(e) + \mathcal{K}[\chi, e]. \tag{4.23b}$$

Expanding the trial function in the form $\chi' = |\xi\rangle \mathbf{c}(e)$, with

$\langle \varphi | \xi \rangle = 0$, we obtain

$$\mathbf{c}(e) = \langle \xi | H_2 - H_2 R_1 H_2 | \xi \rangle^{-1} \langle \xi | H_2 R_1 H_2 | \psi_1(e) \rangle \tag{4.24}$$

and

$$\operatorname{stat}_{\mathbf{c}} \mathcal{K}[\mathbf{c}, e] = \langle \psi_1(e) | H_2 R_1 H_2 | \xi \rangle \mathbf{c}(e). \tag{4.25}$$

Further by choosing $\xi_i = \Phi_i$ for $i = 1, 2, \ldots, N$ it follows that

$$f'(e) = g_0(e) + g_1(e) + g_2(e) + \mathbf{g}_3^{\dagger}(\mathbf{G}_3 - \mathbf{G}_4)^{-1}\mathbf{g}_3 \tag{4.26}$$

and

$$\psi'(e) = \psi_1(e) + R_1(e) H_2 \psi_1(e) + R_1 H_2 |\Phi\rangle (\mathbf{G}_3 - \mathbf{G}_4)^{-1}\mathbf{g}_3. \tag{4.27}$$

Equation (4.26) is the $(N, N+2]$ Padé approximant to the series Eq. (4.17b), and for $\Theta(\delta^2) \geq 0$ one obtains

$$f(e) \geq [N, N+2](e). \tag{4.28}$$

The stationary solutions for $\mathcal{J}$ or $\mathcal{K}$ are identical to the results obtained by Löwdin[1] via an inner projection of the reaction operator $t = H_2 + H_2 R H_2$. Our derivation emphasizes the optimum property provided by the inner projection technique,[4] which is now seen to follow from a variational approach.

A complementary variational principle for χ is obtained from

$$A[\chi'] = \langle \psi_1(e) | H_2 | \chi' \rangle, \tag{4.29a}$$

$$B[\chi'] = \langle \chi' | H_2 | \psi_1(e) \rangle, \tag{4.29b}$$

$$C[\chi', e] = \langle \chi' | e - H | \chi' \rangle, \tag{4.29c}$$

and

$$\mathfrak{M}[\chi', e] = A + B - C. \tag{4.30}$$

Equation (4.30) is stationary for variations of $\psi' = \psi_1 + \chi'$ around the wavefunction given by Eq. (2.13) and then

$$\operatorname{stat}_{\mathbf{c}} \mathfrak{M}[\chi', e] = A[\chi'] = f(e) - g_0(e) - g_1(e) - \Theta(\delta^2), \tag{4.31a}$$

where

$$\Theta(\delta^2) = \langle \delta\chi | e - H | \delta\chi \rangle \tag{4.31b}$$

with $\chi = \psi(e) - \psi_1(e)$ and $\delta\chi = \chi' - \chi$.

Expanding $\chi' = |\xi\rangle \mathbf{c}(e)$ with $\langle \varphi | \xi \rangle = 0$ the variation of $\mathfrak{M}$ gives

$$\mathbf{c}(e) = \langle \xi | e - H | \xi \rangle^{-1} \langle \xi | H_2 | \psi_1(e) \rangle \tag{4.32}$$

and

$$\operatorname{stat}_{\mathbf{c}} \mathfrak{M}[\mathbf{c}, e] = \langle \psi_1(e) | H_2 | \xi \rangle \mathbf{c}(e). \tag{4.33}$$

A further choice $\xi_i = \Phi_i$ for $i = 1, 2, \ldots, N$ leads to

$$f'(e) = g_0(e) + g_1(e) + \mathbf{g}_2^{\dagger}(\mathbf{G}_2 - \mathbf{G}_3)^{-1}\mathbf{g}_2 \tag{4.34}$$

and

$$\psi'(e) = \psi_1(e) + |\Phi\rangle (\mathbf{G}_2 - \mathbf{G}_3)^{-1}\mathbf{g}_2 \tag{4.35}$$

Equation (4.34) is the $[N, N+1]$ Padé approximant to the series (4.17b). Since the result in Eq. (4.34) is identical to that obtained from the Ritz variational method with the basis functions $\psi_1, \Phi_1, \ldots, \Phi_N$, we get an upper bound to the exact energy $E = f(E)$ in accordance with

$$E \leq e = [N, N+1](e) \tag{4.36}$$

In particular, when $H_2 > 0$ and R_1 is negative, we have

$$[N, N+2](e) \leq f(e) \leq [N, N+1](e). \tag{4.37}$$

In fact, it is straightforward to show by constructing the appropriate trial functions that even and odd order approximants, where by definition $[N, N+j]$ is even (odd) whenever $2N + j$ is even (odd), yield a sequence of lower and upper bounds to $f(e)$. The variational principle based on the functional $\mathfrak{M}[\chi', e]$ corresponds to the Hulthen–Kohn[12,13] principle often used in scattering theory.

Returning to Eq. (4), particular applications require calculation of $R_1(e)$. This is simplified when e is real via the bounds previously obtained, which do not involve the solution of the eigenvalue problem for $\bar{H}_1$. Introducing $R_1^{(u)}$ and $R_1^{(l)}$ from Sec. 3, we can write the following bounds to $\mathbf{C}(e)$,

$$\mathbf{C}^{(u)}(e) = \langle \xi | H_2 | \xi \rangle + \langle \xi | H_2 | \mathbf{h} \rangle \langle \mathbf{h} | e - \bar{H}_1 | \mathbf{h} \rangle^{-1} \langle \mathbf{h} | H_2 | \xi \rangle, \tag{4.38}$$

$$\mathbf{C}^{(l)}(e) = \langle \xi | H_2 | \xi \rangle + (e - \bar{E}_1^{(1)})^{-1} [\langle \xi | H_2 P H_2 | \xi \rangle - \langle \xi | H_2 (\bar{E}_1^{(1)} - \bar{H}_1) A' (\bar{E}_1^{(1)} - \bar{H}_1) | \xi \rangle], \tag{4.39}$$

and we obtain

$$\mathbf{a}(e)\{\mathbf{C}^{(l)}(e)\}^{-1}\mathbf{b}(e) \leq g'(e) \leq \mathbf{a}(e)\{\mathbf{C}^{(u)}\}^{-1}\mathbf{b}(e). \quad (4.40)$$

In this way the calculation of upper and lower bounds to E, within $\mathcal{O}(\delta^2)$ may be carried out simply using the inverse of the $N \times N$ matrices in Eqs. (4.38) and (4.39).

Finally a useful application of the general results discussed in previous sections is obtained when H_2 represents a "physical" perturbation μW, such as for instance given by an electromagnetic field. The standard approach in this connection has been to consider an "internal" splitting of H_1 into H_0 and the internal perturbation $V = H_1 - H_0$ and to choose $\varphi = \varphi_0^{(0)}$ to satisfy

$$H_0\varphi_0^{(0)} = E_0^{(0)}\varphi_0^{(0)}. \quad (4.41)$$

In this manner the corrections to the energy and the wavefunction are systematically given within the double perturbation theory framework as developed by Dalgarno *et al.*[14] Variational methods to calculate first- and second-order properties have been derived by Schwarz[15] and Delves.[16] In order to emphasize the connections between the two different formulations, we will consider the perturbations V and μW separately. As a basis for the discussion, we make a Raleigh–Schrödinger expansion of Eq. (4.17b) in powers of H_2 to obtain, indicating with upper indices the orders in μ,

$$E = f(E) = E^{(0)} + \mu\epsilon^{(1)} + \mu^2\epsilon^{(2)} + \mu^3\epsilon^{(3)} + \cdots \quad (4.42)$$

and

$$\psi = \psi_1(E^{(0)}) + \mu\psi^{(1)} + \mu^2\psi^{(2)} + \cdots \quad (4.43)$$

by equating equal powers of μ, using Eq. (4.11) and

$$R_1(e) = R_1(E^{(0)}) + R_1(E^{(0)})R_1(e)(E^{(0)} - e). \quad (4.44)$$

It then follows that the lower-order energy terms, in the strength μ of the physical perturbation, are

$$E^{(0)} = g_0(E^{(0)}) = \langle\varphi | H_1 + H_1R_1(E^{(0)})H_1 | \varphi\rangle = E_0^{(0)} + \langle\varphi_0^{(0)} | V + VR_1(E^{(0)})V | \varphi_0^{(0)}\rangle \quad (4.45a)$$

so that $E^{(0)}$ is the exact eigenenergy for the internal Hamiltonian H_1 and

$$\epsilon^{(1)} = \frac{\langle\psi_1(E^{(0)}) | W | \psi_1(E^{(0)})\rangle}{\langle\psi_1(E^{(0)}) | \psi_1(E^{(0)})\rangle}, \quad (4.45b)$$

$$\epsilon^{(2)} = \frac{\langle\psi_1(E^{(0)}) | (W - \epsilon^{(1)})R_1(E^{(0)})(W - \epsilon^{(1)}) | \psi_1(E^{(0)})\rangle}{\langle\psi_1(E^{(0)}) | \psi_1(E^{(0)})\rangle} \quad (4.45c)$$

and

$$\epsilon^{(3)} = \frac{\langle\psi_1(E^{(0)}) | (W - \epsilon^{(1)})R_1(E^{(0)})(W - \epsilon^{(1)})R_1(E^{(0)})(W - \epsilon^{(1)}) | \psi_1(E^{(0)})\rangle}{\langle\psi_1(E^{(0)}) | \psi_1(E^{(0)})\rangle} - \epsilon^{(2)}\frac{\langle\psi_1(E^{(0)}) | (W - \epsilon^{(1)})R_1(E^{(0)}) + R_1(E^{(0)})(W - \epsilon^{(1)}) | \psi_1(E^{(0)})\rangle}{\langle\psi_1(E^{(0)}) | \psi_1(E^{(0)})\rangle} \quad (4.45d)$$

while the corresponding perturbed states are

$$\psi_1(E^{(0)}) = \varphi_0^{(0)} + R_1(E^{(0)})V\varphi_0^{(0)}, \quad (4.46)$$

$$\psi^{(1)} = R_1(E^{(0)})(W - \epsilon^{(1)})\psi_1(E^{(0)}), \quad (4.47)$$

where $\psi_1(E^{(0)})$ is the exact eigenstate for H_1, and so on.

From the definition of $R_1(E^{(0)})$ we get

$$R_1(E^{(0)}) = R_0(E^{(0)}) + R_0(E^{(0)})VR_1(E^{(0)}) \quad (4.48)$$

with $R_0(E^{(0)}) = P(E^{(0)} - H_0)^{-1}P$. By application of the variational principles discussed earlier, we can write

$$E^{(0)} = \lim_{N\to\infty}[N, N+1](E^{(0)}), \quad (4.49a)$$

where

$$[N, N+1](E^{(0)}) = E_0^{(0)} + \epsilon_1^{(0)} + (\boldsymbol{\epsilon}_2^{(0)})^\dagger(\mathbf{E}_2^{(0)} - \mathbf{E}_3^{(0)})^{-1}\boldsymbol{\epsilon}_2^{(0)} \quad (4.49b)$$

with the lower indices referring to order in V, and

$$\psi_1(E^{(0)}) = \varphi_0^{(0)} + \lim_{N\to\infty}\chi_{[N,N+1]}^{(0)} \quad (4.50a)$$

where

$$\chi_{[N,N+1]}^{(0)} = |\boldsymbol{\varphi}^{(0)}\rangle(\mathbf{E}_2^{(0)} - \mathbf{E}_3^{(0)})^{-1}\boldsymbol{\epsilon}_2^{(0)}. \quad (4.50b)$$

In Eqs. (4.49) and (4.50) we have used the well-known Brillouin–Wigner quantities for the internal problem

$$\epsilon_i^{(0)} = \langle\varphi_0^{(0)} | V | \varphi_{i-1}^{(0)}\rangle, \quad (4.51)$$

$$\varphi_i^{(0)} = (R_0(E^{(0)})V)^i\varphi_0^{(0)}, \quad (4.52)$$

to construct the matrix expressions occurring in Eqs. (49) and (50), i.e.,

$$|\boldsymbol{\varphi}^{(0)}\rangle = |\varphi_1^{(0)}, \varphi_2^{(0)}, \ldots, \varphi_N^{(0)}\rangle, \quad (4.53)$$

$$(\boldsymbol{\epsilon}_i^{(0)})^\dagger = (\epsilon_i^{(0)}, \epsilon_{i+1}^{(0)}, \ldots, \epsilon_{N+i-1}^{(0)}), \quad (4.54)$$

$$\mathbf{E}_i^{(0)} = (\epsilon_i^{(0)}, \boldsymbol{\epsilon}_{i+1}^{(0)}, \ldots, \boldsymbol{\epsilon}_{N+i-1}^{(0)}). \quad (4.55)$$

We have chosen the $[N, N+1]$ approximant as an example, but we could equally well have treated the primary problem by other types of Padé approximants, which we have previously introduced. First-order properties of the system with Hamiltonian H_1 may now be obtained from the energy term linear in μ:

$$\epsilon^{(1)} = \lim_{N\to\infty}\epsilon_{[N,N+1]}^{(1)}, \quad (4.56a)$$

where

$$\epsilon_{[N,N+1]}^{(1)} = \langle\varphi_0^{(0)} | W | \varphi_0^{(0)}\rangle + \frac{\langle\varphi_0^{(0)} | W | \chi_{[N,N+1]}^{(0)}\rangle + \langle\chi_{[N,N+1]}^{(0)} | W | \varphi_0^{(0)}\rangle + \langle\chi_{[N,N+1]}^{(0)} | (W - \langle\varphi_0^{(0)} | W | \varphi_0^{(0)}\rangle) | \chi_{[N,N+1]}^{(0)}\rangle}{1 + \langle\chi_{[N,N+1]}^{(0)} | \chi_{[N,N+1]}^{(0)}\rangle} \quad (4.56b)$$

and similarly for higher-order properties. Due to the construction of $\epsilon^{(1)}_{[N,N+1]}$ from Eqs. (4. 49b), (4. 50b) and (4. 56b) a solution for finite N fulfills the requirements of Delves principle.[16]

5. DISCUSSION

Variational principles of the Lippmann–Schwinger type appear convenient in connection with the wave-operator formalism. In addition to the fact that they apply to quasibound states,[8] they also provide an alternative formulation for bound state problems previously discussed in terms of operator inequalities and inner projections.[1,3,4]

An apparent disadvantage of these variational forms, namely that they include reduced resolvents, may be circumvented by introducing a primary Hamiltonian H_1 that may be chosen to simplify the calculation of R_1. Even when the nature of the physical problem determines H_1, it is still possible to give upper and lower bounds to R_1, and hence to the variational functional as shown in Sec. 4.

A variation-perturbation treatment using the quantities of Brillouin–Winger perturbation theory leads to useful expressions in terms of Padé approximants.[17] They contain previous results using inner projections and continued fraction expansions.[2,11] These appear in our treatment in a simple and natural way.

By considering the sign of the second-order error in the variational principle, it is possible to give upper and lower bounds to the bracketing function $f(e)$, for real e, and therefore bounds to the energy eigenvalues The bounding properties of the $[N, N]$ and $[N, N-1]$ Padé approximants have been recovered in this way. The practical advantage of using trial wavefunctions that lead to Padé approximants have been studied elsewhere,[2,18]

The present approach is further developed in connection with double perturbation theory. A formulation based on a nonlinear Padé summation of the Brillouin–Wigner perturbation series for the internal Hamiltonian combined with a linear Raleigh–Schrödinger in the physical (or external) perturbation is presented as an alternative method. Application within the configuration-interaction framework by means of a reduced partitioning procedure appear promising and are being further studied.

ACKNOWLEDGMENTS

We want to thank Professor Per-Olov Löwdin for helpful discussions and constructive criticism. One of us (E.J.B.) is most grateful for the hospitality of Professor W. M. Jones, Chairman of the Chemistry Department, and Professor S. S. Ballard, Chairman of the Physics Department, at the University of Florida.

---

* This work was sponsored by the Air Force Office of Scientific Research, Grant AF-AFOSR-71-1714, by the National Science Foundation, Grants NSF GP-16666 and GP-23574, and by the American Chemical Society–Petroleum Research Fund.

† Alfred P. Sloan Foundation Fellow.

1 (a) P. O. Löwdin, in *Perturbation Theory and Its Application in Quantum Mechanics*, edited by C. H. Wilcox (Wiley, New York, 1966), p. 255; (b) P. O. Löwdin, Intern. J. Quant. Chem. **2**, 867 (1968); and previous references cited in these articles.

2 (a) O. Goscinski, Intern. J. Quant. Chem. **1**, 769 (1967); (b) E. Brändas and O. Goscinski, Phys. Rev. A, **1**, 552 (1970); (c) O. Goscinski and E. Brändas, Intern. J. Quant. Chem. **5**, 131 (1971).

3 (a) P. O. Löwdin, Phys. Rev. **139**, A357 (1965); J. Chem. Phys. **43**, S175 (1965); (b) The function ψ_1, has been called "the modified reference function" by Löwdin; we are also using the symbol R_0 rather than his T_0 in Sec. V; (c) P. O. Löwdin, J. Math. Phys. **6**, 1341 (1965).

4 P. Lindner and P. O. Löwdin, Intern. J. Quant. Chem. **2S**, 161 (1968); P. O. Löwdin, *ibid.* **4S**, 231 (1971).

5 Y. Hahn, T. F. O'Malley, and L. Spruch, Phys. Rev. **134**, B394 (1964).

6 R. Sugar and R. Blankenbecler, Phys. Rev. **136**, B472 (1964).

7 (a) B. A. Lippmann and J. Schwinger, Phys. Rev. **79**, 469 (1950); (b) B. A. Lippmann, Phys. Rev. **102**, 264 (1956).

8 D. A. Micha and E. J. Brändas, J. Chem. Phys (1971) (in press).

9 (a) G. A. Baker, Jr., Advan. Theoret. Phys. **1**, 1 (1965); (b) G. A. Baker, Jr., in *The Padé Approximant in Theoretical Physics*, edited by G. A. Baker, Jr. and Y. L. Gammel (Academic, New York, 1970), p. 1.

10 J. Nuttall, Phys. Rev. **157**, 1312 (1967).

11 R. C. Young, L. C. Biedenharn, and E. Feenberg, Phys. Rev. **106**, 1151 (1957).

12 L. Hulthen, Arkiv Mat. Fysik Astr. **35a**, no. 25 (1948).

13 W. Kohn, Phys. Rev. **74**, 1763 (1948).

14 (a) A. Dalgarno and J. T. Lewis, Proc. Roy. Soc. **A233**, 70 (1955); (b) A. Dalgarno and A. L. Stewart, Proc. Roy. Soc. **A247**, 245 (1958).

15 C. Schwarz, Ann. Phys. (N.Y.) **6**, 170 (1959).

16 (a) L. M. Delves, Nucl. Physics. **41**, 497 (1963); (b) See also J. O. Hirschfelder, W. B. Brown, and S. T. Epstein, Advan. Quant. Chem. **1**, 255 (1964), for a review and further references.

17 J. Nuttall, in Ref. 9(b), p. 219

18 (a) E. J. Brändas and R. J. Bartlett, Chem. Phys. Letters **8**, 153 (1971); (b) R. J. Bartlett, thesis (University of Florida, 1971); (c) R. J. Bartlett and E. J. Brändas, Intern. J. Quant. **5S**, (1971) (in press).

One-Center Two-Electron Integrals Arising in Electron-Ion Scattering Calculations

David E. Ramaker*
Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455
and
Sandia Laboratories,† Albuquerque, New Mexico 87115
(Received 26 July 1971)

The two-electron integrals which arise in electron-atomic ion scattering theory are expressed as a finite sum of terms each involving a generalized hypergeometric function. We give what is believed to be the first finite expansion known for the exchange integral. The many well-known recursive and transformation features of the hypergeometric functions are then utilized to conveniently and accurately evaluate the integrals with few restrictions on the values of the parameters.

I. INTRODUCTION

We have recently considered the $(n\ell^2)^1S$ auto-ionizing states of helium[1-3] using a generalization of Fano's[4] configuration interaction theory. This approach reduces the problem to one of evaluating matrix elements involving the orbitals of the chosen basis set. These matrix elements require evaluation of a number of basic types of integrals involving a function which will represent the unbound electron, in our case the Coulomb–Bessel functions. Because of an increased interest in many-particle variational methods in electron-atom scattering theory,[5] where these integrals arise, we wish to describe our methods for evaluating these integrals.

The matrix elements we shall consider have recently been treated by Lyons and Nesbet[6] and by Bottcher.[7] Although Lyons and Nesbet particularly treated the matrix elements for the electron-neutral atom case involving the spherical-Bessel functions, their treatment can easily be generalized for the coulomb–bessel functions arising in electron-ion scattering. The reduction of the energy and overlap matrix elements to seven basic types of integrals has been given in detail by Lyons and Nesbet. We wish only to describe our method for evaluating these basic integrals. Hence we shall not present the reduction of the matrix elements here. Our notation follows Lyons and Nesbet. Frequently we shall for the purposes of brevity refer to formulas from a reference given by the United States National Bureau of Standards.[8] This will be indicated by the notation HMF followed by the equation of the handbook.

The seven basic types of integrals are[6]

$$G(\lambda, p\,|\,\eta, k, \alpha) = \int_0^\infty F_\lambda(\eta, kr) r^{p-\lambda} e^{-\alpha r} dr, \tag{1}$$

$$H(\lambda,\mu,p\,|\,\eta, k, \eta', k', \alpha) = \int_0^\infty F_\lambda(\eta, kr) F_\mu(\eta', k'r) r^{p-\lambda-\mu} e^{-\alpha r} dr, \tag{2}$$

$$I(\lambda,\mu,p\,|\,\eta, k, \eta', k') = \int_0^\infty F_\lambda(\eta, kr) F_\mu(\eta', k'r) r^{p-\lambda-\mu} dr, \tag{3}$$

$$B(p, q\,|\,\alpha, \beta) = \int_0^\infty r_1^p e^{-\alpha r_1} \int_{r_1}^\infty r_2^q e^{-\beta r_2} dr_1 dr_2, \tag{4}$$

$$W(\lambda, p, q\,|\,\eta, k, \alpha, \beta) = \int_0^\infty F_\lambda(\eta, kr_1) r_1^{p-\lambda} e^{-\alpha r_1} \int_{r_1}^\infty r_2^q e^{-\beta r_2} dr_1 dr_2, \tag{5}$$

$$V(\lambda,\mu,p,q\,|\,\eta, k, \eta', k', \alpha) = \int_0^\infty F_\lambda(\eta, kr_1) F_\mu(\eta', k'r_1) r_1^{p-\lambda-\mu} \times \int_{r_1}^\infty r_2^q e^{-\alpha r_2} dr_1 dr_2, \tag{6}$$

$$X(\lambda,\mu,p,q\,|\,\eta, k, \eta', k', \alpha, \beta) = \int_0^\infty F_\lambda(\eta, kr_1) r_1^{p-\lambda} e^{-\alpha r_1} \times \int_{r_1}^\infty F_\mu(\eta', k'r_2) r_2^{q-\mu} e^{-\beta r_2} dr_1 dr_2, \tag{7}$$

where for completeness and later reference we have added the elementary integral $B(p, q\,|\,\alpha, \beta)$, which does not contain any continuum functions.

The radial wavefunction in an attractive coulomb field of charge Z takes the form (HMF 14.1.3)

$$F_\lambda(\eta, kr) = C'_\lambda(\eta) e^{-ikr} (2kr)^\lambda {}_1F_1(\lambda + 1 - \eta, 2\lambda + 2, 2ikr), \tag{8}$$

where $\eta = -iZ/k$. When $\eta \to 0$ we have[7] (HMF 14.6.4)

$$\lim_{\eta\to 0} F_\lambda(\eta, kr) = j_\lambda(kr), \tag{9}$$

indicating we can also treat the integrals involving the spherical-Bessel functions for the electron-neutral atom case as considered by Lyons and Nesbet.[6] We will also have need of the integral representation (HMF 13.2.1)

$${}_1F_1(a, c, z) = B(a, c-a)^{-1} \int_0^1 G_{ac}(t) e^{zt} dt, \tag{10}$$

of the confluent hypergeometric function which when inserted in Eq. (8) gives

$$F_\lambda(\eta, kr) = C_\lambda(\eta) r^\lambda e^{-ikr} B(\lambda + 1 - \eta, \lambda + 1 + \eta)^{-1} \times \int_0^1 G_{\lambda+1-\eta, 2\lambda+2}(t) e^{2ikrt} dt, \tag{11}$$

where

$$C_\lambda(\eta) = \sqrt{\frac{2k}{\pi}} \frac{|\Gamma(\lambda + 1 + \eta)|}{(2\lambda + 1)!} (2k)^\lambda e^{i\pi\eta/2} = (2k)^\lambda C'_\lambda(\eta), \tag{12}$$

$$G_{ac}(t) = t^{a-1}(1-t)^{c-a-1}, \tag{13}$$

and $B(a, b)$ and $\Gamma(a)$ are the beta and gamma functions, respectively.[9]

2. THE INTEGRAL G

The integral G, Eq. (1), upon introducing Eq. (11) becomes

$$G(\lambda, p\,|\,\eta, k, \alpha) = \frac{C_\lambda(\eta)}{B(\lambda + 1 - \eta, \lambda + 1 + \eta)} \times \int_0^1 G_{\lambda+1-\eta, 2\lambda+2}(t) \int_0^\infty r^p e^{-(\alpha + ik - 2ikt)r} dr\, dt. \tag{14}$$

Performing the integration over r we obtain

$$G(\lambda, p\,|\,\eta, k, \alpha) = \frac{C_\lambda(\eta)}{B(\lambda + 1 - \eta, \lambda + 1 + \eta)} \frac{p!}{(\alpha + ik)^{p+1}} \times \int_0^1 G_{\lambda+1-\eta, 2\lambda+2}(t) \left(1 - \frac{2ik}{\alpha + ik} t\right)^{-p-1} dt, \tag{15}$$

which reduces to

$$G(\lambda, p\,|\eta, k, \alpha) = C_\lambda(\eta)\frac{p!}{(\alpha + ik)^{p+1}} \times {}_2F_1\left(p+1, \lambda+1-\eta, 2\lambda+2, \frac{2ik}{\alpha+ik}\right) \quad (16)$$

upon identifying the Gaussian hypergeometric function (HMF 15.1.1) by its integral representation (HMF 15.3.1)

$${}_2F_1(a,b,c,z) = B(a,c-a)^{-1}\int_0^1 G_{ac}(t)(1-zt)^{-b}dt. \quad (17)$$

Equation (16), along with the many recursion relations which exist relating the contiguous ${}_2F_1$ functions,[9] (HMF 15.2.10–15.2.27) has been used by Bottcher and Lyons and Nesbet to derive recursion relations for G which allow their evaluation. Earlier techniques[10–12] involved evaluating ${}_2F_1$ directly by summation or by some other related approach. We use the procedure of Callan *et al.*[13]

Using one of the Gauss recursion relations for the contiguous functions (HMF 15.2.11),

$$\begin{aligned} &I(1-z)\,{}_2F_1(a, I+1, b, z) \\ &\quad + [(a-I)(1-z) + (b-I-a)]\,{}_2F_1(a, I, b, z) \\ &\quad + (I-b)\,{}_2F_1(a, I-1, b, z) \\ &\quad = \alpha_I F_+ + B_I F + \gamma_I F_- = 0, \end{aligned} \quad (18)$$

and the relations

$${}_2F_1(a, o, b, z) = 1 \quad (19)$$

and

$${}_2F_1(a, b, b, z) = (1-z)^{-a}, \quad (20)$$

which can be obtained directly from the Gauss series (HMF 15.1.1), we must solve an nth order system of linear equations. In matrix notation this consists of solving the equation $|A|\,|F = |B$, where $|A|$ is a nonsingular $N \times N$ matrix consisting of the α_I, β_I, and γ_I coefficients [Eq. (18)] and $|F$ is a column vector containing the needed N ${}_2F_1$ functions, only $N-2$ of them really unknown. $|B$ is a column vector consisting of zero except for two elements which are set equal to one and $(1-z)^{-a}$, respectively, corresponding to the two known ${}_2F_1$ functions in the $|F$ vector [Eqs. (19) and (20)] and also corresponding to the two additional rows in $|A|$, which each have an appropriate element set equal to one. This method is particularly adaptable to a computer because of the efficient computer codes now available to handle matrix algebra, and it minimizes any truncation error arising in summation techniques.

3. THE INTEGRALS H AND I

If we proceed as in Sec. II and insert Eq. (11) into Eq. (2) for H, and integrate over r, we obtain

$$\begin{aligned} &H(\lambda, \mu, p\,|\eta, k, \eta', k', \alpha) \\ &= \frac{C_\lambda(\eta)C_\mu(\eta')p!(\alpha + ik + ik')^{-p-1}}{B(\lambda+1-\eta, \lambda+1+\eta)B(\mu+1-\eta', \mu+1+\eta')} \\ &\quad \times \int_0^1\int_0^1 G_{\lambda+1-\eta, 2\lambda+2}(t)G_{\mu+1-\eta', 2\mu+2}(s) \\ &\quad \times\left(1 - \frac{2ikt}{\alpha+ik+ik'} - \frac{2ik's}{\alpha+ik+ik'}\right)^{-p-1} dt\,ds. \end{aligned} \quad (21)$$

Identifying the double integral as the integral representation of the F_2 hypergeometric function,[9,14]

$$F_2(a,b,b',c,c',x,y) = B(b, c-b)^{-1}B(b', c'-b')^{-1} \times \int_0^1\int_0^1 G_{bc}(t)G_{b'c'}(s)(1-tx-sy)^{-a}dt\,ds, \quad (22)$$

yields[15]

$$\begin{aligned} &H(\lambda, \mu, p\,|\eta, k, \eta', k', \alpha) = \frac{C_\lambda(\eta)C_\mu(\eta')p!}{(\alpha + ik + ik')^{p+1}} \\ &\quad \times F_2\left(p+1, \lambda+1-\eta, \mu+1-\eta', 2\lambda+2, 2\mu+2, \right. \\ &\quad \left. \frac{2ik}{\alpha + i(k+k')}, \frac{2ik'}{\alpha+i(k+k')}\right). \end{aligned} \quad (23)$$

Our method of evaluation of the F_2 function is given in Sec. IV.

As α goes to zero H becomes I,

$$I(\lambda, \mu, p\,|\eta, k, \eta', k') = \lim_{\alpha\to 0} H(\lambda, \mu, p\,|\eta, k, \eta', k', \alpha). \quad (24)$$

The evaluation of I has been considered extensively by Alder *et al.*,[16] Biedenharn *et al.*[17] and others[10,18] and hence will not be considered in detail here. We use their method which involves the well-known analytical continuation of the F_2 function in Eq. (23) yielding an F_3 function[9] and a polynomial. The F_3 function for reasonably similar values of k and k' can be evaluated directly. For $k = k'$ when this procedure breaks down, we use the special case formulas of Reynolds *et al.*[18] and Swamy *et al.*[19] when applicable. A formula for the general case when $k = k'$ to our knowledge has not been given.[18,19] When the formulas of Reynolds are not applicable, a procedure of evaluating the integrals at $k = k' \pm \epsilon$ and averaging the two results was found to be adequate. This procedure was also used for the H, V, and X integrals when $k = k'$.

4. THE INTEGRALS B, V, W, X

Insertion of Eq. (11) for the continuum function into Eqs. (5), (6), and (7) and reversing the order of integration gives

$$\begin{aligned} &W(\lambda, p, q\,|\eta, k, \alpha, \beta) = \frac{C_\lambda(\eta)}{B(\lambda+1-\eta, \lambda+1+\eta)} \\ &\quad \times \int_0^1 G_{\lambda+1-\eta, 2\lambda+2}(t)B(p, q\,|(\alpha + ik - 2ikt), \beta)dt, \end{aligned} \quad (25)$$

$$\begin{aligned} &V(\lambda, \mu, p, q\,|\eta, k, \eta', k', \alpha) \\ &\quad = \frac{C_\lambda(\eta)C_\mu(n')}{B(\lambda+1-\eta, \lambda+1+\eta)B(\mu+1-\eta', \mu+1+\eta')} \\ &\quad \times \int_0^1\int_0^1 G_{\lambda+1-\eta, 2\lambda+2}(t)G_{\mu+1-\eta', 2\mu+2}(s) \\ &\quad \times B(p, q\,|i(k+k'-2kt-2k's), \alpha)ds\,dt \end{aligned} \quad (26)$$

and

$$\begin{aligned} &X(\lambda, \mu, p, q\,|\eta, k, n', k', \alpha, \beta) \\ &\quad = \frac{C_\lambda(\eta)C_\mu(\eta')}{B(\lambda+1-\eta, \lambda+1+\eta)B(\mu+1-\eta', \mu+1+\eta')} \\ &\quad \times \int_0^1\int_0^1 G_{\lambda+1-\eta, 2\lambda+2}(t)G_{\mu+1-\eta', 2\mu+2}(s) \\ &\quad \times B(p, q\,|\alpha + i(k-2kt), \beta + i(k'-2k's))ds\,dt \end{aligned} \quad (27)$$

relating W, V, and X to the integral B which contains

no continuum functions. We can integrate B by elementary means giving

$$B(p,q\,|\,\alpha,\beta)=\sum_{\nu=0}^{q}C_\nu(p,q\,|\,\alpha,\beta), \tag{28}$$

where to reduce the notation we have introduced

$$C_\nu(p,q\,|\,\alpha,\beta)=\frac{\beta^{\nu-q-1}(p+\nu)!\,q!}{(\alpha+\beta)^{p+\nu+1}\nu!}. \tag{29}$$

Substituting Eq. (28) into Eqs. (25), (26), and (27) and identifying the remaining integrals over t and/or s as the integral representations of ${}_2F_1$, Eq. (17), F_2, Eq. (22), and S,

$$\begin{aligned}S(a,a',b,b'c,c',x,y,z)&=B(b,c-b)^{-1}B(b',c'-b')^{-1}\\&\times\int_0^1\int_0^1 G_{bc}(t)G_{b'c'}(s)(1-zs)^{-a'}\\&\times(1-xt-ys)^{-a}\,dt\,ds,\end{aligned} \tag{30}$$

respectively, we obtain the closed finite sum of hypergeometric functions

$$\begin{aligned}W(\lambda,p,q\,|\,\eta,k,\alpha,\beta)&=C_\lambda(\eta)\sum_{\nu=0}^{q}C_\nu(p,q\,|\,(\alpha+ik),\beta)\\&\times{}_2F_1\left(p+\nu+1,\lambda+1-\eta,2\lambda+2,\frac{2ik}{\alpha+\beta+ik}\right),\end{aligned} \tag{31}$$

$$\begin{aligned}V(\lambda,\mu,p,q\,|\,\eta,k,\eta',k',\alpha)&=C_\lambda(\eta)C_\mu(\eta')\sum_{\nu=0}^{q}C_\nu(p,q\,|\,i(k+k'),\alpha)\\&\times F_2\left(p+\nu+1,\lambda+1-\eta,\mu+1-\eta',2\lambda+2,\right.\\&\left.2\mu+2,\frac{2ik}{(\alpha+ik+ik')},\frac{2ik'}{(\alpha+ik+ik')}\right),\end{aligned} \tag{32}$$

and

$$\begin{aligned}X(\lambda,\mu,p,q\,|\,\eta,k,\eta',k',\alpha,\beta)&=C_\lambda(\eta)C_\mu(\eta')\sum_{\nu=0}^{q}C_\nu(p,q\,|\,\alpha+ik,\beta+ik')\\&\times S\left(p+1+\nu,q+1-\nu,\lambda+1-\eta,\mu+1-\eta',\right.\\&2\lambda+2,2\mu+2,\frac{2ik}{\alpha+\beta+i(k+k')},\\&\left.\frac{2ik'}{\alpha+\beta+i(k+k')},\frac{2ik'}{\beta+ik'}\right)\end{aligned} \tag{33}$$

for each integral.

A comparison of the definition of S above with the F_2 function, Eq. (22), reveals the presence of an additional $(1-zs)^{-a'}$ factor in S. This factor arises from the integration over r_1 which necessarily contains factors from one of the continuum orbitals. This was avoided in W and V by integrating over the discrete orbitals first. Its presence bars the identification of S as any of Appel's functions even though it can be defined by a somewhat similar triple series,

$$S(a,a',b,b',c,c',x,y,z)=\sum_{\lambda=0}^{\infty}\sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\frac{a'_\lambda a_{u+v}b_u b'_{\lambda+v}}{c_u c'_{\lambda+v}\lambda!\,u!\,v!}x^u y^v z^\lambda, \tag{34}$$

where a_λ is Pochhammer's symbol for $\Gamma(a+\lambda)/\Gamma(a)$.[8] More revealing of the nature of this function is its series representation in terms of the F_2 functions,

$$S(a,a',b,b',c,c',x,y,z)=\sum_{\lambda=0}^{\infty}\frac{a'_\lambda b'_\lambda}{c'_\lambda\lambda!}z^\lambda F_2(a,b,b'+\lambda,c,c'+\lambda,x,y), \tag{35}$$

in terms of the F_1 functions,[9,14]

$$S(a,a',b,b',c,c',x,y,z)=\sum_{\lambda=0}^{\infty}\frac{a_\lambda b_\lambda}{c_\lambda\lambda!}x^\lambda F_1(b',a',a+\lambda,c',z,y), \tag{36}$$

or in terms of the Gauss hypergeometric functions ${}_2F_1$ (HMF 15.1.1),

$$S(a,a',b,b',c,c',x,y,z)=\sum_{\lambda=0}^{\infty}\sum_{u=0}^{\infty}\frac{a_\lambda a'_u b'_{\lambda+u}}{c'_{\lambda+u}\lambda!\,u!}z^u y^\lambda\,{}_2F_1(a+\lambda,b,c,x). \tag{37}$$

In general, none of these simple series representations are finite. A finite series representation of the S function can be obtained, however, for the range of values of the parameters desired in Eq. (33).

Let us substitute

$$(1-tx-sy)^{-a}=(1-sy)^{-a}\left(1-\frac{x}{1-sy}t\right)^{-a} \tag{38}$$

into the integral expression for S, Eq. (30). Integration over "t" using Eq. (17) gives

$$\begin{aligned}S(a,a',b,b',c,c',x,y,z)&=B(c',c'-b')^{-1}\\&\times\int_0^1 G_{b'c'}(s)(1-ys)^{-a}(1-zs)^{-a'}\\&\times{}_2F_1\left(a,b,c,\frac{x}{1-sy}\right)ds.\end{aligned} \tag{39}$$

We then utilize the well-known analytical continuation of the ${}_2F_1$ series (HMF 15.3.7), which can be written in the form

$$\begin{aligned}{}_2F_1(a,b,c,z)&=\frac{\Gamma(c)\Gamma(b-a)}{\Gamma(b)\Gamma(c-a)}(-z)^{-a}\\&\times{}_2F_1(a,a+1-c,a+1-b,1/z)\\&+\frac{\Gamma(c)\Gamma(a-b)}{\Gamma(a)\Gamma(c-b)}(-z)^{a-c}(1-z)^{c-a-b}\\&\times{}_2F_1(1-a,c-a,b+1-a,1/z),\end{aligned} \tag{40}$$

when used along with one of the ${}_2F_1$ transformations (HMF 15.3.3). Substitution into Eq. (39) along with the series expression for ${}_2F_1$ (HMF 15.1.1) yields the result

$$\begin{aligned}S(a,a',b,b',c,c',x,y,z)&=\\&=\frac{(-1)^a\Gamma(c)\Gamma(b-a)}{\Gamma(c-a)\Gamma(b)}\sum_{\lambda=0}^{\infty}\frac{a_\lambda(a+1-c)_\lambda}{(a+1-b)_\lambda\lambda!}x^{-a-\lambda}\\&\times B(b',c'-b')^{-1}\int_0^1 G_{b'c'}(s)(1-ys)^{\lambda}(1-zs)^{-a'}ds\\&+\frac{(-1)^{a-c}\Gamma(c)\Gamma(a-b)}{\Gamma(c-b)\Gamma(a)}\sum_{\lambda=0}^{\infty}\frac{(1-a)_\lambda(c-a)_\lambda}{(b+1-a)_\lambda\lambda!}\\&\times x^{a-c-\lambda}(1-x)^{c-a-b}B(b',c'-b')^{-1}\int_0^1 G_{b'c'}(s)\\&\times(1-ys)^{-a+b+\lambda}(1-zs)^{-a'}\left(1-\frac{y}{1-x}s\right)^{c-a-b}ds.\end{aligned} \tag{41}$$

Consider now the integral over s in the second summation. Inserting the transformation

$$s = \frac{t}{(1-z)+tz} \tag{42}$$

gives

$$(1-z)^{c+c'+\lambda-a'-2a-b'}\int_0^1 G_{b'c'}(t) \times\left(1-t\frac{y-z}{1-z}\right)^{-a+b+\lambda}\left(1-t\frac{y+zx-z}{(1-z)(1-x)}\right)^{c-a-b} \times(1-z(1-t))^{2a-c-c'-\lambda+a'}dt. \tag{43}$$

The exponent of $(1-z(1-t))$, $2a-c-c'-\lambda+a'$ will now always be a positive integer for the S functions needed in Eq. (33) allowing us to expand the $(1-z(1-t))$ factor as a polynomial in $z(1-t)$ giving

$$(1-z)^{c'+c-2a-b'+\lambda-a'}\sum_{j=0}^{2a-c-c'-\lambda+a'}\frac{(\lambda+c+c'-2a-a')_j}{j!}z^j \times\int_0^1 G_{b',c'+j}(t)\left(1-t\frac{y-z}{1-z}\right)^{-a+b+\lambda} \times\left(1-t\frac{y+zx-z}{(1-z)(1-x)}\right)^{c-a-b}dt. \tag{44}$$

The remaining integral in Eq. (44) and the first integral in Eq. (41) are representable by an F_1 function,[9]

$$F_1(a,b,b',c,x,y) = B(a,c-a)^{-1}\int_0^1 G_{ac}(t)(1-tx)^{-b}(1-ty)^{-b'}dt,$$

$$\text{Real}\,(a)>0,\ \text{Real}\,(c-a)>0. \tag{45}$$

The final result for S is given by

$$S(a,a',b,b',c,c',x,y,z) = (-1)^a\frac{\Gamma(c)\Gamma(b-a)}{\Gamma(c-a)\Gamma(b)}\sum_{\lambda=0}^{\infty}\frac{a_\lambda(a+1-c)_\lambda}{(a+1-b)_\lambda\lambda!}x^{-a-\lambda} \times F_1(b',a',-\lambda,c',z,y) + (-1)^{a-c}\frac{\Gamma(c)\Gamma(a-b)}{\Gamma(c-b)\Gamma(a)} \times\sum_{\lambda=0}^{\infty}\frac{(1-a)_\lambda(c-a)_\lambda}{(b+1-a)_\lambda\lambda!}x^{a-c-\lambda}(1-x)^{c-a-b} \times(1-z)^{c+c'+\lambda-a'-2a-b'} \times\sum_{j=0}^{a'+2a-c-c'-\lambda}\frac{(c'-b')_j(c+c'+\lambda-a'-2a)_j}{j!c'_j}z^j \times F_1\left(b',a-b-\lambda,b+a-c,c'+j,\frac{z-y}{z-1},\frac{y+zx-z}{(1-z)(1-x)}\right). \tag{46}$$

In the limit of z approaching zero,

$$\lim_{z\to 0} S(a,a',b,b',c,c',x,y,z) = F_2(a,b,b',c,c',x,y), \tag{47}$$

S becomes F_2 giving an analogous expression for F_2,

$$F_2(a,b,b',c,c',x,y) = \frac{(-1)^a\Gamma(c)\Gamma(b-a)}{\Gamma(c-a)\Gamma(b)}\sum_{\lambda=0}^{\infty}\frac{a_\lambda(a+1-c)_\lambda}{(a+1-b)_\lambda\lambda!}x^{-a-\lambda} \times {}_2F_1(b',-\lambda,c',y) + (-1)^{a-c}\frac{\Gamma(c)\Gamma(a-b)}{\Gamma(c-b)\Gamma(a)} \times\sum_{\lambda=0}^{\infty}\frac{(1-a)_\lambda(c-a)_\lambda}{(b+1-a)_\lambda\lambda!}x^{a-c-\lambda}(1-x)^{c-a-b} \times F_1(b',a-b-\lambda,b+a-c,c',y,y/(1-x)). \tag{48}$$

It should be noted that if both summations over λ were infinite the result would be useless for the purpose here intended; but in this work, a, c, and c' are always integers in Eqs. (46) and (48). Consequently, in the first series either $a+1-c$ is a negative integer reducing the sum to $c-a$ terms, or $c-a$ is a negative integer or zero and $\Gamma(c-a)$ is infinite. In the latter case the entire term is zero. The analogous situation follows for the second series involving $1-a$ and $\Gamma(a)$. Hence S and the F_2 functions in the integrals H, I, and V can be expressed as a finite sum of ${}_2F_1$ and F_1 functions. As mentioned earlier we use the method of Alder[16] and Biedenharn,[17] however, to evaluate the F_2 functions in the integral I.

We have thus expressed all of the basic integrals in closed form as a finite sum of terms, the terms involving the ${}_2F_1$ functions for the integrals G and W, F_2 (or ${}_2F_1$ and F_1) in I, F_1 in X, and ${}_2F_1$ and F_1 in H and V. The procedure for evaluating the ${}_2F_1$ functions has been described in Sec. II. The procedure for evaluating the F_1 functions is described in the following section.

5. EVALUATION OF THE FUNCTIONS F_1

The advantage of reducing all integrals to finite series of F_1 functions becomes evident when examining the rather large number of contiguous recursion relations obtainable for these functions. The situation is especially favourable because all of the required F_1 functions can be reduced to the form $F_1(a,n,b,m,x,y)$, where n and m are integers. The F_1 functions appearing in F_2, Eq. (48), and those appearing in S, Eq. (46), will have this form upon insertion of Eq. (A3). When n is zero, F_1 reduces to a hypergeometric function,

$$F_1(\alpha,o,\beta',m,x,y) = {}_2F_1(\alpha,\beta',m,y). \tag{49}$$

The necessary ${}_2F_1$ functions, ${}_2F_1(\alpha,\beta'+\min,m,y)$ to ${}_2F_1(\alpha,\beta'+\max,m,y)$, can be evaluated simultaneously by the matrix method, Sec. II, except that the two end functions above must now be obtained by evaluating the definitive series (HMF 15.1.1). (Since β' is complex, no single analytical expressions are known for any of the ${}_2F_1$ functions needed here.) Having obtained a vector A of F_1 functions such that $a_j = F_1(\alpha,O,\beta'+j-1,m,x,y)$, one can use the appropriate recursion relations between the contiguous functions $F_1(a,i,\beta'+j,m,x,y)$ and $F_1(\alpha,i\pm1,\beta'+j\pm1,m,x,y)$ to extend the number of known functions to a two-dimensional array, $a_{ij} = F_1(\alpha,i-1,\beta'+j-1,m,x,y)$.

The necessary recursion relations can be derived from those available in the literature. Appel and Kampé de Fériet[14] give a number of contiguous recursion relations for the F_1 functions. Those applicable as parent relations for the expressions needed are

$$\alpha F_1(\alpha+1)-\beta F_1(\beta+1)-\beta' F_1(\beta'+1) = (\alpha-\beta-\beta')F_1, \tag{50}$$

$$\alpha F_1(\alpha+1) = (\gamma-1)F_1(\gamma-1) + (\alpha+1-\gamma)F_1, \quad (51)$$

$$(\gamma-\alpha)\ xF_1(\gamma+1) = \gamma F_1(\beta-1) + \gamma(x-1)F_1, \quad (52)$$

$$(\gamma-\alpha)yF_1(\gamma+1) = \gamma F_1(\beta'-1) + \gamma(y-1)F_1, \quad (53)$$

where the notation has been suppressed such that $F_1(\alpha+1,\beta,\beta',\gamma,x,y) = F_1(\alpha+1)$. Eliminating $F_1(\alpha+1)$ between Eqs. (50) and (51), and substitution of $\gamma = \gamma+1$, gives

$$(\gamma-\beta-\beta')F_1(\gamma+1) + \beta F_1(\beta+1,\gamma+1) + \beta' F_1(\beta'+1,\gamma+1) - \gamma F_1 = 0. \quad (54)$$

$F_1(\gamma+1)$ can now be eliminated using Eq. (52). Substituting $\beta = \beta+1$ into Eq. (52) allows for the elimination of $F_1(\beta+1,\gamma+1)$, and substituting $\beta' = \beta'+1$ into Eq. (53) allows for the elimination of $F_1(\beta'+1, \gamma+1)$. The final result,

$$\frac{(\gamma-\beta-\beta')}{x}F_1(\beta-1) + \beta'\left(\frac{y-1}{y}\right)F_1(\beta'+1) + \beta\left(\frac{x-1}{x}\right)F_1(\beta+1) + \left[(\alpha-\beta-\beta') + \frac{\beta'}{y} + \frac{\beta}{x} - \left(\frac{\gamma-\beta-\beta'}{x}\right)\right]F_1 = 0 = A_1 F_1(\beta-1) + A_2 \times F_1(\beta'+1) + A_3 F_1(\beta+1) + A_4 F_1, \quad (55)$$

involves only the contiguous β and β' functions, and provides a recursion equation for extending the known functions from the vector to the two-dimensional array.

A second relation can be derived by eliminating $F_1(\gamma+1)$ between Eqs. (52) and (53):

$$(y-x)F_1 + xF_1(\beta'-1) - yF_1(\beta-1) = 0 = B_1 F_1 + B_2 F_1(\beta'-1) + B_3 F_1(\beta-1). \quad (56)$$

In principle only one recursion relation is necessary. However, round-off error tends to accumulate, and the rate of this accumulation depends on the recursion relation involved. It is desirable then to have several contiguous relations utilizing that one which minimizes round-off error. From Eqs. (55) and (56), several others can be obtained by eliminating unwanted terms. We obtain the following additional equations:

$$B_3A_1F_1(\beta-1) + (B_3A_3 - A_2B_2)F_1(\beta+1) + B_3 A_4F_1 - A_2B_1F_1(\beta+1,\beta'+1) = 0, \quad (57)$$

$$B_1A_1F_1(\beta-1) + B_1A_3F_1(\beta+1) + (B_1A_4 - A_2B_2)F_1 - A_2B_3F_1(\beta-1,\beta'+1) = 0, \quad (58)$$

$$-A_1B_2F_1(\beta'-1) + B_3A_2F_1(\beta'+1) + B_3A_3F_1(\beta+1) + (B_3A_4 - A_1B_1)F_1 = 0, \quad (59)$$

$$-A_1B_2^2F_1(\beta'-1) + (B_2B_3A_2 - B_3^2A_3)F_1(\beta'+1) + B_2(B_3A_4 - A_1B_1)F_1 - B_3A_3B_1F_1(\beta+1,\beta'+1) = 0, \quad (60)$$

$$-B_1A_1B_2F_1(\beta'-1) + B_1B_3A_2F_1(\beta'+1) + (B_1(B_3A_4 - B_1A_1) - B_3^2A_3)F_1 - B_3A_3B_2F_1(\beta+1,\beta'-1) = 0. \quad (61)$$

We need now to expand the two-dimensional array to three dimensions to obtain the F_1 functions of different γ required in S, Eq. (46). To this end we must derive some relations involving contiguous functions in γ. Equations (52) and (53) are already of this type and are applicable. The elimination of F_1 between these two relations yields the equation

$$\left(\frac{\gamma-\alpha}{\gamma}\right)\left(\frac{x}{x-1} - \frac{y}{y-1}\right)F_1(\gamma+1) - \left(\frac{1}{x-1}\right)F_1(\beta-1) + \left(\frac{1}{y-1}\right)F_1(\beta'-1) = 0, \quad (62)$$

which can also be utilized. Equation (54), applicable as it is written, leads to a more useful result if we perform in it the same operations we utilized to reduce Eq. (54) to Eq. (55) omitting the elimination of $F_1(\gamma+1)$. The result is

$$-\beta\left(\frac{x-1}{x}\right)F_1(\beta+1) - \beta'\left(\frac{y-1}{y}\right)F_1(\beta'+1) + \left(\frac{\gamma-\alpha}{\gamma}\right)(\beta+\beta'-\gamma)F_1(\gamma+1) + \left(\gamma-\alpha-\frac{\beta}{x}-\frac{\beta'}{y}\right)F_1 = 0 = C_1F_1(\beta+1) + C_2F_1(\beta'+1) + C_3F_1(\gamma+1) + C_4F_1. \quad (63)$$

Eliminating $F_1(\beta'+1)$ between Eqs. (63) and (59), we obtain another useful relation,

$$(B_3A_2C_1 - C_2B_3A_3)F_1(\beta+1) + C_3B_3A_2F_1(\gamma+1) + (B_3A_2C_4 - C_2(B_3A_4 - A_1B_1))F_1 + A_1C_2B_2F_1(\beta'-1) = 0. \quad (64)$$

Substitution of $\gamma = \gamma+1$ in Eq. (55) and elimination of $F_1(\beta'+1,\gamma+1)$ between the result and Eq. (54), and likewise substitution of $\gamma = \gamma+1$ in Eq. (59), and elimination of $F_1(\beta+1,\gamma+1)$ with Eq. (54) lead to two additional contiguous relations which we can utilize.

Our procedure can now best be described graphically by Figs. 1 and 2. Initially, one calculates a column vector of $F_1(\alpha,0,\beta'+j,\gamma,x,y)$ functions using the ${}_2F_1$ matrix method of Sec. II. It is then necessary to calculate one F_1 function by truncating an infinite series, since none of the recursion relations will relate the contiguous function $F_1(\alpha,0,\beta'+j,\gamma,x,y)$ and $F_1(\alpha,1,\beta'+j,\gamma,x,y)$. [One cannot relate a transcendental function in x to a linear combination of rational functions in x.[6] Analysis of the applicable recursion relations will also reveal that the coefficient of $F_1(\alpha,1,\beta'+j,\gamma,x,y)$ is always zero.] We describe in an appendix our method of reducing the double infinite series of Eq. (A1) to a single series with optimized convergence properties. Having calculated $F_1(\alpha,1,\beta'+j,\gamma,x,y)$, we proceed to calculate the entire two-dimensional array always using the recursion relation introducing the least round-off error. The optimum recursion relation is taken as that one with the smallest subtractive error as determined from the relative magnitude of the

resultant F_1 function and the 2 or 3 terms comprising the recursion relation. When the accumulation of round-off error becomes intolerable as determined by periodic comparisons of the F_1 functions calculated by the recursive and direct summation technique (see Appendix), the calculation is renormalized. This requires the evaluation of the next two rows, Fig. 1, by the optimized single series technique, and thereby continuing the recursive technique.

The calculation of the F_2 function, Eq. (48), uses F_1 functions at only one value of γ requiring only the planar procedure. However the S integral, Eq. (46), requires a three-dimensional array. The recursive technique is now continued upward in γ, Fig. 2, utilizing the nine relations involving the contiguous functions in γ. When round-off error is detected, the

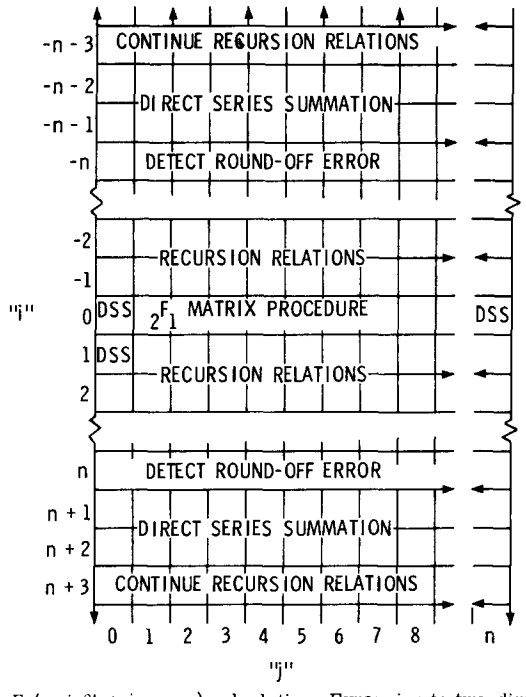

FIG. 1 $F_1(\alpha, i, \beta' + j, \gamma, x, y)$ calculation. Expansion to two-dimensional array. DDS indicates series summation.

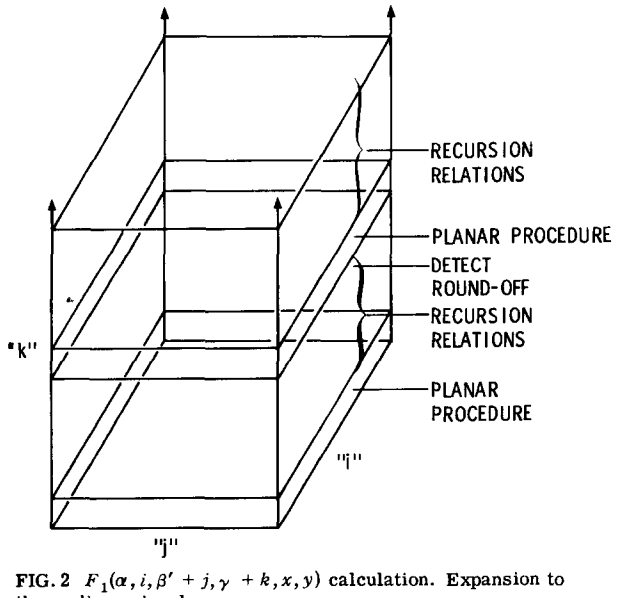

FIG. 2 $F_1(\alpha, i, \beta' + j, \gamma + k, x, y)$ calculation. Expansion to three-dimensional array.

process is renormalized by calculating another planar array of functions exactly as the initial horizontal array was obtained.

The rate of accumulation of round-off error, in addition to the choice of recursion relation, can be influenced by the direction in which the recursion relations are applied. But which direction is optimal is not immediately apparent. Both the magnitude and the sign of x and y, and the remaining parameters must be considered when determining the optimum direction. An example of such considerations is given by Lyons and Nesbet[6] when working with the G integrals involving the $_2F_1$ functions. Such considerations become exceedingly complex when working with the F_1 functions consisting of several more variables. To avoid the computer programming difficulties which result from such considerations, we choose to run consistently in the same direction and to renormalize when necessary. The availability of the $F_1(\alpha, 0, \beta' + j, \gamma, x, y)$ functions as starting points influences our decision to run the recursion relations as indicated in Figs. 1 and 2.

Calculations reveal that the recursive technique works very well when x' and y' are significantly different from each other and different from an absolute value of unity. When this is true a three-dimensional array, containing up to 1000 functions, can be obtained to seven and eight significant figures without renormalizing. On the other hand for less satisfactory values of x' and y', instances occur when renormalization is necessary at every other row in the horizontal arrays and likewise at every other plane when increasing γ. Even in this case, however, only half of the functions are being calculated by the direct series summation, a significant saving in time and effort.

A second point of interest is whether more than one contiguous relation is actually beneficial. Calculations indicate that this is indeed necessary to reduce the round-off error to within tolerable levels. It was found that up to two and three significant figures could be lost due to subtractive error by one recursive equation while another equation would introduce little or no round-off. For any one value of x' and y', it is evident that no one equation is highly preferable over the rest, but rather two or three are equally satisfactory. The three-dimensional matrix is then built up alternating among these three relations.

6. DISCUSSION

A most important feature of the methods outlined in this paper is the single method of evaluation for all the six basic integrals. All of the integrals are reduced to a finite number of similar terms each involving a generalized hypergeometric function. One needs only to program routines for efficiently evaluating the $_2F_1$ and F_1 functions and the integrals can be easily evaluated. Although Lyons and Nesbet have not described their procedure for evaluating the X integrals,[6] they indicate it involves techniques completely different from those of the remaining integrals. Bottcher's[7] method requires a numerical integration for X [given the symbol $\tilde{H}^{ll'}_{nn}(\eta, \eta', k, k', \alpha, \beta)$ in his paper] as opposed to a recursive technique for some of the others.

We have completed our calculations on the $(n\ell^2)^1S$

TABLE I. Integration parameters of the continuum components and estimated significant figures retained in the evaluation of the $H_{nkn'k'}$ integrals.

$n\ell k\ell$	Threshold	ϵ_{max}[a]	k_{max}	Significant figures in $H_{nkn'k'}$[b]					
	(au)	(au)	(au)	1sks	2sks	2pkp	3sks	3pkp	3dkd
1sks	−2.00	2.50	3.00	12(8)	10(6)	8(4)			
2sks	− .500	2.48	2.44		8(4)	8(3)	8(2)	7(2)	6(2)
2pkp	− .500	1.70	2.10			8(3)	7(2)	6(2)	5(2)
3sks	− .222	.320	1.04				5(2)	6(2)	5(2)
3pkp	− .222	.315	1.04					4(2)	5(2)
3dkd	− .222	.312	1.03						4(2)

[a] $\epsilon = k^2/2m_e$

[b] The number of significant figures as estimated from the hermiticity test, the ratio of real and imaginary parts of $H_{nkn'k'}$, and graphical plots of $H_{nkn'k'}$ vs k and k'. The first number of each column is the representative value; the number in parentheses is the estimated minimum number of significant figures retained in some integrals.

autoionizing states of helium[1-3] by using the methods described here for evaluating the necessary integrals. All calculations were performed on the CDC 6600 computer. Thorough testing of the techniques described here revealed two important characteristics concerning production run times and accuracy.

In Table I we give the estimated significant figures retained in the evaluation of the general matrix element,

$$H_{nkn'k'} = \langle (1 - P_{12}) n\ell(1) k\ell(2) \,|\, H - E \,| \times (1 - P_{12}) n'\ell'(1) k'\ell'(2) \rangle , \quad (65)$$

which Lyons and Nesbet have expressed as a finite number of terms involving the seven basic integrals, Eqs. (1)–(7). The table contains two entries for each integral, the first indicating the representative value and the number in parentheses the estimated minimum number of significant figures retained in some integrals, usually those of higher k and k'. Table I also shows a marked decrease in accuracy for integrals involving continua of higher n and n'.

The loss in accuracy for integrals involving continua of higher n and n' can be attributed to cancellation among the terms comprising the expression given by Lyons and Nesbet.[6] This can be reduced by obtaining the basic integrals of Eqs. (1)–(7) more accurately and hence by retaining more precision in the recursive technique. Allowing lesser round-off by renormalization of the recursive technique accomplishes this directly. Round-off for higher k and k' arises because of inadequate optimization of the convergence properties (see Appendix) of the single-series summation technique when evaluating some of the F_1 functions needed in the X integrals. Summing a large number of terms not only introduced round-off error but enormously increases the production run times. Hence the run time for small k and k' (0–2.0 a.u.) which is typically of 1.5–15 secs, depending on n and n', increases to 15–50 secs per integral for larger k and k' (1.0–20 a.u.). For very large k and k', however ($k, k' > 20$ a.u.), the execution time again decreases.

It is apparent our procedures are inconvenient at intermediate energies. This is not a severe limitation, however, since an analysis of resonance states usually is confined to lower k and k'. (The round-off experienced as k and k' becomes larger is not serious in our calculations because one must in any case truncate the continuum at some point k_{max}.) A partial wave expansion in atomic Coulomb (or non-Coulomb) Born scattering, where the identical integrals arise, is confined to very large values of k and k' ($k, k', \gtrsim$ 2000 a.u.) when they can still be conveniently evaluated. The method of Bottcher on the other hand requires a numerical integration over an apparently decreasing oscillatory integrand as k or k' increases. This suggests that the two procedures would complement one another very nicely.

In conclusion, no problems were observed in evaluating the G, H, I, B, V, and W integrals with respect to accuracy or production run times for even remote values of the parameters. Except for the X integrals at intermediate energies and for continua involving large n and n', all integrals can be evaluated conveniently and accurately. We believe the major significance of this work, however, lies in Eqs. (33) and (46) which express the exchange integral X as an heretofore unknown finite sum of F_1 functions. As more convenient means of obtaining the general hypergeometric F_1 functions of Appel[14] become available, these equations along with those for W will be most useful.

ACKNOWLEDGMENTS

The author wishes to thank Professor David M. Schrader for his encouragement and many helpful suggestions throughout this work. The support and splendid cooperation of the University of Minnesota computer center is gratefully acknowledged. Acknowledgment is also made to the donors of the Petroleum Research Fund, administered by the American Chemical Society and to the U.S. Atomic Energy Commission, Sandia Laboratories for partial support of this research.

APPENDIX: OPTIMIZATION OF SERIES CONVERGENCE IN EVALUATION OF F_1

To initiate and renormalize the recursive technique described in Sec. V, the F_1 function must be evaluated by truncating the definitive infinite series,[14]

$$F_1(a, b, b', c, x, y) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \frac{a_{m+n} b_m b'_n}{C_{m+n} m! n!} x^m y^n , \quad |x| < 1, \ |y| < 1 \quad (A1)$$

at some suitable point. To minimize this truncation error and limit the summation to a reasonable number of terms, it is necessary that the arguments x and y be limited to well within the radius of convergence of the series as indicated above. In general this is not true and some appropriate transformation

of the series to some other series which has reduced arguments x' and y' is required.

The possible transformations of Appel's F_1 functions are many. The five most direct and familiar are[9,14]

$$F_1(\alpha,\beta,\beta',\gamma,x,y) = (1-x)^{-\beta}(1-y)^{-\beta'} \times F_1\left(\gamma-\alpha,\beta,\beta',\gamma,\frac{x}{x-1},\frac{y}{y-1}\right), \tag{A2}$$

$$= (1-x)^{-\alpha} F_1\left(\alpha,\gamma-\beta-\beta',\beta',\gamma,\frac{x}{x-1},\frac{x-y}{x-1}\right), \tag{A3}$$

$$= (1-y)^{-\alpha} F_1\left(\alpha,\beta,\gamma-\beta-\beta',\gamma,\frac{y-x}{y-1},\frac{y}{y-1}\right), \tag{A4}$$

$$= (1-x)^{\gamma-\alpha-\beta}(1-y)^{-\beta'} \times F_1\left(\gamma-\alpha,\gamma-\beta-\beta',\beta',\gamma,x,\frac{y-x}{y-1}\right), \tag{A5}$$

$$= (1-x)^{-\beta}(1-y)^{\gamma-\alpha-\beta'} \times F_1\left(\gamma-\alpha,\beta,\gamma-\beta-\beta',\gamma,\frac{x-y}{x-1},y\right). \tag{A6}$$

Via the expression[14]

$$F_1(\alpha,\beta,\beta',\gamma,x,y) = (1-x)^{-\beta} F_3\left(\gamma-\alpha,\alpha,\beta,\beta',\gamma,\frac{x}{x-1},y\right) \tag{A7}$$

and the corresponding expression with arguments x and $y/(y-1)$, twelve transformations involving F_3 become available; but only six are unique. Analagously, applying the relation[14]

$$F_1(\alpha,\beta,\beta',\gamma,x,y) = x^{\beta}y^{-\beta'} F_2\left(\beta+\beta',\alpha,\beta',\gamma,\beta+\beta',x,1-\frac{x}{y}\right) \tag{A8}$$

and the three transformations of F_2,[9,14]

$$F_2(\alpha,\beta,\beta',\gamma,\gamma',x,y) = (1-x)^{-\alpha} F_2\left(\alpha,\gamma-\beta,\beta',\gamma,\gamma',\frac{x}{x-1},\frac{y}{1-x}\right), \tag{A9}$$

$$= (1-y)^{-\alpha} F_2\left(\alpha,\beta,\gamma'-\beta',\gamma,\gamma',\frac{x}{1-y},\frac{y}{y-1}\right), \tag{A10}$$

$$= (1-x-y)^{-\alpha} F_2\left(\alpha,\gamma-\beta,\gamma'-\beta',\gamma,\gamma',\frac{x}{x+y-1},\frac{y}{x+y-1}\right), \tag{A11}$$

we obtain twelve new unique expressions for F_1.

Consider again the six containing F_3 transformations obtained above. By utilizing the analytic continuation formula for F_3,[9,14] and by transforming the resultant F_2 functions to the F_2 functions in Eqs. (A9)–(A11), twenty four additional transformations of F_1 are obtained. These latter transformations are applicable for large starting values of x and y in the desired F_1 function.

Finally, employing the Horn's H_2 function defined by the series[9,14]

$$H_2(\alpha,\beta,\gamma,\delta,\epsilon,x,y) = \sum_{\lambda=0}^{\infty}\sum_{u=0}^{\infty} \frac{\alpha_{\lambda-u}\beta_\lambda\gamma_u\delta_u x^\lambda y^u}{\epsilon_\lambda \lambda!\, u!}, \tag{A12}$$

and the relation this function has with the F_3 function,[20]

$$F_3(\alpha,\alpha',\beta,\beta',\gamma,x,y) = \frac{\Gamma(\beta-\alpha)\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\alpha)}(-x)^{-\alpha} \times H_2\left(1+\alpha-\gamma,\alpha,\alpha',\beta',1+\alpha-\beta,\frac{1}{x},-y\right) + \frac{\Gamma(\alpha-\beta)\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma-\beta)}(-x)^{-\beta} \times H_2\left(1+\beta-\gamma,\beta,\alpha',\beta',1+\beta-\alpha,\frac{1}{x},-y\right), \tag{A13}$$

twelve more unique possibilities exist for expediting the F_1 calculation. Equation (A13) above has the parallel relation[20]

$$F_3(\alpha,\alpha',\beta,\beta',\gamma,x,y) = \frac{\Gamma(\beta'-\alpha')\Gamma(\gamma)}{\Gamma(\beta')\Gamma(\gamma-\alpha')}(-y)^{-\alpha'} \times H_2\left(1+\alpha'-\gamma,\alpha',\alpha,\beta,1+\alpha'-\beta',\frac{1}{y},-x\right) + \frac{\Gamma(\alpha'-\beta')\Gamma(\gamma)}{\Gamma(\alpha')\Gamma(\gamma-\beta')}(-y)^{-\beta'} \times H_2\left(1+\beta'-\gamma,\beta',\alpha,\beta,1+\beta'-\alpha',\frac{1}{y},-x\right) \tag{A14}$$

indicating its utility when either one of the original values of x and y in the F_1 function is large and the remaining argument is small.

We have just outlined means for obtaining 60 transformations of the F_1 series leaving us with 60 unique sets of arguments x' and y' which are analytical functions of the initial arguments x and y. The initial arguments x and y are, of course, fixed by the expression for the integral, either V or X. But there are two different ways of performing V and X, which will give different numerical values of x and y. In Eq. (7) for X one has a choice of integrating over either r_1 or r_2 first. Although indentical analytical expressions for X arise, the numerical values of the x and y arguments in F_1 of Eq. (46) will be different provided we are calculating a nondiagonal matrix element $[(n,\ell,k) \neq (n',\ell',k')$ in Eq. (65)]. Although we must integrate over the bound components first in V, Eq. (6), we have two alternate ways of performing the F_2 to F_1 transformation in Eq. (48). [The second F_2 to F_1 transformation can be obtained by switching the parameters $\beta \rightleftarrows \beta'$, $\gamma \rightleftarrows \gamma'$, and $x \rightleftarrows y$ on the right side of Eq. (48).] This then gives us 120 possible sets of arguments x' and y' for optimizing the convergence of the infinite series.

The hypergeometric functions, although defined by the double infinite series [i.e., Eq. (A1)], can be reduced to a single series of ${}_2F_1$ functions[14]:

$$F_1(\alpha,\beta,\beta',\gamma,x,y) = \sum_{\lambda=0}^{\infty}\frac{\alpha_\lambda\beta_\lambda}{\gamma_\lambda\lambda!}x^\lambda\, {}_2F_1(\alpha+\lambda,\beta',\gamma+\lambda,y), \tag{A15}$$

$$F_2(\alpha,\beta,\beta',\gamma,\gamma',x,y) = \sum_{\lambda=0}^{\infty}\frac{\alpha_\lambda\beta_\lambda}{\gamma_\lambda\lambda!}x^\lambda\, {}_2F_1(\alpha+\lambda,\beta',\gamma',y), \tag{A16}$$

$$F_3(\alpha,\alpha',\beta,\beta',\gamma,x,y) = \sum_{\lambda=0}^{\infty}\frac{\alpha_\lambda\beta_\lambda}{\gamma_\lambda\lambda!}x^\lambda\, {}_2F_1(\alpha',\beta',\gamma+\lambda,y), \tag{A17}$$

$$H_2(\alpha,\beta,\gamma,\delta,\epsilon,x,y)=\sum_{\lambda=0}^{\infty}\frac{\alpha_{-\lambda}\gamma_\lambda\delta_\lambda}{\lambda!}y^\lambda\ {}_2F_1(\alpha-\lambda,\beta,\epsilon,x). \tag{A18}$$

Use of the recursion relation (HMF 15. 2. 12)

$$\begin{aligned}&c(c-1)(z-1)\ {}_2F_1(a,b,c-1,z)\\&\quad+c[c-1-(2c-a-b-1)z]\ {}_2F_1(a,b,c,z)\\&\quad+(c-a)(c-b)z\ {}_2F_1(a,b,c+1,z)=0,\end{aligned} \tag{A19}$$

and Eq. (18) enables us to evaluate all the ${}_2F_1$ functions rapidly, effectively reducing the double infinite series to a single series. Round-off error in the recursion relation requires, however, that the recursive technique be renormalized periodically. We have used these single series expressions to evaluate the Appel functions renormalizing as demanded by round-off accumulation. Calculations reveal that renormalization is required at approximately every five to ten terms indicating that the number of terms have been reduced from the direct double infinite series by a factor of 30 to 100.

The general method of calculating an F_1 function should now be obvious. We simply determine which of the 120 sets of arguments x' and y' is smallest and proceed with the single series summation giving the maximum rate of convergence. It is assumed that with 120 possibilities we will always find an x' and y' combination well within the radius of convergence of one of the infinite series. Calculations revealed, however, that for rather large and different values of x and y, we did not reduce x' and y' sufficiently to insure very fast convergence. Indeed, it is this fact which makes our evaluation techniques inconvenient for integrals in the intermediate energy ranges.

---

* The material in this paper is contained in a Ph.D. dissertation submitted to the faculty of the University of Iowa.
† Present address.

1 D. E. Ramaker, R. B. Stevenson, and D. M. Schrader, J. Chem. Phys. **51**, 5276 (1969).
2 D. E. Ramaker and D. M. Schrader, J. Chem. Phys. **55**, 471 (1971).
3 D. E. Ramaker and D. M. Schrader (unpublished).
4 U. Fano, Phys. Rev. **124**, 1866 (1961); U. Fano and F. Prats, Proc. Natl. Acad. Sci. India **A33**, 553 (1963).
5 F. E. Harris, Phys. Rev. Letters **19**, 173 (1967); H. H. Michels and F. E. Harris, *ibid.*,885 (1967); R. L. Armstead, Phys. Rev. **171**, 91 (1968); R. K. Nesbet, *ibid.* **175**, 134 (1968); F. E. Harris and H. H. Michels, Phys Rev. Letters **22**, 1036 (1969).
6 J. D. Lyons and R. K. Nesbet, J. Comp. Phys. **4**, 499 (1969).
7 C. Bottcher, J. Comp. Phys. **6**, 237 (1970).
8 National Bureau of Standards, Applied Mathematics Series No. 55, *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (U.S. Gov't. Printing Office, Washington, D.C., 1964), Chaps. 13–15.
9 Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdelyi (McGraw-Hill, New York, 1953).
10 W. Gordon, Ann. Physik 2, 1031 (1929).
11 L. D. Landau and E. M. Lifshitz, *Quantum Mechanics, Non-Relativistic Theory* (Addison-Wesley, Reading, Mass., 1958), Appendix F.
12 H. B. Bebb, J. Math. Phys. **7**, 955 (1966).
13 E. J. Callan, P. Nikolai, and W. L. McDavid, *Proceedings of the Third International Conference on the Physics of Electronic and Atomic Collisions*, edited by M. R. C. McDowell (North-Holland, Amsterdam, 1964), p. 348.
14 P. Appell and M. J. Kampé de Fériet, *Functions hypergeometriques et hyperspheriques, polynomes d'Hermite* (Gauthiers, Villars, 1926).
15 When $\alpha = i(\pm k \mp k')$, the F_2 function reduces to a ${}_3F_2$ function as shown by W. W. Gargano and D. S. Onley, J. Math. Phys. **11**, 1191 (1970).
16 K. Alder, A. Bohn, F. Huus, B. Mottelson, and A. Winther, Rev. Mod. Phys. **28**, 432 (1956).
17 L. C. Biedenharn, J. L. McHale, and P. M. Thalen, Phys. Rev. **100**, 376 (1955); L. C. Biedenharn and P. J. Brussard, *Coulomb Excitation* (Oxford U. P., London, 1965).
18 J. T. Reynolds, D. S. Onley, and L. C. Biedenharn, J. Math. Phys. **5**, 411 (1964).
19 N. V. V. J. Swamy, R. G. Kulkarni, and L. C. Biendenharn, J. Math. Phys. **11**, 1165 (1970).
20 A. Erdelyi, Proc. Roy. Soc. (Edinburgh) **A62**, 378 (1948).

Approach to Stochastic Lagrangian Integrals and Their Asymptotic Evaluation for Sound Propagation in Continuous Stochastic Media*

Jerome A. Neubert
Naval Undersea Research and Development Center, Pasadena, California 91107

and

John L. Lumley
Department of Aerospace Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 4 June 1971)

Previously the formal solution of the Eulerian–Lagrangian problem for sound propagation in continuous stochastic media was reframed so that the emphasis on the need for complete knowledge of the statistical nature of the Lagrangian functional of interest is shifted to the need for knowledge of the asymptotic behavior of certain stochastic Lagrangian integrals which result from the application of a central limit theorem for stochastic functionals. In this paper, the asymptotic evaluation of these Lagrangian stochastic integrals is developed and illustrated for both ensemble and subensemble expectations in a statistically isotropic medium. In addition, the relationship between this method of analysis and the comparable Wiener integral is discussed.

INTRODUCTION

In Ref. 1, the search for a *proper asymptotic solution*[2] to the stochastic Helmholtz equation

$$\nabla^2 p + k_0^2\mu^2 p = 0, \tag{1}$$

where p represents the sound pressure wave, k_0 is the free-space wavenumber, and μ is the refractive index, for sound propagation through continuous stochastic media (for example, due to the random effects of the scalar inhomogeneities in a turbulent medium; see Neubert and Lumley[3]) lead to the intrinsically Lagrangian relation $p[\mathbf{X}(s,\boldsymbol{\xi})]$, where $\mathbf{X}(s,\boldsymbol{\xi})$ is a continuous, differentiable path of arc length s from the initial point $\boldsymbol{\xi}$. The one-dimensional solution was given in Ref. 1 (see also Frisch[4]), but the three-dimensional solution resulted in consideration of the Eulerian–Lagrangian problem for sound propagation

$$H_2(\alpha,\beta,\gamma,\delta,\epsilon,x,y)=\sum_{\lambda=0}^{\infty}\frac{\alpha_{-\lambda}\gamma_\lambda\delta_\lambda}{\lambda!}y^\lambda\ {}_2F_1(\alpha-\lambda,\beta,\epsilon,x). \tag{A18}$$

Use of the recursion relation (HMF 15. 2. 12)

$$\begin{aligned}&c(c-1)(z-1)\ {}_2F_1(a,b,c-1,z)\\&\quad+c[c-1-(2c-a-b-1)z]\ {}_2F_1(a,b,c,z)\\&\quad+(c-a)(c-b)z\ {}_2F_1(a,b,c+1,z)=0,\end{aligned} \tag{A19}$$

and Eq. (18) enables us to evaluate all the ${}_2F_1$ functions rapidly, effectively reducing the double infinite series to a single series. Round-off error in the recursion relation requires, however, that the recursive technique be renormalized periodically. We have used these single series expressions to evaluate the Appel functions renormalizing as demanded by round-off accumulation. Calculations reveal that renormalization is required at approximately every five to ten terms indicating that the number of terms have been reduced from the direct double infinite series by a factor of 30 to 100.

The general method of calculating an F_1 function should now be obvious. We simply determine which of the 120 sets of arguments x' and y' is smallest and proceed with the single series summation giving the maximum rate of convergence. It is assumed that with 120 possibilities we will always find an x' and y' combination well within the radius of convergence of one of the infinite series. Calculations revealed, however, that for rather large and different values of x and y, we did not reduce x' and y' sufficiently to insure very fast convergence. Indeed, it is this fact which makes our evaluation techniques inconvenient for integrals in the intermediate energy ranges.

---

* The material in this paper is contained in a Ph.D. dissertation submitted to the faculty of the University of Iowa.
† Present address.

1 D. E. Ramaker, R. B. Stevenson, and D. M. Schrader, J. Chem. Phys. **51**, 5276 (1969).
2 D. E. Ramaker and D. M. Schrader, J. Chem. Phys. **55**, 471 (1971).
3 D. E. Ramaker and D. M. Schrader (unpublished).
4 U. Fano, Phys. Rev. **124**, 1866 (1961); U. Fano and F. Prats, Proc. Natl. Acad. Sci. India **A33**, 553 (1963).
5 F. E. Harris, Phys. Rev. Letters **19**, 173 (1967); H. H. Michels and F. E. Harris, *ibid.*,885 (1967); R. L. Armstead, Phys. Rev. **171**, 91 (1968); R. K. Nesbet, *ibid.* **175**, 134 (1968); F. E. Harris and H. H. Michels, Phys Rev. Letters **22**, 1036 (1969).
6 J. D. Lyons and R. K. Nesbet, J. Comp. Phys. **4**, 499 (1969).
7 C. Bottcher, J. Comp. Phys. **6**, 237 (1970).
8 National Bureau of Standards, Applied Mathematics Series No. 55, *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (U.S. Gov't. Printing Office, Washington, D.C., 1964), Chaps. 13–15.
9 Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdelyi (McGraw-Hill, New York, 1953).
10 W. Gordon, Ann. Physik 2, 1031 (1929).
11 L. D. Landau and E. M. Lifshitz, *Quantum Mechanics, Non-Relativistic Theory* (Addison-Wesley, Reading, Mass., 1958), Appendix F.
12 H. B. Bebb, J. Math. Phys. **7**, 955 (1966).
13 E. J. Callan, P. Nikolai, and W. L. McDavid, *Proceedings of the Third International Conference on the Physics of Electronic and Atomic Collisions*, edited by M. R. C. McDowell (North-Holland, Amsterdam, 1964), p. 348.
14 P. Appell and M. J. Kampé de Fériet, *Functions hypergeometriques et hyperspheriques, polynomes d'Hermite* (Gauthiers, Villars, 1926).
15 When $\alpha = i(\pm k \mp k')$, the F_2 function reduces to a ${}_3F_2$ function as shown by W. W. Gargano and D. S. Onley, J. Math. Phys. **11**, 1191 (1970).
16 K. Alder, A. Bohn, F. Huus, B. Mottelson, and A. Winther, Rev. Mod. Phys. **28**, 432 (1956).
17 L. C. Biedenharn, J. L. McHale, and P. M. Thalen, Phys. Rev. **100**, 376 (1955); L. C. Biedenharn and P. J. Brussard, *Coulomb Excitation* (Oxford U. P., London, 1965).
18 J. T. Reynolds, D. S. Onley, and L. C. Biedenharn, J. Math. Phys. **5**, 411 (1964).
19 N. V. V. J. Swamy, R. G. Kulkarni, and L. C. Biendenharn, J. Math. Phys. **11**, 1165 (1970).
20 A. Erdelyi, Proc. Roy. Soc. (Edinburgh) **A62**, 378 (1948).

Approach to Stochastic Lagrangian Integrals and Their Asymptotic Evaluation for Sound Propagation in Continuous Stochastic Media*

Jerome A. Neubert
Naval Undersea Research and Development Center, Pasadena, California 91107

and

John L. Lumley
Department of Aerospace Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 4 June 1971)

Previously the formal solution of the Eulerian–Lagrangian problem for sound propagation in continuous stochastic media was reframed so that the emphasis on the need for complete knowledge of the statistical nature of the Lagrangian functional of interest is shifted to the need for knowledge of the asymptotic behavior of certain stochastic Lagrangian integrals which result from the application of a central limit theorem for stochastic functionals. In this paper, the asymptotic evaluation of these Lagrangian stochastic integrals is developed and illustrated for both ensemble and subensemble expectations in a statistically isotropic medium. In addition, the relationship between this method of analysis and the comparable Wiener integral is discussed.

INTRODUCTION

In Ref. 1, the search for a *proper asymptotic solution*[2] to the stochastic Helmholtz equation

$$\nabla^2 p + k_0^2\mu^2 p = 0, \tag{1}$$

where p represents the sound pressure wave, k_0 is the free-space wavenumber, and μ is the refractive index, for sound propagation through continuous stochastic media (for example, due to the random effects of the scalar inhomogeneities in a turbulent medium; see Neubert and Lumley[3]) lead to the intrinsically Lagrangian relation $p[\mathbf{X}(s,\boldsymbol{\xi})]$, where $\mathbf{X}(s,\boldsymbol{\xi})$ is a continuous, differentiable path of arc length s from the initial point $\boldsymbol{\xi}$. The one-dimensional solution was given in Ref. 1 (see also Frisch[4]), but the three-dimensional solution resulted in consideration of the Eulerian–Lagrangian problem for sound propagation

in continuous stochastic media which was resolved in Ref. 5 in terms of stochastic Lagrangian integrals. In Appendix A, Eq. (14) of Ref. 5 is related to the comparable Wiener integral for a stationary Markov process. However, the main concern of this paper is the treatment of non-Markovian, continuous stochastic Lagrangian integrals. A realistic model for continuous, weakly inhomogeneous, stochastic media is constructed by inference from sound propagation studies[1,4,6-8] and turbulent diffusion studies[9-13] so that these integrals can be treated, and understood, analytically. In the phenomenon of turbulent particle diffusion, the Lagrangian autopath correlations of particle velocity fluctuations are stationary, but the analogous crosspath correlations are not *nearly stationary* (i.e., not mainly a function of the transit time difference $t_2 - t_1$ and not only a weak function of $t_2 + t_1$) because the paths wander apart too rapidly over a characteristic turbulent memory time. However, in steady-state sound propagation from a highly directional transducer through a continuous, weakly inhomogeneous medium, there exists a predominant direction of transit so that the paths wander apart only slightly over a characteristic memory interval. This permits the stochastic crosspath Lagrangian integrals which occur in such problems to be reformulated in such a manner that the effects of the curvilinear statistical inhomogeneities are minimized and the integrals asymptotically assume analytically convenient forms. In practice, stochastic crosspath Lagrangian integrals appear in sound propagation intensity calculations in continuous stochastic media; see, for example, Refs. 6-8 and Sec. VI.

It is convenient to have some measure of the rms inhomogeneity of the medium. Since this paper is restricted to statistically isotropic media,[14] let α represent the rms variation from a uniform medium. For example, α can be defined such that

$$\mu(\mathbf{x}) = 1 + \alpha n(\mathbf{x}), \tag{2}$$

where

$$\langle \mu(\mathbf{x}) \rangle = 1, \quad \langle n^2(\mathbf{x}) \rangle = 1, \tag{2'}$$

and

$$\alpha \ll 1 \tag{3}$$

denotes a *weakly inhomogeneous medium*.

In this paper (as in Ref. 5), the ensemble expectation of a bounded, continuous Lagrangian functional $F[\mathbf{X}(s, \xi)]$ will be expressed by $E\{F[\mathbf{X}(s, \xi)]\}$ in contrast to the ensemble expectation of the corresponding Eulerian function $F(\mathbf{x})$ which will be designated by $\langle F(\mathbf{x}) \rangle$. This notational distinction is advisable because the behavior of these two stochastic quantities differs in very significant ways. For example, contrast $E\{\mu[\mathbf{X}(s, \xi)]\}$, which can be represented by $E\{\mu(s)\}$, with $\langle \mu(\mathbf{x}) \rangle$. When it is stated that the continuous refractive index field $\mu(\mathbf{x})$ is statistically homogeneous (in the Eulerian sense), it is implied that for any two different points $\mathbf{x}_1, \mathbf{x}_2$, including the arbitrary origin 0 of the coordinate system,

$$\langle \mu(\mathbf{x}_1) \rangle = \langle \mu(\mathbf{x}_2) \rangle = \langle \mu(\mathbf{0}) \rangle, \tag{4}$$

i.e., there exists no intrinsic origin or spacial reference point in the sense that $\langle \mu(\mathbf{x}) \rangle$ behaves the same at all points $\mathbf{x}$. On the other hand, $\mu(s)$ is a bounded, continuous, differentiable functional of the particular continuous, differentiable path $\mathbf{X}(s, \xi)$ of arc length s from ξ which can terminate at a different point $\mathbf{X} = \mathbf{X}(s, \xi)$ for each field realization $\mu_\beta(\mathbf{x}) \in \{\mu_\beta\}$. Therefore, for each $s \in (0, \infty)$, $\mu(s)$ has a volume spread (over all $\mu_\beta \in \{\mu_\beta\}$) of terminal locations $\mathbf{X}$ and the total ensemble of Eulerian fields $\{\mu_\beta(\mathbf{x})\}$ generates an ensemble of paths $\{\mathbf{X}_\beta(s, \xi)\}$ for each pair s, ξ. Thence, while (see Sec. V)

$$E\{\mu(s = 0)\} = E\{\mu(0)\} = E\{\mu(\xi)\} = \langle \mu(\xi) \rangle, \tag{4'}$$

so that $E\{\mu(0)\}$ is still Eulerian, the Lagrangian ensemble expectation $E\{\mu(s > 0)\}$ encompasses a continuum of terminal points and need not equal $E\{\mu(s = 0)\}$. Thus, in general, it is not expected that Lagrangian ensemble expectations will behave like their analogous Eulerian ensemble expectations because of the Lagrangian spreading of the terminal position. Still it might be hoped, that over a statistically homogeneous field, as s becomes long enough $E\{\mu(s)\}$ would not continue to increase, decrease, or vacillate but would settle down, after some initial adjustments, to a constant value. In fact, however, it shall be shown in Sec. IV that, in the case of Fermat paths, $E\{\mu(s)\}$ evolves continuously and does not reach an asymptotic value even if the refractive field is statistically isotropic (it does, however, obtain an asymptotic form). This occurs because the Fermat paths are continually seeking regions of lower sound velocity c so that, since $\mu \propto 1/c$, $E\{\mu(s > 0, \alpha > 0\} > E\{\mu(0)\} = \langle \mu(\xi) \rangle$ and increasingly so as s and/or α increase. However, some additional stochastic Lagrangian concepts must first be developed before $E\{\mu(s)\}$ can be evaluated. It was shown in Ref. 5 that consideration of the Lagrangian spreading and its implications, as well as the nature of the path $\mathbf{X}(s, \xi)$, can often serve as a suitable bridge from the known, *or assumed*, Eulerian statistical concepts to the unknown Lagrangian statistical concepts.[1] However, the full significance of this spreading will only be revealed gradually throughout this paper.

I. THE ASYMPTOTIC EVALUATION OF STOCHASTIC AUTOPATH LAGRANGIAN INTEGRALS

Let $F[\mathbf{X}(s_1', \xi_1)]$, which can be represented by $F(s_1')$, be an arbitrary bounded, continuous Lagrangian functional of the continuous path $\mathbf{X}(s_1', \xi_1)$ of arc length s_1' from a point ξ_1 on the bounded, continuous initial surface S_0 and let $G[\mathbf{X}(s_2', \xi_2)]$, which can be represented by $G(s_2')$, be an arbitrary bounded, continuous Lagrangian functional of the continuous path $\mathbf{X}(s_2', \xi_2)$ of arc length s_2' from a point ξ_2 on the same initial surface S_0. Each realization $\mu_\beta(\mathbf{x})$ generates a pair of such paths and the paths coincide for $s_1' = s_2'$ if and only if $\xi_1 = \xi_2$. It can be assumed without loss of generality that

$$E\{F[\mathbf{X}(s_1', \xi_1)]\} = 0 \tag{5a}$$

and

$$E\{G[\mathbf{X}(s_2', \xi_2)]\} = 0, \tag{5b}$$

since otherwise the deviation from a nonzero expectation may be considered.

There are two ways of viewing integral ensemble expectations like

$$E\left\{\int_0^{s_1} ds_1' \int_0^{s_2} ds_2' \, F(s_1')G(s_2')\right\}:$$

1. *The Lagrangian viewpoint:* Assume s_1, s_2 given and then $\mathbf{X}(s_1, \xi_1), \mathbf{X}(s_2, \xi_2)$ are stochastic, i.e., the arc lengths s_1, s_2 traveled are known but not the terminal points $\mathbf{X}(s_1, \xi_1), \mathbf{X}(s_2, \xi_2)$, which vary from realization to realization. This permits

$$E\left\{\int_0^{s_1} ds_1' \int_0^{s_2} ds_2' F(s_1')G(s_2')\right\} = \int_0^{s_1} ds_1' \int_0^{s_2} ds_2' E\{F(s_1')G(s_2')\}, \quad (6)$$

where $E\{F(s_1')G(s_2')\}$ has some convenient properties, which will be discussed later in this section, that permit treatment of this integral. The problems which arise from an unknown terminal location are resolved in Ref. 5

2. *The Eulerian viewpoint*: Assume $\mathbf{x}_1, \mathbf{x}_2$ given and then $s_1 = s(\mathbf{x}_1, \xi), s_2 = (\mathbf{x}_2, \xi_2)$ are stochastic, i.e., the terminal points $\mathbf{x}_1, \mathbf{x}_2$ are known but the arc lengths s_1, s_2 traveled vary from realization to realization (erogodicity is assumed); this requires

$$E\left\{\int_0^{s_1} ds_1' \int_0^{s_2} ds_2' F(s_1')G(s_2')\right\} = \int_0^{x_{1i}} dx_{1i}' \int_0^{x_{2j}} ds_{2j}' E\left\{\left(F\frac{ds_1}{dX_i}\right)[X(s_1', \xi_1)] \times \left(G\frac{ds_2}{dX_j}\right)[X(s_2', \xi_2)]\right\}, \quad (7)$$

no summation on i, j. Unfortunately, there seems to be no convenient method for resolving the integrand in Eq. (7) [see also the comments concerning Eq. (29) of Ref. 1].

Thus, the Lagrangian viewpoint is chosen since, in every case which arises in this study, it proves possible, in principle, to express and evaluate the stochastic Lagrangian functionals in terms of the stochastic field $\mu(s)$, its derivatives, and the initial values dX_i/ds (0). This obviates the difficulty, associated with the Eulerian viewpoint, of treating stochastic sinusoidal functionals dX_i/ds (s'), and even their derivatives, along the Lagrangian paths $\mathbf{X}(\mathbf{s}_1', \xi_1)$, $\mathbf{X}(\mathbf{s}_2', \xi_2)$ and not being able to express them conveniently in terms of the known field $\mu(\mathbf{x})$.

In the treatment of phenomena such as sound propagation through continuous stochastic media, the evaluation of integrals of the type given by Eq. (6) is often of primary importance.[5] The integrand of Eq. (6)

$$\hat{g}(s_1', s_2', \xi_1, \xi_2) \equiv E\{F[\mathbf{X}(s_1', \xi_1)]G[\mathbf{X}(s_2', \xi_2)]\} \quad (8)$$

is called a *Lagrangian crosspath correlation* when $\xi_1 \neq \xi_2$ (in general, a *caret* will be used to denote a crosspath quantity). Consider first the simpler case of $g(s_1', s_2', \xi)$, where $\xi_1 = \xi_2 = \xi$; $g(s_1', s_2', \xi)$ may be called a *Lagrangian autopath correlation* to distinguish it from the corresponding crosspath correlation. Statistical isotropy eliminates dependence on ξ and the Lagrangian spreading does not permit dependence on a definite terminal point for $s_1, s_2 > 0$, so that

$$g(s_1', s_2', \xi) \equiv E\{F[\mathbf{X}(s_1', \xi)]G[\mathbf{X}(s_2', \xi)]\} \quad (9)$$

$$= g(s_1', s_2') \quad (10a)$$

$$= g(\sigma, S), \quad (10b)$$

where Eq. (10b) follows from the coordinate transformation

$$\sigma = s_1' - s_2', \qquad S = s_1' + s_2'. \quad (11)$$

The advantage of the form $g(\sigma, S)$ is that σ pertains to the local properties of the correlation while $S/2$ serves as a convenient center-of-mass type coordinate. For a *long enough* curvilinear separation σ, $F[\mathbf{X}(s_1', \xi)]$ and $G[\mathbf{X}(s_2', \xi)]$ become uncorrelated so that

$$g(\sigma, S) \xrightarrow[\sigma \to \infty]{} 0 \quad (12)$$

in a boundless, weakly inhomogeneous medium [i.e., $g(\sigma, S)$ becomes *unrestrictedly weak* as σ increases]. Note that

$$g(\sigma, S) \geq \langle F(\mathbf{x}_1)G(\mathbf{x}_2)\rangle \quad (13)$$

for monotonically decreasing $g(\sigma, S)$ and $\langle F(\mathbf{x}_1)G(\mathbf{x}_2)\rangle$, when $\sigma \leq |x_1 - x_2|$ and $\alpha > 0$, due to the path curvature in almost all realizations. The larger α is in Eq. (2), the faster $F(s_1')$ and $G(s_2')$ become effectively uncorrelated with $|x_1 - x_2|$. The equality applies in Eq. (13) only in the nonstochastic limit $\alpha \to 0$.

Since the statistical characteristics of two functionals along any path should be unchanged in a statistically isotropic medium with the curvilinear distance traveled (this property may be referred to as *autopath stochastic invariance*), a Lagrangian autopath correlation should be *locally independent* of S, i.e.,

$$g(\sigma, S) = g(\sigma). \quad (14)$$

Furthermore, for *reversible* paths (in the sense that there exists no favored direction of travel along a path between two points, e.g., Fermat paths) it makes no difference which functional precedes the other. In other words, $s_1' > s_2'$ gives the same result as $s_2' > s_1'$, so that when μ is statistically isotropic,

$$g(\sigma) = g(-\sigma) \quad (15)$$

(this property may be referred to as *autopath correlation symmetry*). Note that, in the absence of path reversibility, Eq. (15) is true if and only if $F(s) \equiv G(s)$. Equations (14) and (15) represent statistical isotropy in the Lagrangian sense. Therefore, the Lagrangian stochastic integral of Eq. (6) can be evaluated as follows, in a statistically isotropic medium, when $s_1 = s_2 = s$ [see Eq. (40) for a method of evaluation when the upper limits are not equal] and $\xi_1 = \xi_2 = \xi$:

$$\int_0^s ds_1' \int_0^s ds_2' g(s_1', s_2') = \tfrac{1}{2}\int_0^s d\sigma \int_\sigma^{2s-\sigma} dS\, g(\sigma, S) + \tfrac{1}{2}\int_{-s}^0 d\sigma \int_{-\sigma}^{2s+\sigma} dS\, g(\sigma, S) \quad (16)$$

$$= 2\int_0^s d\sigma\, (s - \sigma)g(\sigma) \quad (17a)$$

$$= 2s\, g_0 \int_0^s d\sigma \left(1 - \frac{\sigma}{s}\right)\frac{g(\sigma)}{g_0} \quad (17b)$$

$$\sim 2s g_0 L_g. \quad (18)$$

Equation (17b) goes asymptotically to Eq. (18) via autopath stochastic invariance and $g(\sigma)$ unrestrictedly weak, where the Lagrangian integral scale L_g corresponding to the correlation $g(\sigma)$ is defined by

$$L_g \equiv \lim_{s\to\infty} \int_0^s d\sigma \left(1 - \frac{\sigma}{s}\right) \frac{g(\sigma)}{g_0} \tag{19}$$

and where the intensity factor g_0 corresponding to $g(\sigma)$ is defined by

$$g_0 \equiv E\{F(\xi)G(\xi)\} \tag{20a}$$

$$= \langle F(\xi)G(\xi)\rangle \tag{20b}$$

$$= g(0) = g(s_1' = s_2'), \tag{20c}$$

and normalizes $g(\sigma)$ in Eq. (19). Equation (77) justifies Eq. (20b) *for all* s *of physical interest*. The integral scale L_g is assumed to exist and cannot be a function of s. In Eq. (19), L_g is defined as a Cesaro-1 integral; for even more general definitions of integral scales, see Lumley.[15] In essence, L_g represents a measure of the curvilinear range of strong correlation between $F[\mathbf{X}(s_1', \xi)]$ and $G[\mathbf{X}(s_2', \xi)]$; when $\sigma \gg L_g$, they are essentially uncorrelated. Thus, L_g serves as a stochastic scaling factor for the *curvilinear memory* of $g(\sigma, S)$ and is particularly important in interpreting integrals involving $g(\sigma, S)$.

In practice, the result of Eq. (18) is *physically useful* if and only if this asymptotic form occurs for s small enough to satisfy the conditions imposed by the problem of interest. This empirical point will be developed further as this study evolves. Obviously the nature of the convergence assumed for Eq. (12) determines the class of integral scales L_g that are *physically useful* in Eq. (18), and especially the rate of convergence of Eq. (17b) to the asymptotic form of Eq. (18). Note that the faster $g(\sigma)/g_0$ declines to a negligible value, the smaller L_g becomes. However, the analogous Eulerian integral scale L_G,

$$L_G \equiv \int_0^\infty d|\mathbf{x}_1 - \mathbf{x}_2| \frac{\langle F(\mathbf{x}_1)G(\mathbf{x}_2)\rangle}{\langle F(\xi)G(\xi)\rangle} \leq L_g, \tag{21}$$

represents a lower bound for L_g; equality occurs only in the nonstochastic limit $\alpha \to 0$, where there exists no path curvature. Monin and Yaglom[11] observe (see Sec. 9.3) from empirical studies that large variations in the value of $\int_0^s ds_1' \int_0^s ds_2' g(s_1', s_2')$ for $\sigma = O(L_g)$, due to variations in the form of $g(s_1', s_2')$, occur only if $g(\sigma, S)$ is permitted to assume negative values and to change sign frequently with increasing σ and that when $g(\sigma, S)$ remains nonnegative for all σ, the dependence of $\int_0^s ds_1' \int_0^s ds_2' g(s_1', s_2')$ on the specific form of $g(\sigma, S)$ is very weak and the asymptotic form of Eq. (18) is rather well satisfied for all $\sigma \gtrsim 5L_g$. For a fluid mechanics analog of Eq. (18), see Eqs. (9.30)–(9.36) of Monin and Yaglom[11] [also consider Eq. (7.1.3) of Tennekes and Lumley[13]].

II. THE ASYMPTOTIC EVALUATION OF STOCHASTIC CROSSPATH LAGRANGIAN INTEGRALS

The asymptotic evaluation of integrals involving Lagrangian crosspath correlations is more complicated than for autopath correlations but is analogous.

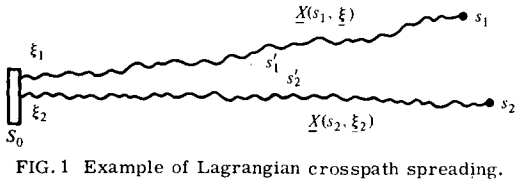

FIG. 1 Example of Lagrangian crosspath spreading.

When $\xi_1 \neq \xi_2$, statistical isotropy gives

$$\hat{g}(s_1', s_2', \xi_1, \xi_2) = \hat{g}(s_1', s_2', |\xi_1 - \xi_2|) = \hat{g}(s_1', s_2', \delta), \tag{22}$$

as can be seen by rotating and translating the source, with

$$\delta \equiv \xi_1 - \xi_2 \tag{23a}$$

and

$$\delta \equiv |\xi_1 - \xi_2|. \tag{23b}$$

Furthermore,

$$\hat{g}(s_1', s_2', \delta) < g(s_1', s_2'), \qquad \delta > 0, \tag{24a}$$

$$\xrightarrow[\delta\to 0]{} g(s_1', s_2') \tag{24b}$$

and

$$\hat{g}(s_1', s_2', \delta) \xrightarrow[\delta\to\infty]{} 0, \qquad s_1', s_2' \text{ finite;} \tag{25}$$

note that, in general, the limit $\delta \to 0$ reduces crosspath quantities to the corresponding autopath quantities. Again, in a boundless, weakly inhomogeneous medium,

$$\hat{g}(s_1', s_2', \delta) = \hat{g}(\sigma, S, \delta) \tag{26}$$

$$\xrightarrow[\sigma\to\infty]{} 0. \tag{27}$$

However, in contrast to the Lagrangian autopath correlation $g(\sigma, S)$, the Lagrangian crosspath correlation $\hat{g}(\sigma, S, \delta)$ cannot be exactly independent of S because the path spreading must affect the statistical characteristics of the product of the two Lagrangian functionals; see Fig. 1. However, if the path divergence is negligible over some suitable measure of the curvilinear range of significant correlation between $F[X(s_1', \xi_1)]$ and $G[X(s_2', \xi_2)]$, $\hat{g}(\sigma, S, \delta)$ should be at most a weak function of S [this property may be referred to as the *crosspath stochastic invariance approximation;* see Eq. (69) and Appendix B]. Likewise, since the paths from ξ_1 and ξ_2 are continually diverging, $\hat{g}(\sigma, S, \delta)$ will not, in general, equal $\hat{g}(-\sigma, S, \delta)$ unless the *mean path divergence* occurs so slowly over a region of significant correlation between $F[X(s_1', \xi_1)]$ and $G[X(s_2', \xi_2)]$ that $\hat{g}(\sigma, S, \delta)$ is essentially unaware of this gradual spreading of paths (this property may be referred to as the *crosspath correlation symmetry approximation*). Since the amount of mean path divergence can be kept small by making the inhomogeneity factor α as small as necessary [see Eq. (69)]; it will be assumed that the mean path divergence is negligible over a curvilinear range of several $\hat{L}_g(\delta)$, as defined by Eq. (33), *for all* s *of physical interest* so that

$$\hat{g}(\sigma, S, \delta) = \hat{g}(-\sigma, S, \delta) \tag{28}$$

$$\xrightarrow[\delta\to 0]{} g(\sigma), \tag{14'}$$

with $\hat{g}(\sigma, S, \delta)$ a weak function of S. Note that for reversible paths, the mean path divergence can be the only reason for the dissymmetry of the correlation $\hat{g}(\sigma, S, \delta)$ since, in the absence of spreading, there exists no means of inferring a distinction between $s_1' > s_2'$ and $s_2' > s_1'$ from any statistical characteristic when $\mu(\mathbf{x})$ is statistically isotropic.

Therefore, the integral of Eq. (6) can be evaluated as follows when $s_1 = s_2 = s$:

$$\int_0^s ds_1' \int_0^s ds_2' \, \hat{g}(s_1', s_2', \delta)$$

$$= \int_0^s d\sigma \int_\sigma^{2s-\sigma} dS \, \hat{g}(\sigma, S, \delta) \tag{29}$$

$$\equiv 2 \int_0^s d\sigma \, (s - \sigma) \bar{g}(\sigma, s, \delta) \tag{30a}$$

$$= 2s\bar{g}(0, s, \delta) \int_0^s d\sigma \, (1 - \sigma/s) \frac{\bar{g}(\sigma, s, \delta)}{\bar{g}(0, s, \delta)} \tag{30b}$$

$$\sim 2s\bar{g}_a(\delta)\hat{L}_g(\delta) \tag{31}$$

$$\xrightarrow{\delta \to 0} 2sg_0 L_g \tag{18'}$$

$$\xrightarrow{\delta \to \infty} 0, \quad s \text{ finite}, \tag{27'}$$

where the convenient definition

$$\bar{g}(\sigma, s, \delta) \equiv \frac{1}{s - \sigma} \frac{1}{2} \int_\sigma^{2s-\sigma} dS \, \hat{g}(\sigma, S, \delta) \tag{32}$$

$$\xrightarrow{\delta \to 0} g(\sigma) \tag{14''}$$

gives Eq. (30a) and Eq. (31) follows asymptotically via Eqs. (33) and (34). The new function $\bar{g}(\sigma, s, \delta)$ represents a sort of curvilinear average of $\hat{g}(\sigma, S, \delta)$. Since $\hat{g}(\sigma, S, \delta)$ is a weak function of S for sufficiently slowly diverging paths, it is expected that $\bar{g}(\sigma, s, \delta)$ is an even weaker function of s. The Lagrangian integral scale $\hat{L}_g(\delta)$ corresponding to $\bar{g}(\sigma, s, \delta)$, i.e.,

$$\hat{L}_g(\delta) \equiv \lim_{s \to \infty} \int_0^s d\sigma \left(1 - \frac{\sigma}{s}\right) \frac{\bar{g}(\sigma, s, \delta)}{\bar{g}(0, s, \delta)} \tag{33}$$

$$\xrightarrow{\delta \to 0} L_g \tag{19'}$$

is assumed to exist and cannot be a function of s. The intensity factor $\bar{g}(0, s, \delta)$, where

$$\bar{g}(0, s, \delta) = \lim_{\sigma \to 0} \bar{g}(\sigma, s, \delta) \tag{32'}$$

$$\sim \bar{g}_a(\delta) \tag{34}$$

$$\xrightarrow{s \to 0} \hat{g}_0(\delta) \equiv E\{F(\xi_1)G(\xi_2)\} \tag{35a}$$

$$= \langle F(\xi_1)G(\xi_2) \rangle \geq \bar{g}_a(\delta) \tag{35b}$$

$$\xrightarrow{\delta \to 0} g_0 \tag{20'}$$

$$\xrightarrow{\delta \to \infty} 0, \quad s \text{ finite}, \tag{27'}$$

serves as a running measure of intensity and produces the necessary normalized correlation $\bar{g}(\sigma, s, \delta)/\bar{g}(0, s, \delta)$ in Eq. (33). Equation (35a) follows in the same manner as Eq. (20b). Again, $\hat{L}_g(\delta)$ represents a curvilinear memory interval in the sense that when $F[\mathbf{X}(s_1', \xi_1)]$ and $G[\mathbf{X}(s_2', \xi_2)]$ are sufficiently separated in *almost all* realizations, they forget about each other [see Eq. (27)].

It would not be surprising to find [when $\bar{g}(\sigma, s, S)$ is unrestrictedly weak] that $\hat{L}_g(\delta)$ is largely independent of δ as well as of s *for all* s *of physical interest* because of the normalizing presence of $\bar{g}(0, s, \delta)$ in Eq. (33). Furthermore, since

$$\frac{\bar{g}(\sigma, s, \delta)}{\bar{g}(0, s, \delta)} \xrightarrow{\delta \to 0} \frac{g(\sigma)}{g_0}, \tag{36a}$$

$\hat{L}_g(\delta)$ is probably nearly equal to L_g *for all* s *of physical interest* and δ less than, say, $3L_g$, i.e.,

$$\hat{L}_g(\delta) \simeq L_g, \quad \delta < 3L_g. \tag{36b}$$

The main analytical restriction on the usefulness of Lagrangian crosspath asymptotic integral evaluations like Eq. (31) proves to be the δ dependence of the intensity factors. This difficulty can be obviated in practice by averaging δ over the surface S_0 of the acoustic source.[7]

The factor $\bar{g}_a(\delta)$ has been introduced into Eq. (31) to replace $\bar{g}(0, s, \delta)$ since, although $\bar{g}(0, s, \delta)$ is a weak function of s, it may not be a negligible function of s because of the cumulative (spreading) effects of S in Eq. (34) over large s. On the other hand, as discussed in Appendix B, $g(0, s, \delta)$ probably varies little over the s range of experimental interest even though it may be significantly less than $\hat{g}_0(\delta)$ of Eq. (35a). Therefore, $\bar{g}_a(\delta)$ represents the curvilinear average of $\bar{g}(0, s, \delta)$ over the s range of experimental interest.

This study will not be concerned with the satisfaction in particular experiments of explicit conditions for the asymptotic validity of relations like Eqs. (18) and (31), since the resulting integral scales are usually poorly known or not at all.[7,11,14,16] Instead, an empirical point of view will be adopted in which it will be assumed that relations like Eqs. (18) and (31) are valid, with the integral scales only negligibly dependent on s *for all* s *of physical interest*. Therefore, in practice, after carrying through all of the analysis, the final results must be compared with experimental results to determine the region of validity. At this juncture, it does not appear possible to devise a completely general method for determining the s dependence of Lagrangian integrals like Eq. (6).[10] In fact, a completely general Lagrangian method of attack on problems of the type considered in this study does not appear to be presently available.[7,9]

Equation (6) can now be considered in its full generality under the assumptions of the crosspath stochastic invariance and correlation symmetry approximations. Arbitrarily choose $s_2 \geq s_1$ and consider the situation illustrated in Fig. 1 for a typical realization μ_β with $\xi_1 \neq \xi_2$:

$$\int_0^{s_1} ds_1' \int_0^{s_2} ds_2' \, \hat{g}(s_1', s_2', \delta)$$

$$= \tfrac{1}{2} \int_0^{s_1} d\sigma \int_\sigma^{2s_1-\sigma} dS \, \hat{g}(\sigma, S, \delta) + \tfrac{1}{2} \int_0^{s_1} d\sigma \int_\sigma^{2s_2-\sigma} dS \, \hat{g}(\sigma, S, \delta) - \tfrac{1}{2} \int_0^{s_2-s_1} d\sigma \int_{2s_1+\sigma}^{2s_2-\sigma} dS \, \hat{g}(\sigma, S, \delta) \tag{37}$$

$$\xrightarrow{s_1 \to s_2 = s} \int_0^s d\sigma \int_\sigma^{2s-\sigma} dS \, \hat{g}(\sigma, S, \delta) \tag{29'}$$

$$\xrightarrow{\delta \to 0} 2 \int_0^s d\sigma \, (s - \sigma) g(\sigma) \tag{17'}$$

$$\equiv s_1\,\bar{g}(0,s_1,\delta)\int_0^{s_1} d\sigma\left(1-\frac{\sigma}{s_1}\right)\frac{\bar{g}(\sigma,s_1,\delta)}{\bar{g}(0,s_1,\delta)} + s_2\,\bar{g}(0,s_2,\delta)\int_0^{s_2} d\sigma\left(1-\frac{\sigma}{s_2}\right)\frac{\bar{g}(\sigma,s_2,\delta)}{\bar{g}(0,s_2,\delta)} - \int_0^{s_2-s_1} d\sigma\,(s_2-s_1-\sigma)\hat{g}_2(\sigma,s_1,s_2,\delta) \quad (38)$$

$$\sim s_1\,\bar{g}_a(\delta)\left[\hat{L}_g(\delta)+\int_0^{s_2-s_1} d\sigma\left(1+\frac{\sigma}{2s_1}\right)\frac{\hat{g}_2(\sigma,s_1,s_2,\delta)}{\bar{g}(0,s_1,\delta)}\right] + s_2\,\bar{g}_a(\delta)\left[\hat{L}_g(\delta)-\int_0^{s_2-s_1} d\sigma\left(1-\frac{\sigma}{2s_2}\right)\frac{\hat{g}_2(\sigma,s_1,s_2,\delta)}{\bar{g}(0,s_2,\delta)}\right] \quad (39)$$

$$\xrightarrow[s_1\to s_2=s]{} 2s\bar{g}_a(\delta)\hat{L}_g(\delta) \quad (31')$$

$$\xrightarrow[\delta\to 0]{} 2sg_0\hat{L}_g \quad (18'')$$

$$\equiv \bar{g}_a(\delta)[s_1\hat{L}_1(s_1,s_2,\delta)+s_2\hat{L}_2(s_1,s_2,\delta)]. \quad (40)$$

If this derivation is repeated with $s_1 \geq s_2$, the last term in Eq. (37) is replaced by

$$-\tfrac{1}{2}\int_0^{s_1-s_2} d\sigma\int_{2s_2+\sigma}^{2s_1-\sigma} dS\,\hat{g}(\sigma,S,\delta) = -\tfrac{1}{2}\int_{s_2-s_1}^{0} d\sigma\int_{2s_2-\sigma}^{2s_1+\sigma} dS\,\hat{g}(\sigma,S,\delta) = -\tfrac{1}{2}\int_0^{s_2-s_1} d\sigma\int_{2s_1+\sigma}^{2s_2-\sigma} dS\,\hat{g}(\sigma,S,\delta) \quad (41a)$$

$$\xrightarrow{\delta\to 0} \int_0^{s_2-s_1} d\sigma(s_2-s_1-\sigma)g(s) = 0 \quad (41b)$$

$$\xrightarrow[s_1\to s_2=s]{} 0. \quad (41c)$$

Equation (41a) shows that Eq. (37) is valid for all pairs $s_1, s_2 \in [0,\infty)$. The limit of Eq. (41c) shows the last term in Eq. (37) is a measure of the effect of the curvilinear length disparity $|s_2 - s_1|$. Equation (32) and

$$\hat{g}_2(\sigma,s_1,s_2,\delta) \equiv \frac{1}{(s_2-s_1-\sigma)}\,\tfrac{1}{2}\int_{2s_1+\sigma}^{2s_2-\sigma} dS\,\hat{g}(\sigma,S,\delta) \quad (42a)$$

$$\xrightarrow{\delta\to 0} g(\sigma), \quad (42b)$$

which should be only a very weak function of s_1 and of s_2, give Eq. (38) from which Eq. (33) yields Eq. (39). It is convenient to define

$$\hat{L}_1(s_1,s_2,\delta) \equiv \hat{L}_g(\delta)+\int_0^{s_2-s_1} d\sigma\left(1+\frac{\sigma}{2s_1}\right)\frac{\hat{g}_2(\sigma,s_1,s_2,\delta)}{\bar{g}(0,s_1,\delta)} \quad (43)$$

$$\not\equiv \hat{L}_2(s_1,s_2,\delta) \quad (44a)$$

$$\xrightarrow[s_1\to s_2=s]{} \hat{L}_g(\delta) \quad (45a)$$

$$\xrightarrow{\delta\to 0} L_g \quad (46)$$

and

$$\hat{L}_2(s_1,s_2,\delta) \equiv \hat{L}_g(\delta)-\int_0^{s_2-s_1} d\sigma\left(1-\frac{\sigma}{2s_2}\right)\frac{\hat{g}_2(\sigma,s_1,s_2,\delta)}{\bar{g}(0,s_2,\delta)} \quad (47)$$

$$\not\equiv \hat{L}_1(s_1,s_2,\delta) \quad (44b)$$

$$\xrightarrow[s_1\to s_2=s]{} \hat{L}_g(\delta) \quad (45b)$$

$$\xrightarrow{\delta\to 0} L_g. \quad (48)$$

Equations (43) and (47) give Eq. (40). Note that the limits $s_1 \to s_2 = s$ and/or $\delta \to 0$ reduce the new expressions to the earlier, less complicated results. When s_1, s_2 are both large, which is the only case of interest in this study,

$$\hat{L}_1(s_1,s_2,\delta) \simeq \hat{L}_g(\delta)+\int_0^{s_2-s_1} d\sigma\,\frac{\hat{g}_2(\sigma,s_1,s_2,\delta)}{\bar{g}(0,s_1,\delta)} \quad (49)$$

and

$$\hat{L}_2(s_1,s_2,\delta) \simeq \hat{L}_g(\delta)-\int_0^{s_2-s_1} d\sigma\,\frac{\hat{g}_2(\sigma,s_1,s_2,\delta)}{\bar{g}(0,s_2,\delta)}. \quad (50)$$

Since $|s_2 - s_1|$ can still be quite large, $\hat{L}_1(s_1,s_2,\delta)$ and $\hat{L}_2(s_1,s_2,\delta)$ can differ considerably; for example, when $s_2 - s_1 >$ several L_g, $\hat{L}_1(s_1,s_2,\delta) \approx 2L_g$ and $\hat{L}_2(s_1,s_2,\delta) \approx 0$ for δ small. Therefore, relations like Eq. (40) could present considerable analytical difficulty because the exact nature of the s_1, s_2 dependence cannot be discerned. Fortunately, a saddle-point evaluation of $\langle|p(\mathbf{x})|^2\rangle$ (via Ref. 5) results[7] in $s_1 = s_2 = s_0$, the saddle-point, and, consequently, the limit of Eq. (31') occurs.

III. SOME CONSEQUENCES OF STATISTICAL ISOTROPY

Appendix 4 of Lumley[15] contains a thorough discussion of invariant theory and the consequences of statistical isotropy. From this discussion, the following four theorems are adopted for the purpose of this study. Theorems Ia and IIa are straightforward, but Theorems Ib and IIb need some special consideration.

Theorem Ia: When $\mu(s_1'), \mu(s_2'), \ldots, \mu(s_3'), \mu_{,i}(s_4')$ are Lagrangian functionals with the arc lengths s_1', $s_2', \ldots, s_3', s_4'$ along the same Fermat path from the initial point $\boldsymbol{\xi}$ in each realization,

$$E\{\mu^m(s_1')\mu^n(s_2')\cdots\mu^l(s_3')\mu_{,i}(s_4')\} = 0, \quad (51)$$

where $m, n, \ldots, l$ are finite integers, and the refractive index field $\mu(\mathbf{x})$ is statistically isotropic. Likewise for all such tensor functionals of odd finite rank.

Only the case of

$$f_i(s_1', s_2') \equiv E\{\mu_{,i}[\mathbf{X}(s_1', \xi)]/\mu[\mathbf{X}(s_2', \xi)]\} = 0 \qquad (51')$$

need be considered; the more general situation can be demonstrated in an analogous manner. When $\mu(\mathbf{x})$ is a statistically isotropic field, the vector $f_i(s_1', s_2')$ can only depend on $s_1', s_2' \in [0, \infty)$. Let B_i be an arbitrary vector. Then $f_i B_i = g(B_i B_i) = 0$, since $f_i B_i$ must be linear in B_i and yet it can only be a function, denoted by $g(B_i B_i)$, of the only possible transformation invariant $B_i B_i$. Therefore, $f_i(s_1', s_2') = 0$. Equation (51'), and, in general, Eq. (51) can be viewed as a consequence of the Lagrangian spread of terminal locations when $s_1', s_2' > 0$ since no distinct terminal separation vector means that there can exist no such dependence.

Theorem Ib: When $\mu(s_1'), \mu(s_2'), \ldots, \mu(s_3'), \mu_{,i}(s_4')$ are Lagrangian functionals with arc lengths $s_1', s_2', \ldots, s_3', s_4'$ along different Fermat paths from a continuous initial surface S_0 in each realization and $\alpha > 0$,

$$E\{\mu^m(s_1')\mu^n(s_2') \cdots \mu^p(s_3')\mu_{,i}(s_4')\} \sim 0, \qquad (52)$$

where $m, n, \ldots, p$ are finite integers, and the refractive index field $\mu(\mathbf{x})$ is statistically isotropic. Likewise for all such tensor functionals of odd finite rank.

In contrast to Eq. (51'), consider

$$\hat{f}(s_1', s_2', \delta) \equiv E\{\mu_{,i}[\mathbf{X}(s_1', \xi_1')]/\mu[\mathbf{X}(s_2', \xi_2')]\} \qquad (52')$$

$$\xrightarrow{\delta \to 0} f_i(s_1', s_2') = 0 \qquad (51'')$$

$$\xrightarrow{\delta \to \infty} 0 \quad (s_1', s_2' \text{ finite}) \qquad (53)$$

$$\xrightarrow{s_1', s_2' \to 0} A_e(\delta)\delta_i. \qquad (52'')$$

Note that $\hat{f}_i(s_1', s_2', \delta)$ can depend on δ and, in fact, is given by Eq. (52'') for $s_1', s_2' \to 0$. In the case of turbulent particle diffusion,[10-13] the Lagrangian cross-path correlations of particle velocity fluctuations are a function of δ. However, in case of sound propagation, when all the paths from S_0 share a dominant initial direction, they leave the initial surface rapidly as they diffuse and their memory of δ fades. When their memory of the directionality of δ is lost due to the Lagrangian path spreading,

$$\hat{f}_i(s_1', s_2', \delta) \sim \hat{f}_i(s_1', s_2', \delta), \quad \alpha > 0, \qquad (54)$$

i.e., $\delta = |\boldsymbol{\delta}|$ may still be important. However, Eqs. (24a) and (51') give

$$0 \le \hat{f}_i(s_1', s_2', \delta) \le f_i(s_1', s_2') = 0,$$

so that

$$\hat{f}_i(s_1', s_2', \delta) \sim 0, \ \alpha > 0. \qquad (54')$$

Theorem IIa: When $\mu(s_1'), \ldots, \mu(s_2'), \mu_{,i}(s_3')$, $\mu_{,j}(s_4')$ are Lagrangian functionals with arc lengths $s_1', \ldots, s_2', s_3', s_4'$ along the same Fermat path from the initial point ξ in each realization,

$$E\{\mu^m(s_1') \cdots \mu^n(s_2')\mu_{,i}(s_3')\mu_{,j}(s_4')\} = B_0\delta_{ij} \not\equiv 0, \qquad (55)$$

where B_0 is a scalar function of $s_1', \ldots, s_2', s_3', s_4' \in [0, \infty)$ and δ; $m, \ldots, n$ are finite integers, and the refractive index field $\mu(\mathbf{x})$ is statistically isotropic.

Note that this theorem can be extended to include all such autopath Lagrangian functionals of finite even rank; for a rigorous treatment of this theorem and its extension, consult Lumley.[15]

Theorem IIb: When $\mu(s_1'), \ldots, \mu(s_2'), \mu_{,i}(s_3')$, $\mu_{,j}(s_4')$ are Lagrangian functionals with arc lengths $s_1', \ldots, s_2', s_3', s_4'$ along different Fermat paths from a continuous surface S_0 in each realization,

$$E\{\mu^m(s_1') \cdots \mu^n(s_2')\mu_{,i}(s_3')\mu_{,j}(s_4')\} \sim \hat{B}_0\delta_{ij} \not\equiv 0, \qquad (56)$$

where $\hat{B}_0$ is a scalar function of $s_1', \ldots, s_2', s_3', s_4' \in [0, \infty)$ and all the initial separations $\xi_{pq} \equiv |\xi_p - \xi_q|$, with $p, q = 1, \ldots, 2, 3, 4$; $m, \ldots, n$ are finite integers and the refractive field $\mu(\mathbf{x})$ is statistically isotropic.

This theorem follows by a direct extension of the previous methods.

IV. SOME STOCHASTIC LAGRANGIAN INTEGRALS ALONG FERMAT PATHS

Now that most of the basic concepts have been discussed, some representative Lagrangian ensemble expectations, which are dependent upon stochastic Lagrangian integrals, may be investigated analytically for the case of Fermat paths. Of particular significance are $E\{n(s)\}$ and $\Delta X_i(s, \xi)$ [see Eq. (66)]. This latter quantity represents a useful analytical measure of the Lagrangian spreading and proves to be in agreement with the discussion given in Sec. III of Ref. 5.

When $\mu(s) = \mu[\mathbf{X}(s, \xi)]$ is a functional of a Fermat path, i.e., $\mathbf{X}(s, \xi)$ is governed by[1,7,14]

$$\mu_{,i} = \frac{d}{ds}\left(\mu \frac{dX_i}{ds}\right) \qquad (57)$$

for all realizations of a statistically isotropic ensemble $\{\mu_\beta\}$, $E\{\mu(s)\}$ can be related to $E\{\mu(0)\}$ of Eq. (4') as follows:

$$E\{\mu(s) - \mu(0)\} = \alpha[E\{n(s)\} - E\{n(0)\}] = \alpha E\{n(s)\} \qquad (58a)$$

$$= E\left\{\int_0^s ds'\, \mu_{,i}(s') \frac{dX_i}{ds'}(s')\right\} \qquad (58b)$$

$$= \frac{dX_i}{ds}(0) \int_0^s ds' E\left\{\frac{\mu(0)}{\mu(s')}\, \mu_{,i}(s')\right\} + \alpha^2 \int_0^s ds' \int_0^{s'} ds'' E\left\{\frac{n_{,i}(s')n_{,i}(s'')}{\mu(s')}\right\} \qquad (58c)$$

$$= \alpha^2 \int_0^s ds' \int_0^{s'} ds'' E\{n_{,i}(s')n_{,i}(s'') - O(\alpha)\} \qquad (59a)$$

$$= \alpha^2 s R_0 \int_0^s d\sigma \left(1 - \frac{\sigma}{s}\right) \frac{R(\sigma)}{R_0} - O(\alpha^3 s)$$

$$\sim \alpha^2 s R_0 A \qquad (60a)$$

$$\xrightarrow{s \to \infty} \infty \qquad (60b)$$

$$\xrightarrow{\alpha \to 0} 0. \qquad (60c)$$

Equations (2') and (4') yield Eq. (58a); Eq. (59a) follows from expanding $[1 + \alpha n(s')]^{-1}$, with $0 < \alpha \ll 1$,

and from Eq. (51). Defining

$$R(\sigma) = R(s', s'') \equiv E\{n_{,i}(s')n_{,i}(s'')\} \tag{61}$$

via Eq. (14), gives Eq. (59b) which goes asymptotically to Eq. (60a), where the Lagrangian integral scale corresponding to the correlation $R(\sigma)$ is defined by

$$A \equiv \lim_{s\to\infty} \int_0^s d\sigma \; 1\left(-\frac{\sigma}{s}\right)\frac{R(\sigma)}{R_0} \;, \tag{62}$$

with the intensity factor

$$R_0 \equiv E\{n_{,i}(\xi)n_{,i}(\xi)\} \tag{63a}$$

$$= \langle n_{,i}(\xi)n_{,i}(\xi)\rangle ; \tag{63b}$$

A is assumed to exist and cannot be a function of s. The limit of Eq. (60b) is a direct consequence of unlimited Lagrangian terminal spreading over Fermat paths and Eq. (60c) results because, in the nonstochastic limit,

$$E\{\mu(s)\} \xrightarrow[\alpha\to 0]{} 1, \tag{64}$$

as required in a uniform medium. Note that

$$E\{\mu(s)\} = 1 + \alpha E\{n(s)\} \sim 1 = \langle\mu(\xi)\rangle = E\{\mu(0)\} \tag{65a}$$

if and only if

$$1 \gg \alpha^2 s R_0 A \geq 0. \tag{65b}$$

Equation (65b) is the condition for Lagrangian statistical homogeneity for this case in the sense that

$$E\{\mu(s_1)\} \simeq E\{\mu(s_2)\} \sim E\{\mu(0)\} \tag{65c}$$

is analogous to Eq. (4) when both s_1 and s_2 satisfy Eq. (65b).

The Lagrangian spread of Eulerian terminal points for Fermat paths can now be investigated in order to understand better how it affects Lagrangian ensemble expectations over Fermat paths. An appropriate measure of the Lagrangian spread is

$$[\Delta X_i(s,\xi)]^2 \equiv E\{[X_i(s,\xi) - E\{X_i(s,\xi)\}]^2\} \tag{66}$$

$$= E\left\{\left[\frac{dX_i}{ds}(0)\int_0^s ds'\left[\frac{\mu(0)}{\mu(s')} - E\left\{\frac{\mu(0)}{\mu(s')}\right\}\right] + \int_0^s ds'\int_0^{s'} ds''\,\frac{\mu_{,i}(s'')}{\mu(s')}\right]^2\right\}$$

$$= \int_0^s ds_1 \int_0^s ds_2 \left[E\left\{\int_0^{s_1} ds_1'\,\frac{\mu_{,i}(s_1')}{\mu(s_1)} \times \int_0^{s_2} ds_2'\,\frac{\mu_{,i}(s_2')}{\mu(s_2)}\right\} + \left(E\left\{\frac{\mu^2(0)}{\mu(s_1)\mu(s_2)}\right\} - E\left\{\frac{\mu(0)}{\mu(s_1)}\right\}E\left\{\frac{\mu(0)}{\mu(s_2)}\right\}\right) \times \frac{dX_i}{ds}(0)\frac{dX_i}{ds}(0)\right] \tag{67}$$

$$= \alpha^2 s^3 R_0 \left[\int_0^s d\sigma\left(\frac{2}{3} - \frac{\sigma}{s} - \frac{\sigma^3}{2s^3}\right)\frac{1}{3}\frac{R(\sigma)}{R_0} - O(\alpha s^3)\right]\left[1 + \frac{dX_i}{ds}(0)\frac{dX_i}{ds}(0)\right] \tag{68}$$

$$\sim \alpha^2 s^3 R_0 A_3 \left[1 + \frac{dX_i}{ds}(0)\frac{dX_i}{ds}(0)\right], \tag{69}$$

no summation on i, which represents the diagonal terms of the *Lagrangian spreading matrix*, Eq. (20) of Ref. 5, for Fermat paths. Equation (69) has a fluid mechanics analog in Richardson's "nearest-neighbor diffusion" (cf. Eq. 113 on page 48 of Corrsin[12]). Equation (67) follows from

$$X_i(s,\xi) = \xi_i + \int_0^s ds'\,\frac{dX_i}{ds}(s',\xi)$$

$$= \xi_i + \frac{dX_i}{ds}(0)\int_0^s ds'\,\frac{\mu(0)}{\mu(s')} + \int_0^s ds'\int_0^{s'} ds''\,\frac{\mu_{,i}(s'')}{\mu(s')} \tag{70}$$

and Eq. (51′). The expression

$$\frac{\mu(0)}{\mu(s)} = 1 - \frac{dX_i}{ds}(0)\int_0^s ds'\,\frac{\mu(0)\mu_{,i}(s')}{\mu(s)\mu(s')} - \int_0^s ds'\int_0^{s'} ds''\,\frac{\mu_{,i}(s')\mu_{,i}(s'')}{\mu(s)\mu(s')} \tag{71}$$

produces Eq. (68) which goes asymptotically to Eq. (69), where

$$A_3 \equiv \lim_{s\to\infty} \tfrac{2}{9}\int_0^s d\sigma\left(1 - \frac{3\sigma}{2s} + \frac{3\sigma^3}{4s^3}\right)\frac{R(\sigma)}{R_0} \tag{72}$$

is assumed to exist and cannot be a function of s.

V. THE LAGRANGIAN SUBENSEMBLE EXPECTATION

In Ref. 5, the *Lagrangian subensemble expectation* of $F(\mathbf{x})$,

$$\tilde{E}\{F[\mathbf{X}(s,\xi)]\} = E\{F[\mathbf{X}(s,\xi)]|\mathbf{X}(s,\xi) \approx \mathbf{x}\}, \tag{73a}$$

where the rigorous interpretation of $\mathbf{X}(s,\xi) \approx \mathbf{x}$ is $\mathbf{x} + d\mathbf{x} > \mathbf{X}(s,\xi) \geq \mathbf{x}$, was introduced. For the purposes of stochastic Lagrangian analysis, it would be convenient if integral ensemble expectations like Eq. (6) were analogous in their asymptotic behavior for full ensemble and subensemble expectations, even though the respective integral scales and intensity factors are not expected to be identical. Before investigating this hypothesis, it is important first to understand the essential difference between $\tilde{E}\{F[\mathbf{X}(s,\xi)]\}$ and $E\{F[\mathbf{X}(s,\xi)]\}$. In Secs. I and IV, it was emphasized that Lagrangian spreading distinguished the Lagrangian ensemble expectation $E\{F[\mathbf{X}(s,\xi)]\}$ from the Eulerian ensemble expectation $\langle F(\mathbf{x})\rangle$. However, by definition

$$\tilde{E}\{F[\mathbf{X}(s,\xi)]\} \equiv E\{F[\mathbf{X}(s,\xi)]|\mathbf{X}(s,\xi) \approx \mathbf{x}\} \tag{73a}$$

$$= E\{F(\mathbf{x})\} \tag{73b}$$

$$\neq E\{F[\mathbf{X}(s,\xi)]\}, \quad s,\alpha > 0, \tag{74}$$

$$\neq \langle F(\mathbf{x})\rangle = \int_0^\infty ds \int_{S_0} d\xi\, \tilde{E}\{F[\mathbf{X}(s,\xi_2)] \times B(\mathbf{X},\xi|s), \quad (75)$$

i.e., $\tilde{E}\{F[\mathbf{X}(s,\xi)]\}$ permits no spread $\Delta\mathbf{X}(s,\xi)$ of the terminal location [Eq. (73b)] and, therefore, it differs from $E\{F[\mathbf{X}(s,\xi)]\}$ [Eq. (74)]. Equation (75) is Eq. (13) of Ref. 5 and shows that $\tilde{E}\{F[\mathbf{X}(s,\xi)]\}$ is only "quasi-Eulerian" and does not equal $\langle F(\mathbf{x})\rangle$ [unless the joint probability density $B(\mathbf{x},\xi|s)$ acts like a δ function; see Sec. III of Ref. 5] because it includes only one value of the pair $s \in [0,\infty), \xi \in S_0$. In the case of $E\{F[\mathbf{X}(s,\xi)]|\mathbf{X}(s,\xi) \approx \xi\}$, it is reasonable to assume that, for statistically isotropic media, only $s = 0$, $\mathbf{X}(0,\xi) \approx \xi$ can contribute for almost all μ_β, i.e., the measure of the set of μ_β which produce paths that double back from S_0 to ξ is zero for $\alpha \ll 1$ and *for all s of physical interest*. Thence,

$$E\{F[\mathbf{X}(s,\xi)]|\mathbf{X}(s,\xi) \approx \xi\}$$

$$= E\{F[\mathbf{X}(0,\xi)]|\mathbf{X}(0,\xi) = \xi\} = E\{F(\xi)\} \quad (76a)$$

$$= \tilde{E}\{F[\mathbf{X}(0,\xi)]\} = \tilde{E}\{F(\xi)\} \quad (76b)$$

$$= \langle F(\xi)\rangle. \quad (77)$$

However, between the initial location $\mathbf{X}(0,\xi) = \xi$ and the terminal location $\mathbf{X}(s,\xi) \approx \mathbf{x}$,

$$\xi \neq \mathbf{X}(s',\xi) \neq \mathbf{x} \quad (78)$$

for almost all $\mu_\beta \in \{\mu_\beta\}_s$, where

$$\{\mu_\beta\}_s \equiv \{\mu_\beta|\mathbf{X}(s,\xi) \approx \mathbf{x}\} \subset \{\mu_\beta\} \quad (79)$$

and $0 < s' < s$. Actually, there exists a Lagrangian subensemble spread for each $s' \in (0,s)$, which may be represented by $\tilde{\Delta}\mathbf{X}(s',\xi)$ in direct analogy with $\Delta\mathbf{X}(s,\xi)$ of Eq. (66). (This is in direct contrast to Eulerian ensemble expectations where a spread never occurs.) The Lagrangian subensemble spread $\tilde{\Delta}\mathbf{X}(s',\xi)$ must observe the limits

$$\lim_{s'\to 0} \tilde{\Delta}\mathbf{X}(s',\xi) = \tilde{\Delta}\mathbf{X}(0,\xi) = 0 \quad (80)$$

and

$$\lim_{s'\to s} \tilde{\Delta}\mathbf{X}(s',\xi) = \tilde{\Delta}\mathbf{X}(s,\xi) = 0. \quad (81)$$

Thus, $\tilde{E}\{F[\mathbf{X}(s',\xi)]\}$ and $E\{F[\mathbf{X}(s',\xi)]\}$ behave alike as s' increases from zero because their spreading behavior is exactly similar; also $\tilde{E}\{F[\mathbf{X}(s',\xi)]\}$ behaves in this same manner as s' decreases from s. (See Figs. 2 and 3.) In fact, $\tilde{E}\{F[\mathbf{X}(s',\xi)]\}$ and $\tilde{E}\{F[\mathbf{X}(s-s',\xi)]\}$ exhibit a sort of spreadwise symmetry about $s' = s/2$. However, since $\{\mu_\beta\}_s \subset \{\mu_\beta\}$ and $\tilde{\Delta}\mathbf{X}(s,\xi) = 0$, while $\Delta\mathbf{X}(s,\xi) > 0$ for all $\alpha, s > 0$, $\tilde{E}\{F[\mathbf{X}(s',\xi)]\}$ must spread less rapidly as s' increases than $E\{F[\mathbf{X}(s',\xi)]\}$, i.e.,

$$\tilde{\Delta}\mathbf{X}(s',\xi) < \Delta\mathbf{X}(s',\xi), \quad \alpha > 0,\ 0 < s' \leq s. \quad (82)$$

The consequence of this should be that $\tilde{E}\{F[\mathbf{X}(s',\xi)]\}$ *settles down* to the same asymptotic form as $E\{F[\mathbf{X}(s',\xi)]\}$ at somewhat smaller values of s'. Also, there exists a random spread in the terminal angles at $\mathbf{x}$, which may be represented by $\tilde{\Delta}\theta(s,\xi)$, when $\tilde{\Delta}\mathbf{X}(s,\xi) = 0$, while the initial angle $\theta(\xi)$ is constant when $\tilde{\Delta}\mathbf{X}(0,\xi) = 0$; this distorts the symmetry of $\tilde{E}\{F[\mathbf{X}(s',\xi)]\}$ and $\tilde{E}\{F[\mathbf{X}(s-s',\xi)]\}$ about $s' = s/2$ and means that $\tilde{E}\{F[\mathbf{X}(s-s',\xi)]\}$ spreads more rapidly with increasing $s - s'$ than does $\tilde{E}\{F[\mathbf{X}(s',\xi)]\}$ with increasing s'. Thus, the continuity of $\tilde{E}\{F[\mathbf{X}(s',\xi)]\}$ guarantees that $\tilde{E}\{F[\mathbf{X}(s-s',\xi)]\}$ obtains the same asymptotic form for reversible paths as $\tilde{E}\{F[\mathbf{X}(s',\xi)]\}$ (although more slowly) if s is large enough. Therefore, $\tilde{E}\{F[\mathbf{X}(s',\xi)]\}$ should obtain the same asymptotic form as $E\{F[\mathbf{X}(s',\xi)]\}$, although more rapidly, and $\tilde{E}\{F[\mathbf{X}(s',\xi)]\}$ should retain this asymptotic form as long as it remains asymptotically far from both $s' = 0$ and $s' = s$. The behavior described above is illustrated in two dimensions in Fig. 2, where the bounding area for $\Delta\mathbf{X}(s,\xi)$ is given, and in Fig. 3, where the bounding area for $\tilde{\Delta}\mathbf{X}(s,\xi)$ is shown. Note that the whole question of the evaluation of $\tilde{E}\{F[\mathbf{X}(s',\xi)]\}$, as well as of $E\{F[\mathbf{X}(s,\xi)]\}$, has been reduced to the problem of determining the asymptotic form of the pertinent stochastic Lagrangian integrals that results from the Lagrangian spreading.

It will now be shown that the asymptotic evaluations of Lagrangian ensemble and subensemble stochastic integrals are exactly similar when s_1, s_2 are *large enough*. The subensemble analog of Eq. (6) is

$$\tilde{E}\left\{\int_0^{s_1} ds_1' \int_0^{s_2} ds_2' F(s_1')G(s_2')\right\}$$

$$\equiv E\left\{\int_0^{s_1} ds_1' \int_0^{s_2} ds_2' F[\mathbf{X}(s_1',\xi_1)] \times G[\mathbf{X}(s_2',\xi_2)]|\mathbf{X}(s_1,\xi_1) \approx \mathbf{x}, \mathbf{X}(s_2,\xi_2) \approx \mathbf{x}\right\} \quad (83)$$

$$= \int_0^{s_1} ds_1' \int_0^{s_2} ds_2'\, \tilde{g}(s_1',s_2',\xi_1,\xi_2,\mathbf{x}), \quad (84)$$

where the Lagrangian subensemble crosspath correlation

$$\tilde{g}(s_1',s_2',\xi_1,\xi_2,\mathbf{x}) \equiv E\{F[\mathbf{X}(s_1',\xi_1)]G[\mathbf{X}(s_2',\xi_2)]|\mathbf{X}(s_1,\xi_1) \approx \mathbf{x}, \mathbf{X}(s_2,\xi_2) \approx \mathbf{x}\} \quad (85)$$

should behave like the corresponding full ensemble

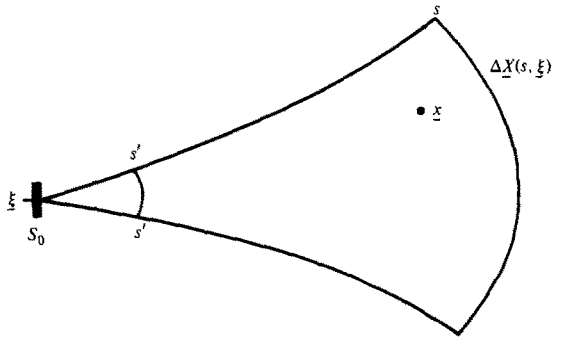

FIG. 2 Lagrangian autopath full ensemble spreading volume.

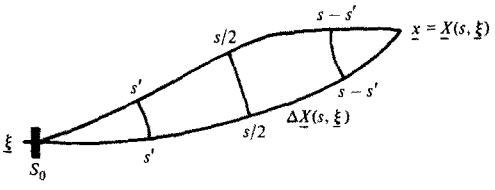

FIG. 3 Lagrangian autopath subensemble spreading volume.

correlation of Eq. (6), with respect to its curvilinear memory lapse behavior, when s_1' is asymptotically far from both $0, s_1$ and s_2' is asymptotically far from both $0, s_2$. This follows from the material of Sec. I and II by noting, as discussed above, that the Lagrangian ensemble and subensemble spreading is exactly similar. (See Figs. 4 and 5.) The methods of Secs. I and II will now be applied in detail to Eqs. (85) and (83). In general, a tilde will be used to denote a Lagrangian subensemble quantity and to distinguish it from the corresponding full ensemble quantity which is not necessarily identical in value; retaining the $\delta = |\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2|$ in the function argument indicates that it is also a crosspath stochastic quantity.

Consider first the simpler case of the Lagrangian subensemble autopath correlation $\tilde{g}(s_1', s_2', \boldsymbol{\xi}, \mathbf{x})$. Statistical homogeneity reduces this to $\tilde{g}(s_1', s_2', \boldsymbol{\xi} - \mathbf{x})$ and statistical isotropy further reduces this to $\tilde{g}(s_1', s_2', |\boldsymbol{\xi} - \mathbf{x}|)$. When s_1' is asymptotically far from both $0, s_1$ and s_2' is asymptotically from both $0, s_2$ [which requires $|\mathbf{x} - \boldsymbol{\xi}| \gg \tilde{L}_g$ of Eq. (95) when $0 < \alpha \ll 1$], $\tilde{g}(s_1', s_2', \boldsymbol{\xi}, \mathbf{x})$ reduces to $\tilde{g}(s_1', s_2')$. In short,

$$\tilde{g}(s_1', s_2', \boldsymbol{\xi}, \mathbf{x}) = \tilde{g}(s_1', s_2', \boldsymbol{\xi} - \mathbf{x}) \tag{86}$$

$$= \tilde{g}(s_1', s_2', |\boldsymbol{\xi} - \mathbf{x}|) \tag{87}$$

$$\sim \tilde{g}(s_1', s_2'). \tag{88}$$

In contrast to this, it should be observed that although $B(\mathbf{x}, \boldsymbol{\xi} | s)$ of Eq. (75) reduces to $B(\mathbf{x} - \boldsymbol{\xi} | s)$ for statistical homogeneity, statistical isotropy does not give $B(|\mathbf{x} - \boldsymbol{\xi}| \, | s)$ because the orientation of $\boldsymbol{\xi} - \mathbf{x}$ with respect to the initial path direction strongly affects the magnitude of $B(\mathbf{x} - \boldsymbol{\xi} | s)$; for example, as $\boldsymbol{\xi} - \mathbf{x}$ is rotated away from the initial path direction with $|\boldsymbol{\xi} - \mathbf{x}|$ kept constant, $B(\mathbf{x} - \boldsymbol{\xi} | s)$ declines in value in a statistically isotropic medium since the measure of $\{\mu_\beta\}_s$ decreases. However, statistical isotropy does permit considerable simplification in the asymptotic form of $B(\mathbf{x} - \boldsymbol{\xi} | s)$; see Sec. III of Ref. 5.

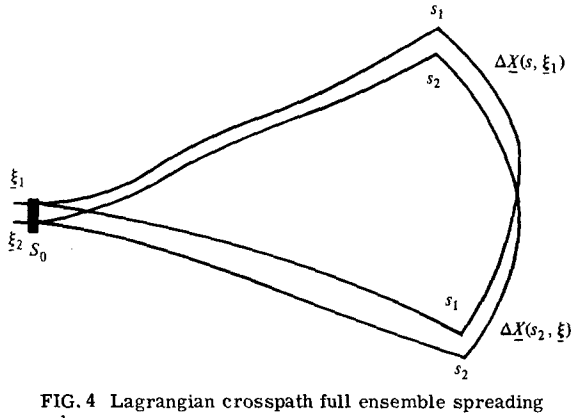

FIG. 4 Lagrangian crosspath full ensemble spreading volume.

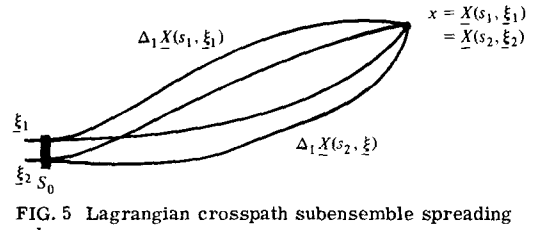

FIG. 5 Lagrangian crosspath subensemble spreading volume.

Assuming autopath stochastic invariance, for the subensemble $\{\mu_\beta\}_s$ in a statistically isotropic medium, and reversible paths gives

$$\tilde{g}(s_1', s_2', \boldsymbol{\xi}, \mathbf{x}) = \tilde{g}(\sigma, S, \boldsymbol{\xi}, \mathbf{x}) \tag{89}$$

$$= \tilde{g}(\sigma, \boldsymbol{\xi}, \mathbf{x}) \tag{90}$$

$$= \tilde{g}(-\sigma, \boldsymbol{\xi}, \mathbf{x}) \tag{91}$$

$$\xrightarrow[\sigma \to \infty]{} 0 \tag{92}$$

in direct analogy with Eqs. (10b), (14), (15), and (12), respectively. Therefore, when $s_1 = s_2 = s$ and $\boldsymbol{\xi}_1 = \boldsymbol{\xi}_2 = \boldsymbol{\xi}$, Eq. (84) can be evaluated in the same manner as Eq. (18):

$$\int_0^s ds_1' \int_0^s ds_2' \, \tilde{g}(s_1', s_2', \boldsymbol{\xi}, \mathbf{x}) = 2s\tilde{g}_0 \int_0^s d\sigma \left(1 - \frac{\sigma}{s}\right) \frac{\tilde{g}(\sigma, \boldsymbol{\xi}, \mathbf{x})}{\tilde{g}_0} \tag{93}$$

$$\sim 2s\tilde{g}_0 \tilde{L}_g, \tag{94}$$

where the Lagrangian integral scale $\tilde{L}_g$, corresponding to the correlation $g(\sigma)$, is defined by

$$\tilde{L}_g \equiv \lim_{s \to \infty} \int_0^s d\sigma \left(1 - \frac{\sigma}{s}\right) \frac{\tilde{g}(\sigma)}{\tilde{g}_0}, \tag{95}$$

considering Eq. (88), and where the intensity factor is

$$\tilde{g}_0 \equiv \tilde{E}\{F[\mathbf{X}(0, \boldsymbol{\xi})] G[\mathbf{X}(0, \boldsymbol{\xi})]\} \tag{96}$$

$$= \langle F(\boldsymbol{\xi}) G(\boldsymbol{\xi}) \rangle = g_0 \tag{97}$$

via Eqs. (20b) and (77); $\tilde{L}_g$ is assumed to exist and cannot be a function of s.

When $\boldsymbol{\xi}_1 \neq \boldsymbol{\xi}_2$, statistical isotropy gives

$$\tilde{g}(s_1', s_2', \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x}) = \tilde{g}(s_1', s_2', \boldsymbol{\delta}, \boldsymbol{\xi}_1 - \mathbf{x}, \boldsymbol{\xi}_2 - \mathbf{x}) \tag{98}$$

$$= \tilde{g}(s_1', s_2', \delta, |\boldsymbol{\xi}_1 - \mathbf{x}|, |\boldsymbol{\xi}_2 - \mathbf{x}|) \tag{99}$$

$$\xrightarrow[\delta \to 0]{} g(s_1', s_2', \boldsymbol{\xi}, \mathbf{x}) \tag{100}$$

$$\xrightarrow[\delta \to \infty]{} 0, \quad (s_1', s_2' \text{ finite}) \tag{101}$$

$$\sim \tilde{g}(s_1', s_2', \delta) \tag{102}$$

[cf. Eqs. (22)–(25)]. Assume that the path divergence is negligible over a curvilinear range of several $\tilde{L}_g(\delta)$ [this Lagrangian integral is defined in Eq. (116)] for all s of interest so that

$$\tilde{g}(s_1', s_2', \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x}) = \tilde{g}(\sigma, S, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x}) \tag{103}$$

$$= \tilde{g}(-\sigma, S, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x}) \tag{104}$$

$$\xrightarrow[\delta \to 0]{} \tilde{g}(\sigma, \boldsymbol{\xi}, \mathbf{x}) \tag{90'}$$

$$\xrightarrow[\sigma \to \infty]{} 0, \tag{105}$$

and $\tilde{g}(\sigma, S, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x})$ is a weak function of S, in direct analogy with Eqs. (26), (28), (14′), and (27), respectively. Therefore, Eq. (84) can be considered in its full generality by the same procedure that gave Eq. (40):

$$\int_0^{s_1} ds_1' \int_0^{s_2} ds_2' \, \tilde{g}(s_1', s_2', \xi_1, \xi_2, \mathbf{x})$$

$$\xrightarrow[s_1 \to s_2 = s]{} \int_0^s d\sigma \int_\sigma^{2s-\sigma} dS \, \tilde{g}(\sigma, S, \xi_1, \xi_2, \mathbf{x}) \tag{106}$$

$$\xrightarrow[\delta \to 0]{} 2 \int d\sigma \, (s - \sigma) \tilde{g}(\sigma, \xi_1, \mathbf{x}) \tag{93'}$$

$$\equiv s_1 \tilde{g}_3(0, s_1, \xi_1, \xi_2, \mathbf{x}) \int_0^{s_1} d\sigma \left(1 - \frac{\sigma}{s_1}\right) \times \frac{\tilde{g}_3(\sigma, s_1, \xi_1, \xi_2, \mathbf{x})}{\tilde{g}_3(0, s_1, \xi_1, \xi_2, \mathbf{x})} + s_2 \tilde{g}_3(0, s_2, \xi_1, \xi_2, \mathbf{x}) \int_0^{s_2} d\sigma \left(1 - \frac{\sigma}{s_2}\right) \times \frac{\tilde{g}_3(\sigma, s_2, \xi_1, \xi_2, \mathbf{x})}{\tilde{g}_3(0, s_2, \xi_1, \xi_2, \mathbf{x})} - \int_0^{s_2 - s_1} d\sigma \, (s_2 - s_1 - \sigma) \tilde{g}_4(\sigma, s_1, s_2, \xi_1, \xi_2, \mathbf{x}) \tag{107}$$

$$\sim s_1 \tilde{g}_a(\delta) \left(\tilde{L}_g(\delta) + \int_0^{s_2 - s_1} d\sigma \, \frac{\tilde{g}_4(\sigma, s_1, s_2, \delta)}{\tilde{g}_3(0, s_1, \delta)} \right) + s_2 \tilde{g}_a(\delta) \left(\tilde{L}_g(\delta) - \int_0^{s_2 - s_1} d\sigma \, \frac{\tilde{g}_4(\sigma, s_1, s_2, \delta)}{\tilde{g}_3(0, s_2, \delta)} \right) \tag{108}$$

$$\xrightarrow[s_1 \to s_2 = s]{} 2s \tilde{g}_a(\delta) \tilde{L}_g(\delta) \tag{109}$$

$$\xrightarrow[\delta \to 0]{} 2s \tilde{g}_0 \tilde{L}_g \tag{110}$$

$$\equiv \tilde{g}_a(\delta) [s_1 \tilde{L}_1(s_1, s_2, \delta) + s_2 \tilde{L}(s_1, s_2, \delta)]. \tag{111}$$

The convenient definitions

$$\tilde{g}_3(\sigma, s, \xi_1, \xi_2, \mathbf{x}) \equiv \frac{1}{s - \sigma} \frac{1}{2} \int_\sigma^{2s-\sigma} dS \, \tilde{g}(\sigma, S, \xi_1, \xi_2, \mathbf{x}) \tag{112}$$

$$\sim \tilde{g}_3(\sigma, s, \delta) \tag{113a}$$

$$\xrightarrow[\delta \to 0]{} \tilde{g}(\sigma, \xi, \mathbf{x}) \tag{113b}$$

(these should be only a very weak function of s) and

$$\tilde{g}_4(\sigma, s_1, s_2, \xi_1, \xi_2, \mathbf{x}) \equiv \frac{1}{(s_2 - s_1 - \sigma)} \times \tfrac{1}{2} \int_{2s_1 + \sigma}^{2s_2 - \sigma} dS \, \tilde{g}(\sigma, S, \xi_1, \xi_2, \mathbf{x}) \tag{114}$$

$$\sim \tilde{g}_4(\sigma, s_1, s_2, \delta) \tag{115a}$$

$$\xrightarrow[\delta \to 0]{} \tilde{g}(\sigma, \xi, \mathbf{x}) \tag{115b}$$

(these should be only a very weak function of s_1 and s_2) give Eq. (107) which goes asymptotically to Eq. (108), where the Lagrangian integral scale $\tilde{L}_g(\delta)$ corresponding to the correlation $\tilde{g}_3(\sigma, s, \xi_1, \xi_2, \mathbf{x})$, i.e.,

$$\tilde{L}_g(\delta) \equiv \lim_{s \to \infty} \int_0^s d\sigma \left(1 - \frac{\sigma}{s}\right) \frac{\tilde{g}_3(\sigma, s, \delta)}{\tilde{g}_3(0, s, \delta)} \tag{116}$$

$$\xrightarrow[\delta \to 0]{} \tilde{L}_g, \tag{95'}$$

is assumed to exist, and cannot be a function of s, and where the intensity factor is

$$\tilde{g}_3(0, s, \delta) = \lim_{\sigma \to 0} \tilde{g}_3(\sigma, s, \delta) \sim \tilde{g}_a(\delta); \tag{117}$$

$\tilde{g}_a(\delta)$ represents the curvilinear average of $\tilde{g}_3(0, s, \delta)$ over the s range of experimental interest and is analogous to $\bar{g}_a(\delta)$ of Eq. (34). Note that the intensity factor at $s = 0$ is

$$\tilde{g}_0(\delta) \equiv \tilde{E}\{F[\mathbf{X}(0, \xi_1)] G[\mathbf{X}(0, \xi_2)]\} \tag{118}$$

$$= \langle F(\xi_1) G(\xi_2) \rangle \gtrsim g_a(\delta) \tag{119}$$

$$\xrightarrow[\delta \to 0]{} \tilde{g}_0 \tag{96'}$$

$$\xrightarrow[\delta \to \infty]{} 0. \tag{120}$$

Because of the normalizing presence of $\tilde{g}_3(0, s, \delta)$ in Eq. (116), it would not be surprising to find that $\tilde{L}_g(\delta)$ is largely independent of δ, as well as of s, for all s of interest. Furthermore, since

$$\frac{\tilde{g}_3(\sigma, s, \delta)}{\tilde{g}_3(0, s, \delta)} \xrightarrow[\delta \to 0]{} \frac{\tilde{g}(\sigma)}{\tilde{g}_0}, \tag{121a}$$

$\tilde{L}_g(\delta)$ is probably nearly equal to $\tilde{L}_g$ *for all* s *of physical interest* and δ less than, say, $3\tilde{L}_g$, i.e.,

$$\tilde{L}(\delta) \simeq \tilde{L}_g, \quad \delta < 3\tilde{L}_g. \tag{121b}$$

It is convenient to define

$$\tilde{L}_1(s_1, s_2, \delta) \equiv \tilde{L}_g(\delta) + \int_0^{s_2 - s_1} d\sigma \, \frac{\tilde{g}_4(\sigma, s_1, s_2, \delta)}{\tilde{g}_3(0, s_1, \delta)} \tag{122}$$

$$\not\equiv \tilde{L}_2(s_1, s_2, \delta) \tag{123}$$

$$\xrightarrow[s_1 \to s_2 = s]{} \tilde{L}_g(\delta) \tag{116'a}$$

$$\xrightarrow[\delta \to 0]{} \tilde{L}_g \tag{124}$$

and

$$\tilde{L}_2(s_1, s_2, \delta) \equiv \tilde{L}_g(\delta) - \int_0^{s_2 - s_1} d\sigma \, \frac{\tilde{g}_4(\sigma, s_1, s_2, \delta)}{\tilde{g}_3(0, s_2, \delta)} \tag{125}$$

$$\not\equiv L_1(s_1, s_2, \delta) \tag{123}$$

$$\xrightarrow[s_1 \to s_2 = s]{} \tilde{L}_g(\delta) \tag{116'b}$$

$$\xrightarrow[\delta \to 0]{} \tilde{L}_g. \tag{126}$$

Equations (122) and (125) give Eq. (111) which is analogous to Eq. (40).

Furthermore, when $s_1', \ldots, s_4'$ are all asymptotically far from 0 and from $s_1, \ldots, s_4$, respectively, of $\mathbf{X}(s_1, \xi_1) \approx \mathbf{x}, \ldots, \mathbf{X}(s_4, \xi_4) \approx \mathbf{x}$, theorems Ia–IIb of Sec. III should be asymptotically valid for the Lagrangian subensemble expectation. Likewise, the central limit theorem for stochastic Lagrangian functionals that is cited in Sec. II of Ref. 5 can be assumed valid for Lagrangian subensemble expectations in the same sense that it was adopted for full ensemble expectations. From the preceding discussion, it is apparent that the results of Sec. IV are also valid for Lagrangian subensemble expectations.

It proves interesting to contrast the behavior of $\tilde{E}\{(dX_i/ds)(s)\}$ and $\langle (dX_i/ds)(x) \rangle$. Consider a straight line extending in the initial direction, as indicated by $(dX_i/ds)(0)$, from a point ξ on the source S_0; all the paths that cross this line at each point $\mathbf{x}$ contribute to $\langle (dX_i/ds)(\mathbf{x}) \rangle = 1$, where μ is statistically isotropic. In contrast to this, Eqs. (57) and (51') yield

$$\tilde{E}\left\{\frac{dX_i}{ds}(s)\right\} = \tilde{E}\left\{\frac{\mu(0)}{\mu(s)}\right\} \frac{dX_i}{ds}(0) \tag{127a}$$

$$= \frac{dX_i}{ds}(0)[1 - \alpha^2 s\tilde{R}_0\tilde{A}] \tag{127b}$$

$$\xrightarrow[s\to 0]{} \frac{dX_i}{ds}(0) \tag{128}$$

$$\xrightarrow[\alpha\to 0]{} \frac{dX_i}{ds}(\mathbf{x}), \tag{129}$$

where Eqs. (71), (60a), and (62) have been employed to obtain Eq. (127b). In the first situation, $\langle(dX_i/ds)(\mathbf{x})\rangle$ represents the Eulerian ensemble expectation of $(dX_i/ds)(\mathbf{x})$ at each point $\mathbf{x}$ on the prescribed path; the result of unity implies that there exists no net angular bias at any such $\mathbf{x}$ when averaging over $\{\mu_\beta\}$. Equation (75) shows that many paths from the initial surface can contribute, at some arc length s, to $\langle(dX_i/ds)(\mathbf{x})\rangle$. In the second situation

$$\tilde{E}\left\{\frac{dX_i}{ds}(s)\right\} \equiv E\left\{\frac{dX_i}{ds}(s)\,\middle|\,\mathbf{X}(s,\boldsymbol{\xi}) \approx \mathbf{x}\right\} \tag{130}$$

represents the approximate contribution at $\mathbf{x}$ of all the paths of arc length s that reach $\mathbf{x}$ reach from $\boldsymbol{\xi} \in S_0$; from Eq. (75), it is seen that the weighed summation of all the $\tilde{E}\{(dX_i/ds)(s)\}$ for all $s \in [0,\infty)$ and for all $\boldsymbol{\xi} \in S_0$ give $\langle(dX_i/ds)(\mathbf{x})\rangle$. Due to the Lagrangian spreading, $\tilde{E}\{(dX_i/ds)(s)\}$ becomes increasingly smaller than $(dX_i/ds)(0)$ as s increases since $\tilde{E}\{(dX_i/ds)(s)$ represents the mean x_i-axis length per unit arc length for $\{\mu_\beta\,|\,\mathbf{X}(s,\boldsymbol{\xi}) \approx \mathbf{x}\}$. In the nonstochastic limit, the medium becomes uniform and Eq. (129) shows that the paths $\mathbf{X}(s,\boldsymbol{\xi})$ become straight lines. As an interesting contrast to Eq. (127b), Eq. (57) yields

$$E\left\{\mu(s)\frac{dX_i}{ds}(s)\right\} = \frac{dX_i}{ds}(0)E\{\mu(0)\} = \frac{dX_i}{ds}(0), \tag{131}$$

via Eq. (4'), which means that $\mu dX_i/ds$ is a Fermat invariant under the operation of taking the Lagrangian ensemble expectation.

VI. SOME FERMAT EXAMPLES OF SUBENSEMBLE INTENSITY CALCULATIONS

At this point, it is appropriate to consider some examples of subensemble intensity calculations. Consider the Lagrangian sound-pressure wave over a continuous $\mu(\mathbf{x})$[1,7,14]:

$$p[\mathbf{X}(s,\boldsymbol{\xi})] = p_0(\boldsymbol{\xi})\exp\left(-\tfrac{1}{2}\int_0^s \frac{ds'}{\mu(s')} \times \int_0^{s'} ds''\mu_{,ii}(s'') + ik_0\int_0^s ds'\mu(s')\right). \tag{132}$$

Case (1): Assume that the phase factor dominates in Eq. (132). Then[7]

$$p[\mathbf{X}(s,\boldsymbol{\xi})] = p(s) \simeq p_0(\boldsymbol{\xi})\exp\left(ik_0\int_0^s ds_2'\,\mu(s_2')\right) \tag{133}$$

and

$$\tilde{E}\{p(s_1)p^*(s_2)\} \simeq |p_0|^2 E\left\{\exp\left(ik_0\int_0^{s_1} ds_1'\,\mu(s_1') - ik_0\int_0^{s_2} ds_2'\,\mu(s_2')\right)\right\} \tag{134}$$

$$\sim |E\{p(s_0)\}|^2\exp[2\alpha^2k_0^2s_0\tilde{P}_a(\delta)\tilde{H}(\delta)] \tag{135}$$

$$\xrightarrow[\alpha\to 0]{} |p_0|^2 \tag{136}$$

(when $s_1 = s_2 = s_0$), where

$$\tilde{P}(s_1',s_2',\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\mathbf{x}) \equiv \tilde{E}\{[n(s_1') - \tilde{E}\{n(s_1')\}] \times [n(s_2') - \tilde{E}\{n(s_2')\}]\} \tag{137}$$

$$= \tilde{P}(\sigma,S,\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\mathbf{x}), \tag{138}$$

$$\int_0^{s_0} ds_1' \int_0^{s_0} ds_2'\,\tilde{P}(s_1',s_2',\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\mathbf{x}) = 2s_0\tilde{P}_0(0,s_0,\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\mathbf{x})\int_0^{s_0} d\sigma\,\frac{1}{2s_0} \times \int_\sigma^{2s_0-\sigma} dS\,\frac{\tilde{P}(\sigma,S,\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\mathbf{x})}{\tilde{P}_0(0,s_0,\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\mathbf{x})} \tag{139}$$

$$\sim 2s_0\tilde{P}_a(\delta)\tilde{H}(\delta) \tag{140}$$

and

$$\tilde{P}_0(0,s_0,\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\mathbf{x}) \equiv \frac{1}{2s_0}\int_0^{2s_0} dS\,\tilde{P}(0,S,\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\mathbf{x}) \tag{141}$$

$$\sim \tilde{P}_0(0,s_0,\delta) \geq \tilde{P}_a(\delta). \tag{142}$$

Equation (139) goes asymptotically to Eq. (140), where the Lagrangian subensemble crosspath phase integral scale $\tilde{H}(\delta)$,

$$\tilde{H}(\delta) \equiv \lim_{s_0\to\infty}\int_0^{s_0} d\sigma\,\frac{1}{2s_0}\int_\sigma^{2s_0-\sigma} dS\,\frac{\tilde{P}(\sigma,S,\delta)}{\tilde{P}_0(0,s_0,\delta)}, \tag{143}$$

is assumed to exist and cannot be a function of s_0 and where $\tilde{P}_a(\delta)$ represents the curvilinear average of $\tilde{P}_0(0,s_0,\delta)$ over the s_0 range of experimental interest. Therefore, the coefficient of intensity variation $\tilde{V}_p$ can be considered

$$\tilde{V}_p^2 \equiv \frac{\tilde{E}\{|p(s_0)|^2\} - |E\{p(s_0)\}|^2}{|E\{p(s_0)\}|^2} \tag{144}$$

$$\sim \exp[2\alpha^2k_0^2s_0\tilde{P}_a(\delta)\tilde{H}(\delta)] - 1 \tag{145}$$

$$\simeq 2\alpha^2k_0^2s_0\tilde{P}_a(\delta)\tilde{H}(\delta) \tag{146}$$

for

$$1 \gg 2\alpha^2k_0^2s_0\tilde{P}_a(\delta)\tilde{H}(\delta). \tag{147}$$

Equation (146) reduces, in the Eulerian limit, to result of Mintzer,[8]

$$V^2 \simeq 2\alpha^2k_0^2r\int_0^\infty d\rho_1 N(\rho_1), \tag{148}$$

where

$$N(\rho) = \langle n(\mathbf{x})n(\mathbf{x}+\boldsymbol{\rho})\rangle, \tag{149}$$

$$\rho = (\rho_1^2 + \rho_2^2 + \rho_3^2)^{1/2} \tag{150}$$

is the separation of the points $\mathbf{x}$ and $\mathbf{x}+\boldsymbol{\rho}$, ρ_1 is taken along the line between the source and the receiver, and r is the source-to-receiver separation. The one-dimensional correlation function $N(\rho_1)$ is the Eulerian equivalent of $\tilde{P}(\sigma,S,\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\mathbf{x})$ in the sense that when $\boldsymbol{\xi}_1 \to \boldsymbol{\xi}_2$ and the Lagrangian spreading (see Fig. 5) is neglected, $\tilde{P}(\sigma,S,\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\mathbf{x})$ reduces to $N(\rho_1)$ with $s_0 = r$. Furthermore, in the Eulerian limit, Eq. (147) reduces to

$$1 \gg 2\alpha^2k_0^2r\int_0^\infty d\rho_1 N(\rho_1), \tag{151}$$

which is analogous to Mintzer's[17] validity condition for Eq. (148).[1] In Eq. (146), $\tilde{P}_a(\delta)\tilde{H}(\delta)$ replaces $\int_0^\infty d\rho_1 N(\rho_1)$ of Eq. (148). Although these factors are analogous, it should be observed that Eq. (148) is strictly Eulerian, since $n(\mathbf{x})$ and $n(\mathbf{x}+\boldsymbol{\rho})$ are along the same known source-to-receiver path in every realization μ_β. In addition, Mintzer[8] derived Eq. (148) using a

single-scattering assumption which does not consider the continuous path fluctuations due to the continuous variation in $\mu(\mathbf{x})$. However, in one dimension, the Debye (continuous fluctuation) and Born (single-scattering) pressure wave relations coincide when the cumulative phase effects are small.[1,7]

Case (2): Assume that amplitude factor dominates in Eq. (132). Then[7]

$$p[\mathbf{X}(s,\boldsymbol{\xi})] = p(s) \simeq p_0(\boldsymbol{\xi}) \exp\left(-\tfrac{1}{2}\int_0^s \frac{ds'}{\mu(s')}\int_0^{s'} ds''\mu_{,ii}(s'')\right) \tag{152}$$

and

$$\tilde{E}\{p(s_1)p^*(s_2)\} \simeq |p_0|^2 E\left\{\exp\left(-\tfrac{1}{2}\int_0^{s_1} ds_1'\int_0^{s_1'} ds_1''\,\mu_{,ii}(s_1'') - \tfrac{1}{2}\int_0^{s_2} ds_2'\int_0^{s_2'} ds_2''\,\mu_{,ii}(s_2'')\right)\right\} \tag{153}$$

$$\sim |E\{p(s_0)\}|^2 \exp[2\alpha^2 s_0^3 \tilde{Q}_a(\delta)\tilde{J}(\delta)/12] \tag{154}$$

$$\xrightarrow[\alpha\to 0]{} |p_0|^2 \tag{155}$$

(when $s_1 = s_2 = s_0$), where

$$\mu_{,ii}(s'')/\mu(s') \doteq \alpha n_{,ii}(s'') \tag{156a}$$

via Eq. (3),

$$\tilde{Q}(s_1'', s_2'', \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x}) \equiv \tilde{E}\{[n_{,ii}(s_1'') - \tilde{E}\{n_{,ii}(s_1'')\}] \times [n_{,ii}(s_2'') - \tilde{E}\{n_{,ii}(s_2'')\}]\} \tag{156b}$$

$$= \tilde{Q}(\sigma, S, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x}), \tag{157}$$

$$\int_0^{s_0} ds_1'\int_0^{s_1'} ds_1''\int_0^{s_0} ds_2\int_0^{s_2'} ds_2''\,\tilde{Q}(s_1'', s_2'', \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x}) = \int_0^{s_0} ds_1''(s_0 - s_1'')\int_0^{s_0} ds_2''\,(s_0 - s_2'') \times \tilde{Q}(s_1'', s_2'', \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x}) \tag{158}$$

$$\doteq s_0^3 \tilde{Q}_3(0, s_0, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x})\tfrac{2}{3}\int_0^{s_0} d\sigma \times\left(1 - \frac{\sigma}{2s_0} - \frac{\sigma^2}{2s_0^2} + \frac{3\sigma^2}{4s_0^3}\right)\frac{\tilde{Q}_3(\sigma, s_0, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x})}{\tilde{Q}_3(0, s_0, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x})} \sim \tfrac{2}{3} s_0^3 \tilde{Q}_a(\delta)\tilde{J}(\delta), \tag{159}$$

and

$$[\tfrac{2}{3}s_0^3 - \tfrac{1}{3}\sigma s_0^2 - \tfrac{1}{3}\sigma^2 s_0 + \tfrac{1}{2}\sigma^3]\tilde{Q}_3(\sigma, s_0, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x}) \equiv \int_\sigma^{2s_0-\sigma} dS[s - \tfrac{1}{2}(S+\sigma)] \times [s - \tfrac{1}{2}(S-\sigma)]\tilde{Q}(\sigma, S, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x}) \tag{160}$$

$$\sim \tfrac{2}{3} s_0^3 \tilde{Q}_3(\sigma, s_0, \delta). \tag{161}$$

Equation (158) goes asymptotically to Eq. (159), where the Lagrangian subensemble crosspath amplitude integral scale $\tilde{J}(\delta)$,

$$\tilde{J}(\delta) \equiv \lim_{s_0\to\infty}\int_0^{s_0} d\sigma\left(1 - \frac{\sigma}{2s_0} - \frac{\sigma^2}{2s_0^2} + \frac{3\sigma^2}{4s_0^3}\right)\frac{\tilde{Q}_3(\sigma, s_0, \delta)}{\tilde{Q}_3(0, s_0, \delta)}, \tag{162}$$

is assumed to exist and cannot be a function of s_0 and where $\tilde{Q}_a(\delta)$ represents the curvilinear average of $\tilde{Q}_3(0, s_0, \delta)$ over the s_0 range of experimental interest. Therefore, the coefficient of intensity variation $\tilde{V}_A$ can be considered:

$$\tilde{V}_A^2 \equiv \frac{\tilde{E}\{|p_0(s_0)|^2\} - |E\{p(s_0)\}|^2}{|E\{p(s_0)\}|^2} \tag{163}$$

$$\sim \exp[2\alpha^2 s_0^3 Q_a(\delta)\tilde{J}(\delta)/12] - 1 \tag{164}$$

$$\simeq 2\alpha^2 s_0^3[\tilde{Q}_a(\delta)\tilde{J}(\delta)/12] \tag{165}$$

for

$$1 \gg 2\alpha^2 s_0^3[\tilde{Q}_a(\delta)\tilde{J}(\delta)/12]. \tag{166}$$

Equation (165) reduces, in the Eulerian limit, to

$$V^2 \simeq 2\alpha^2 r^3\,\tfrac{1}{120}\int_0^\infty d\rho_1[\nabla^2\nabla^2 N(\rho)]_{\rho=\rho_1}, \tag{167}$$

which was derived by Bergmann[6] from the eikonal and transport equations[7] via a variational approach. The one-dimensional correlation function $[\nabla^2\nabla^2 N(\rho)]_{\rho_1}$ is the Eulerian equivalent of $\tilde{Q}(\sigma, S, \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \mathbf{x})$.

VII. CONCLUSION

Thus, the stochastic tools for employing Lagrangian ensemble and subensemble expectations are, in practice, the same. The basic stochastic concepts and assumptions utilized in this study for both Lagrangian subensemble and full ensemble expectations can now be summarized as follows:

(i) The central limit theorem for stochastic Lagrangian functionals from Sec. II of Ref. 5;

(ii) Eq. (14) of Ref. 5 in which $B(x_i - \xi_i|s)$ and $\tilde{E}\{F[\mathbf{X}(s,\boldsymbol{\xi})]\}$ are expressed asymptotically in terms of Lagrangian full ensemble and subensemble, respectively, stochastic integrals via the central limit theorem;

(iii) the consequences of statistical isotropy as discussed in Sec. III and extended in Sec. V, i.e., Theorems Ia–IIb and their extension;

(iv) the asymptotic evaluation of Lagrangian stochastic integrals as developed in Secs. I, II, and V, in particular, Eqs. (40) and (111).

APPENDIX A: THE WIENER INTEGRAL

Since the completion of Ref. 5, the first author has found that Eq. (14) of Ref. 5 reduces in a natural way to the corresponding *Wiener integral*[4,18,19] for a stationary Markov process. Equations (14), (19a), (20) and (27) of Ref. 5 give

$$\langle F(\mathbf{x})\rangle = \int_0^\infty ds\int_{S_0} d\boldsymbol{\xi}\; B(x_i - \xi_i|s)\tilde{E}\{F[\mathbf{X}(s,\boldsymbol{\xi})]\}, \tag{A1}$$

where

$$B(x_i - \xi_i|s) = \prod_{i=1}^{3}\frac{\exp[-\tfrac{1}{2}(x_i - \xi_i - m_i)^2/U_{ii}]}{(2\pi U_{ii})^{1/2}}, \tag{A2}$$

$$U_{ii} \equiv E\{[X_i(s,\boldsymbol{\xi}) - E\{X_i(s,\boldsymbol{\xi})\}][X_i(s,\boldsymbol{\xi}) - E\{X_i(s,\boldsymbol{\xi})\}]\}, \tag{A3}$$

and

$$m_i \equiv E\{X_i(s,\boldsymbol{\xi})\} \tag{A4}$$

(no summation on ii), for a statistically isotropic medium when the Lagrangian spreading matrix $[U_{ij}]$ is expressed in terms of its principal axes. Assume a single initial point $\boldsymbol{\eta} = (\xi, \eta, \zeta)$ so that

$$\langle F(\mathbf{x})\rangle = \int_0^\infty ds\, p(x_i - \eta_i \mid s)\tilde{E}\{F[\mathbf{X}(s,\eta)]\}$$
$$= E\left\{\int_0^\infty ds\, p(x_i - \eta_i \mid s)F[\mathbf{X}(s,\eta)]\delta[\mathbf{X}(s,\eta) - \mathbf{x}]\right\},$$

where (A5)

$$B(x_i - \xi_i \mid s) \equiv p(x_i - \eta_i \mid s)\delta(\xi_i - \eta_i) \quad \text{(A6)}$$

and $\delta[\mathbf{X}(s,\eta) - \mathbf{x}]$ represents the Lagrangian subensemble condition. Define

$$\int_0^\infty F[\mathbf{X}(s,\eta)]d_{p(\mathbf{x})}s$$
$$\equiv \int_0^\infty F[\mathbf{X}(s,\eta)]\delta[\mathbf{X}(s,\eta) - \mathbf{x}]p(x_i - \eta_i \mid s)\, ds \quad \text{(A7)}$$
$$= \int_0^\infty F[\mathbf{X}(s,\eta)]\delta[\mathbf{X}(s,\eta) - \mathbf{x}]$$
$$\times \prod_{i=1}^{3} \frac{\exp[-\frac{1}{2}(x_i - \eta_i - m_i)^2/U_{ii}]}{(2\pi U_{ii})^{1/2}} ds$$

via Eq. (A2), so that

$$\langle F(\mathbf{x})\rangle = E\left\{\int_0^\infty F[\mathbf{X}(s,\eta)]d_{p(\mathbf{x})}s\right\}. \quad \text{(A9)}$$

When the continuous path $\mathbf{X}(s,\eta)$ becomes a stationary Markov processes, it takes on a three-dimensional *zigzag* appearance and is not necessarily differentiable for each $\mathbf{X}(s,\eta) \in (-\infty,\infty)$. Assume that there exists *complete randomness* with respect to changes in direction (e.g., Brownian motion under the influence of random impulses in the absence of any kind of systematic forces and neglecting any particle inertia), so that $m_i = 0$. The velocity $c(s)$ along $\mathbf{X}(s,\eta)$ is then constant between the abrupt changes in path direction, but changes randomly and discontinuously in magnitude with path direction. For such a random process, it is conceptually easier to first consider Eq. (A8) in one dimension and then generalize the result to three dimensions. Even then it is not possible to merely integrate the one-dimensional form of Eq. (A8) in one step since $X(s)$ is not known. However, the motion from η to x can be paramatrized by the transit time t which is discretized such that

$$0 = t_0 < t_1 < t_2 < \cdots < t_n = t \quad \text{(A10a)}$$

(where the points $t_1, t_2, \ldots, t_{n-1}$ divide the time interval $[0,t]$ into n equal parts of length $\Delta t = t/n$), so that the curve $X(t)$ satisfies

$$\eta = x(t_0) \equiv x_0, \quad x(t_1) \equiv x_1, \quad \ldots, \quad x(t_{n-1}) \equiv x_{n-1},$$
$$x(t_n) \equiv x_n = x, \quad \text{(A10b)}$$

where each x_i can take on any value $(-\infty,\infty)$. It is, therefore, natural to assume that U_{ii} of Eq. (A3) is proportional to the transit time interval involved, i.e.,

$$U_{11}(t_i - t_{i-1}) = \tfrac{1}{2}D(t_i - t_{i-1}). \quad \text{(A11)}$$

Since Eq. (A8) will now yield a Wiener process [see Eqs. (1.7) and (1.4) of Ref. 18], $4D$ represents a diffusion coefficient and it is convenient to choose a system of units such that $D = 1$. Thus, as the number of *steps* n in going from η to x is increased, Eq. (A8) yields, successively, in one dimension

$$F(x,\eta)\exp[-(x-\eta)^2/t]/(\pi t)^{1/2},$$
$$\int_{-\infty}^{\infty} dx_1 F(x,x_1,\eta)\frac{\exp[-(x-x_1)^2/(t-t_1) - (x_1-\eta)^2/(t_1-t_0)]}{[\pi^2(t-t_1)(t_1-t_0)]^{1/2}},$$
$$\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 F(x,x_2,x_1,\eta)$$
$$\times \frac{\exp[-(x-x_2)^2/(t-t_2) - (x_2-x_1)^2/(t_2-t_1) - (x_1-\eta)^2/(t_1-t_0)]}{[\pi^3(t-t_2)(t_2-t_1)(t_1-t_0)]^{1/2}},$$
$$\vdots$$
$$\int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_{n-1} F(x,x_{n-1},\ldots,x_1,\eta)\prod_{i=1}^{n}\frac{\exp[-(x_i - x_{i-1})^2/(t_i - t_{i-1})]}{[\pi(t_i - t_{i-1})]^{1/2}}. \quad \text{(A12)}$$

In the limit of $n \to \infty$, Eqs. (A12) reduces to the *conditional Wiener* integral of Ref. 18: $\int_{c_{t,x}} F[x(t)]d_{w(t,x)}x$ and $\mathbf{X}(s,\eta)$ becomes a Wiener process. Note that Eq. (1.7) of Ref. 18 gives the Lagrangian spreading matrix for a Wiener process.

APPENDIX B: COMMENTS CONCERNING $\bar{g}_a(\delta)$

The factor $\bar{g}_a(\delta)$ has been introduced in Eq. (31) to replace $\bar{g}(0,s,\delta)$, which may not be a negligible function of s because of cumulative spreading effects over large s. However, $\bar{g}(0,s,\delta)$ may vary insignificantly over the s range of experimental interest [even though it may be much less than $\hat{g}_0(\delta)$ of Eq. (35a)].[7] Therefore, $\bar{g}_a(\delta)$ represents the curvilinear average value of $\bar{g}(0,s,\delta)$ over the s range of experimental interest. As was pointed out in Sec. II, the s dependence of $\bar{g}(0,s,\delta)$ can result only from the Lagrangian path spreading. The general situation is illustrated in Fig. 6, where the outer boundaries of the ensemble expected region of spreading, as represented by $\Delta\mathbf{X}(s,\boldsymbol{\xi}_1)$ and $\Delta\mathbf{X}(s,\boldsymbol{\xi}_2)$ from Eq. (66), and where the two paths for a typical realization are shown in two dimensions. When all the paths are initially perpendicular to the surface S_0, Eq. (69) gives the traverse Lagrangian spreading to be

$$[\Delta X_t(s)]^2 \sim \alpha^2 s^3 A_3 R_0 \quad \text{(B1)}$$

for Fermat paths and, on the average, $F(s)$ and $G(s)$ are separated by $\delta = |\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2|$. Even then, $\bar{g}(0,s,\delta)$ decreases as $s_1 = s = s_2$ increases since $F(s)$ and $G(s)$ are on different, and generally diverging, paths in each realization and, therefore, tend to be found

farther apart than δ by, on the average, the amount $\Delta X_t(s)$. Thence,

$$\bar{g}(0,s,\delta) \approx \bar{g}[0,0,\delta + \Delta X_t(s)]. \tag{B2}$$

This is in direct contrast with the analogous autopath situation where $g(\sigma)$ is independent of $\Delta X_t(s)$; note that

$$\bar{g}_a(\delta) \xrightarrow{\delta \to 0} g_0. \tag{20'}$$

For simplicity, consider the $\Delta X_t(s)$ dependence of the running intensity factor $\hat{g}(0,S,\delta) = \hat{g}(0,2s,\delta)$ for $s_1 = s_2 = s$; if this is reasonably insensitive to $\Delta X_t(s)$ over a given s range, then, by Eq. (32), $\bar{g}(0,s,\delta)$ is even more insensitive. The scaling factor[3] for the resultant decrease in correlation (due to path spreading) with increase in $\Delta X_t(s)$ is the Eulerian integral scale L_G of Eq. (21) so that

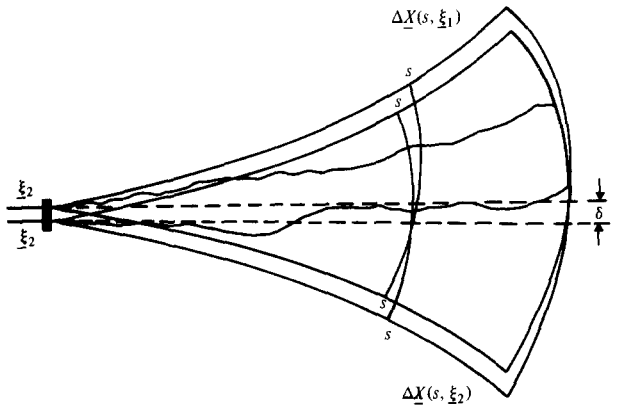

FIG. 6 Lagrangian crosspath spreading volume.

$$\begin{aligned}\hat{g}(0,s_{II},\delta) &\approx \hat{g}[0,0,\delta + \Delta X_t(s_{II})] \\ &= \hat{g}(0,0,[\delta + \Delta X_t(s_I)] \\ &\quad + [\Delta X_t(s_{II}) - \Delta X_t(s_I)]) \end{aligned} \tag{B3}$$

$$\approx \hat{g}[0,0,\delta + \Delta X_t(s_I)] \times \left[1 - \frac{\Delta X_t(s_{II}) - \Delta X_t(s_I)}{L_G}\right] \tag{B4}$$

$$\approx \hat{g}(0,s_I,\delta)\left[1 - \frac{\Delta X_t(s_{II}) - \Delta X_t(s_I)}{L_G}\right], \tag{B5}$$

where $0 \ll s_I < s_{II}$. Thus,

$$\frac{\Delta X_t(s_{II}) - \Delta X_t(s_I)}{L_G} = \frac{\hat{g}(0,s_I,\delta) - \hat{g}(0,s_{II},\delta)}{\hat{g}(0,s_I,\delta)} \tag{B6}$$

gives a measure of the relative decrease in correlation as s increases from s_I to s_{II} and shows that if the mean spreading (relative to L_G) is small, $\hat{g}(0,s,\delta)$ varies slightly with $s \in [s_I, s_{II}]$.

---

* Taken from the doctoral dissertation of J. A. Neubert, "Sound Propagation in Continuous Stochastic Media," under the Interdisciplinary Program in Engineering Acoustics, The Pennsylvania State University, University Park, Pa., 1970.

1 J. A. Neubert, J. Acoust. Soc. Am. **48**, 1203 (1970).
2 J. D. Cole, *Perturbation Methods in Applied Mathematics* (Blaisdell, Waltham, Mass., 1968), Chap. 3.
3 J. A. Neubert and J. L. Lumley, J. Acoust. Soc. Am. **48**, 1212 (1970).
4 U. Frisch, *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968).
5 J. A. Neubert, J. Math. Phys. **12**, 1030 (1971).
6 P. G. Bergmann, Phys. Rev. **70**, 486 (1946).
7 J. A. Neubert, "Sound Propagation in Continuous Stochastic Media," Ph.D. Thesis, Dept. of Engineering Acoustics, The Pennsylvania State University at University Park, 1970 (unpublished).
8 D. Mintzer, J. Acoust. Soc. Am. **25**, 922 (1953).
9 J. L. Lumley, J. Math. Phys. **3**, 309 (1962).
10 A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics* (MIT Press, Cambridge, Mass., 1971), Chap. 5.
11 J. L. Lumley, in *Mécanique de la turbulence*, Colloques Internationaux du Centre National de la Recherche Scientifique No. 108 (CNRS, Paris, 1962), pp. 18–26.
12 S. Corrsin, in Ref. 11, pp. 27–52.
13 H. Tennekes and J. L. Lumley, *A First Course in Turbulence* (MIT Press, Cambridge, 1971), Chap. 7.
14 The extension of the analytical methods of this paper to a practical problem in a statistically anisotropic medium is illustrated in J. A. Neubert, "Multipath Summability in Ray Theory Intensity Calculations in the Real Ocean," J. Acoust. Soc. Am. (to be published).
15 J. L. Lumley, *Stochastic Tools in Turbulence* (Academic, New York, 1970).
16 J. L. Lumley and H. A. Panofsky, *The Structure of Atmospheric Turbulence* (Wiley, New York, 1964).
17 D. Mintzer, J. Acoust. Soc. Am. **25**, 1107 (1953).
18 I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. **1**, 48 (1960).
19 N. Wiener, *Nonlinear Problems in Random Theory* (Wiley, New York, 1958).

$SU(n-2) \times SU(2) \times U(1)$ Bases for $SU(n)$*

R. T. Sharp

Department of Physics, McGill University, Montreal, Canada

(Received 9 August 1971)

Explicit basis dates for the general IR of $SU(n)$ are defined in the scheme $SU(n-2) \times SU(2) \times U(1)$; the approach makes use of elementary multiplets.

1. INTRODUCTION

Explicit basis states for $SU(n)$, reduced according to the subgroup $SU(n-1) \times U(1)$, were constructed many years ago by Gel'fand and Zetlin[1]; the generator matrix elements for these states were written down in Ref. 1, and derivations were later given by Baird and Biedenharn[2] and by Nagel and Moshinsky.[3]

In recent years there has been some interest in the reduction $SU(n) \supset SU(n-2) \times SU(2) \times U(1)$ especially for the case $n = 6$, which is the $SU(6)$ group of particle physics reduced according to the supermultiplet scheme.[4] Perelomov, Popov, and Malkin[5] give a method based on Young patterns for determining the $SU(4) \times SU(2) \times U(1)$ content of an IR of $SU(6)$; Hagen and Macfarlane[6] show how to determine the $SU(m) \times SU(n) \times U(1)$ content of an IR of $SU(m+n)$ by expanding the character function for the IR as a sum of products of $SU(m)$ and $SU(n)$ characters; more recently Mickelsson[7] has achieved the same object by examin-

farther apart than δ by, on the average, the amount $\Delta X_t(s)$. Thence,

$$\bar{g}(0,s,\delta) \approx \bar{g}[0,0,\delta + \Delta X_t(s)]. \tag{B2}$$

This is in direct contrast with the analogous autopath situation where $g(\sigma)$ is independent of $\Delta X_t(s)$; note that

$$\bar{g}_a(\delta) \xrightarrow{\delta \to 0} g_0. \tag{20'}$$

For simplicity, consider the $\Delta X_t(s)$ dependence of the running intensity factor $\hat{g}(0,S,\delta) = \hat{g}(0,2s,\delta)$ for $s_1 = s_2 = s$; if this is reasonably insensitive to $\Delta X_t(s)$ over a given s range, then, by Eq. (32), $\bar{g}(0,s,\delta)$ is even more insensitive. The scaling factor[3] for the resultant decrease in correlation (due to path spreading) with increase in $\Delta X_t(s)$ is the Eulerian integral scale L_G of Eq. (21) so that

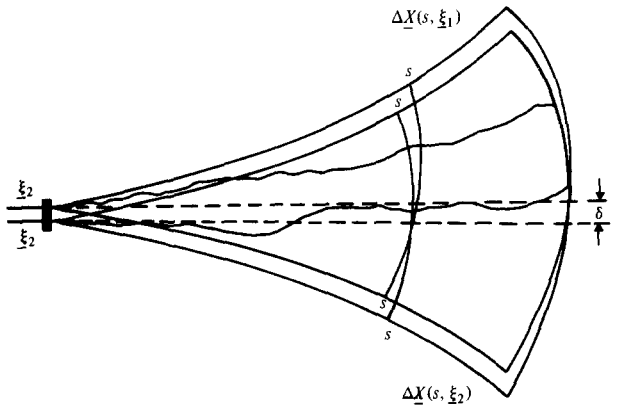

FIG. 6 Lagrangian crosspath spreading volume.

$$\begin{aligned}\hat{g}(0,s_{II},\delta) &\approx \hat{g}[0,0,\delta + \Delta X_t(s_{II})] \\ &= \hat{g}(0,0,[\delta + \Delta X_t(s_I)] \\ &\quad + [\Delta X_t(s_{II}) - \Delta X_t(s_I)]) \end{aligned} \tag{B3}$$

$$\approx \hat{g}[0,0,\delta + \Delta X_t(s_I)] \times \left[1 - \frac{\Delta X_t(s_{II}) - \Delta X_t(s_I)}{L_G}\right] \tag{B4}$$

$$\approx \hat{g}(0,s_I,\delta)\left[1 - \frac{\Delta X_t(s_{II}) - \Delta X_t(s_I)}{L_G}\right], \tag{B5}$$

where $0 \ll s_I < s_{II}$. Thus,

$$\frac{\Delta X_t(s_{II}) - \Delta X_t(s_I)}{L_G} = \frac{\hat{g}(0,s_I,\delta) - \hat{g}(0,s_{II},\delta)}{\hat{g}(0,s_I,\delta)} \tag{B6}$$

gives a measure of the relative decrease in correlation as s increases from s_I to s_{II} and shows that if the mean spreading (relative to L_G) is small, $\hat{g}(0,s,\delta)$ varies slightly with $s \in [s_I, s_{II}]$.

---

* Taken from the doctoral dissertation of J. A. Neubert, "Sound Propagation in Continuous Stochastic Media," under the Interdisciplinary Program in Engineering Acoustics, The Pennsylvania State University, University Park, Pa., 1970.

1 J. A. Neubert, J. Acoust. Soc. Am. **48**, 1203 (1970).
2 J. D. Cole, *Perturbation Methods in Applied Mathematics* (Blaisdell, Waltham, Mass., 1968), Chap. 3.
3 J. A. Neubert and J. L. Lumley, J. Acoust. Soc. Am. **48**, 1212 (1970).
4 U. Frisch, *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968).
5 J. A. Neubert, J. Math. Phys. **12**, 1030 (1971).
6 P. G. Bergmann, Phys. Rev. **70**, 486 (1946).
7 J. A. Neubert, "Sound Propagation in Continuous Stochastic Media," Ph.D. Thesis, Dept. of Engineering Acoustics, The Pennsylvania State University at University Park, 1970 (unpublished).
8 D. Mintzer, J. Acoust. Soc. Am. **25**, 922 (1953).
9 J. L. Lumley, J. Math. Phys. **3**, 309 (1962).
10 A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics* (MIT Press, Cambridge, Mass., 1971), Chap. 5.
11 J. L. Lumley, in *Mécanique de la turbulence*, Colloques Internationaux du Centre National de la Recherche Scientifique No. 108 (CNRS, Paris, 1962), pp. 18–26.
12 S. Corrsin, in Ref. 11, pp. 27–52.
13 H. Tennekes and J. L. Lumley, *A First Course in Turbulence* (MIT Press, Cambridge, 1971), Chap. 7.
14 The extension of the analytical methods of this paper to a practical problem in a statistically anisotropic medium is illustrated in J. A. Neubert, "Multipath Summability in Ray Theory Intensity Calculations in the Real Ocean," J. Acoust. Soc. Am. (to be published).
15 J. L. Lumley, *Stochastic Tools in Turbulence* (Academic, New York, 1970).
16 J. L. Lumley and H. A. Panofsky, *The Structure of Atmospheric Turbulence* (Wiley, New York, 1964).
17 D. Mintzer, J. Acoust. Soc. Am. **25**, 1107 (1953).
18 I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. **1**, 48 (1960).
19 N. Wiener, *Nonlinear Problems in Random Theory* (Wiley, New York, 1958).

$SU(n-2) \times SU(2) \times U(1)$ Bases for $SU(n)$*

R. T. Sharp

Department of Physics, McGill University, Montreal, Canada

(Received 9 August 1971)

Explicit basis dates for the general IR of $SU(n)$ are defined in the scheme $SU(n-2) \times SU(2) \times U(1)$; the approach makes use of elementary multiplets.

1. INTRODUCTION

Explicit basis states for $SU(n)$, reduced according to the subgroup $SU(n-1) \times U(1)$, were constructed many years ago by Gel'fand and Zetlin[1]; the generator matrix elements for these states were written down in Ref. 1, and derivations were later given by Baird and Biedenharn[2] and by Nagel and Moshinsky.[3]

In recent years there has been some interest in the reduction $SU(n) \supset SU(n-2) \times SU(2) \times U(1)$ especially for the case $n = 6$, which is the $SU(6)$ group of particle physics reduced according to the supermultiplet scheme.[4] Perelomov, Popov, and Malkin[5] give a method based on Young patterns for determining the $SU(4) \times SU(2) \times U(1)$ content of an IR of $SU(6)$; Hagen and Macfarlane[6] show how to determine the $SU(m) \times SU(n) \times U(1)$ content of an IR of $SU(m+n)$ by expanding the character function for the IR as a sum of products of $SU(m)$ and $SU(n)$ characters; more recently Mickelsson[7] has achieved the same object by examin-

ing the related problem of finding the Clebsch-Gordan series for appropriate IR's of $SU(m+n)$. Devi and Venkatarayudu[8] have treated the special case $SU(4) \supset SU(2) \times SU(2) \times U(1)$ by the use of elementary multiplets; they have not only solved the counting problem, but specified explicit bases.

The method of elementary multiplets (Devi, Venkatarayudu, and Moshinksy's elementary permissible diagrams[8,9]) has been used to solve a number of internal labeling problems.[8-12] It is based on the fact that the stretched product of subgroup multiplets from one or more IR's defines a multiplet of a higher IR. The labeling problem is solved when it is demonstrated that all subgroup multiplets of all IR's are obtained in this way in terms of a finite number of elementary multiplets; relations connecting the elementary multiplets make certain combinations of them redundant in general. For general discussions of the method see Ref. 12. In the next section a complete set of elementary multiplets, with redundant combinations, is presented for $SU(n) \supset SU(n-2) \times SU(2) \times U(1)$. They define a complete, independent set of subgroup multiplets for all IR's of $SU(n)$ and thus pave the way to the direct calculation of generator and transformation matrix elements and Clebsch-Gordan coefficients with respect to the basis states in question.

The $SU(n) \supset SU(n-2) \times SU(2) \times U(1)$ scheme is defined by specifying the subgroup decomposition of the first FIR (fundamental irreducible representation) of $SU(n)$. The n states are broken into two sets of $n-2$ and 2, respectively. The first set transforms according to the first FIR of $SU(n-2)$ (and is an $SU(2)$ scalar); the second is an $SU(2)$ doublet (and an $SU(n-2)$ scalar). A "hypercharge" quantum number Z is assigned the value $2/n$ for the $(n-2)$-plet, $-(n-2)/n$ for the doublet.

In the $SU(6)$ scheme of particle physics, the six states of the first FIR are those of three quarks, each of spin $\frac{1}{2}$; the first four refer to the nonstrange quarks with spin and isospin each $\frac{1}{2}$, and the other two are the spin states of the strange quark. In a general $SU(4) \times SU(2) \times U(1)$ multiplet of an IR of $SU(6)$, the $SU(4)$ labels describe the nonstrange quarks—in the $SU(2) \times SU(2)$ decomposition of $SU(4)$, the two $SU(2)$ subgroups describe their spin and isospin, respectively, in the manner of the Wigner supermultiplet scheme; the $SU(2)$ labels refer to the spin of the strange quarks; the $U(1)$ label Z is the total hypercharge. Although this scheme is sometimes called the Wigner supermultiplet decomposition of $SU(6)$, it should be stressed that it generally assigns nuclear states to $SU(4)$ multiplets different from those of the conventional supermultiplet model; thus in $SU(6)$ the nucleon states, together with the states of the (3, 3) resonance form a basis for the (300) IR of $SU(4)$ while in the usual scheme the nucleon states transform according to (100).

The $SU(4) \supset SU(2) \times SU(2) \times U(1)$ scheme provides bases (but not the conventional ones) for the Wigner supermultiplet nuclear model. Under it, the $SU(2)$ subgroups describe, respectively, the proton and neutron spins; the "hypercharge" is the 3-component of isospin.

For $n = 3$, the scheme reduces to the canonical $SU(3) \supset SU(2) \times U(1)$ decomposition applicable to elementary particles; $SU(2)$ describes isospin, $U(1)$ hypercharge.

2. ELEMENTARY MULTIPLETS

According to Racah,[13] the number of internal (state) labels required is $\frac{1}{2}(r_G - l_G)$, where r_G and l_G are the order (number of generators) and rank (number of IR labels, or number of FIR's) of the Group G. If r_H and l_H are the corresponding quantities for the subgroup H, it follows that there are $\frac{1}{2}(r_G - l_G - r_H - l_H)$ missing labels. In the elementary multiplet method of defining states, the labels are the exponents of the powers of the elementary multiplets whose product defines the subgroup multiplet in question; the labels provide the IR labels for group and subgroup, plus the "missing" labels. It follows that the number of independent elementary multiplets is $\frac{1}{2}(r_G + l_G - r_H + l_H)$. For the scheme $SU(n) \supset SU(n-2) \times SU(2) \times U(1)$ we have $r_G = n^2 - 1, l_G = n - 1, r_H = n^2 - 4n + 7, l_H = n - 1$. Hence there are $n - 3$ missing labels and $3n - 5$ independent elementary multiplets.

We first list the elementary multiplets; they are found by examining low-lying IR's of $SU(n)$. The proof that they solve the labeling problem in general is outlined in the next section. The IR labels are those of Cartan (the ith label is the number of columns containing i boxes in the Young diagram of the IR; it is the number of times the ith FIR occurs in the stretched product which defines the IR in question); λ_i, α_i are the Cartan labels for $SU(n)$, $SU(n-2)$, respectively; $\alpha (= 2j)$ is the IR label for $SU(2)$; Z is the hypercharge or $U(1)$ label. The ith fundamental FIR of $SU(n)$ is that with $\lambda_i = 1$ and all other λ's equal to zero. For $1 < i < n-1$, it contains three multiplets of $SU(n-2) \times SU(2) \times U(1)$, each of which is an elementary multiplet. The first, A_1^i, has $\alpha_i = 1$, all other α's equal to zero, $Z = 2i/n$; the second, A_2^i, has $\alpha_{i-1} = 1$, $\alpha = 1$, all other α's equal to 0, $Z = (2i/n) - 1$; the third, A_3^i, has $\alpha_{i-2} = 1$, all other α's equal to zero, $Z = (2i/n) - 2$ (A_1^{n-2} and A_3^2 have all α's equal to zero, i.e., are $SU(n-2) \times SU(2)$ scalars). The first and $(n-1)$th FIR's of $SU(n)$ each contain just two (elementary) multiplets. For the first, A_1^1 is as above with $i = 1$; A_2^1 has $\alpha = 1$ and all other α's equal to zero, $Z = (2/n) - 1$; there is no A_3^1. For the $(n-1)$th, there is no A_1^{n-1}; A_2^{n-1} has $\alpha = 1$ and all other α's equal to zero, $Z = -(2/n) + 1$; A_3^{n-1} is as above with $i = n = 1$. Thus there are $3n - 5$ elementary multiplets

$$\begin{aligned} &A_1^i \ (i = 1, \ \ldots, n-2), \\ &A_2^i \ (i = 1, \ \ldots, n-1), \\ &A_3^i \ (i = 2, \ \ldots, n-1). \end{aligned} \tag{1}$$

For $n = 6$, they are (the notation is $(\lambda_1\lambda_2\lambda_3\lambda_4\lambda_5; \alpha_1\alpha_2\alpha_3, \alpha, Z)$):

$A_1^1 = (10000; 100, 0, \frac{1}{3})$, $A_3^3 = (00100; 100, 0, -1)$,
$A_2^1 = (10000; 000, 1, -\frac{2}{3})$, $A_1^4 = (00010; 000, 0, \frac{4}{3})$,
$A_1^2 = (01000; 010, 0, \frac{2}{3})$, $A_2^4 = (00010; 001, 1, \frac{1}{3})$,
$A_2^2 = (01000; 100, 1, -\frac{1}{3})$, $A_3^4 = (00010; 010, 0, -\frac{2}{3})$,
$A_3^2 = (01000; 000, 0, -\frac{4}{3})$, $A_2^5 = (00001; 000, 1, \frac{2}{3})$,
$A_1^3 = (00100; 001, 0, 1)$, $A_3^5 = (00001; 001, 0, -\frac{1}{3})$.
$A_2^3 = (00100; 010, 1, 0)$,

Although the elementary multiplets (1) are correct in number, they do not, unfortunately, by themselves solve the labeling problem; certain additional composite elementary factors A^{ij} $(1 \leq i \leq j-2 \leq n-3)$, $\frac{1}{2}(n-2)(n-3)$ in number, are required. A^{ij} belongs to the IR of $SU(n)$ with $\lambda_i = \lambda_j = 1$ and all other λ's equal to zero. As an $SU(n-2) \times SU(2) \times U(1)$ multiplet, it has $\alpha_{i-1} = \alpha_{j-1} = 1$ and all other α's equal to zero (A^{1j} has $\alpha_{j-1} = 1$ and all other α's equal to zero; $A^{i\,n-1}$ has $\alpha_{i-1} = 1$ and all other α's equal to zero; $A^{1\,n-1}$ has all α's equal to zero); its hypercharge is $Z = 2(i+j-n)/n$.

The A^{ij} are of course not independent of the $A^k_{1,2,3}$. They satisfy relations of the form $(i < k < j)$

$$a_1 A^{ij}A^k_2 + a_2 A^{ik}A^j_2 + a_3 A^{kj}A^i_2 = 0, \tag{2}$$

where a_1, a_2, a_3 are nonzero constants. In (2) if $i = k-1, A^{ik}$ should be replaced by $A^i_1 A^k_3$ and similarly for A^{kj} if $k = j-1$. In addition, the A^{ij} satisfy relations among themselves of the form $(i < k < j < h)$

$$b_1 A^{ij}A^{kh} + b_2 A^{ik}A^{jh} + b_3 A^{ik}A^{kj} = 0 \tag{3}$$

with nonzero b's. If $i = k-1, k = j-1$, or $j = h-1$ then A^{ik} should be replaced $A^i_1 A^k_3$, etc. Equations (2) and (3) can be used to express the first term on the left-hand side in terms of the other two; the eliminated terms must be regarded as redundant combinations of elementary multiplets. We may say that A^{ij} is incompatible with A^k_2 for $i < k < j$ and with A^{kh} if $i < k < j < h$ or $k < i < h < j$; incompatible pairs must not appear together for the purpose of forming higher multiplets.

For $SU(6)$ the composite elementary multiplets are

$A^{13} = (10100; 010; 0; -\frac{2}{3})$, $A^{14} = (10010; 001; 0; -\frac{1}{3})$,
$A^{15} = (10001; 000; 0; 0)$, $A^{24} = (01010; 101; 0; 0)$,
$A^{25} = 01001; 100; 0; \frac{1}{3})$, $A^{35} = (00101; 010; 0; \frac{2}{3})$.

The following pairs of elementary factors are incompatible: A^{13} with A^2_2, A^{24}, A^{25}; A^{14} with $A^2_2, A^3_2, A^{25}, A^{35}$; A^{15} with A^2_2, A^3_2, A^4_2; A^{24} with A^3_2, A^{35}; A^{25} with A^3_2, A^4_2; A^{35} with A^4_2.

The adjoint IR of $SU(n)$, by which the generators transform, is $(10\cdots01)$. Its $SU(n-2) \times SU(2) \times U(1)$ multiplets are defined by the elementary multiplet products $A^1_1 A^{n-1}_3$, $A^1_2 A^{n-1}_2$, $A^{1\,n-1}$, $A^1_1 A^{n-1}_2$, $A^1_2 A^{n-1}_3$. In such products all the quantum numbers are additive so we find the subgroup multiplets $(10\cdots01, 0, 0)$, $(0\cdots0; 2; 0)$, $(0\cdots0; 0; 0)$, $(10\cdots0; 1; 1)$, $(0\cdots01; 1; -1)$. Interpreting them as generators, we note that the first two multiplets comprise the generators of $SU(n-2)$ and $SU(2)$, respectively; the third, an $SU(n-2) \times SU(2)$ scalar, is the $U(1)$ generator, proportional to the hypercharge Z. The other two multiplets consist of the generators which link different subgroup multiplets in the general IR of $SU(n)$. It would be of interest to work out their reduced matrix elements in the general case. Our results reduce to those of Devi and Venkatarayudu[8] in the special case $n = 4$.

3. JUSTIFICATION OF SCHEME

The set of elementary multiplets for $SU(n) \supset SU(n-2) \times SU(2) \times U(1)$ and their compatibility rules are remarkably similar to those for $Sp(2n) \supset Sp(2n-2) \times SU(2)$.[11] Since the justification of their validity is formally also much the same, it is not reproduced here in full. Only the differences are pointed out, which are chiefly in the interpretation of the symbols.

According to Weyl[14] the characters of the IR's of $SU(n)$ are (see also Ref. 6)

$$\chi(l_1, \ldots, l_{n-1}) = \begin{vmatrix} P_{l_1} & \cdots & P_{l_1+n-2} \\ \cdot & & \\ \cdot & & \\ P_{l_{n-1}} & \cdots & P_{l_{n-1}+n-2} \end{vmatrix}, \tag{4}$$

where

$$l_i = \left(\sum_{j=i}^{n} \lambda_j\right) - i + 1 = l_i^W - n + 1;$$

l_i^W is Weyl's l_i; $P_l(\epsilon_1, \ldots, \epsilon_n)$ is defined as the coefficient of z^l in the expansion

$$\prod_{i=1}^{n} (1 - \epsilon_i z)^{-1} = \sum_{l=0}^{\infty} P_l(z) z^l.$$

Because of the unimodular condition, the ϵ are not independent but satisfy $\Pi_i \epsilon_i = 1$.

To effect the $SU(n-2) \times SU(2) \times U(1)$ reduction, it is convenient to write

$$\epsilon_n = \epsilon_Z^{-(n-2)Z/n} \epsilon^{-1}, \quad \epsilon_{n-1} = \epsilon_Z^{-(n-2)Z/n} \epsilon,$$
$$\epsilon_i = \epsilon_i' \epsilon_Z^{2Z/n}, \quad i = 1, \ldots, n-2.$$

Then it is not hard to show that

$$P_l = \sum_{a=0}^{l} p'_a P'_{l-a} \epsilon_Z^{Z(2l/n-a)} \tag{5}$$

where $p'_a(\epsilon)$ and $P'_{l-a}(\epsilon'_1, \ldots, \epsilon'_{n-2})$ are the P appropriate to $SU(2)$ and $SU(n-2)$, respectively. Equations (4) and (5) are the analogs of Eqs. (3.5, 8) of Ref. 11.

The proof that the conjectured elementary multiplets give the correct subgroup content proceeds by induction. With the assumption that they are correct for IR's with $\lambda_i = 0, i > m$, it can be shown with the help of (4) and (5) that they continue to give the correct subgroup multiplets when λ_m is increased by unity. Since the proof corresponds line by line to that for $Sp(2n) \supset Sp(2n-2) \times Sp(2)$, we refer the reader to Ref. 11 for the details. The only further difference is that the determinants which vanish here, in analogy with (3.12) of Ref. 11, are those of the form

$$\begin{vmatrix} P_{l_1} & \cdots & P_{l_1+n-2} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ P_{l_{n-1}} & \cdots & P_{l_{n-1}+n-2} \end{vmatrix}$$

with $l + n - 2 \geq 1$, $i = 1, \ldots, n-1$.

---

* Supported by the National Research Council of Canada.

1 I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR 71, 825, 1017 (1950).
2 G. E. Baird and L. C. Biedenharn, J. Math. Phys. 4, 1449 (1963).
3 J. G. Nagel and M. Moshinsky, J. Math. Phys. 6, 682 (1965).
4 F. Gürsey, A. Pais, and L. A. Radicati, Phys. Rev. Letters 13, 299 (1964).

[5] A. M. Perelomov, V. S. Popov, and I. A. Malkin, J. Nucl. Phys. (U.S.S.R.) **2**, 533 (1965).
[6] C. R. Hagen and A. J. Macfarlane, J. Math. Phys. **6**, 1366 (1965).
[7] J. Mickelsson, J. Math Phys. **11**, 2803 (1970).
[8] V. S. Devi and T. Venkatarayudu, J. Math Phys. **9**, 1057 (1968).
[9] V. S. Devi and M. Moshinksy, J. Math. Phys. **10**, 455 (1969).
[10] The earliest use of the method appears to be by V. Bargmann and M. Moshinksy, Nucl. Phys. **23**, 177 (1961); recently the method was applied to the Wigner supermultiplet scheme by M. Brunet and M. Resnikoff, J. Math. Phys. **11**, 1474 (1970), and by K. Ahmed and R. T. Sharp, Ann. Phys. (N.Y.) (to be published)
[11] R. T. Sharp, Proc. Cambridge Phil. Soc. **68**, 571 (1970).
[12] R. T. Sharp and C. S. Lam, J. Math. Phys. **10**, 2033 (1969); V. S. Devi, Nuovo Cimento **68A**, 1 (1970); J. Math. Phys. **12**, 1732 (1971).
[13] G. Racah, Lecture Notes "Group Theory and Spectroscopy," Institute for Advance Study (1951), reprinted in Ergeb. Exact. Naturw. **37**, 28 (1965).
[14] H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, N.J., 1946), see Eq. (6.5), p. 203.

On the Asymptotic Behavior of Certain Dynamical Systems

Reese T. Prosser
Department of Mathematics, Dartmouth College, Hanover, New Hampshire 03755
(Received 16 August 1971)

The asymptotic behavior at large times for certain dynamical systems arising in the Hamiltonian formulation of classical mechanics is investigated. It is shown that for potentials which die out sufficiently fast at large distances the unbounded states of the system are asymptotically free. This result complements the corresponding result for quantum mechanical systems, and is obtained by analogous methods. In addition, the existence, differentiability, and asymptotic completeness of the associated wave mappings is established under appropriate further assumptions by classical methods.

In this paper, we investigate the asymptotic behavior at large times of those dynamical systems arising in the Hamiltonian formulation of classical mechanics. For such systems, we show quite generally that for potentials which die out sufficiently fast at large distances, the unbound states of the system are asymptotically free. This result parallels the corresponding result for the corresponding systems of quantum mechanics, first obtained by J. M. Cook[1] and is obtained here by analogous methods. In particular, we make use of the increasingly familiar device of converting a nonlinear problem to a linear one by passing from the solution manifold to a suitably chosen function space over the manifold.

1. PRELIMINARIES

By a classical system we mean a classical mechanical system consisting of a finite number N of point masses moving in a three-dimensional Euclidean space according to the laws of classical mechanics.[2] At any instant of time, the state of the system is completely determined by specifying a suitable set of canonical coordinates, consisting of the positions x_i and the conjugate momenta y_i of the individual point masses. These coordinates then locate a point $(\mathbf{x}, \mathbf{y})$ in the phase space E_{2n} of $2n$ dimensions, where $2n = 6N$.

The development of the system in time is completely determined by specifying an everywhere continuously differentiable Hamiltonian function $H(x_1, \cdots, x_n, y_1, \cdots, y_n) = H(\mathbf{x}, \mathbf{y})$ of these canonical variables, and imposing the Hamiltonian equations of motion:

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial y_i}, \qquad \frac{dy_i}{dt} = -\frac{\partial H}{\partial x_i}, \qquad i = 1, 2, \cdots, n \tag{1.1}$$

Every solution $(\mathbf{x}(t), \mathbf{y}(t))$ of these equations of motion starting from a prescribed initial point $(\mathbf{x}, \mathbf{y}) = (\mathbf{x}(0), \mathbf{y}(0))$ determines a trajectory of the system in the phase space. Moreover, we know that the *motion* $w(t)$: $E_{2n} \to E_{2n}$ of the phase space given by

$$w(t)(\mathbf{x}, \mathbf{y}) = (\mathbf{x}(t), \mathbf{y}(t)) \tag{1.2}$$

is defined and continuously differentiable in both $(\mathbf{x}, \mathbf{y})$ and t, and preserves the volume element of the phase space E_{2n} for each t.[2]

The derivative $w'(t) = (d/dt)w(t)$ of the motion $w(t)$ then assigns to each point $(\mathbf{x}, \mathbf{y})$ in E_{2n} a *tangent vector* $w'(t)(\mathbf{x}, \mathbf{y})$, giving the direction at $(\mathbf{x}(t), \mathbf{y}(t))$ of the motion along the trajectory starting from $(\mathbf{x}, \mathbf{y})$. The tangent vector is readily computed from the equations of motion (1.1), and is found to be

$$w'(t)(\mathbf{x}, \mathbf{y}) = \frac{d}{dt}(\mathbf{x}(t), \mathbf{y}(t)) = (\mathbf{u}(t), \mathbf{v}(t)), \tag{1.3}$$

where $(\mathbf{u}, \mathbf{v})$ denotes the tangent vector field whose components are given by

$$u_i = \frac{\partial H}{\partial y_i}, \qquad v_i = -\frac{\partial H}{\partial x_i}, \qquad i = 1, \cdots, n. \tag{1.4}$$

It is known that $w'(t)(\mathbf{x}, \mathbf{y})$ depends continuously upon both $(\mathbf{x}, \mathbf{y})$ and t, and that $w(t)(\mathbf{x}, \mathbf{y})$ may be recovered from $w'(t)(\mathbf{x}, \mathbf{y})$ by performing a quadrature:

$$w(t)(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y}) + \int_0^t w'(\tau)(\mathbf{x}, \mathbf{y})d\tau. \tag{1.5}$$

The motion $w(t)$ induces a covariant motion $w_*(t)$: $\mathcal{C}_0(E_{2n}) \to \mathcal{C}_0(E_{2n})$ on the space of all continuous functions with compact support on E_{2n}. This induced motion is defined by

$$w_*(t)f(\mathbf{x}, \mathbf{y}) = f(w(-t)(\mathbf{x}, \mathbf{y})). \tag{1.6}$$

It is clear from the definition that $w_*(t)f$ is linear in f, and depends continuously upon both f (with the topology of uniform convergence) and t. Moreover, since $w(t)$ preserves the volume element in E_{2n}, it follows that $w_*(t)$ preserves the $\mathcal{L}_p$ norm $\| \ \|_p$ in $\mathcal{C}_0(E_{2n})$, and hence admits by uniform continuity an isometric extension $W(t)$ to each of the spaces $\mathcal{L}_p(E_{2n})$, $1 \le p < \infty$, and to the space $\mathcal{C}_\infty(E_{2n})$, the closure of $\mathcal{C}_0(E_{2n})$ in $\mathcal{L}_\infty(E_{2n})$. The operators $W(t)$, so defined, form for each fixed p a strongly continuous one-parameter group of linear isometries of $\mathcal{L}_p(E_{2n})$.

If $f \in \mathcal{C}_0(E_{2n})$ happens to be a continuously differentiable function, then $w_*(t)f$ is continuously differentiable in t. The derivative $w_*'(t) = (d/dt)w_*(t)$ then assigns to the differentiable function f a continuous

[5] A. M. Perelomov, V. S. Popov, and I. A. Malkin, J. Nucl. Phys. (U.S.S.R.) **2**, 533 (1965).
[6] C. R. Hagen and A. J. Macfarlane, J. Math. Phys. **6**, 1366 (1965).
[7] J. Mickelsson, J. Math Phys. **11**, 2803 (1970).
[8] V. S. Devi and T. Venkatarayudu, J. Math Phys. **9**, 1057 (1968).
[9] V. S. Devi and M. Moshinksy, J. Math. Phys. **10**, 455 (1969).
[10] The earliest use of the method appears to be by V. Bargmann and M. Moshinksy, Nucl. Phys. **23**, 177 (1961); recently the method was applied to the Wigner supermultiplet scheme by M. Brunet and M. Resnikoff, J. Math. Phys. **11**, 1474 (1970), and by K. Ahmed and R. T. Sharp, Ann. Phys. (N.Y.) (to be published)
[11] R. T. Sharp, Proc. Cambridge Phil. Soc. **68**, 571 (1970).
[12] R. T. Sharp and C. S. Lam, J. Math. Phys. **10**, 2033 (1969); V. S. Devi, Nuovo Cimento **68A**, 1 (1970); J. Math. Phys. **12**, 1732 (1971).
[13] G. Racah, Lecture Notes "Group Theory and Spectroscopy," Institute for Advance Study (1951), reprinted in Ergeb. Exact. Naturw. **37**, 28 (1965).
[14] H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, N.J., 1946), see Eq. (6.5), p. 203.

On the Asymptotic Behavior of Certain Dynamical Systems

Reese T. Prosser
Department of Mathematics, Dartmouth College, Hanover, New Hampshire 03755
(Received 16 August 1971)

The asymptotic behavior at large times for certain dynamical systems arising in the Hamiltonian formulation of classical mechanics is investigated. It is shown that for potentials which die out sufficiently fast at large distances the unbounded states of the system are asymptotically free. This result complements the corresponding result for quantum mechanical systems, and is obtained by analogous methods. In addition, the existence, differentiability, and asymptotic completeness of the associated wave mappings is established under appropriate further assumptions by classical methods.

In this paper, we investigate the asymptotic behavior at large times of those dynamical systems arising in the Hamiltonian formulation of classical mechanics. For such systems, we show quite generally that for potentials which die out sufficiently fast at large distances, the unbound states of the system are asymptotically free. This result parallels the corresponding result for the corresponding systems of quantum mechanics, first obtained by J. M. Cook[1] and is obtained here by analogous methods. In particular, we make use of the increasingly familiar device of converting a nonlinear problem to a linear one by passing from the solution manifold to a suitably chosen function space over the manifold.

1. PRELIMINARIES

By a classical system we mean a classical mechanical system consisting of a finite number N of point masses moving in a three-dimensional Euclidean space according to the laws of classical mechanics.[2] At any instant of time, the state of the system is completely determined by specifying a suitable set of canonical coordinates, consisting of the positions x_i and the conjugate momenta y_i of the individual point masses. These coordinates then locate a point $(\mathbf{x}, \mathbf{y})$ in the phase space E_{2n} of $2n$ dimensions, where $2n = 6N$.

The development of the system in time is completely determined by specifying an everywhere continuously differentiable Hamiltonian function $H(x_1, \cdots, x_n, y_1, \cdots, y_n) = H(\mathbf{x}, \mathbf{y})$ of these canonical variables, and imposing the Hamiltonian equations of motion:

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial y_i}, \qquad \frac{dy_i}{dt} = -\frac{\partial H}{\partial x_i}, \qquad i = 1, 2, \cdots, n \tag{1.1}$$

Every solution $(\mathbf{x}(t), \mathbf{y}(t))$ of these equations of motion starting from a prescribed initial point $(\mathbf{x}, \mathbf{y}) = (\mathbf{x}(0), \mathbf{y}(0))$ determines a trajectory of the system in the phase space. Moreover, we know that the *motion* $w(t)$: $E_{2n} \to E_{2n}$ of the phase space given by

$$w(t)(\mathbf{x}, \mathbf{y}) = (\mathbf{x}(t), \mathbf{y}(t)) \tag{1.2}$$

is defined and continuously differentiable in both $(\mathbf{x}, \mathbf{y})$ and t, and preserves the volume element of the phase space E_{2n} for each t.[2]

The derivative $w'(t) = (d/dt)w(t)$ of the motion $w(t)$ then assigns to each point $(\mathbf{x}, \mathbf{y})$ in E_{2n} a *tangent vector* $w'(t)(\mathbf{x}, \mathbf{y})$, giving the direction at $(\mathbf{x}(t), \mathbf{y}(t))$ of the motion along the trajectory starting from $(\mathbf{x}, \mathbf{y})$. The tangent vector is readily computed from the equations of motion (1.1), and is found to be

$$w'(t)(\mathbf{x}, \mathbf{y}) = \frac{d}{dt}(\mathbf{x}(t), \mathbf{y}(t)) = (\mathbf{u}(t), \mathbf{v}(t)), \tag{1.3}$$

where $(\mathbf{u}, \mathbf{v})$ denotes the tangent vector field whose components are given by

$$u_i = \frac{\partial H}{\partial y_i}, \qquad v_i = -\frac{\partial H}{\partial x_i}, \qquad i = 1, \cdots, n. \tag{1.4}$$

It is known that $w'(t)(\mathbf{x}, \mathbf{y})$ depends continuously upon both $(\mathbf{x}, \mathbf{y})$ and t, and that $w(t)(\mathbf{x}, \mathbf{y})$ may be recovered from $w'(t)(\mathbf{x}, \mathbf{y})$ by performing a quadrature:

$$w(t)(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y}) + \int_0^t w'(\tau)(\mathbf{x}, \mathbf{y})d\tau. \tag{1.5}$$

The motion $w(t)$ induces a covariant motion $w_*(t)$: $\mathcal{C}_0(E_{2n}) \to \mathcal{C}_0(E_{2n})$ on the space of all continuous functions with compact support on E_{2n}. This induced motion is defined by

$$w_*(t)f(\mathbf{x}, \mathbf{y}) = f(w(-t)(\mathbf{x}, \mathbf{y})). \tag{1.6}$$

It is clear from the definition that $w_*(t)f$ is linear in f, and depends continuously upon both f (with the topology of uniform convergence) and t. Moreover, since $w(t)$ preserves the volume element in E_{2n}, it follows that $w_*(t)$ preserves the $\mathcal{L}_p$ norm $\| \ \|_p$ in $\mathcal{C}_0(E_{2n})$, and hence admits by uniform continuity an isometric extension $W(t)$ to each of the spaces $\mathcal{L}_p(E_{2n})$, $1 \le p < \infty$, and to the space $\mathcal{C}_\infty(E_{2n})$, the closure of $\mathcal{C}_0(E_{2n})$ in $\mathcal{L}_\infty(E_{2n})$. The operators $W(t)$, so defined, form for each fixed p a strongly continuous one-parameter group of linear isometries of $\mathcal{L}_p(E_{2n})$.

If $f \in \mathcal{C}_0(E_{2n})$ happens to be a continuously differentiable function, then $w_*(t)f$ is continuously differentiable in t. The derivative $w_*'(t) = (d/dt)w_*(t)$ then assigns to the differentiable function f a continuous

function $w_*'(t)f$ according to the formula

$$w_*'(t)f(\mathbf{x},\mathbf{y}) = \frac{d}{dt} f(\mathbf{x}(-t),\mathbf{y}(-t)) = -\sum \frac{\partial H}{\partial y_i}\frac{\partial f}{\partial x_i} - \frac{\partial H}{\partial x_i}\frac{\partial f}{\partial y_i} = -\{H,f\}, \tag{1.7}$$

where $\{\ ,\ \}$ denotes the usual Poisson bracket.[2] It is a straightforward matter to show that this derivative, evaluated at $t = 0$, is essentially the generator of the one-parameter group of isometries $W(t)$ in each of the $\mathcal{L}_p$ spaces over E_{2n}. Specifically, we have

$$W(t) = \exp iLt \tag{1.8}$$

with the generator L given by

$$L = +\, i\{H, \cdot\}. \tag{1.9}$$

We shall refer to this operator L as the *Liouville operator* of the system. The Liouville operator determines the motion of the function spaces over the phase space of the system.

2. THE MOTION OF NONINTERACTING PARTICLES

In this section we shall obtain a precise description of the various operators introduced in the preceding section under the assumption that the Hamiltonian function H is independent of the position coordinates of the phase space. This assumption holds for the Hamiltonian functions describing every system of noninteracting particles, as well as those systems reducible to such systems through a suitable canonical transformation.

We shall assume, then, that $H(\mathbf{x},\mathbf{y})$ is everywhere twice continuously differentiable and that

$$\frac{\partial H}{\partial x_i} \equiv 0, \quad 1 \le i \le n. \tag{2.1}$$

Under these assumptions, the equations of motion become

$$\frac{d}{dt}(\mathbf{x},\mathbf{y}) = (\mathbf{u}, 0), \tag{2.2}$$

where

$$u_i = u_i(\mathbf{y}) = \frac{\partial H}{\partial y_i}. \tag{2.3}$$

The solutions of these equations of motion then all have the form

$$w(t)(\mathbf{x},\mathbf{y}) = (\mathbf{x}(t),\mathbf{y}(t)) = (\mathbf{x} + \mathbf{u}t, \mathbf{y}). \tag{2.4}$$

It follows that the motion $w^*(t)$ induced in $\mathcal{C}_0(E_{2n})$ has the form

$$w_*(t)(f)(\mathbf{x},\mathbf{y}) = f(\mathbf{x} - \mathbf{u}t, \mathbf{y}), \tag{2.5}$$

while the Liouville operator L has the form

$$L = +\, i\{H, \cdot\} = +\, i\sum_{i=1}^{n} u_i \frac{\partial}{\partial x_i}. \tag{2.6}$$

In order to analyze the structure of L, we need only introduce the Fourier transform F in the first n coordinates x_i. If $f \in \mathcal{C}_0(E_n)$, then

$$(Ff)(\mathbf{k},\mathbf{y}) = (2\pi)^{-n/2} \int_{E_n} \exp i(\mathbf{k}\cdot\mathbf{x}) f(\mathbf{x},\mathbf{y})\, d\mathbf{x}. \tag{2.7}$$

Then from (2.6) we have

$$(FLF^{-1}f)(\mathbf{k},\mathbf{y}) = -(\mathbf{u}\cdot\mathbf{k}) f(\mathbf{k},\mathbf{y}). \tag{2.8}$$

Thus we see that L is equivalent, via the Fourier transform F, to the operation of multiplication by the real-valued function $-(\mathbf{u}\cdot\mathbf{k})$.

It follows now by standard methods of spectral analysis that the spectrum of L in $\mathcal{L}_2(E_{2n})$ consists precisely of all real numbers in the range of $-(\mathbf{u}\cdot\mathbf{k})$ as the u_i take on all values of the form $\partial H/\partial y_i$ and the k_i range over all real numbers. We conclude in this case that the spectral measure of L is absolutely continuous and that the spectrum of L consists of the whole real line.

3. THE MOTION OF INTERACTING PARTICLES

In this section we shall investigate the structure of the various operators introduced in Sec. 1 under the assumption that the Hamiltonian can be expressed as the sum of two terms, one of which satisfies the requirements of Sec. 2, and the other of which dies out in a suitable sense at large distances from the origin in the phase space. Under these assumptions, the first term dominates the motion at large distances from the origin, and the analysis of Sec. 2 can be expected to apply. The behavior of the motion near the origin is then determined by following the motion in from large distances. This procedure has already been used successfully in analyzing scattering problems in quantum mechanics; we have only to adapt those techniques to the present problem.[1,3]

We shall assume, then, that the Hamiltonian function has the following form:

$$H = H_0 + V, \tag{3.1}$$

where H_0 is everywhere twice continuously differentiable and satisfies

$$\frac{\partial H}{\partial x_i} \equiv 0, \quad 1 \le i \le n \tag{3.2}$$

and V is also everywhere twice continuously differentiable, and satisfies

$$|\operatorname{grad} V(\mathbf{x},\mathbf{y})| \le \text{const}\ |\mathbf{x}|^{-2-\eta} \tag{3.3}$$

for some $\eta > 0$. For convenience we shall also assume that the gradient of H_0 vanishes only on a set Z_0 of measure zero in E_{2n}.

We now fix a value of p, $1 \le p \le \infty$, and consider the motions $W_0(t)$ and $W(t)$ induced on $\mathcal{L}_p(E_{2n})$ by the Hamiltonians H_0 and H, respectively. We form the *transition operators*

$$W(s,t) = W_0(-s)W(s-t)W_0(t). \tag{3.4}$$

These operators describe the relative motion of the space $\mathcal{L}_p(E_{2n})$ obtained by moving the system according to H_0 for time t, then according to H for time $(s-t)$, and then according to H_0 again for time $-s$. These operators obviously are isometries, and satisfy the following relations,

$$W(s,s) = \mathrm{I}, \qquad W(r,s)W(s,t) = W(r,t). \tag{3.5}$$

Of particular interest is the behavior of these operators at $t \to -\infty$.

Lemma 3.1: Suppose that $\lim W(s,t)$ as $t \to -\infty$ converges strongly to the operator $W(s,-\infty)$. Then $W(s,-\infty)$ is an isometry for each s, and satisfies

$$W(s,-\infty) = W_0(-s)W(0,-\infty)W_0(+s), \tag{3.6}$$

$$LW(0,-\infty), \; = W(0,-\infty)L_0. \tag{3.7}$$

Proof: We know that strong limits of isometries are always isometries. From the definition (3.4) we obtain immediately the identities

$$W_0(s)W(s,t+s) = W(0,t)W_0(s) = W(s)W(0,t+s). \tag{3.8}$$

the first of these identities leads to (3.6) as $t \to -\infty$, while the second leads to

$$W(s)W(0,-\infty) = W(0,-\infty)W_0(s), \tag{3.9}$$

which then leads to (3.7) by standard operator methods.

It follows from this Lemma that $W_- = W(0,-\infty)$ is an isometry which effects a similarity between L and L_0. It does not follow from this Lemma that the similarity is invertible, i.e., that the range of W_- is all of $\mathfrak{L}_p(E_{2n})$. In general, we know only that the range of W_- reduces L, and that the restriction of L to the range of W_- is isometrically equivalent with L_0.

Thus we are led to investigate the convergence of the transition operators as $t \to -\infty$.

A convergence criterion useful for this purpose was developed by Cook[3] in his study of the corresponding problem in quantum mechanics. This criterion has a far wider scope, however, and applies to the present situation as well.

We first introduce the difference operator M, defined by

$$L = L_0 + M \tag{3.10}$$

and in terms of M we introduce the one-parameter family $M(t)$ defined by

$$M(t) = W_0(-t)MW_0(t). \tag{3.11}$$

Lemma 3.2: Let $\mathfrak{D}$ be any set of functions dense in $\mathfrak{L}_p(E_{2n})$ on which L_0 and L are both defined, and which is invariant under $W_0(t)$ for all t. Suppose that for each f in $\mathfrak{D}$ we have

$$\int_{-\infty}^{0} \|M(t)f\|_p \, dt < \infty. \tag{3.12}$$

Then $\lim W(s,t)$ as $t \to -\infty$ converges strongly in $\mathfrak{L}_p(E_{2n})$, for each s, and on $\mathfrak{D}$ we have

$$W(s,-\infty) = \mathrm{I} + i\int_{-\infty}^{s} W(s,t)M(t)dt. \tag{3.13}$$

Proof: According to (3.6) it suffices to consider the case $s = 0$. The hypotheses imply that $W(0,t)f$ is strongly continuously differentiable in t for all f in $\mathfrak{D}$, and we have

$$\frac{d}{dt}W(0,t)f = -iW(-t)(L - L_0)W_0(t)f = -iW(0,t)M(t)f. \tag{3.14}$$

Integrating both sides from $t' < 0$ to 0, we obtain

$$f - W(0,t')f = -i\int_{t'}^{0} W(0,\tau)M(\tau)f\,d\tau. \tag{3.15}$$

Hence we have, for $t' < t'' < 0$,

$$W(0,t')f - W(0,t'')f = i\int_{t'}^{t''} W(0,\tau)M(\tau)f\,d\tau. \tag{3.16}$$

It follows that

$$\|W(0,t')f - W(0,t'')f\|_p \le \int_{t'}^{t''} \|W(0,\tau)M(\tau)f\|_p d\tau = \int_{t'}^{t''} \|M(\tau)f\|_p \, d\tau. \tag{3.17}$$

Now if (3.12) holds, then the expression in (3.17) tends to zero as t' and t'' tend to $-\infty$. This means that $W(0,t)f$ converges strongly as $t \to -\infty$. If the set $\mathfrak{D}$ of such f is dense in $\mathfrak{L}_p(E_{2n})$, then our conclusion follows from the uniform continuity of $W(0,t)$ by the Banach–Steinhaus theorem.

Theorem 3.3: Suppose that H_0 and V satisfy the conditions (3.2) and (3.3), respectively. Then $\lim W(s,t)$ as $t \to -\infty$ converges strongly to an isometry W satisfying the relations (3.6) and (3.7).

Proof: Let $\mathfrak{D}$ be the set of all everywhere continuously differentiable functions whose support is compact and does not intersect the manifold Z_0 on which the gradient of H_0 vanishes. Then it is clear from (1.9) that both L_0 and L are defined on $\mathfrak{D}$, and from (2.5) that $\mathfrak{D}$ is invariant under $W_0(t)$. Moreover, it is clear that $\mathfrak{D}$ is dense in $\mathfrak{L}_p(E_{2n})$, $1 \le p < \infty$, provided that the manifold Z_0 on which the gradient of H_0 vanishes has measure zero in E_{2n}, which we are here assuming. It remains to show that (3.12) holds for all f in $\mathfrak{D}$.

Now from the definitions of L and L_0 it follows that for any f in $\mathfrak{D}$ we have

$$\begin{aligned} Mf &= Lf - L_0 f \\ &= +i\{H,f\} - i\{H_0,f\} \\ &= +i\{V,f\}, \end{aligned} \tag{3.18}$$

where $V = H - H_0$. From (3.11) and (2.5) it follows that

$$MW_0(t)f = i\{V, W_0(t)f\} = i\sum \frac{\partial V}{\partial y_i}\frac{\partial W_0(t)f}{\partial x_i} - \frac{\partial V}{\partial x_i}\frac{\partial W_0(t)f}{\partial y_i}. \tag{3.19}$$

Hence we have

$$\|MW_0(t)f\|_p \le \| \, |\text{grad } V| \, |\text{grad } W_0(t)f| \, \|_p \le \|(\text{grad } V)(W_0(t)\chi)\|_\infty \|\text{grad } (W_0(t)f)\|_p, \tag{3.20}$$

where χ is the characteristic function of the (compact) support of f.

Now we observe that

$$\frac{\partial W_0(t)f}{\partial \mathbf{x}}(\mathbf{x},\mathbf{y}) = \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x} - \mathbf{u}_0 t, \mathbf{y}), \tag{3.21}$$

while

$$\frac{\partial W_0(t)f}{\partial \mathbf{y}}(\mathbf{x},\mathbf{y}) = \frac{\partial f}{\partial \mathbf{y}}(\mathbf{x}-\mathbf{u}_0 t,\mathbf{y}) - \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}-\mathbf{u}_0 t,\mathbf{y})\frac{\partial \mathbf{u}_0}{\partial \mathbf{y}}t. \quad (3.22)$$

It follows that

$$\|\text{grad } W(t)f\|_p \le \text{const}(1+|t|)\,\|\text{grad } f\|_p. \quad (3.23)$$

Moreover, if we choose numbers r and R such that $f(\mathbf{x},\mathbf{y})$ vanishes whenever $|x|>R$, $|y|>R$ and $|\text{grad } H_0|<r$, then clearly we have

$$\begin{aligned}\|\text{grad } V(W_0(t)\chi)\|_\infty &= \|W_0(-t)(\text{grad } V)\chi\|_\infty \\ &\le \sup\{|(\text{grad } V)(\mathbf{x}+\mathbf{u}_0 t,\mathbf{y}): |\mathbf{x}|\le R, |\mathbf{y}|\le R, \\ &\quad |\mathbf{u}_0|\ge r\} \\ &\le \sup\{\text{const}\,|\mathbf{x}+\mathbf{u}_0 t|^{-2-\eta}: |\mathbf{x}|\le R, \\ &\quad |\mathbf{y}|\le R, |\mathbf{u}_0|\ge r\}. \end{aligned} \quad (3.24)$$

But if $|\mathbf{x}|\le R$, $|\mathbf{y}|\le R$, $|\mathbf{u}_0|\ge r$, then clearly for large $|t|$,

$$|\mathbf{x}+\mathbf{u}_0 t| \ge |\mathbf{u}_0|\,|t| - |\mathbf{x}| \ge r|t| - R. \quad (3.25)$$

It follows that for large $|t|$,

$$\|\text{grad } VW_0(t)\chi\|_\infty \le \text{const}\,(r|t|-R)^{-2-\eta}. \quad (3.26)$$

Combining (3.23) and (3.26) we find

$$\begin{aligned}\|M(t)f\|_p &= \|MW_0(t)f\|_p \\ &\le \text{const}\,(1+|t|)(r|t|-R)^{-2-\eta}\,\|\text{grad } f\|_p, \end{aligned} \quad (3.27)$$

where the constant depends only on H_0 and V. Thus we see that $\|M(t)f\|_p$ satisfies the condition (3.12) for all p, $1\le p\le\infty$, and hence that $W(s,t)f$ converges strongly in the $\mathfrak{L}_p$ norm as $t\to-\infty$ for all f in $\mathfrak{D}$.

We have already noted that $\mathfrak{D}$ is dense in $\mathfrak{L}_p(E_{2n})$ for $1\le p<\infty$. When $p=\infty$, however, this is no longer so. It is easy to see, though, that the closure of $\mathfrak{D}$ in $\mathfrak{L}_\infty(E_{2n})$ is the subspace $\mathfrak{C}_\infty(E_{2n}-Z_0)$, consisting of all everywhere continuous functions which vanish at ∞ *and* on the manifold Z_0 of zeros of grad H_0.

Thus we have established the existence of the wave operators $W(s,-\infty)$ with the properties (3.6) and (3.7) in each of the spaces $\mathfrak{L}_p(E_{2n})$, $1\le p<\infty$, and in the space $\mathfrak{C}_\infty(E_{2n}-Z_0)$.

If the measure of the zero set Z_0 is positive in E_{2n}, then our results hold only for the subspaces $\mathfrak{L}_p(E_{2n}-Z_0)$, now no longer dense in $\mathfrak{L}_p(E_{2n})$. On the complementary subspaces $\mathfrak{L}_p(Z_0)$, the operator $W_0(t)$ reduces to the identity and the wave operators exist in the sense of strong convergence only if $W(t)$ also reduces to the identity, i.e., only if H as well as H_0 vanishes on Z_0.

Now we recall that the transition operators $W(s,t)$ acting on $\mathfrak{C}_0(E_{2n})$ are induced by transition mappings $w(t,s)$ acting on E_{2n}, where

$$w(t,s) = w_0(-t)w(t-s)w_0(s). \quad (3.28)$$

It is therefore tempting to conjecture that the wave operators $W(s,-\infty)$ are similarly induced by mappings of the form $w(-\infty,s)$, which are limits in a suitable sense of the transition mappings $w(t,s)$ as $t\to-\infty$. Our next result makes this conjecture precise.

Theorem 3.4: Suppose that H_0 and V satisfy the conditions (3.2) and (3.3), respectively. Then the transition mappings $w(t,s)$ defined by (3.28) converge pointwise on an open subset S_- of E_{2n} as $t\to-\infty$. The limit mappings $w(-\infty,s)$ map this open subset S continuously onto $E_{2n}-Z_0$, and induce the wave operators $W(s,-\infty)$ on $\mathfrak{C}_\infty(E_{2n}-Z_0)$.

Proof: We fix s and consider the transition operator $W(s,t)$: $\mathfrak{C}_\infty(E_{2n}-Z_0)\to\mathfrak{C}_\infty(E_{2n})$. The adjoint operator $W(s,t)^*$: $\mathfrak{C}_\infty(E_{2n}-Z_0)^*\leftarrow\mathfrak{C}_\infty(E_{2n})^*$, then maps the adjoint space $\mathfrak{C}_\infty(E_{2n})^*$, consisting of all finite Borel measures on E_{2n}, into the adjoint space $\mathfrak{C}_\infty(E_{2n}-Z_0)^*$, consisting of all finite Borel measures on $E_{2n}-Z_0$. Since $W(s,t)$ is isometric into, we know that $W(s,t)^*$ is continuous onto. Moreover, since $W(s,t)$ converges strongly on $\mathfrak{C}_\infty(E_{2n}-Z_0)$ to $W(s,-\infty)$, we know that if μ is any measure on E_{2n} and f any function in $\mathfrak{C}_\infty(E_{2n}-Z_0)$, then $W(s,t)^*\mu(f)$ converges to $W(s,-\infty)^*\mu(f)$.

Now if μ is any measure on E_{2n} which is multiplicative, in the sense that $\mu(fg)=\mu(f)\mu(g)$, then $W(s,t)^*\mu$ is also multiplicative, and hence $W(s,-\infty)^*\mu$ is also multiplicative. This conclusion follows immediately from the fact that the transition operators $W(s,t)$, and hence the wave operators $W(s,-\infty)$, all preserve products in $\mathfrak{C}_\infty(E_{2n})$, and hence $(W(s,t)^*\mu)(fg) = \mu(W(s,t)(fg)) = \mu(W(s,t)fW(s,t)g) = \mu(W(s,t)f)\mu(W(s,t)g) = (W(s,t)^*\mu)(f)(W(s,t)^*\mu)(g)$.

But the only multiplicative measures on $E_{2n}-Z_0$ are the unit point measures and the zero measure.[4] We concluded, therefore, that if μ is a point measure on E_{2n}, then $W(s,t)^*\mu$ is either a point measure on $E_{2n}-Z_0$ or else vanishes on $\mathfrak{C}_\infty(E_{2n}-Z_0)$. The same holds true for $W(s,-\infty)^*\mu$. If we now associate with the point $(\mathbf{x},\mathbf{y})$ in E_{2n}, the point measure $\mu(x,y)$ in $\mathfrak{C}_\infty(E_{2n})^*$ and observe that $W(s,t)^*\mu(\mathbf{x},\mathbf{y})$ is then associated with $w(t,s)(\mathbf{x},\mathbf{y})$, then we see that $w(t,s)(\mathbf{x},\mathbf{y})$ either converges to a point in $E_{2n}-Z_0$ or else eventually leaves and remains outside of every compact subset of $E_{2n}-Z_0$, as $t\to-\infty$, since $\lim f(w(t,s)(x,y))$ as $t\to-\infty$ either $=f(w(-\infty,s)(\mathbf{x},\mathbf{y}))$ or else $=0$ for all functions f in $\mathfrak{C}_\infty(E_{2n}-Z_0)$.

In order to distinguish between the two possibilities, we note that if $\lim f(w(t,s)\mathbf{x},\mathbf{y})=0$ as $t\to-\infty$ for all f in $\mathfrak{C}_\infty(E_{2n}-Z_0)$, then $\lim W(s,t)f(\mathbf{x},\mathbf{y}) = W(s,-\infty)f(\mathbf{x},\mathbf{y})=0$ as $t\to-\infty$ for all f in $\mathfrak{C}_\infty(E_{2n}-Z_0)$. Hence if we define S_- as the complement of the closed set in E_{2n} where all functions of the form $W(s,-\infty)f$ vanish, then we see that $\lim w(t,s)(\mathbf{x},\mathbf{y})$ as $t\to-\infty$ converges to a point in $E_{2n}-Z_0$ if $(\mathbf{x},\mathbf{y})\in S_-$, and eventually leaves every compact subset of $E_{2n}-Z_0$ otherwise.

Hence $\lim w(t,s)(\mathbf{x},\mathbf{y})$ as $t\to-\infty$ converges to a point $w(-\infty,s)(\mathbf{x},\mathbf{y})$ in $E_{2n}-Z_0$ (in the topology of E_{2n}) if $(\mathbf{x},\mathbf{y})$ lies in S_- and does not define $w(-\infty,s)(\mathbf{x},\mathbf{y})$ otherwise. It is clear that this limit mapping $w(-\infty,s)$ is continuous and onto from S_- to $E_{2n}-Z_0$, but we cannot conclude that it is one to one.

We can, however, conclude the following:

Corollary 3.5: If K is any compact subset of $E_{2n} - Z_0$, then $w(-\infty, s)^{-1}(K)$ is a compact subset of S_- of the same measure.

Proof: Choose a function $f \in \mathcal{C}_0(E_{2n} - Z_0)$ such that $0 \le f \le 1$ on $E_{2n} - Z_0$, and $f = 1$ on K. Then for all t, $W(s,t)f$ has compact support, and for $|t|$ sufficiently large we know that $\|W(s,-\infty)f - W(s,t)f\|_\infty < \frac{1}{2}$. It follows that the set $K_1 = \{(\mathbf{x},\mathbf{y})\colon W(s,-\infty)f(\mathbf{x},\mathbf{y}) = 1\}$ is contained in the set $K_2 = \{(\mathbf{x},\mathbf{y})\colon W(s,t)f(\mathbf{x},\mathbf{y}) > \frac{1}{2}\}$ for all sufficiently large $|t|$, and K_2 is compact. Since $w(-\infty, s)^{-1}(K)$ is contained in K_1, it is also contained in K_2 and hence is compact.

Now let g be the characteristic function of K, and note that $W(s,-\infty)g$ is then the characteristic function of $w(-\infty, s)^{-1}(K)$. Since $W(s,-\infty)$ is an isometry, we have

$$\begin{aligned}\mu(w(s,-\infty)^{-1}K) &= \|W(s,-\infty)g\|_1 \\ &= \|g\|_1 = \mu(K)\end{aligned}$$

as required.

If we know in addition that the mapping $w(-\infty, s)$ is continuously differentiable, then we can conclude that the Jacobian is nonzero, and hence the mapping $w(-\infty, s)$ is nonsingular. The differentiability of $w(-\infty, s)$ will be established in the next section.

We remark here that the set S_- is invariant under the motion $w(t)$. In fact, $(\mathbf{x},\mathbf{y}) \in E_{2n} - S_-$ if and only if $W(0,-\infty)f(\mathbf{x},\mathbf{y}) = 0$ for all $f \in \mathcal{C}_\infty(E_{2n} - Z_0)$, and hence if and only if $W(0,-\infty)W_0(t)f(\mathbf{x},\mathbf{y}) = W(t)W(0,-\infty)f(\mathbf{x},\mathbf{y}) = W(0,-\infty)f(w(t)(\mathbf{x},\mathbf{y})) = 0$ for all $f \in \mathcal{C}_\infty(E_{2n} - Z_0)$, i.e., if and only if $w(t)(\mathbf{x},\mathbf{y}) \in E_{2n} - S_-$.

In problems of physical interest the points in the set S_- describe the unbound states of the system, i.e., the states which behave asymptotically like noninteracting states as $t \to -\infty$. To show this, it suffices to show that as $t \to \infty$, $w(t)(\mathbf{x}_1,\mathbf{y}_1) \sim w_0(t)(\mathbf{x}_0,\mathbf{y}_0)$ for all $(\mathbf{x}_1,\mathbf{y}_1) \in S_-$ and $(\mathbf{x}_0,\mathbf{y}_0) = w(-\infty, s)(\mathbf{x}_1,\mathbf{y}_1)$.

Theorem 3.5: If H_0 and V satisfy (3.2) and (3.3), respectively, with $\eta > 0$, then as $t \to -\infty$,

$$|w(t)(\mathbf{x}_1,\mathbf{y}_1) - w_0(t)(\mathbf{x}_0,\mathbf{y}_0)| \le \text{const}\ |t|^{1-\eta/2} \qquad (3.29)$$

for all $(\mathbf{x}_1,\mathbf{y}_1) \in S$ and $(\mathbf{x}_0,\mathbf{y}_0) = w(-\infty, 0)(\mathbf{x}_1,\mathbf{y}_1)$.

Proof: Fix $\varphi \in \mathcal{C}_0(E_{2n} - Z_0)$ so that φ is twice continuously differentiable, $0 \le \varphi(\mathbf{x},\mathbf{y}) \le 1$, and for some $\delta > 0$,

$$\varphi(\mathbf{x},\mathbf{y}) = \begin{cases} 1 - |(\mathbf{x},\mathbf{y}) - (\mathbf{x}_0,\mathbf{y}_0)|^2, & \text{if } |(\mathbf{x},\mathbf{y}) - (\mathbf{x}_0,\mathbf{y}_0)| \le \delta \\ < 1 - \delta^2, & \text{if } |(\mathbf{x},\mathbf{y}) - (\mathbf{x}_0,\mathbf{y}_0)| > \delta \end{cases}. \qquad (3.30)$$

Then we know that if $\varphi(\mathbf{x},\mathbf{y}) \ge 1 - \epsilon^2 \ge 1 - \delta^2$, we necessarily have $|(\mathbf{x},\mathbf{y}) - (\mathbf{x}_0,\mathbf{y}_0)| \le \epsilon$.

Now put $\psi = W_-\varphi$, and note that

$$W_0(+t)\varphi(w_0(t)(\mathbf{x}_0,\mathbf{y}_0)) = \varphi(\mathbf{x}_0,\mathbf{y}_0) = 1 \qquad (3.31)$$

and

$$W(+t)\psi(w(t)(\mathbf{x}_1,\mathbf{y}_1)) = \psi(\mathbf{x}_1\mathbf{y}_1) = \varphi(\mathbf{x}_0,\mathbf{y}_0) = 1. \qquad (3.32)$$

Hence if

$$\|W(+t)\psi - W_0(+t)\varphi\|_\infty < \text{const}\ |t|^{-\eta}, \qquad (3.33)$$

then

$$\begin{aligned}&W(+t)\psi(w(t)(\mathbf{x}_1,\mathbf{y}_1)) - W_0(+t)\varphi(w(t)(\mathbf{x}_1,\mathbf{y}_1)) \\ &\quad = 1 - \varphi(w_0(-t)w(t)(\mathbf{x}_1,\mathbf{y}_1)) < \text{const}\ |t|^{-\eta} \qquad (3.34)\end{aligned}$$

from which we conclude that, for $|t|$ sufficiently large,

$$|w_0(-t)w(t)(\mathbf{x}_1,\mathbf{y}_1) - (\mathbf{x}_0,\mathbf{y}_0)| < \text{const}\ |t|^{-\eta/2}. \qquad (3.35)$$

To obtain (3.33), we use (3.27):

$$\begin{aligned}&\|W(+t)\psi - W_0(+t)\varphi\|_\infty \\ &\quad = \|W(0,-\infty)\varphi - W(-t)W_0(+t)\varphi\|_\infty \\ &\quad \le \int_{-\infty}^{t} \|M(\tau)\varphi\|d\tau \\ &\quad \le \text{const}\ |t|^{-\eta} < \delta^2 \text{ for } |t| \text{ large.} \qquad (3.36)\end{aligned}$$

It follows that we have, for all sufficiently large $|t|$,

$$|w_0(-t)w(t)(\mathbf{x}_1,\mathbf{y}_1) - (\mathbf{x}_0,\mathbf{y}_0)| \le \text{const}\ |t|^{-\eta/2}. \qquad (3.37)$$

Finally, we recall that $|w_0(t)(\mathbf{x},\mathbf{y}) - w_0(t)(\mathbf{x}',\mathbf{y}')| \le \text{const}\ |t|\ |(\mathbf{x},\mathbf{y}) - (\mathbf{x}',\mathbf{y}')|$, from which

$$|w(t)(\mathbf{x}_1,\mathbf{y}_1) - w_0(t)(\mathbf{x}_0,\mathbf{y}_0)| \le \text{const}\ |t|^{1-\eta/2} \qquad (3.38)$$

as required.

We are now free to *define* $W(0,-\infty)f$ for any measurable function f on $E_{2n} - Z_0$ by the formula

$$W(0,-\infty)f(\mathbf{x},\mathbf{y}) = f(w(-\infty,0)(\mathbf{x},\mathbf{y})), \qquad \mathbf{x},\mathbf{y} \in S. \qquad (3.39)$$

$W(0,-\infty)f$ is then a measurable function on S_-. In particular, if f is a constant of the (noninteracting) motion determined by H_0, then $W(0,-\infty)f$ is a constant of the (interacting) motion determined by H, since $W(0,-\infty)f(w(t)(\mathbf{x},\mathbf{y})) = f(w(-\infty,0)w(t)(\mathbf{x},\mathbf{y})) = f(w_0(t)w(-\infty,0)(\mathbf{x},\mathbf{y})) = f(w(-\infty,0)(\mathbf{x},\mathbf{y})) = W(0,-\infty)f(\mathbf{x},\mathbf{y})$. Now any function of the momentum coordinates alone is a constant of the noninteraction motion, and in fact such functions form a complete set. It follows that $W(0,-\infty)f$ is then a constant of the interacting motion for all such f, and, when $w(-\infty,0)$ is nonsingular, such functions form a complete set for all intereacting motions in S_-. In particular, we have $W(0,-\infty)H_0 = H$.

4. DIFFERENTIABILITY

In this section we shall prove that the mapping $w(-\infty, 0)$ defined in the last section is differentiable, and hence nonsingular. For this purpose it will suffice to show that the Jacobian matrix $\mathbf{J}(t,0)$ of the mapping $w(t,0)$ converges uniformly on compact subsets of S_- as $t \to -\infty$.

We assume again that $H = H_0 + V$ where H_0 and V satisfy (3.2) and (3.3), respectively, and in addition V is assumed to be $(m+1)$-fold continuously differentiable, and such that

$$|D^k V(\mathbf{x},\mathbf{y})| \le \text{const}(|x|^{-k-1-\eta}), \qquad (4.1)$$

where D^kV denotes any k-fold derivative of V, $k \leq m + 1$, $\eta > 0$.

We begin by writing for the motion $w(t)$,

$$\mathbf{J}(t) = \frac{\partial(\mathbf{x}(t), \mathbf{y}(t))}{\partial(\mathbf{x}(0), \mathbf{y}(0))} = (\mathbf{a}_{ij}(t)), \tag{4.2}$$

where, to ease the notational burden, we have put

$$\mathbf{a}_{11}(t) = \frac{\partial \mathbf{x}(t)}{\partial \mathbf{x}(0)}, \quad \mathbf{a}_{12}(t) = \frac{\partial \mathbf{x}(t)}{\partial \mathbf{y}(0)},$$

$$\mathbf{a}_{21}(t) = \frac{\partial \mathbf{y}(t)}{\partial \mathbf{x}(0)}, \quad \mathbf{a}_{22}(t) = \frac{\partial \mathbf{y}(t)}{\partial \mathbf{y}(0)}.$$

If $\mathbf{J}_0(t)$ is the Jacobian matrix of $w_0(t)$ and $\mathbf{J}(t, s)$ of $w(t, s)$, then evidently $\mathbf{J}(t, s) = \mathbf{J}_0(-t)\mathbf{J}(t - s)\mathbf{J}_0(s)$.

Now from Sec. 2 we know that $w_0(t)(\mathbf{x}, \mathbf{y}) = (\mathbf{x} + \mathbf{u}_0 t, \mathbf{y})$, and hence

$$\mathbf{J}_0(t) = \begin{pmatrix} 1 & \frac{\partial \mathbf{u}_0}{\partial \mathbf{y}} t \\ 0 & 1 \end{pmatrix}, \tag{4.3}$$

where $\mathbf{u}_0 = \partial H_0/\partial \mathbf{y}$. The time dependence of $\mathbf{J}_0(t)$ is given by

$$\frac{d}{dt}\mathbf{J}_0(t) = \mathbf{A}_0 = \begin{pmatrix} 0 & \frac{\partial \mathbf{u}_0}{\partial \mathbf{y}} \\ 0 & 0 \end{pmatrix}. \tag{4.4}$$

From Sec. 3 we know that $w(t)(\mathbf{x}, \mathbf{y}) = (\mathbf{x}(t), \mathbf{y}(t))$, where

$$\begin{aligned}(\mathbf{x}(t), \mathbf{y}(t)) &= (\mathbf{x}(0), \mathbf{y}(0)) + \int_0^t (\mathbf{x}(\tau), \mathbf{y}(\tau))d\tau \\ &= (\mathbf{x}(0), \mathbf{y}(0)) + \int_0^t \left(\frac{\partial H}{\partial \mathbf{y}(\tau)}, -\frac{\partial H}{\partial \mathbf{x}(\tau)}\right) d\tau.\end{aligned} \tag{4.5}$$

It follows that

$$\begin{aligned}\mathbf{a}_{11}(t) = \frac{\partial \mathbf{x}(t)}{\partial \mathbf{x}(0)} &= 1 + \int_0^t \frac{\partial}{\partial \mathbf{x}(0)} \frac{\partial H}{\partial \mathbf{y}(\tau)} d\tau \\ &= 1 + \int_0^t \left(\frac{\partial^2 H}{\partial \mathbf{x}(\tau)\partial \mathbf{y}(\tau)} \mathbf{a}_{11}(\tau) + \frac{\partial^2 H}{\partial \mathbf{y}(\tau)^2} \mathbf{a}_{21}(\tau)\right) d\tau\end{aligned} \tag{4.6}$$

and similarly for the other $\mathbf{a}_{ij}(t)$. The time-dependence of $\mathbf{a}_{11}(t)$ is given by

$$\frac{d}{dt}\mathbf{a}_{11}(t) = \frac{\partial^2 H}{\partial \mathbf{x}(t)\partial \mathbf{y}(t)} \mathbf{a}_{11}(t) + \frac{\partial^2 H}{\partial \mathbf{y}(t)^2} \mathbf{a}_{21}(t) \tag{4.7}$$

and similarly for the other $\mathbf{a}_{ij}(t)$. These computations can be summarized in the form

$$\frac{d}{dt}\mathbf{J}(t) = \mathbf{A}(t)\mathbf{J}(t) \tag{4.8}$$

with $\mathbf{A}(t)$ given by

$$\mathbf{A}(t) = \begin{pmatrix} \frac{\partial^2 H}{\partial \mathbf{x}(t)\partial \mathbf{y}(t)} & \frac{\partial^2 H}{\partial \mathbf{y}(t)^2} \\ -\frac{\partial^2 H}{\partial \mathbf{x}(t)^2} & -\frac{\partial^2 H}{\partial \mathbf{y}(t)\partial \mathbf{x}(t)} \end{pmatrix}. \tag{4.9}$$

Note that (4.9) reduces to (4.4) when H reduces to H_0.

Combining (4.4) and (4.9), we find that the time dependence of $\mathbf{J}(t, 0) = \mathbf{J}_0(-t)\mathbf{J}(t)$ is given by

$$\begin{aligned}\frac{d}{dt}\mathbf{J}(t, 0) &= \frac{d}{dt}(\mathbf{J}_0(-t)\mathbf{J}(t)) \\ &= \mathbf{J}_0(-t)(-\mathbf{A}_0 + \mathbf{A}(t))\mathbf{J}(t) \\ &= \mathbf{B}(t)\mathbf{J}(t, 0)\end{aligned} \tag{4.10}$$

with $\mathbf{B}(t) = \mathbf{J}_0(-t)(\mathbf{A}(t) - \mathbf{A}_0)\mathbf{J}_0(t)$. Accordingly, we can write $\mathbf{J}(t, 0)$ as the unique solution of the matrix differential equation (4.10), subject to the initial condition $\mathbf{J}(0, 0) = \mathbf{I}$:

$$\mathbf{J}(t, 0) = \mathbf{I} + \int_0^t \mathbf{B}(\tau)\mathbf{J}(\tau, 0)d\tau. \tag{4.11}$$

Now if K is any compact subset of S_-, we denote by $\|\mathbf{B}(t)\|_K$ the norm

$$\|\mathbf{B}(t)\|_K = \sup\{\|\mathbf{B}(t)\|: (\mathbf{x}(0), \mathbf{y}(0)) \in K\}, \tag{4.12}$$

where $\mathbf{B}(t)$ is the matrix introduced in (4.10), regarded as a function of the initial point $(\mathbf{x}(0), \mathbf{y}(0))$ of the motion $w(t)$, and $\|\mathbf{B}(t)\|$ is its matrix norm.

Theorem 4.1: If the matrix $B(t)$ of (4.9) satisfies

$$\int_{-\infty}^0 \|\mathbf{B}(t)\|_K dt < \infty, \tag{4.13}$$

then $\lim \mathbf{J}(t, 0) = \mathbf{J}(-\infty, 0)$ as $t \to -\infty$ converges uniformly on every compact subset K of S_-.

Proof: For $t < 0$, we rewrite (4.11) as

$$\mathbf{J}(t, o) = \mathbf{I} - \int_t^0 \mathbf{B}(\tau)\mathbf{J}(\tau, 0)d\tau. \tag{4.14}$$

An n-fold iteration of (4.14) gives

$$\begin{aligned}\mathbf{J}(t, 0) = \mathbf{I} &- \int_t^0 \mathbf{B}(\tau_1)d\tau_1 \\ &+ \int_t^0 \int_\tau^0 \mathbf{B}(\tau_1)\mathbf{B}(\tau_2)d\tau_2 d\tau_1 + \cdots \\ &+ (-1)^n \int_t^0 \int_{\tau_1}^0 \cdots \int_{\tau_n}^0 \mathbf{B}(\tau_1)\mathbf{B}(\tau_2) \\ &\quad \cdots \mathbf{B}(\tau_n)d\tau_n \cdots d\tau_2 d\tau_1 \\ &+ (-1)^{n+1} \int_t^0 \int_{\tau_1}^0 \cdots \int_{\tau_n}^0 \mathbf{B}(\tau_1)\mathbf{B}(\tau_2) \\ &\quad \cdots \mathbf{B}(\tau_n)\mathbf{J}(\tau_{n+1}, 0)d\tau_{n+1} \cdots d\tau_2 d\tau_1.\end{aligned} \tag{4.15}$$

Introducing the time-ordering operator P by

$$P(\mathbf{B}(\tau_1) \cdots \mathbf{B}(\tau_n)) = \mathbf{B}(\tau_{j(1)}) \cdots \mathbf{B}(\tau_{j(n)}), \tag{4.16}$$

where $\tau_{j(i)}$ is any permutation of τ_i for which $t \leq \tau_{j(1)} \leq \cdots \leq \tau_{j(n)} \leq 0$, we may express (4.15) as

$$\begin{aligned}\mathbf{J}(t, 0) = \mathbf{I} &- \int_t^0 \mathbf{B}(\tau_1)d\tau_1 \\ &+ \frac{1}{2!} \int_t^0 \int_t^0 P((\mathbf{B}(\tau_1)\mathbf{B}(\tau_2))d\tau_2 d\tau_1 + \cdots \\ &+ \frac{(-1)^n}{n!} \int_t^0 \int_t^0 \cdots \int_t^0 P(\mathbf{B}(\tau_1)\mathbf{B}(\tau_2) \\ &\quad \cdots \mathbf{B}(\tau_n))d\tau_n \cdots d\tau_2 d\tau_1 \\ &+ \frac{(-1)^{n+1}}{(n+1)!} \int_t^0 \int_t^0 \cdots \int_t^0 P(\mathbf{B}(\tau_1) \\ &\quad \cdots \mathbf{B}(\tau_n)\mathbf{J}(\tau_{n+1}, 0))d\tau_{n+1} \cdots d\tau_2 d\tau_1\end{aligned} \tag{4.17}$$

It is clear that for fixed t the matrix norm of the remainder term $\mathbf{R}_n(t)$ of this expansion is bounded uniformly on compact subsets of S_- by

$$\|\mathbf{R}_n(t)\|_K \leq \frac{t^{n+1}}{(n+1)!} \left(\sup_{t \leq \tau \leq 0} \|\mathbf{B}(\tau)\|_K\right)^n \sup_{t \leq \tau \leq 0} \|\mathbf{J}(\tau, 0)\|_K, \tag{4.18}$$

from which it follows that $\|\mathbf{R}_n(t)\|_K \to 0$ as $n \to \infty$, so that the associated series expansion converges in matrix norm uniformly on K to $\mathbf{J}(t, 0)$.

Now for any t, the matrix norm of the integrand of the nth term in the resulting series is bounded by

$$\|P(\mathbf{B}(\tau_1) \cdots \mathbf{B}(\tau_n))\|_K \leq \|\mathbf{B}(\tau_1)\|_K \cdots \|\mathbf{B}(\tau_n)\|_K. \quad (4.19)$$

In view of (4.13) it follows that the nth term is absolutely integrable for all t, and is bounded by

$$\|\mathbf{J}_n(t, 0)\|_K \leq \frac{1}{n!} \left(\int_{-\infty}^{0} \|B(\tau)\|_K d\tau\right)^n, \quad (4.20)$$

which is independent of t. Hence we may compute the lim $\mathbf{J}(t, 0)$ as $t \to -\infty$ term by term in the expansion (4.17), knowing that the resulting series will converge to $\mathbf{J}(-\infty, 0)$ uniformly on K:

$$\mathbf{J}(-\infty, 0) = \mathbf{I} - \int_{-\infty}^{0} \mathbf{B}(\tau)d\tau + \cdots + \frac{(-1)^n}{n!} \int_{-\infty}^{0} \cdots \times \int_{-\infty}^{0} P(\mathbf{B}(\tau_1) \cdots \mathbf{B}(\tau_n))d\tau_n \cdots d\tau_1 + \cdots \quad (4.21)$$

In particular, it follows that

$$\|\mathbf{J}(-\infty, 0)\|_K \leq \exp \int_{-\infty}^{0} \|\mathbf{B}(\tau)\|_K d\tau. \quad (4.22)$$

A similar argument will establish the limit as $t \to +\infty$.

Theorem 4.2: Suppose $H = H_0 + V$, where H_0 satisfies (3.2) and V satisfies (3.3) and (4.1) with $m = 1$, $\eta > 0$. Then lim $\mathbf{J}(t, 0) = \mathbf{J}(-\infty, 0)$ as $t \to -\infty$ converges uniformly on every compact subset K of S_-.

Proof: Under these assumptions, we have

$$\mathbf{A}_0 = \begin{pmatrix} 0 & \frac{\partial^2 H_0}{\partial \mathbf{y}(t)^2} \\ 0 & 0 \end{pmatrix} \quad (4.23)$$

and

$$\mathbf{A}(t) = \begin{pmatrix} \frac{\partial^2 V}{\partial \mathbf{x}(t)\partial \mathbf{y}(t)} & \frac{\partial^2 H_0}{\partial \mathbf{y}(t)^2} + \frac{\partial^2 V}{\partial \mathbf{y}(t)^2} \\ -\frac{\partial^2 V}{\partial \mathbf{x}(t)^2} & -\frac{\partial^2 V}{\partial \mathbf{y}(t)\partial \mathbf{x}(t)} \end{pmatrix}, \quad (4.24)$$

where $(\mathbf{x}(t), \mathbf{y}(t)) = w(t)(\mathbf{x}, \mathbf{y})$. Hence

$$\mathbf{B}(t) = \mathbf{J}_0(-t)(\mathbf{A}(t) - \mathbf{A}_0)\mathbf{J}(t)$$

$$= \begin{pmatrix} 1 & -\frac{\partial H}{\partial \mathbf{y}(t)} t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial^2 V}{\partial \mathbf{x}(t)\mathbf{y}(t)} & \frac{\partial^2 V}{\partial \mathbf{y}(t)^2} \\ -\frac{\partial^2 V}{\partial \mathbf{x}(t)^2} & -\frac{\partial^2 V}{\partial \mathbf{y}(t)\partial \mathbf{x}(t)} \end{pmatrix} \times \begin{pmatrix} 1 & \frac{\partial H}{\partial \mathbf{y}(t)} t \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \left(\frac{\partial^2 V}{\partial \mathbf{x}(t)\partial \mathbf{y}(t)} + \frac{\partial^2 V}{\partial \mathbf{x}(t)^2} t\right) & \left(\frac{\partial^2 V}{\partial \mathbf{y}(t)^2} - \frac{\partial^2 V}{\partial \mathbf{x}(t)^2} t^2\right) \\ -\frac{\partial^2 V}{\partial \mathbf{x}(t)^2} & -\left(\frac{\partial^2 V}{\partial \mathbf{y}(t)\partial \mathbf{x}(t)} + \frac{\partial^2 V}{\partial \mathbf{x}(t)^2} t\right) \end{pmatrix}. \quad (4.25)$$

Now if $(\mathbf{x}, \mathbf{y}) \in K \cap \mathbf{S}$, then by (3.29) we have $|w(t)(\mathbf{x}, \mathbf{y}) - w_0(t)(\mathbf{x}_0\mathbf{y}_0)| < \text{const}(|t|^{1-\eta/2})$, where $\eta > 0$, $(\mathbf{x}_0\mathbf{y}_0) = w(-\infty, 0)(\mathbf{x}, \mathbf{y})$, and the constant depends only on K. Since $w_0(t)(\mathbf{x}_0, \mathbf{y}_0) = (\mathbf{x}_0 + \mathbf{u}_0 t, \mathbf{y}_0)$, it follows that $|\mathbf{x}(t) - (\mathbf{x}_0 + \mathbf{u}_0 t)| < \text{const}(|t|^{1-\eta/2})$, so that $|x(t)| > |\mathbf{u}_0|\,|t| - \text{const}(|t|^{1-\eta/2}) > \text{const}\,|t|$, for all sufficiently large $|t|$.

Now if V satisfies (4.1), then clearly the entries $\mathbf{b}_{ij}|t|$ of the matrix $\mathbf{B}(t)$ of (4.25) all satisfy

$$|\mathbf{b}_{ij}(t)| \leq \text{const}(|t|^{-1-\eta}) \quad (4.26)$$

from which we derive the hypothesis, and hence the conclusion of Theorem 4.1.

A variation of the preceding argument will show that, under the assumption (4.1) for all $k \leq m + 1$, the mapping $w(-\infty, 0)$ is in fact m times differentiable. For this it suffices to show that all the partial derivatives up to order m converge uniformly on all compact subsets of S_-. We shall content ourselves here with a proof for the case $m = 2$; the general result follows by induction on m.

We begin with the expansion (4.21) for $\mathbf{J}(t, 0)$, and compute $\text{grad}_0\, \mathbf{J}(t, 0)$, the vector matrix whose entries are the gradients of the entries of $\mathbf{J}(t, 0)$ with respect to the initial point of the motion $w(t)$. For any fixed t, $w(t, 0)$ is twice differentiable and

$$\text{grad}_0\, \mathbf{J}(t, 0) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \times \int_t^0 \cdots \int_t^0 \text{grad}_0\, P(\mathbf{B}(\tau_1) \cdots \mathbf{B}(\tau_n))\, d\tau_n \cdots d\tau_1, \quad (4.27)$$

the sum being uniformly convergent as before. But

$$\|\text{grad}_0\, P(\mathbf{B}(\tau_1) \cdots \mathbf{B}(\tau_n))\|_K \leq \sum_{i=1}^{n} \|P(\mathbf{B}(\tau_1) \cdots \text{grad}_0 \mathbf{B}(\tau_i) \cdots \mathbf{B}(\tau_n))\|_K \leq \sum_{i=1}^{n} \|\mathbf{B}(\tau_1)\|_K \cdots \|\text{grad}_0 \mathbf{B}(\tau_i)\|_K \cdots \|\mathbf{B}(\tau_n)\|_K. \quad (4.28)$$

Hence the nth term in the expansion (4.27) is bounded by

$$\|\text{grad}_0\, \mathbf{J}_n(t, 0)\|_K \leq \frac{n}{n!} \left(\int_t^0 \|\mathbf{B}(\tau)\|_K d\tau\right)^{n-1} \times \int_t^0 \|\text{grad}_0\, \mathbf{B}(\tau)\|_K d\tau. \quad (4.29)$$

We know that $\int_t^0 \|\mathbf{B}(\tau)\|_K d\tau$ is bounded for all t, and we assert that the same holds for $\int_t^0 \|\text{grad}_0 \mathbf{B}(\tau)\|_K d\tau$. In fact, we have

$$\text{grad}_0\, \mathbf{B}(t) = \text{grad}_0(\mathbf{b}_{ij}(t)) = ((\text{grad}\, \mathbf{b}_{ij})(t))\mathbf{J}(t) = ((\text{grad}\, \mathbf{b}_{ij})(t))\mathbf{J}_0(t)\mathbf{J}(t, 0), \quad (4.30)$$

where the entries $\mathbf{b}_{ij}(t)$ are given in (4.25). It follows that grad $\mathbf{b}_{ij}(t)$ consists of linear combinations of third partial derivatives of V, all of which, according to (4.1) are bounded by const $|t|^{-4-\eta}$. The entries of $\mathbf{J}_0(t)$ are bounded by const $|t|$, while the entries of $\mathbf{J}(t,0)$ are bounded by const. Hence,

$$\|\text{grad}_0\, \mathbf{B}(\tau)\|_K \leq \text{const}\ |t|^{-1-\eta} \tag{4.31}$$

and our assertion follows.

Thus we see that the expansion for $\text{grad}_0\, \mathbf{J}(t,0)$ converges term by term uniformly on K as $t \to -\infty$, and that the resulting series converges uniformly on K to $\text{grad}_0\, \mathbf{J}(-\infty, 0)$, as required. We summarize these results as follows.

Corollary 4.3: If V satisfies (4.1) for $0 \leq k \leq m+1$, then $w(-\infty, 0)$ is m times continuously differentiable.

Using Theorem 4.2, we can now strengthen the conclusion of Theorem 3.3, as follows. We denote by $C_0^m(E_{2n} - Z_0)$ the space of all m-fold continuously differentiable functions with compact support in $E_{2n} - Z_0$, and introduce the norm $\|\ \|_{\infty,m}$ by

$$\|f\|_{\infty,m} = \sup\{\|D^k f\|_\infty : 0 \leq k \leq m\}, \tag{4.32}$$

where $D^k f$ denotes any k-fold partial derivative of f. We put $C_\infty^m(E_{2n} - Z_0)$ for the completion of $C_0^m(E_{2n} - Z_0)$ under the norm (4.32).

Corollary 4.4: If V satisfies (4.1) for all $k \leq m+1$, then $\lim W(0,t) = W(0,-\infty)$ as $t \to -\infty$ converges strongly in $C_0^m(E_{2n} - Z_0)$.

Proof: We present the proof for $m = 1$; the general case follows by induction on m. Consider

$$\begin{aligned} \text{grad}\, W(0,t) f(\mathbf{x},\mathbf{y}) &= \text{grad}\, f(w(t,0)(\mathbf{x},\mathbf{y})) \\ &= (\text{grad}\, f)(w(t,0)\mathbf{x},\mathbf{y})\mathbf{J}(t,0) \\ &= \mathbf{J}(t,0) W(0,t)(\text{grad}\, f). \end{aligned} \tag{4.33}$$

It follows that if $f \in C_0^2(E_{2n} - Z_0)$,

$$\begin{aligned} \|W(0,-\infty)f - W(0,t)f\|_{\infty,1} &= \|D(W(0,-\infty)f - W(0,t)f)\|_\infty \\ &\leq \|\text{grad}(W(0,-\infty)f - W(0,t)f)\|_\infty \\ &\quad \times \|(\mathbf{J}(-\infty,0)W(0,-\infty) - \mathbf{J}(t,0)W(0,t))\text{grad}\, f\|_\infty \\ &\leq \text{const}\, \|(W(0,-\infty) - W(0,t))\text{grad}\, f\|_\infty \\ &\leq \text{const}\ |t|^{-\eta}. \end{aligned} \tag{4.34}$$

Here we have used the fact that $\mathbf{J}(t,0)$ is bounded as $t \to -\infty$; the rest follows as in Sec. 3.

In particular, if f is m times differentiable, so is $W(0,-\infty)f$. The results of this section are equally valid, of course, for $t \to +\infty$.

5. BOUND STATES

It is useful to provide an independent characterization of the set $S_- \cup S_+$ of unbound states. In this section, we shall show that under suitable additional hypotheses, the complementary set $B = E_{2n} - S_- \cup S_+$ of bound states has a compact intersection with each level surface of the total energy function H.

For this purpose, we shall assume again that $H = H_0 + V$, where H_0 and V satisfy (3.1) and (3.2). In addition, we shall assume that $H_0(\mathbf{y}) \geq 0$, and that *the level surfaces of* $H_0(\mathbf{y})$ *are compact* in $E_n - Z_0$. For this it suffices to assume, for instance, that the Hessian matrix $\partial^2 H_0/\partial y_i \partial y_j$ is everywhere a positive definite quadratic form bounded away from zero, as it is in the classical case $H_0 = \frac{1}{2}\mathbf{y}^2$.

Throughout this section, we shall write $(\mathbf{x}(t), \mathbf{y}(t))$ for $w(t)(\mathbf{x},\mathbf{y})$, $(\mathbf{u}(t), \mathbf{v}(t))$ for $(d/dt)(\mathbf{x}(t), \mathbf{y}(t))$ and E for $H(\mathbf{x}(t), \mathbf{y}(t))$. We recall that E is the total energy associated with the trajectory $(\mathbf{x}(t), \mathbf{y}(t))$ and is a constant of the motion.

We shall say that the point $(\mathbf{x},\mathbf{y}) \in E_{2n}$ is an *incoming (outgoing) point* if the inner product $\mathbf{x}\cdot(\partial H/\partial \mathbf{y}) < 0$ (> 0). Evidently if $(\mathbf{x}(t), \mathbf{y}(t))$ is an incoming (outgoing) point then we have $(d/dt)|\mathbf{x}(t)|^2 = 2(\mathbf{x}(t)\cdot(\partial H/\partial \mathbf{y})(t)) < 0$ (> 0).

Lemma 5.1: For each $E > 0$ there exists a constant $R > 1$ such that if $(\mathbf{x}(t), \mathbf{y}(t))$ is any trajectory with $E \leq H(\mathbf{x}(t), \mathbf{y}(t)) \leq 2E$ and $(\mathbf{x}(0), \mathbf{y}(0))$ is an incoming point with $|\mathbf{x}(0)| > 2R$, then for all $t < 0$ we have

$$|x(t)| \geq R + \text{const}\ |t|. \tag{5.1}$$

Proof: Given $E > 0$, we define

$$m = \min\left\{\left|\frac{\partial H_0}{\partial \mathbf{y}}\right| : \tfrac{1}{2}E < H_0(\mathbf{y}) < \tfrac{5}{2}E\right\}, \tag{5.2}$$

$$M = \max\left\{\left|\frac{\partial^2 H}{\partial \mathbf{y}\partial \mathbf{y}}\right| : \tfrac{1}{2}E < H_0(\mathbf{y}) < \tfrac{5}{2}E\right\}. \tag{5.3}$$

Our assumptions on H_0 ensure that $m > 0$ and $M < \infty$.

It follows directly from (3.3) that

$$|V(\mathbf{x},\mathbf{y})| \leq \text{const}\ |\mathbf{x}|^{-1-\eta}, \quad \eta > 0. \tag{5.4}$$

Hence we may choose $R \geq 1$ so that, if $|\mathbf{x}| > R$,

$$\begin{aligned} &\text{(a)}\ |V(\mathbf{x})| < \tfrac{1}{2}E, \\ &\text{(b)}\ \left|\frac{\partial V}{\partial \mathbf{y}}\right| < \tfrac{1}{8}m, \\ &\text{(c)}\ \left|\frac{\partial V}{\partial \mathbf{x}}\right| < \frac{1}{8\pi}\frac{m^2}{M}|\mathbf{x}|^{-2}. \end{aligned} \tag{5.5}$$

Then for all points $(\mathbf{x}(t), \mathbf{y}(t))$ with $E < H(\mathbf{x}(t), \mathbf{y}(t)) < 2E$ and $|\mathbf{x}(t)| > R$, we know that

$$\begin{aligned} |H_0(\mathbf{y}(t)) - E| &= |H_0(\mathbf{y}(t)) - H(\mathbf{x}(t), \mathbf{y}(t))| \\ &\leq |V(\mathbf{x}(t), \mathbf{y}(t))| < \tfrac{1}{2}E \end{aligned} \tag{5.6}$$

and hence that

$$\left|\frac{\partial H_0}{\partial \mathbf{y}(t)}\right| \geq m, \tag{5.7}$$

$$\left|\frac{\partial^2 H_0}{\partial \mathbf{y}(t)\partial \mathbf{y}(t)}\right| \leq M. \tag{5.8}$$

Now we introduce the *rectilinear* trajectory $(\mathbf{x}_0(t), \mathbf{y}_0(t))$ by

$$\begin{aligned} \mathbf{x}_0(t) &= \mathbf{x}(0) + t\mathbf{u}(0), \\ \mathbf{y}_0(t) &= \mathbf{y}(0) \end{aligned} \tag{5.9}$$

and define the set

$$I=\{t\le 0:\ |\mathbf{x}(t)-\mathbf{x}_0(t)|<\tfrac{1}{2}m|t| \quad \text{and} \quad |\mathbf{y}(t)-\mathbf{y}_0(t)|<\tfrac{1}{4}m/M\}. \tag{5.10}$$

Evidently $t=0\in I$, and I is *open* in the left half line. We shall show that I is also *closed*.

It suffices to assume that $\tau\in I$ for all τ such that $t<\tau\le 0$, and prove that then $t\in I$. In this case we know that for all such τ,

$$\begin{aligned}|\mathbf{x}(\tau)|&>|\mathbf{x}_0(\tau)|-\tfrac{1}{2}m|\tau|\\&=|\mathbf{x}(0)+\mathbf{u}(0)\tau|-\tfrac{1}{2}m|\tau|\\&>(|\mathbf{x}(0)|^2+|\mathbf{u}(0)|^2\tau^2)^{1/2}-\tfrac{1}{2}m|\tau|\\&>\tfrac{1}{2}(|\mathbf{x}(0)|^2+|\mathbf{u}(0)|^2\tau^2)^{1/2}>\tfrac{1}{2}|\mathbf{x}(0)|>R,\end{aligned} \tag{5.11}$$

where we have used the hypothesis that $\mathbf{x}(0)\cdot\mathbf{u}(0)<0$ and $t<0$. It follows that for all such τ conditions (5.5), (5.7), and (5.8) hold at $(\mathbf{x}(\tau),\mathbf{y}(\tau))$. Hence

$$\begin{aligned}|\mathbf{x}(t)-\mathbf{x}_0(t)|&\le\int_t^0\Big|\frac{\partial H}{\partial\mathbf{y}(\tau)}-\frac{\partial H}{\partial\mathbf{y}(0)}\Big|d\tau\\&\le\int_t^0\Big|\frac{\partial H_0}{\partial\mathbf{y}(\tau)}-\frac{\partial H_0}{\partial\mathbf{y}(0)}\Big|d\tau\\&+\int_t^0\Big|\frac{\partial V}{\partial\mathbf{y}(\tau)}-\frac{\partial V}{\partial\mathbf{y}(0)}\Big|d\tau\\&<\int_t^0 M|\mathbf{y}(\tau)-\mathbf{y}(0)|d\tau+\int_t^0\tfrac{1}{4}m d\tau\\&<\tfrac{1}{2}m|t|.\end{aligned} \tag{5.12}$$

Similarly,

$$\begin{aligned}|\mathbf{y}(t)-\mathbf{y}_0(t)|&\le\int_t^0\Big|\frac{\partial V}{\partial\mathbf{x}(\tau)}\Big|d\tau\\&<\frac{1}{8\pi}\frac{m^2}{M}\int_t^0\frac{d\tau}{|\mathbf{x}(\tau)|^2}\\&<\frac{1}{8\pi}\frac{m^2}{M}\int_t^0\frac{4d\tau}{|\mathbf{x}(0)|^2+|\mathbf{u}(0)|^2\tau^2}\\&=\frac{1}{8\pi}\frac{m^2}{M}\Big(\frac{2\pi}{|\mathbf{x}(0)|\,|\mathbf{u}(0)|}\Big)<\frac{1}{4}\frac{m}{M}\end{aligned} \tag{5.13}$$

Thus we see that if $\tau\in I$ for all τ, $t<\tau\le 0$, then $t\in I$, and hence $t\in I$ for all $t<0$. In particular, (5.11) holds for all t, so that

$$\begin{aligned}|x(t)|&\ge\tfrac{1}{2}(|\mathbf{x}(0)|^2+|\mathbf{u}(0)|^2t^2)^{1/2}\\&\ge R+\tfrac{1}{2}(\sqrt{2}|\mathbf{u}(0)|\,|t|)\ge R+\tfrac{1}{4}m|t|,\end{aligned} \tag{5.14}$$

as required.

We can now give a sufficient condition that a point $(\mathbf{x},\mathbf{y})\in E_{2n}-Z_0$ lie in S_-.

Theorem 5.2: Every incoming point $(\mathbf{x}_0,\mathbf{y}_0)\in E_{2n}$ of energy $H(\mathbf{x}_0,\mathbf{y}_0)>0$ lies in S_- if $|\mathbf{x}_0|>2R$, where R depends only on E.

Proof: Given $(\mathbf{x}_0,\mathbf{y}_0)\in E_{2n}$ with energy $H(\mathbf{x}_0\mathbf{y}_0)>0$, choose $E=\frac{2}{3}H(\mathbf{x}_0,\mathbf{y}_0)$ and R by Lemma 5.1. Now we can find a neighborhood U of $(\mathbf{x}_0,\mathbf{y}_0)$ such that every point in U is an incoming point of energy lying between E and $2E$, for which Lemma 5.1 holds. In particular, if $(\mathbf{x},\mathbf{y})\in U$ then $|\mathbf{x}(t)|>R+\frac{1}{4}m|t|$ for all $t<0$, where m depends only on E.

Now choose $\varphi\in C_0^2(E_{2n})$, any function positive at (x_0,y_0) and vanishing off U, and consider $W(t,0)\varphi$. As in Sec. 3, we have

$$\begin{aligned}\frac{d}{dt}W(t,0)\varphi&=\frac{d}{dt}W_0(-t)W(t)\varphi\\&=iM(t)W(t,0)\varphi.\end{aligned} \tag{5.15}$$

Hence

$$W(t,0)=i\int_t^0 M(\tau)W(\tau,0)\varphi d\tau \tag{5.16}$$

and

$$\begin{aligned}\|W(t',0)\varphi-W(t'',0)\varphi\|_p&\le\int_{t'}^{t''}\|M(\tau)W(\tau,0)\varphi\|_p d\tau\\&=\int_{t'}^{t''}\|MW(\tau)\varphi\|_p d\tau.\end{aligned} \tag{5.17}$$

Now

$$\begin{aligned}MW(\tau)\varphi&=i\{V,W(\tau)\}\varphi\\&=i\sum\frac{\partial V}{\partial\mathbf{y}}\frac{\partial W(\tau)\varphi}{\partial\mathbf{x}}-\frac{\partial V}{\partial\mathbf{x}}\frac{\partial W(\tau)\varphi}{\partial\mathbf{y}}.\end{aligned} \tag{5.18}$$

So

$$\begin{aligned}\|MW(\tau)\varphi\|_p&\le\|\,|\mathrm{grad}V|\,|\mathrm{grad}W(\tau)\varphi|\,\|_p\\&\le\|(\mathrm{grad}V)(W(\tau)\chi\|_\infty\|\mathrm{grad}W(\tau)\varphi\|_p,\end{aligned} \tag{5.19}$$

where χ is the characteristic function of the (compact) support K of φ. Now (cf 3.21)

$$\mathrm{grad}W(\tau)\varphi=J(\tau)W(\tau)\,\mathrm{grad}\varphi, \tag{5.20}$$

so that

$$\|\mathrm{grad}W(\tau)\varphi\|_p\le\|J(\tau)\chi\|_\infty\|W(\tau)\,\mathrm{grad}\varphi\|_p. \tag{5.21}$$

But

$$\|W(\tau)\,\mathrm{grad}\varphi\|_p=\|\mathrm{grad}\varphi\|_p, \tag{5.22}$$

and

$$\begin{aligned}\|J(\tau)\chi\|_\infty&=\|J_0(\tau)J(\tau,0)\chi\|_\infty\\&\le\|J_0(\tau)W(\tau,0)\chi\|_\infty\|J(\tau,0)\chi\|_\infty\\&\le\mathrm{const}(1+|\tau|),\end{aligned} \tag{5.23}$$

where we have used the fact that $J(\tau,0)$ is uniformly bounded on compact sets for all τ, and $J_0(\tau)$ is uniformly bounded on compact sets by const$(1+|\tau|)$.

On the other hand, we have

$$\begin{aligned}\|(\mathrm{grad}V)(W(\tau)\chi\|_\infty&=\|W(-\tau)(\mathrm{grad}V)\chi\|_\infty\\&=\sup\{|(\mathrm{grad}V)(\mathbf{x}(\tau),\mathbf{y}(\tau))|:\ (\mathbf{x},\mathbf{y})\in K\}\\&\le\sup\{\mathrm{const}\,|\mathbf{x}(\tau)|^{-2-\eta}:\ (\mathbf{x},\mathbf{y})\in K\}\\&\le\mathrm{const}(R+\tfrac{1}{4}m|\tau|)^{-2-\eta},\end{aligned} \tag{5.24}$$

where we have used (3.2) and (5.14). Combining (5.21), (5.23), and (5.24), we find

$$\|MW(\tau)\|_p\le\mathrm{const}(1+|\tau|)(R+\tfrac{1}{4}m|\tau|)^{-2-\eta}\|\mathrm{grad}\varphi\|_p \tag{5.25}$$

from which it follows that $W(t,0)\varphi$ converges to $W(-\infty,0)\varphi$ as $t\to-\infty$, in such a way that

$$\|W(-\infty,0)\varphi-W(t,0)\varphi\|\le\mathrm{const}\,|t|^{-\eta}. \tag{5.26}$$

Since $W(t,0)$ is the inverse of $W(0,t)$ for all t, evidently $W(-\infty,0)$ is the inverse of $W(0,-\infty)$ wherever it

exists. Thus if we put $\psi = W(-\infty, 0)\varphi$, then $\varphi = W(0, -\infty)\psi$, so that $\varphi \in$ range $W(0, -\infty)$, and since φ is positive at $(\mathbf{x}_0, \mathbf{y}_0)$, it follows that $(\mathbf{x}_0, \mathbf{y}_0) \in S_-$.

A similar result holds for the limit $t \to +\infty$. In this case we find that every outgoing state $(\mathbf{x}_0, \mathbf{y}_0) \in E_{2n}$ of energy $H(\mathbf{x}_0, \mathbf{y}_0)$ lying between E and $2E$ lies in S_+ if $|\mathbf{x}_0| > 2R$, where R depends only on E.

If we define the set B of bound states as the complement $E_{2n} - (S_+ \cup S_-)$, then we can prove

Theorem 5.3: Every bound state $(\mathbf{x}_0, \mathbf{y}_0)$ of energy $H(\mathbf{x}_0, \mathbf{y}_0)$ lying between $E > 0$ and $2E$ must satisfy $|\mathbf{x}| \leq R$, $|\mathbf{y}| \leq R$ for some R depending only on E.

Proof: Choose $R = \max\{2R_-, 2R_+\}$, where R_- is given by Theorem 5.2 and R_+ is given by the companion result for the limit $t \to +\infty$. Then if $|\mathbf{x}_0| > R$ and $(\mathbf{x}_0, \mathbf{y}_0)$ is an incoming or outgoing state, then $(\mathbf{x}_0, \mathbf{y}_0)$ lies in S_- or S_+ by Theorem 5.2. If $(\mathbf{x}_0, \mathbf{y}_0)$ is neither, then $(\mathbf{x}_0, \mathbf{u}_0) = 0$, and we can write $(\mathbf{x}_0, \mathbf{y}_0)$ as the limit of a sequence of incoming states $(\mathbf{x}_n, \mathbf{y}_n)$ for which $(\mathbf{x}_n, \mathbf{u}_n) < 0$. Now from Lemma 5.1 we know that $|\mathbf{x}_n(t)| > R + \frac{1}{4}m|t|$ for all $t < 0$, where m is independent of n for n sufficiently large. It follows that $|\mathbf{x}_0(t)| > R + \frac{1}{4}m|t| > |\mathbf{x}_0|$ if $|t|$ is sufficiently large. Hence we know that for some τ, $t < \tau < 0$, $|\mathbf{x}(\tau)| > R$ and $(d/d\tau)|\mathbf{x}(\tau)| < 0$. Hence $(\mathbf{x}(\tau), \mathbf{y}(\tau))$ is an incoming state of the same energy as $(\mathbf{x}_0, \mathbf{y}_0)$, and so lies in S_-. Since S_- is invariant under the motion, $(\mathbf{x}_0, \mathbf{y}_0)$ also lies in S_-.

Corollary: The set of all bound states of given total energy is compact.

6. SCATTERED STATES

We have defined in Sec. 3 the set $S = S_-$ so that S_- is the complete inverse image of $E_{2n} - Z_0$ under the mapping $w(-\infty, 0)$. We have shown in Sec. 4 that $w(-\infty, 0)$ is nonsingular and hence invertible. It is clear that the inverse $w(0, -\infty)$ can also be realized as the limit as $t \to -\infty$ of the inverses $w(0, t)$ of $w(t, 0)$.

We have also defined the set S_+ as the complete inverse image of $E_{2n} - Z_0$ under the mapping $w(+\infty, 0)$, which is also invertible, and can be realized as the limit as $t \to +\infty$ of the inverses $w(0, t)$ of $w(t, 0)$.

Now on the set $S = S_+ \cap S_-$ we can form the mapping

$$w(\infty, -\infty) = \lim_{t \to +\infty} \lim_{t \to -\infty} w(s, t)\colon E_{2n} - Z_0 \to E_{2n} - Z_0$$

and interpret $w(+\infty, -\infty)$ as a *scattering mapping* which compares the asymptotic behavior as $t \to \pm\infty$ of the incoming and outgoing trajectories $(\mathbf{x}(t), \mathbf{y}(t))$ of the *scattered* states $(\mathbf{x}(0), y(0)) \in S$.

It need not be true, in general, that $S = S_+ = S_-$. We shall show in this section, however, that under the assumptions of the last section the symmetric difference $(S_+ \cup S_-) - S = E_{2n} - (B \cup S)$ has measure zero. It follows that S is not empty. It also follows that if $1 \leq p < \infty$ then $\mathcal{L}^p(S_+) \cong \mathcal{L}^p(S_-) \cong \mathcal{L}^p(S)$. Since $\mathcal{L}^p(S_\pm) = W(0, \pm\infty)\mathcal{L}^p(E_{2n})$, we have $W(0, \pm\infty)\mathcal{L}^p(E_{2n}) \cong \mathcal{L}^p(S)$. This means that in $\mathcal{L}^p(E_{2n})$, $1 \leq p < \infty$, we may define the *scattering operator*

$$W(+\infty, -\infty) = \lim_{s \to +\infty} \lim_{t \to -\infty} W(s, t)$$

and prove that this operator maps $\mathcal{L}^p(E_{2n})$ isometrically onto itself.

Now if $S_- - S$ has positive measure, then it contains a compact subset K of positive measure together with all of the translates $w(t)K$ of K under the motion $w(t)$. Hence to show $S_- - S$ has measure zero, it will suffice to show that if K is any compact subset of S_- of positive measure, then for some t, $w(t)K \cap S_+$ also has positive measure.

Lemma 6.1: If K is any compact subset of S_- of positive measure, then (under the assumptions of Section 5) $w(t)K \cap S_+$ has positive measure for some $t > 0$.

Proof: Let K_1 and K_2 be any two compact subsets of E_{2n}, and χ_1 and χ_2 their characteristic functions. If $K_2 \subset S_-$, then $\chi_2 = W(0, -\infty)\chi_3$, where χ_3 is the characteristic function of $K_3 = w(-\infty, 0)K_2$. Hence

$$\begin{aligned}\|\chi_1 W(-t)\chi_2\|_1 &= \|\chi_1 W(-t)W(0, -\infty)\chi_3\|_1 \\ &= \|\chi_1 W(0, -\infty)W_0(-t)\chi_3\|_1 \\ &= (\chi_1, W(0, -\infty)W_0(-t)\chi_3)_2 \\ &= (W(0, -\infty)^*\chi_1, W_0(-t)\chi_3)_2. \qquad (6.1)\end{aligned}$$

Now this last term is of the form $(f, W_0(-t)g)_2$ for $f, g \in \mathcal{L}^2(E_{2n})$. Since the group of isometries $W_0(-t)$ is generated by an operator L_0 of absolutely continuous spectrum (see Sec. 2), we know from the Lemma of Riemann–Lebesgue that $(f, W_0(-t)g)_2 \to 0$ as $t \to -\infty$ for all $f, g \in \mathcal{L}^2(E_{2n})$. Hence for all sufficiently large $t > 0$,

$$\|\chi_1 W(-t)\chi_2\|_1 \leq \tfrac{1}{2}\|\chi_1\chi_2\|_1 \qquad (6.2)$$

or equivalently

$$\mu(K_1 \cap w(t)K_2) \leq \tfrac{1}{2}\mu(K_1 \cap K_2). \qquad (6.3)$$

Now given $E > 0$, choose R by Lemma 5.1 and T so that if $|x| < 2R$ and $H(\mathbf{x}, \mathbf{y}) < 2E$, then $|y| < 2T$.

Put $K_1 = \{(\mathbf{x}, \mathbf{y})\colon |\mathbf{x}| < 2R, |\mathbf{y}| < 2T\}$ and $K_2 = K$. Then for all sufficiently large $t > 0$ the measure $\mu(K_1^c \cap w(t)K_2)$ of the set of states $(\mathbf{x}(t), \mathbf{y}(t))$ which lie in $w(t)K_2$ but not in K_1 is at least $\frac{1}{2}\mu(K_1 \cap K_2)$. But every such point is outgoing as it crosses $|\mathbf{x}| = 2R$, and hence must belong to S_+ by Theorem 5.2. For such t, then, $(w(t)K \cap S_+) > \frac{1}{2}\mu(K \cap K_1)$. If $\mu(K \cap K_1) = 0$ for all choices of E, R and T defining K_1, then $\mu(K) = 0$; hence if $\mu(K) > 0$, then $\mu(w(t)K \cap S_+) > 0$, as required.

The same argument applies to $S_+ - S$. Hence,

Corollary 6.2: Under the same assumptions, the difference $E_{2n} - (S \cup B)$ has measure zero.

Corollary 6.3: Under the same assumptions, we have for all p, $1 \leq p < \infty$

$$W(0, \pm\infty)\mathcal{L}^p(E_{2n}) = \mathcal{L}^p(S_\pm) \cong \mathcal{L}^p(S). \qquad (6.4)$$

Finally, we know that on $\mathcal{L}^p(S_\pm)$ the limit $\lim_{s \to +\infty} W(s, 0)$ converges strongly to an isometry inverse to $W(0, \pm\infty)$. In fact, if $W(0, +\infty)f = g$, then

$$\|f - W(s, 0)g\|_p = \|f - W(s, 0)W(0, +\infty)f\|_p$$

$$= \|W(0, s)f - W(0, +\infty)f\|_p,$$
$$\to 0 \text{ as } s \to +\infty, \qquad (6.5)$$

and similarly for $s \to -\infty$.

Hence in view of Corollary 6.3 we have

Corollary 6.4: Under the same assumptions the

$$\lim_{s\to+\infty} \lim_{t\to-\infty} W(s, t)$$

converges strongly to an isometry $W(+\infty, -\infty) = W(+\infty, 0)W(0, -\infty)$ of $\mathcal{L}^p(E_{2n})$ onto itself for $1 \le p < \infty$.

7. DISCUSSION

A few comments are now in order.

(7.1.) The results of Sec. 3 may be regarded as a contribution to the study of the stability of differential systems under perturbations. We are dealing here with "incompressible" autonomous systems, for which the standard methods of stability theory do not apply. The method developed here provides another approach to the study of such systems, but depends essentially on the incompressible character of both the perturbed and unperturbed motions. It may be possible, however, to extend this method to suitably defined "almost incompressible" or "asymptotically incompressible" motions.

(7.2.) Condition (3.3) on the perturbation $V(\mathbf{x})$ is essential, in view of the fact that the results of Sec. 3 are known to fail for Coulomb forces, for which $V(\mathbf{x}) = \text{const}|\mathbf{x}|^{-1}$.

(7.3.) It is tempting to conjecture that the results of Sec. 6 can be improved in various ways. For instance, it seems plausible that $S_+ = S_-$. But there are systems with trajectories $(\mathbf{x}, (t), \mathbf{y}(t))$ for which $\lim|\mathbf{x}(t)| = \infty$ as $t \to -\infty$, but $\lim|\mathbf{x}(t)| = 0$ as $t \to +\infty$. Such a trajectory describes the motion of an incoming particle which spends its whole lifetime climbing a potential hill, ultimately coming to rest at the top. All the points on this trajectory then lie in S_-, but not in S_+. We have shown that the set of all such points must have measure zero under the hypotheses of Sec. 5.

It is also plausible that $|\mathbf{x}(t)| \ge \text{const}(1 + |t|)$ as $t \to +\infty$ uniformly on compact subsets of S_-. But there are systems in which an incoming particle may enter a potential cavity with a very narrow entrance, and then spend an arbitrary long time finding its way out again; for such a particle $|\mathbf{x}(t)| \le \text{const}$ for arbitrarily large t. It may be possible to establish a uniform asymptotic behavior for $t \to +\infty$ on compact subsets of S_- in the absence of potential cavities, a possibility suggested by analogous results recently obtained for acoustical scattering from convex boundaries.

Note added in manuscript: After completing this manuscript, the author learned of a paper by W. Hunziker[5] dealing with the same subject. Hunziker considers the much more difficult many-body problem, but restricts his attention, as does Cook,[2] to finite-range forces. He indicates in Sec. 5, however, how to extend his proof of the existence of the wave operators to include infinite-range forces; our proof of Theorem 3.3 is then a special case of his. He does not indicate how to prove the existence, differentiability, or asymptotic completeness of the wave mappings for such forces; perhaps our methods will provide a starting point for an attack on these questions.

---

[1] J. M. Cook, J. Math. Phys. **36**, 82 (1957).

[2] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U.P., Cambridge, 1937); H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1950).

[3] J. M. Cook, "Banach Algebras and Asymptotic Mechanics," Institute D'Etudes Scientifiques de Cargese, 1965.

[4] C. E. Rickart, *General Theory of Banach Algebras* (Nostrand, New York, 1960).

[5] W. Hunziker, Commun. Math. Phys. **8**, 282 (1968).

Contractions of the Low-Dimensional Real Lie Algebras

C. W. Conatser

Department of Mathematics, Texas Tech University, Lubbock, Texas 79409

(Received 8 May 1970)

A complete and detailed classification and analysis of all the Inonu-Wigner contractions of all the real Lie algebras of dimension 1, 2, and 3 is presented. Starting with a more natural classification of the algebras, many corrections and completions are made to the earlier results of Sharp. Among other things, a proper contraction is constructed between two three-dimensional Lie algebras both of which have two-dimensional derived algebras—a phenomenon previously claimed impossible.

INTRODUCTION

The purpose of this paper is to present a complete and detailed classification and analysis of all the Inonu-Wigner contractions (IWC's) of all the real Lie algebras of dimension 1, 2, and 3.

Despite the growing interest in contractions in both mathematics and physics, the only analysis of this type is in a somewhat obscure report by Sharp.[1] Unfortunately Sharp's analysis contains a number of errors in and omissions of some of the most interesting cases. Surprisingly, Sharp's work has been referenced throughout the literature with no recognition of its errors and has thus led to numerous incomplete and, in certain instances even inaccurate results. The author will be dealing with some of these in another paper.

A repertoire of accurate and detailed (e.g., the dimensionality of the various subalgebras involved, which

$$= \|W(0, s)f - W(0, +\infty)f\|_p,$$
$$\to 0 \text{ as } s \to +\infty, \qquad (6.5)$$

and similarly for $s \to -\infty$.

Hence in view of Corollary 6.3 we have

Corollary 6.4: Under the same assumptions the

$$\lim_{s\to+\infty} \lim_{t\to-\infty} W(s, t)$$

converges strongly to an isometry $W(+\infty, -\infty) = W(+\infty, 0)W(0, -\infty)$ of $\mathcal{L}^p(E_{2n})$ onto itself for $1 \le p < \infty$.

7. DISCUSSION

A few comments are now in order.

(7.1.) The results of Sec. 3 may be regarded as a contribution to the study of the stability of differential systems under perturbations. We are dealing here with "incompressible" autonomous systems, for which the standard methods of stability theory do not apply. The method developed here provides another approach to the study of such systems, but depends essentially on the incompressible character of both the perturbed and unperturbed motions. It may be possible, however, to extend this method to suitably defined "almost incompressible" or "asymptotically incompressible" motions.

(7.2.) Condition (3.3) on the perturbation $V(\mathbf{x})$ is essential, in view of the fact that the results of Sec. 3 are known to fail for Coulomb forces, for which $V(\mathbf{x}) = \text{const}|\mathbf{x}|^{-1}$.

(7.3.) It is tempting to conjecture that the results of Sec. 6 can be improved in various ways. For instance, it seems plausible that $S_+ = S_-$. But there are systems with trajectories $(\mathbf{x}, (t), \mathbf{y}(t))$ for which $\lim|\mathbf{x}(t)| = \infty$ as $t \to -\infty$, but $\lim|\mathbf{x}(t)| = 0$ as $t \to +\infty$. Such a trajectory describes the motion of an incoming particle which spends its whole lifetime climbing a potential hill, ultimately coming to rest at the top. All the points on this trajectory then lie in S_-, but not in S_+. We have shown that the set of all such points must have measure zero under the hypotheses of Sec. 5.

It is also plausible that $|\mathbf{x}(t)| \ge \text{const}(1 + |t|)$ as $t \to +\infty$ uniformly on compact subsets of S_-. But there are systems in which an incoming particle may enter a potential cavity with a very narrow entrance, and then spend an arbitrary long time finding its way out again; for such a particle $|\mathbf{x}(t)| \le \text{const}$ for arbitrarily large t. It may be possible to establish a uniform asymptotic behavior for $t \to +\infty$ on compact subsets of S_- in the absence of potential cavities, a possibility suggested by analogous results recently obtained for acoustical scattering from convex boundaries.

Note added in manuscript: After completing this manuscript, the author learned of a paper by W. Hunziker[5] dealing with the same subject. Hunziker considers the much more difficult many-body problem, but restricts his attention, as does Cook,[2] to finite-range forces. He indicates in Sec. 5, however, how to extend his proof of the existence of the wave operators to include infinite-range forces; our proof of Theorem 3.3 is then a special case of his. He does not indicate how to prove the existence, differentiability, or asymptotic completeness of the wave mappings for such forces; perhaps our methods will provide a starting point for an attack on these questions.

---

[1] J. M. Cook, J. Math. Phys. **36**, 82 (1957).

[2] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U.P., Cambridge, 1937); H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1950).

[3] J. M. Cook, "Banach Algebras and Asymptotic Mechanics," Institute D'Etudes Scientifiques de Cargese, 1965.

[4] C. E. Rickart, *General Theory of Banach Algebras* (Nostrand, New York, 1960).

[5] W. Hunziker, Commun. Math. Phys. **8**, 282 (1968).

Contractions of the Low-Dimensional Real Lie Algebras

C. W. Conatser

Department of Mathematics, Texas Tech University, Lubbock, Texas 79409

(Received 8 May 1970)

A complete and detailed classification and analysis of all the Inonu-Wigner contractions of all the real Lie algebras of dimension 1, 2, and 3 is presented. Starting with a more natural classification of the algebras, many corrections and completions are made to the earlier results of Sharp. Among other things, a proper contraction is constructed between two three-dimensional Lie algebras both of which have two-dimensional derived algebras—a phenomenon previously claimed impossible.

INTRODUCTION

The purpose of this paper is to present a complete and detailed classification and analysis of all the Inonu-Wigner contractions (IWC's) of all the real Lie algebras of dimension 1, 2, and 3.

Despite the growing interest in contractions in both mathematics and physics, the only analysis of this type is in a somewhat obscure report by Sharp.[1] Unfortunately Sharp's analysis contains a number of errors in and omissions of some of the most interesting cases. Surprisingly, Sharp's work has been referenced throughout the literature with no recognition of its errors and has thus led to numerous incomplete and, in certain instances even inaccurate results. The author will be dealing with some of these in another paper.

A repertoire of accurate and detailed (e.g., the dimensionality of the various subalgebras involved, which

subalgebras lead to which IWC's, and which algebras can be obtained as IWC's) examples is essential in "guessing" general theorems and checking counter-examples. In order to provide such a repertoire and to stop the perpetuation of the errors mentioned above, the author has carried out the analysis presented in the following sections.

1. TERMINOLOGY

Let $G = (V, [\ ,\ ])$ denote a Lie algebra with underlying vector space V and Lie bracket $[\ ,\ ]$. We then, except where specified otherwise, let G' and Z (or Z_G if G needs emphasis) denote the derived algebra of G and the center of G, respectively. If $W \subset G$ is a Lie subalgebra of G then let $n = \dim(G)$, $n' = \dim(G')$, $n_Z = \dim(Z)$, and $n_W = \dim(W)$.

For $t \in [0, 1]$, let A_t be a linear operator on V, with A_t nonsingular for $t \in (0, 1]$. Define, for each $t \in (0, 1]$, a new Lie bracket on V by

$$[X, Y]_t = A_t^{-1}[A_t X, A_t Y]$$

for each $X, Y \in V$ and let $G_t = (V, [\ ,\ ]_t)$. Suppose there are complementary subspaces W and U of V such that the A_t "block decompose", re (i.e., with respect to) the decomposition $V = W \dotplus U$, into the form

$$A_t = \begin{pmatrix} I_1 & 0 \\ 0 & 0 \end{pmatrix} + t \begin{pmatrix} D & 0 \\ 0 & I_2 \end{pmatrix} \tag{1.1}$$

for all $t \in [0, 1]$, where I_1 and I_2 are identity operators on W and U, and D is an operator on W.

The limit

$$\lim_{t \to 0} [X, Y]_t = [X, Y]_\infty$$

exists for all $X, Y \in V$ iff $[\ ,\ ]_\infty$ is a Lie bracket on V. Whenever this occurs, the resulting Lie algebra $G_\infty = (V, [\ ,\ ]_\infty)$ is called an *Inonu-Wigner contraction* (IWC) of the original Lie algebra $G = (V, [\ ,\ ])$. Inonu and Wigner[2,3] discovered that this happens iff W is a subalgebra of G. Moreover the outcome of an IWC is not affected by the operator D, so there is no loss in generality by our considering only $D = 0$.

It can also be shown[4] that, for a fixed subalgebra W, any other complementary (to W in V) subspace U_1 together with the new operators of the form (1) re the decomposition $V = W \dotplus U_1$ yields the same (i.e., isomorphic) Lie algebra. Hence we may unambiguously write G_∞^W and call this *the IWC of G re the subalgebra W*. Furthermore, every IWC of a given Lie algebra G is obtained if we (i) take each subalgebra W of G; (ii) fix a basis $\{X_1, \ldots, X_{n_W}\}$ for W; (iii) complete this to a basis $\{X_1, \ldots, X_n\}$ for V; (iv) define $A_t X_i = X_i$ for $1 \le i \le n_W$, $A_t X_i = tX_i$ for $n_W < i \le n$, and extend by linearity; and (v) determine the resulting G_∞^W.

Following the usual convention, we define an algebra by specifying the commutation relations on a fixed basis with only the nonzero ones being actually written down, except for special emphasis. Since any proper (i.e., $G_\infty \not\cong G$) IWC results in a semidirect sum, the semidirect decompositions of the various algebra are of special interest and are hence pointed out in the classification. We omit repeating that the k-dimensional Abelian Lie algebra A^k is the direct sum of 1-dimensional (Abelian) algebras. Some other notation used is now listed.

$sp\{X_1, \ldots, X_k\}$ = the vector subspace with basis $\{X_1, \ldots, X_k\}$.

$L\{X_1, \ldots, X_k\}$ = the Lie subalgebra with basis $\{X_1, \ldots, X_k\}$.

$P \oplus K$ = Lie algebra direct sum of the Lie algebras P and K.

$P \oplus_\pi K$ = Lie algebra semidirect sum of the Lie algebras P and K, where π is a representation of K in P by derivations.

2. REAL LIE ALGEBRAS OF LOW DIMENSION

For purposes of uniformity and convenience, we now classify and establish notation for the real Lie algebras with $n = 1, 2$, and 3. The bases for these algebras are chosen to best facilitate our later calculations of all the IW contractions of these algebras and, at the same time, to minimize the number of nonzero structure constants. For other discussions of low-dimensional Lie algebras, the reader is referred to Jacobson[5] and Talman.[6]

A. $n = 1$

The only one-dimensional Lie algebra we denote by $A_1 (\simeq A^1)$. For any basis $\{X_1\}$, A_1 is defined by

$$[X_1, X_1] = 0 \tag{2.1}$$

and, of course, $n_z = n = 1$ and $n' = 0$.

B. $n = 2$

We classify these according to value of n'. Since $n' = 2$ is clearly impossible with $n = 2$ we have only the following two cases.

1. $n' = 0$: ($n_Z = 2$) Denote it $B_1 (\simeq A^2$, since Abelian). For any basis $\{X_1, X_2\}$, B_1 is defined by

$$[X_1, X_2] = 0. \tag{2.2}$$

2. $n' = 1$: ($n_Z = 0$) Denote it B_2. Let $X_1 \in B_2'$ ($X_1 \neq 0$) so that $B_2' = L\{X_1\}$ and choose X_2 so that $B_2 = L\{X_1, X_2\}$ and scaled so that B_2 is defined by

$$[X_2, X_1] = X_1, \tag{2.3}$$

sometimes called the "$ax + b$ (or affine) algebra", since it is the Lie algebra of the "$ax + b$ (or affine) group" in one dimension, B_2 is noncompact (i.e., it is the Lie algebra of no compact Lie group) and can be represented by 2×2 matrices by letting

$$X_1 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad X_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

We also note that B_2 is complete,[5] is solvable but not nilpotent, has $L\{X_1\}$ as its only proper ideal, and is the semidirect sum

$$B_2 = A_1 \oplus_\pi A_1 = L\{X_1\} \oplus_\pi L\{X_2\},$$

where π is the scalar representation.

C. $n = 3$

Again analyzing re the value of n', we have the following cases:

1. $n' = 0$: ($n_Z = 3$) Denote it $C_1 (\simeq A^3$, since Abelian). For any basis $\{X_1, X_2, X_3\}$, C_1 is defined by

$$[X_i, X_j] = 0. \tag{2.4}$$

2. $n' = 1$: ($n_Z = 1$, by a simple dimension argument.) Choose X_1 so that $G' = L\{X_1\}$. Any choice of X_2 and X_3 so that $G = L\{X_1, X_2, X_3\}$ gives the form $[X_1, X_2] = aX_1, [X_1, X_3] = bX_1, [X_2, X_3] = cX_1$ and the following two cases arise.

Case (a). $G' = Z$: ($a = 0 = b$) Denote it C_2 and scale so that $c = -1$ so that C_2 is defined by

$$[X_1, X_2] = 0 = [X_1, X_3], \quad [X_3, X_2] = X_1. \tag{2.5}$$

Often called the Heisenberg algebra, C_2 is the only non-Abelian nilpotent algebra with $n \leqq 3$ and can be represented by 3×3 matrices by letting

$$X_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Notice that the only one-dimensional ideal is G', all the two-dimensional ideals are Abelian ($\simeq B_1$) and are of the form $L\{X_1, aX_2 + bX_3 = Y_2\}$, with $a \neq 0$ or $b \neq 0$, and C_2 is of the form $C_2 = B_1 \oplus_\pi A_1 = (L\{X_1\} \oplus L\{Y_2\}) \oplus_\pi L\{Y_3\}$, where $Y_3 = cX_2 + dX_3$, with $ad - bc \neq 0$, and π is the representation defined by $\pi(Y_3)$: $X_1 \to 0$; $Y_2 \to (ad\text{-}bc)X_1$.

Case (b) $G' \neq Z$: Here choose X_2 and X_3 so that $Z = L\{X_2\}$, so that $a = 0 = c$, and scaled so that $b = -1$. Denoting this unique algebra by C_3, it is defined by

$$[X_1, X_2] = 0 = [X_2, X_3], \quad [X_3, X_1] = X_1 \tag{2.6}$$

and can be represented by 3×3 matrices by letting

$$X_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

or by 2×2 matrices by letting

$$X_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

In contrast to C_2, C_3 has two one-dimensional ideals G' and Z. All the two-dimensional ideals of C_3 can be written in exactly the same form as for C_2, but, unlike that case, $L\{X_1, X_2\}$ alone is Abelian ($\simeq B_1$) and all the others, of the form $L\{X_1, aX_2 + X_3 = Y_3\}$, are non-Abelian ($\simeq B_2$). Thus we have

$$\begin{aligned} C_3 &= B_1 \oplus_{\pi_1} A_1 = (L\{X_1\} \oplus L\{X_2\}) \oplus_{\pi_1} L\{X_3\} \\ &= B_2 \oplus A_1 = (L\{X_1\} \oplus_\pi L\{Y_3\}) \oplus L\{X_2\}, \end{aligned}$$

where π was defined re B_2 and π_1 is the representation defined by

$$\pi_1(X_3): X_1 \to X_1 : X_2 \to 0.$$

3. $n' = 2$: ($n_Z = 0$) Since G' is an ideal in G, if $G' = B_2$ then its completeness[5] would lead to $G = B_2 \oplus A_1$ and thus to $G' = B_2'$, contrary to the dimensionalities of G' being 2 and of B_2' being 1. Hence $G' = B_1$ and we see that every algebra in this class is solvable, but not nilpotent. Moreover, it is now clear that, for any choice of X_1, X_2, and X_3 so that $G' = L\{X_1, X_2\}$ and $G = L\{X_1, X_2, X_3\}$, the adjoint action of X_3 is that of a nonsingular linear operator on the two-dimensional space $G' = L\{X_1, X_2\}$. Letting λ_1 and λ_2 be the roots of the characteristic equation of this operator (note: $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$ always), we classify these algebras in the following way:

(a) λ_1 *and* λ_2 *real and their eigenspaces span* G': In this case choose X_1 and X_2 to be eigenvectors corresponding to λ_1 and λ_2, respectively, scale X_3 so that $\lambda_1 = 1$, and denote the resulting λ_2 by λ and the algebra by $C_4(\lambda)$. Then $C_4(\lambda)$ is defined by

$$[X_1, X_2] = 0, \quad [X_3, X_1] = X_1, \quad [X_3, X_2] = \lambda X_2 \tag{2.7}$$

and can be represented by letting

$$X_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

It should be pointed out that $C_4(\lambda) \simeq C_4(1/\lambda)$ [let $X_1' = X_2, X_2' = X_1$, and $X_3' = (1/\lambda)X_3$].

Each of these algebras only has proper ideals of the form:

$$L\{X_1\} \simeq A_1, \quad L\{X_2\} \simeq A_1, \quad \text{and } L\{X_1, X_2\} = C_4(\lambda)' \simeq B_1,$$

except for $C_4(1)$ which, since here every element of $C_4(1)'$ is an eigenvector, also has all the subalgebras of the form $L\{aX_1 + bX_2\} \simeq A_1, a \neq 0$ or $b \neq 0$, as ideals. Thus we have $C_4(\lambda) = B_1 \oplus_\pi A_1 = (L\{X_1\} \oplus L\{X_2\}) \oplus_\pi L\{X_3\}$, where each $\pi = \pi(\lambda)$ is the direct sum of two scalar representations.

(b) $\lambda_1 = \lambda_2$ *real but not diagonalizable:* In this case choose X_1 and X_2 so that ad X_3 is triangular with X_1 as eigenvector, scale X_3 so that $\lambda_1 = \lambda_2 = 1$, and scale X_1 so that we finally have

$$[X_1, X_2] = 0, \quad [X_3, X_1] = X_1, \quad [X_3, X_2] = X_1 + X_2. \tag{2.8}$$

Denoting this algebra by C_5, we see that it can be represented by setting

$$X_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Notice that the only proper ideals of C_5 are $L\{X_1\} \simeq A_1$, and $L\{X_1, X_2\} = C_5' \simeq B_1$ and we have

$$C_5 = B_1 \oplus_\pi A_1 = (L\{X_1\} \oplus L\{X_2\}) \oplus_\pi L\{X_3\},$$

where π is clear by now.

(c) λ_1 *and* λ_2 *complex:* ($\lambda_1 = \bar{\lambda}_2$) In this case, choose X_1 and X_2 as the "real" and "imaginary" components of the eigenvector (in the complexification of G') corresponding to $\lambda_1 = a + ib$ ($b \neq 0$) so

that $[X_3, X_1 + iX_2] = (a + ib)(X_1 + iX_2)$. Taking real and imaginary parts of this equation, we get

$$[X_3, X_1] = aX_1 - bX_2, \quad [X_3, X_2] = bX_1 + aX_2.$$

We then scale X_3 so that $b = -1$, denote the resulting a by λ, and denote the resulting algebra by $C_6(\lambda)$, which is thus defined by

$$[X_1, X_2] = 0, \quad [X_3, X_1] = X_2 + \lambda X_1, \quad [X_3, X_2] = -X_1 + \lambda X_2. \tag{2.9}$$

Moreover $C_6(\lambda)$ can be represented by letting

$$X_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad X_4 = \begin{pmatrix} \lambda & -1 & 0 \\ 1 & \lambda & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

These $C_6(\lambda)$ contain no one-dimensional ideals and only one two-dimensional one, namely $L\{X_1, X_2\} = C_6(\lambda)' \simeq B_1$, and we have

$$C_6(\lambda) = B_1 \oplus_\pi A_1 = (L\{X_1\} \oplus L\{X_2\}) \oplus_\pi L\{X_3\}.$$

We should note that $C_6(0)$ is the Lie algebra of the group of Euclidean motions in the plane.

4. $n' = 3$: ($n_Z = 0$) In this case X_1, X_2, X_3 can be chosen so that $[X_1, X_2] = X_3$ and $G = L\{X_1, X_2, X_3\}$. Although the details are not interesting for our purposes and are hence omitted, a straightforward calculation shows that there are exactly two distinct algebras in this class. Both these algebras are simple (i.e., have no proper ideals) and, by additional judiciousness in the choice of X_1, X_2, X_3, we are led to the following two cases:

Case (a): Denoted by C_7 and defined by

$$[X_1, X_2] = X_3, \quad [X_2, X_3] = X_1, \quad [X_3, X_1] = X_2. \tag{2.10}$$

This is the Lie algebra of the three-dimensional rotation group $SO(3)$ and the special two-dimensional unitary group $SU(2)$ and is hence also denoted $so(3)$ and $su(2)$ and can be represented by letting

$$X_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \quad X_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

It should be noted that C_7 is the compact real form of the Lie algebra $sl(2, C)$.

Case (b): Denoted by C_8 and defined by

$$[X_1, X_2] = X_3, \quad [X_2, X_3] = -X_1, \quad [X_3, X_1] = -X_2. \tag{2.11}$$

This is the Lie algebra of the two-dimensional Lorentz group $SO(2, 1)$ and is hence also denoted $so(2, 1)$ and can be represented by

$$X_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

It should be pointed out that C_8 is also the Lie algebra $sl(2, R)$ of the two-dimensional special linear group $SL(2, R)$ and is the noncompact (contains the noncompact B_2 as a subalgebra) real form of $sl(2, C)$. When viewed as $sl(2, R)$, the "standard" basis and representation chosen are

$$h = 2X_1, \quad e = X_2 + X_3, \quad f = X_2 - X_3,$$

so that

$$[e, f] = h, \quad [h, e] = 2e, \quad [h, f] = -2f,$$

and

$$h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad e = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad f = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

3. INONU-WIGNER CONTRACTIONS

We now determine all of the Inonu-Wigner contractions (IWC's) of the real Lie algebras with $n \leq 3$, using the notation established in Sec. 2. In defining the maps A_t by telling what they do to a specified basis, we are repeatedly using the fact that an IWC is completely determined by the subalgebra W, which the contraction is with respect to, in the sense that G_∞^W is independent of the complementary subspace (of W in G) chosen. We now list (without proof here) a few other basic theorems about IWC's which are used repeatedly and without comment throughout this section:

(i) $G_\infty^G = G$ for any Lie algebra G.
(ii) $G_\infty^{\{0\}} = A^n (n = \dim G)$ for any Lie algebra G.
(iii) If W is a central subalgebra of G (i.e., $W \subset Z \subset G$), then $G_\infty^W = A^n (n = \dim G)$.
(iv) $n'_\infty \leq n'$ for any IWC.[4]
(v) $(A^n)_\infty^W = A^n$ for any subalgebra $W \subset A^n$.

Without further refreshers we proceed directly to the calculations at hand.

A_1: Since $A_1 = A^1$, $A_{1\infty} = A_1$ always.

B_1: Since $B_1 = A^2$, $B_{1\infty} = B_1$ always.

B_2: Two cases occur here:

$W = L\{X_1\} = B_2'$:
$A_t: X_1 \to X_1; X_2 \to tX_2$,
$[X_2, X_1]_t = tA_t^{-1}(X_1) = tX_1 \to 0 : (B_2)_\infty^W = B_1$.

$W = W(a) = L\{aX_1 + X_2 = Y_2\} \neq B_2'$:
$A_t: X_1 \to tX_1; Y_2 \to Y_2$,
$[Y_2, X_1]_t = tA_t^{-1}(X_1) = X_1 \to X_1 : B_{2\infty}^{W(a)} = B_2$.

C_1: Since $C_1 = A^3$, $C_{1\infty} = C_1$ always.

C_2: We consider first $n_W = 1$ and then $n_W = 2$:

$W = L\{X_1\} = C_2' = Z$ yields $C_{2\infty} = C_1$.
$W = W(a, b, c) = L\{aX_1 + bX_2 + cX_3 = Y_2\}$ with $d = b^2 + c^2 \neq 0$.

Let $Y_3 = (-c/d)X_2 + (b/d)X_2$:

$$\left.\begin{aligned} &A_t: X_1 \to tX_1; Y_2 \to Y_2; Y_3 \to tY_3; \\ &[X_1, Y_2]_t = [X_1, Y_3]_t = 0 \to 0 \\ &[Y_3, Y_2]_t = tA_t^{-1}(X_1) = X_1 \to X_1 \end{aligned}\right\} : C_{2\infty} = C_2.$$

$W = W(a,b) = L\{X_1, aX_2 + bX_3 = Y_2\}$ with $d = a^2 + b^2 \neq 0$.

Let $Y_3 = (-b/d)X_2 + (a/d)X_3$:

$A_t: X_1 \to X_1; Y_2 \to Y_2; Y_3 \to tY_3;$
$[X_1, Y_2]_t = [X_1, Y_3]_t = 0 \to 0;$
$[Y_3, Y_2]_t = tA_t^{-1}(X_1) = tX_1 \to 0.$

Hence $C_{2\infty}^{W} = C_1$ whenever $n_W = 2$.

Remark: Since IWC with respect to W leaves W unchanged and makes its complementary subspace Abelian, one might make the following:

Conjecture: If $W_1 \subset W_2 \subset G$ are subalgebras and $G_\infty^{W_1} = G$, then $G_\infty^{W_2} = G$.

Counterexample: Let $G = C_2$, $W_1 = W(a,b,c)$, and $W_2 = W(b,c)$, as above. Then $W_1 \subset W_2 \subset G = C_2$ and $C_\infty^{W_1} = C_2$ but $C_{2\infty}^{W_2} \neq C_2$.

C_3: Consider first $n_W = 1$ and then $n_W = 2$:

$W = L\{X_2\} = Z$ yields $C_{3\infty} = C_1$.
$W = W(a) = L\{X_1 + aX_2 = Y_1\},$
$A_t: Y_1 \to Y_1; X_2 \to tX_2; X_3 \to tX_3;$
$[X_2, Y_1]_t = [X_2, X_3]_t = 0 \to 0;$
$[X_3, Y_1]_t = tA_t^{-1}(X_1)$
$= t(X_1 + a(t-1)t^{-1}X_2)$
$= tX_1 + a(t-1)X_2 \to -aX_2.$

Hence, for each $a \neq 0$, $C_{3\infty}^{W(a)} = C_2$, the Heisenberg algebra, and for $a = 0$, $C_{3\infty}^{W(0)} = C_1$ [note: $W(0) = C_3'$]:

$W = W(a,b) = L\{aX_1 + bX_2 + X_3 = Y_3\},$
$A_t: X_1 \to tX_1; X_2 \to tX_2; Y_3 \to Y_3;$
$[X_2, X_1]_t = [X_2, Y_3]_t = 0 \to 0;$
$[Y_3, X_1]_t = A_t^{-1}(X_1) = X_1 \to X_1.$

Hence, for each a and b, $C_{3\infty} = C_3$.

$W = L\{X_1, X_2\}$ $(\approx B_1)$,
$A_t: X_1 \to X_1; X_2 \to X_2; X_3 \to tX_3,$
$[X_2, X_1]_t = [X_2, X_3]_t = 0 \to 0,$
$[X_3, X_1]_t = tA^{-1}(X_1) = tX_1 \to 0.$

Hence $C_{3\infty} = C_1$.

$W = W(a) = L\{X_1, aX_2 + X_3 = Y_3\}$ $(\approx B_2)$,
$A_t: X_1 \to X_1; X_2 \to tX_2; Y_3 \to Y_3,$
$[X_2, X_1]_t = [X_2, Y_3]_t = 0 \to 0,$
$[Y_3, X_1]_t = A_t^{-1}(X_1) = X_1 \to X_1.$

Hence, for each a, $C_{3\infty} = C_3$.

The remaining two-dimensional subspaces are of the form $sp\{aX_1 + X_2, bX_1 + X_3\}$, which are subalgebras iff $a = 0$, so that the only remaining case is the following:

$W = W(b) = L\{X_2, bX_1 + X_3 = Y_3\}$ $(\approx B_1)$,
$A_t: X_1 \to tX_1; X_2 \to X_2; Y_3 \to Y_3,$
$[X_2, X_1]_t = [X_2, Y_3]_t = 0 \to 0,$
$[Y_3, X_1]_t = tA_t^{-1}(X_1) = X_1 \to X_1.$

Hence, for each b, $C_{3\infty} = C_3$.

$C_4(\lambda)$: Again consider $n_W = 1$ and then $n_W = 2$:

$W = L\{X_1\},$
$A_t: X_1 \to X_1; X_2 \to tX_2; X_3 \to tX_3,$
$[X_1, X_2]_t = tA_t^{-1}(0) = 0 \to 0,$
$[X_3, X_1]_t = tA_t^{-1}(X_1) = tX_1 \to 0,$
$[X_3, X_2]_t = t^2 A_t^{-1}(\lambda X_2) = t\lambda X_2 \to 0.$

Hence $C_4(\lambda)_\infty = C_1$.

$W = W(a) = L\{aX_1 + X_2 = Y_2\},$
$A_t: X_1 \to tX_1; Y_2 \to Y_2; X_3 \to tX_3,$
$[X_1, Y_2]_t = tA_t^{-1}(0) = 0 \to 0,$
$[X_3, X_1]_t = t^2 A_t^{-1}(X_1) = tX_1 \to 0,$
$[X_3, Y_2]_t = tA_t^{-1}(aX_1 + \lambda X_2)$
$= aX_1 + t\lambda(a(t-1)t^{-1}X_1 + X_2)$
$= (a + \lambda a(t-1))X_1 + t\lambda X_2 \to a(1-\lambda)X_1.$

For $\lambda = 1$ this gives $C_4(1)_\infty^{W(a)} = C_1$.
For $\lambda \neq 1$ this gives $C_4(\lambda \neq 1)_\infty^{W(0)} = C_1$ for $a = 0$ and $C_4(\lambda \neq 1)_\infty^{W(a)} = C_2$ for $a \neq 0$.

$W = W(a,b) = L\{aX_1 + bX_2 + X_3 = Y_3\},$
$A_t: X_1 \to tX_1; X_2 \to tX_2; Y_3 \to Y_3,$
$[X_1, X_2]_t = t^2 A_t^{-1}(0) = 0 \to 0,$
$[Y_3, X_1]_t = tA_t^{-1}(X_1) = X_1 \to X_1,$
$[Y_3, X_2]_t = tA_t^{-1}(\lambda X_2) = \lambda X_2 \to \lambda X_2.$

Hence $C_4(\lambda)_\infty = C_4(\lambda)$, same λ.

$W = L\{X_1, X_2\} = C_4(\lambda)',$
$A_t: X_1 \to X_1; X_2 \to X_2; X_3 \to tX_3,$
$[X_1, X_2]_t = A_t^{-1}(0) = 0 \to 0,$
$[X_3, X_1]_t = tA_t^{-1}(X_1) = tX_1 \to 0,$
$[X_3, X_2]_t = tA_t^{-1}(\lambda X_2) = t\lambda X_2 \to 0.$

Hence $C_4(\lambda)_\infty^{W} = C_1$.

$W = W(a) = L\{X_1, aX_2 + X_3 = Y_3\},$
$A_t: X_1 \to X_1; X_2 \to tX_2; Y_3 \to Y_3,$
$[X_1, X_2]_t = tA_t^{-1}(0) = 0 \to 0,$
$[Y_3, X_1]_t = A_t^{-1}(X_1) = X_1 \to X_1,$
$[Y_3, X_2]_t = tA_t^{-1}(\lambda X_2) = \lambda X_2 \to \lambda X_2.$

Hence, for each a, $C_4(\lambda)_\infty = C_4(\lambda)$, same λ.

The remaining two-dimensional subspaces are of the form $sp\{aX_1 + X_2 = Y_2, bX_1 + X_3 = Y_3\}$, which are

subalgebras iff $a = 0$ or $\lambda = 1$. Thus the two remaining cases are as follows:

Case $a = 0$:

$$W = W(b) = L\{X_2, bX_1 + X_3 = Y_3\},$$
$$A_t: X_1 \to tX_1; X_2 \to X_2; Y_3 \to Y_3,$$
$$[X_1, X_2]_t = tA_t^{-1}(0) = 0 \to 0,$$
$$[Y_3, X_1]_t = tA_t^{-1}(X_1) = X_1 \to X_1,$$
$$[Y_3, X_2]_t = A_t^{-1}(\lambda X_2) = \lambda X_2 \to \lambda X_2.$$

Hence, for each b, $C_4(\lambda)_\infty = C_4(\lambda)$, same λ.

Case $\lambda = 1$:

$$W = W(a,b) = L\{aX_1 + X_2 = Y_2, bX_1 + X_3 = Y_3\},$$
$$A_t: X_1 \to tX_1; Y_2 \to Y_2; Y_3 \to Y_3,$$
$$[X_1, Y_2]_t = tA_t^{-1}(0) = 0 \to 0,$$
$$[Y_3, X_1]_t = tA_t^{-1}(X_1) = X_1 \to X_1,$$
$$[Y_3, Y_2]_t = A_t^{-1}(aX_1 + 1 \cdot X_2)$$
$$= A_t^{-1}(Y_2) = Y_2 \to 1 \cdot Y_2.$$

Hence, for each a and b, $C_4(1)_\infty = C_4(1)$.

C_5: Again consider $n_W = 1$ and then $n_W = 2$:

$$W = L\{X_1\},$$
$$A_t: X_1 \to X_1; X_2 \to tX_2; X_3 \to tX_3,$$
$$[X_1, X_2]_t = tA_t^{-1}(0) = 0 \to 0,$$
$$[X_3, X_1]_t = tA_t^{-1}(X_1) = tX_1 \to 0,$$
$$[X_3, X_2]_t = t^2 A_t^{-1}(X_1 + X_2)$$
$$= t^2 X_1 + tX_2 \to 0.$$

Hence $C_{5\infty} = C_1$.

$$W = W(a) = L\{aX_1 + X_2 = Y_2\},$$
$$A_t: X_1 \to tX_1; Y_2 \to Y_2; X_3 \to tX_3,$$
$$[X_1, Y_2]_t = tA_t^{-1}(0) = 0 \to 0,$$
$$[X_3, X_1]_t = t^2 A_t^{-1}(X_1) = tX_1 \to 0,$$
$$[X_3, Y_2]_t = tA_t^{-1}(aX_1 + X_1 + X_2)$$
$$= tA_t^{-1}(X_1 + Y_2)$$
$$= X_1 + tY_2 \to X_1.$$

Hence, for each a, $C_{5\infty} = C_2$, the Heisenberg algebra.

$$W = W(a,b) = L\{aX_1 + bX_2 + X_3 = Y_3\},$$
$$A_t: X_1 \to tX_1; X_2 \to tX_2; Y_3 \to Y_3,$$
$$[X_1, X_2]_t = t^2 A_t^{-1}(0) = 0 \to 0,$$
$$[Y_3, X_1]_t = tA_t^{-1}(X_1) = X_1 \to X_1,$$
$$[Y_3, X_2]_t = tA_t^{-1}(X_1 + X_2)$$
$$= X_1 + X_2 \to X_1 + X_2.$$

Hence, for each a and b, $C_{5\infty} = C_5$.

$$W = L\{X_1, X_2\} = C_5',$$
$$A_t: X_1 \to X_1; X_2 \to X_2; X_3 \to tX_3,$$
$$[X_1, X_2]_t = A_t^{-1}(0) = 0 \to 0,$$
$$[X_3, X_1]_t = tA_t^{-1}(X_1) = tX_1 \to 0,$$
$$[X_3, X_2]_t = tA_t^{-1}(X_1 + X_2) = tX_1 + tX_2 \to 0.$$

Hence $(C_5)_\infty^{C_5'} = C_1$.

$$W = W(a) = L\{X_1, aX_2 + X_3 = Y_3\},$$
$$A_t: X_1 \to X_1; X_2 \to tX_2; Y_3 \to Y_3,$$
$$[X_1, X_2]_t = tA_t^{-1}(0) = 0 \to 0,$$
$$[Y_3, X_1]_t = A_t^{-1}(X_1) = X_1 \to X_1,$$
$$[Y_3, X_2]_t = tA_t^{-1}(X_1 + X_2)$$
$$= tX_1 + X_2 \to X_2.$$

Hence, for each a, $C_{5\infty}^{W(a)} = C_4(1)$.

Remark: The derived algebras of both C_5 and $C_4(1)$ have dimension 2, so the contraction just obtained is an example of a type contraction which is claimed by Sharp[1] to be impossible.

All other two-dimensional subspaces are of the form $sp\{aX_1 + X_2, bX_1 + X_3\}$, none of which are subalgebras, so the analysis is complete for C_5

$C_6(\lambda)$: Again consider $n_W = 1$ and then $n_W = 2$:

$$W = L\{X_1\},$$
$$A_t: X_1 \to X_1; X_2 \to tX_2; X_3 \to tX_3,$$
$$[X_1, X_2]_t = tA_t^{-1}(0) = 0 \to 0,$$
$$[X_3, X_1]_t = tA_t^{-1}(X_2 + \lambda X_1)$$
$$= X_2 + t\lambda X_1 \to X_2,$$
$$[X_3, X_2]_t = t^2 A_t^{-1}(-X_1 + \lambda X_2)$$
$$= -t^2 X_1 + t\lambda X_2 \to 0.$$

Hence $C_6(\lambda)_\infty = C_2$, the Heisenberg algebra.

$$W = W(a) = L\{aX_1 + X_2 = Y_2\},$$
$$A_t: X_1 \to tX_1; Y_2 \to Y_2; X_3 \to tX_3,$$
$$[X_1, Y_2]_t = tA_t^{-1}(0) = 0 \to 0$$
$$[X_3, X_1]_t = t^2 A_t^{-1}(X_3 + \lambda X_1)$$
$$= t^2 A_t^{-1}(Y_2 + (\lambda - a)X_1)$$
$$= t^2 Y_2 + t(\lambda - a)X_1 \to 0,$$
$$[X_3, Y_2]_t = tA_t^{-1}(a(X_2 + \lambda X_1) + (-X_1 + \lambda X_2))$$
$$= tA_t^{-1}((a + \lambda)X_2 + (a\lambda - 1)X_1)$$
$$= tA_t^{-1}((a + \lambda)Y_2 + (-a^2 - 1)X_1)$$
$$= t(a + \lambda)Y_2 - (a^2 + 1)X_1$$
$$\to -(a^2 + 1)X_1.$$

Hence, $C_6(\lambda)_\infty = C_2$, the Heisenberg algebra.

$$W = W(a,b) = L\{aX_1 + bX_2 + X_3 = Y_3\},$$
$$A_t: X_1 \to tX_1; X_2 \to tX_2; Y_3 \to Y_3,$$
$$[X_1, X_2]_t = t^2 A_t^{-1}(0) = 0 \to 0,$$
$$[Y_3, X_1]_t = tA_t^{-1}(X_2 + \lambda X_1)$$
$$= X_2 + \lambda X_1 \to X_2 + \lambda X_1,$$

$$[Y_3,X_2]_t = tA_t^{-1}(-X_1 + \lambda X_2)$$
$$= -X_1 + \lambda X_2 \to -X_1 + \lambda X_2.$$

Hence, for each a and b, $C_6(\lambda)_\infty = C_6(\lambda)$, same λ.

$$W = L\{X_1,X_2\} = C_6(\lambda)',$$
$$A_t: X_1 \to X_1; X_2 \to X_2; X_3 \to tX_3,$$
$$[X_1,X_2]_t = A_t^{-1}(0) = 0 \to 0,$$
$$[X_3,X_1]_t = tA_t^{-1}(X_2 + \lambda X_1)$$
$$= tX_2 + t\lambda X_1 \to 0,$$
$$[X_3,X_2]_t = tA_t^{-1}(-X_1 + \lambda X_2)$$
$$= -tX_1 + t\lambda X_2 \to 0.$$

Hence $C_6(\lambda)_\infty^W = C_1$.

All the other two-dimensional subspaces can be put in one of the following two forms:

(i) $sp\{X_1, aX_2 + X_3\}$,
(ii) $sp\{aX_1 + X_2, bX_1 + X_3\}$.

It is easily checked that none of these is closed under the Lie product and are hence not subalgebras. Thus the derived algebra is the only two-dimensional subalgebra and the analysis of $C_6(\lambda)$ is complete.

C_7: ($\approx so(3)$) Besides the properties already pointed out about C_7, it has a very useful symmetry about it which can be described by saying that given any nonzero element of C_7, by properly scaling this element, two other elements can be chosen so that the canonical commutation relations (2.10) hold for these three elements. Thus it suffices to consider any one one-dimensional subalgebra.

$$W = L\{X_3\},$$
$$A_t: X_1 \to tX_1; X_2 \to tX_2; X_3 \to X_3,$$
$$[X_1,X_2]_t = t^2A_t^{-1}(X_3) = t^2X_3 \to 0,$$
$$[X_3,X_1]_t = tA_t^{-1}(X_2) = X_2 \to X_2,$$
$$[X_3,X_2]_t = tA_t^{-1}(-X_1) = -X_1 \to -X_1.$$

Hence $C_{7\infty} = C_6(0)$, the Lie algebra of the group of Euclidean motions in the plane.

C_7 has the additional characterizing (among three-dimensional algebras, at least) property that it has no two-dimensional subalgebras whatsoever, due to the fact that any two linearly independent elements generate (as a Lie algebra) all of C_7. Thus the analysis of C_7 is complete.

C_8: Again consider $n_W = 1$ and then $n_W = 2$:

$$W = L\{aX_1 + bX_2 = Y_1\} \quad \text{with } d^2 = a^2 + b^2 \neq 0.$$

Let $Y_2 = (-b/d^2)X_1 + (a/d^2)X_2$:

$$A_t: Y_1 \to Y_1; Y_2 \to tY_2; X_3 \to tX_3,$$
$$[Y_1,Y_2]_t = tA_t^{-1}(X_3) = X_3 \to X_3,$$
$$[Y_1,X_3]_t = tA_t^{-1}(aX_2 + b(-X_1))$$
$$= tA_t^{-1}(d^2Y_2) = d^2Y_2 \to d^2Y_2,$$
$$[Y_2,X_3]_t = t^2A_t^{-1}((-b/d^2)X_2$$
$$+ (a/d^2)(-X_1))$$
$$= t^2A_t^{-1}(-d^{-2}Y_1)$$
$$= -t^2d^{-2}Y_1 \to 0.$$

Hence we can see (by setting $X_1' = Y_2 + d^{-1}X_3$, $X_2' = Y_2 - d^{-1}X_3, X_3' = d^{-1}Y_1$) that, for all admissible a and b, $C_{8\infty} = C_4(-1)$.

$$W = W(a,b) = L\{aX_1 + bX_2 + X_3 = Y_3\},$$
$$A_t: X_1 \to tX_1; X_2 \to tX_2; Y_3 \to Y_3,$$
$$[X_1,X_2]_t = t^2A_t^{-1}(X_3)$$
$$= t^2A_t^{-1}(Y_3 - aX_1 - bX_2)$$
$$= t^2Y_3 - taX_1 - tbX_2 \to 0,$$
$$[Y_3,X_1]_t = tA_t^{-1}(b(-X_3) + (-X_2))$$
$$= tA_t^{-1}(-bY_3 + abX_1 + (b^2 - 1)X_2)$$
$$= -tbY_3 + abX_1 + (b^2 - 1)X_2$$
$$\to abX_1 + (b^2 - 1)X_2,$$
$$[Y_3,X_2]_t = tA_t^{-1}(aX_3 + X_1)$$
$$= tA_t^{-1}(aY_3 - abX_2 - (a^2 - 1)X_1)$$
$$= taY_3 - abX_2 + (1 - a^2)X_1$$
$$\to (1 - a^2)X_1 - abX_2.$$

Hence it is clear that $ad_\infty Y_3$ acts on $L\{X_1,X_2\}$ as a linear operator whose determinant is $1 - (a^2 + b^2)$ and whose characteristic polynomial has roots $\pm(a^2 + b^2 - 1)^{1/2}$. In case $a^2 + b^2 = 1$, the operator is easily seen to have rank 1, so the derived algebra is one-dimensional, and to have roots ± 0, so that $ad_\infty Y_3$ is nilpotent, and hence we have C_2, the Heisenberg algebra. Hence we have:

If $a^2 + b^2 = 1$, then $C_{8\infty} = C_2$.

If $a^2 + b^2 > 1$, then the roots are real and of equal magnitude so that $C_{8\infty} = C_4(-1)$.

If $a^2 + b^2 < 1$, then the roots are pure imaginary so that $C_{8\infty} = C_6(0)$. This completes the analysis for the one-dimensional subalgebras.

Of all the two-dimensional subspaces of the form $sp\{X_2, aX_1 + bX_3\}$ (i.e., those containing X_2), only two are algebras, namely the following: $W_\pm = L\{X_2, X_1 \pm X_3 = Y_\pm\}$. These can be treated simultaneously by *consistently* taking the upper (for W_+) or the lower (for W_-) sign in the following analysis:

$$W_\pm = L\{X_2, X_1 \pm X_3 = Y_\pm\},$$
$$A_t: Y_\pm \to Y_\pm; X_2 \to X_2; X_3 \to tX_3,$$
$$[X_3,Y_\pm]_t = tA_t^{-1}(-X_2)$$
$$= -tX_2 \to 0,$$
$$[X_2,Y_\pm]_t = A_t^{-1}(-X_3 \pm (-X_1))$$
$$= A_t^{-1}(\mp Y_\pm) = \mp Y_\pm \to \mp Y_\pm,$$
$$[X_2,X_3]_t = tA_t^{-1}(-X_1)$$
$$= tA_t^{-1}(\pm X_3 - Y_\pm)$$
$$= \pm X_3 - tY_\pm \to \pm X_3.$$

Hence we see (by setting $X_1' = Y_+$, $X_2' = X_3$, and $X_3' = \mp X_2$) that $(C_8)_\infty^{W_\pm} = C_4(-1)$.

The remaining (i.e., those not containing X_2) two-dimensional subspaces are of the form $sp\{aX_2 + X_3 =$

$Y_3, X_1 + bX_2 = Y_1\}$. It is easily checked that such subspaces are subalgebras iff $a = \pm (1 + b^2)^{1/2}$. Thus we have, by letting $c = (1 + b^2)^{1/2}$, the following two infinite classes of subalgebras:

$$
\begin{aligned}
W_{b+} &= L\{X_1 + bX_2 = Y_1, X_3 + cX_2 = Y_3\},\\
&\quad A_t: Y_1 \to Y_1; X_2 \to tX_2; Y_3 \to Y_3,\\
[Y_1, X_2]_t &= tA_t^{-1}(X_3)\\
&= tA_t^{-1}(Y_3 - cX_2)\\
&= tY_3 - cX_2 \to -cX_2,\\
[Y_1, Y_3]_t &= A_t^{-1}(X_2 + cX_3 - bX_1)\\
&= A_t^{-1}(cY_3 - bY_1)\\
&= cY_3 - bY_1 \to cY_3 - bY_1,\\
[X_2, Y_3]_t &= tA_t^{-1}(-X_1)\\
&= tA_t^{-1}(bX_2 - Y_1)\\
&= bX_2 - tY_1 \to bX_2.
\end{aligned}
$$

By letting $X_1' = cX_3 + X_2 - bX_1 = cY_3 - bY_1$, $X_2' = X_2$, and $X_3' = c^{-1}X_1 + bc^{-1}X_2 = c^{-1}Y_1$, we can see that $C_{8\infty} = C_4(-1)$.

$$
\begin{aligned}
W_{b-} &= L\{X_1 + bX_2 = Y_1, X_3 - cX_2 = Y_3\},\\
&\quad A_t: Y_1 \to Y_1; X_2 \to tX_2; Y_3 \to Y_3,\\
[Y_1, X_2]_t &= tA_t^{-1}(X_3)\\
&= tA_t^{-1}(Y_3 + cX_2)\\
&= tY_3 + cX_2 \to cX_2,\\
[Y_1, Y_3]_t &= A_t^{-1}(X_2 - cX_3 - bX_1)\\
&= A_t^{-1}(-bY_1 - cY_3)\\
&= -bY_1 - cY_3 \to -b_1Y_1 - cY_3,\\
[X_2, Y_3]_t &= tA_t^{-1}(-X_1)\\
&= tA_t^{-1}(bX_2 - Y_1)\\
&= bX_2 - tY_1 \to bX_2.
\end{aligned}
$$

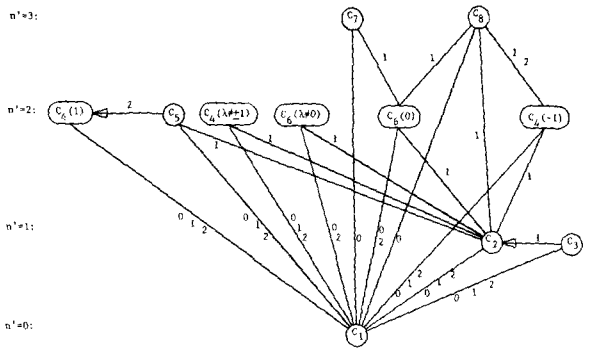

FIG. 1. All Inonu-Wigner contractions of all the real Lie algebras of dimension 3.

By letting $X_1' = cX_3 - X_2 + bX_1 = cY_3 + bY_1$, $X_2' = X_2$, and $X_3' = -c^{-1}X_1 - bc^{-1}X_2 = -c^{-1}Y_1$, we can see that $C_{8\infty} = C_4(-1)$.

This concludes the complete analysis of the IWC's for the real Lie algebras with $n \leqq 3$. For facility in use and understanding of this material we have constructed the following chart for the $n = 3$ case. Recalling that $n'_\infty \leqq n'$, the progress is never upward in this chart and where $n'_\infty = n'$ we have indicated the direction of the IWC by an arrow. The possible dimensions n_W of W by which it is possible to obtain each result is indicated along the line representing the IWC.

ACKNOWLEDGMENTS

The analysis presented here was done as part of the author's doctoral dissertation at the University of Illinois. The author would like to thank Dr. R. R. Rao, who first interested him in contractions, for both his encouragement and his direction. The author also acknowledges partial support of his work by the National Science Foundation.

1 W. T. Sharp, "Racah Algebras and the Contraction of Groups," CRT-935, AECL-1098, 1960.
2 E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. (U.S.) **39**, 510 (1953).
3 E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. (U.S.) **40**, 119 (1954).
4 E. J. Saletan, J. Math. Phys. **2**, 1 (1961).
5 N. Jacobson, *Lie Algebras* (Wiley, New York, 1962).
6 J. D. Talman, *Special Functions: A Group Theoretic Approach* (Benjamin, New York, 1968).

Particle Transport in Spherical Media with a Central Black Cavity

O. J. Sheaks
Department of Chemical Engineering, University of Maryland, College Park, Maryland 20742

(Received 8 September 1971)

A transform procedure is applied to the integral form of the Boltzmann transport equation to obtain the particle density in a spherically symmetric medium surrounding a central black absorber. The singular eigenfunction expansion technique is applied to provide a general solution, and boundary conditions are derived for the general equation for arbitrary multiplication and source distributions. Specific applications to the spherical Milne problem and a uniform source distribution are presented to demonstrate the application of the general transform technique.

1. INTRODUCTION

The analysis of the particle density in a spherical medium containing a central black absorber is relevant to several practical problems, e.g., preliminary estimates for resonance escape probabilities in systems where fuel lumps are dispersed in a moderating medium, unit-cell flux analysis for coated spherical fuel pellets, shielding calculations for annular sources of gamma radiation, and the analysis of fuel elements with included burnable poisons in pellet form.

$Y_3, X_1 + bX_2 = Y_1\}$. It is easily checked that such subspaces are subalgebras iff $a = \pm (1 + b^2)^{1/2}$. Thus we have, by letting $c = (1 + b^2)^{1/2}$, the following two infinite classes of subalgebras:

$$
\begin{aligned}
W_{b+} &= L\{X_1 + bX_2 = Y_1, X_3 + cX_2 = Y_3\},\\
&\quad A_t: Y_1 \to Y_1; X_2 \to tX_2; Y_3 \to Y_3,\\
[Y_1, X_2]_t &= tA_t^{-1}(X_3)\\
&= tA_t^{-1}(Y_3 - cX_2)\\
&= tY_3 - cX_2 \to -cX_2,\\
[Y_1, Y_3]_t &= A_t^{-1}(X_2 + cX_3 - bX_1)\\
&= A_t^{-1}(cY_3 - bY_1)\\
&= cY_3 - bY_1 \to cY_3 - bY_1,\\
[X_2, Y_3]_t &= tA_t^{-1}(-X_1)\\
&= tA_t^{-1}(bX_2 - Y_1)\\
&= bX_2 - tY_1 \to bX_2.
\end{aligned}
$$

By letting $X_1' = cX_3 + X_2 - bX_1 = cY_3 - bY_1$, $X_2' = X_2$, and $X_3' = c^{-1}X_1 + bc^{-1}X_2 = c^{-1}Y_1$, we can see that $C_{8\infty} = C_4(-1)$.

$$
\begin{aligned}
W_{b-} &= L\{X_1 + bX_2 = Y_1, X_3 - cX_2 = Y_3\},\\
&\quad A_t: Y_1 \to Y_1; X_2 \to tX_2; Y_3 \to Y_3,\\
[Y_1, X_2]_t &= tA_t^{-1}(X_3)\\
&= tA_t^{-1}(Y_3 + cX_2)\\
&= tY_3 + cX_2 \to cX_2,\\
[Y_1, Y_3]_t &= A_t^{-1}(X_2 - cX_3 - bX_1)\\
&= A_t^{-1}(-bY_1 - cY_3)\\
&= -bY_1 - cY_3 \to -b_1Y_1 - cY_3,\\
[X_2, Y_3]_t &= tA_t^{-1}(-X_1)\\
&= tA_t^{-1}(bX_2 - Y_1)\\
&= bX_2 - tY_1 \to bX_2.
\end{aligned}
$$

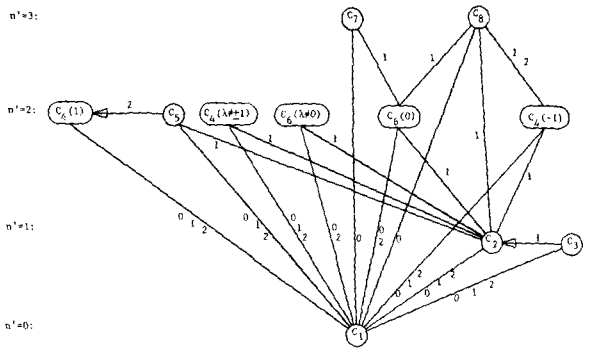

FIG. 1. All Inonu-Wigner contractions of all the real Lie algebras of dimension 3.

By letting $X_1' = cX_3 - X_2 + bX_1 = cY_3 + bY_1$, $X_2' = X_2$, and $X_3' = -c^{-1}X_1 - bc^{-1}X_2 = -c^{-1}Y_1$, we can see that $C_{8\infty} = C_4(-1)$.

This concludes the complete analysis of the IWC's for the real Lie algebras with $n \leqq 3$. For facility in use and understanding of this material we have constructed the following chart for the $n = 3$ case. Recalling that $n'_\infty \leqq n'$, the progress is never upward in this chart and where $n'_\infty = n'$ we have indicated the direction of the IWC by an arrow. The possible dimensions n_W of W by which it is possible to obtain each result is indicated along the line representing the IWC.

ACKNOWLEDGMENTS

The analysis presented here was done as part of the author's doctoral dissertation at the University of Illinois. The author would like to thank Dr. R. R. Rao, who first interested him in contractions, for both his encouragement and his direction. The author also acknowledges partial support of his work by the National Science Foundation.

1 W. T. Sharp, "Racah Algebras and the Contraction of Groups," CRT-935, AECL-1098, 1960.
2 E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. (U.S.) **39**, 510 (1953).
3 E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. (U.S.) **40**, 119 (1954).
4 E. J. Saletan, J. Math. Phys. **2**, 1 (1961).
5 N. Jacobson, *Lie Algebras* (Wiley, New York, 1962).
6 J. D. Talman, *Special Functions: A Group Theoretic Approach* (Benjamin, New York, 1968).

Particle Transport in Spherical Media with a Central Black Cavity

O. J. Sheaks
Department of Chemical Engineering, University of Maryland, College Park, Maryland 20742

(Received 8 September 1971)

A transform procedure is applied to the integral form of the Boltzmann transport equation to obtain the particle density in a spherically symmetric medium surrounding a central black absorber. The singular eigenfunction expansion technique is applied to provide a general solution, and boundary conditions are derived for the general equation for arbitrary multiplication and source distributions. Specific applications to the spherical Milne problem and a uniform source distribution are presented to demonstrate the application of the general transform technique.

1. INTRODUCTION

The analysis of the particle density in a spherical medium containing a central black absorber is relevant to several practical problems, e.g., preliminary estimates for resonance escape probabilities in systems where fuel lumps are dispersed in a moderating medium, unit-cell flux analysis for coated spherical fuel pellets, shielding calculations for annular sources of gamma radiation, and the analysis of fuel elements with included burnable poisons in pellet form.

For planar geometries, the method of singular eigenfunction expansion developed by Case[1] has proven quite successful in providing expressions for the particle density in terms of a set of expansion coefficients of the normal modes of the separable integro-differential transport equation for a variety of problems. For all but very idealized problems, however, the expansion coefficients are solutions to Fredholm integral equations. The advantage of the method lies in the insight gathered from the mathematical structure of the solution, the practical information derived from various approximations, and rapidly converging numerical techniques developed by several authors including Mitsis[2], Mendelson[3], and Bond[4].

The straightforward extension of Case's method to nonplanar geometries has been attempted by several investigators with little success. Bareiss and Abu-Shumays[5] have presented solutions to the separable Boltzmann equation; Mitsis[2] and Davison[6,7] have given the normal modes to a nonseparable form in spherical and cylindrical geometries with the appropriate symmetry conditions. These normal modes have proven useful for a specific application by Erdmann and Siewert[8], where they were able to deduce the expansion coefficients by an intuitive approach. However, no method has been developed by which the expansion coefficients may be determined in general.

The most promising method of mathematically analyzing nonplanar problems has been the development of suitable transforms which reduce the integral equation to a new integro-differential equation amenable to solution by the singular eigenfunction expansion method. Leonard and Mullikan[9] were the first to suggest this idea in an application involving neutron transport. However, the utility of the concept was demonstrated by Mitsis[2] in solutions to single region critical problems in spherical and cylindrical geometries. The mathematical sophistication of the method was finalized by Gibbs[10] in a general formulation of the transform method for arbitrary homogeneous, convex bodies.

The purpose of this paper is to extend the work of Mitsis and Gibbs to include a nonhomogeneous medium consisting of a moderator surrounding a central black absorber. The development of the transform procedure is presented in Sec. 2. The application of the method to a homogeneous problem, i.e., the classic spherical Milne problem in Sec. 3, and a solution to the nonhomogeneous problem of a uniform source distribution are presented in Sec. 4.

2. TRANSFORM TECHNIQUE

We consider a moderating medium surrounding a central "black" (infinite absorption) cavity. Under the assumptions of isotropic scattering and sources and a uniform, homogeneous medium, the mono-energetic particle density satisfies the equation[11]

$$n(\mathbf{r}) = \int_v [c\,n(\mathbf{r}') + S(\mathbf{r}')]K(|\mathbf{r}' - \mathbf{r}|)dV, \qquad \mathbf{r} \in V, \quad (2.1)$$

where $n(\mathbf{r})$ is the particle density at position vector $\mathbf{r}$, c is the mean number of secondaries per collision, $S(\mathbf{r})$ is the volumetric source distribution, V is the domain of the moderating medium, $\mathbf{r}$ is measured in number of mean free paths, and

$$K|\mathbf{r}| = e^{-|\mathbf{r}|}/4\pi|\mathbf{r}|^2. \quad (2.2)$$

Under the assumptions of symmetry, Eq. (2.1) reduces to

$$rn(r) = \int_0^1 \frac{ds}{s}\int_a^\infty dr'\, r'\, \tfrac{1}{2}[cn(r') + S(r')] \times \left[\exp\left(\frac{-|r-r'|}{s}\right) - \exp\left(\frac{-(r^2-a^2)^{1/2}}{s} + \frac{(r'^2-a^2)^{1/2}}{s}\right)\right]. \quad (2.3)$$

The straightforward application of the transform techniques employed by Mitsis[2] or Gibbs[9] is not possible, since the exponentials in the kernel are not compatible, in the sense they are not annihilated by the same differential operator, or from the Gibbs approach, the limits on the domain of definition V, Eq. (2.1), are a function of the variables of the null operator.

Familiarity with the linear exponential part of the kernel leads to an examination of the integral

$$I = \int_0^1 \frac{ds}{s}\int_a^\infty dr'\, \tfrac{1}{2}[cn(r') + S(r')]r' \times \exp\left(\frac{-(r^2-a^2)^{1/2}}{s} + \frac{(r'^2-a^2)^{1/2}}{s}\right) \quad (2.4)$$

in, an attempt to convert I to a form which will be compatible with the integrals over the first part of the kernel. To accomplish this we make use of the identity[12]

$$\exp\left(\frac{-(r^2-a^2)^{1/2}}{s}\right) = \int_0^s \frac{dv}{v^2}\exp\left(\frac{-r}{v}\right)G(v,s), \quad (2.5)$$

where

$$G(v,s) = v^2\,\delta(v-s) + av\frac{I_1((a/vs)\sqrt{s^2-v^2})}{\sqrt{s^2-v^2}} \quad (2.6)$$

and $I_1(\chi)$ is the modified Bessel function of the first kind. We substitute Eq. (2.5) in Eq. (2.4), interchange the order of integration over s and v, and write Eq. (2.4) as

$$rn(r) = \int_0^1 d\mu\left[\frac{1}{\mu}\int_a^r dr'\, r'\, \tfrac{1}{2}[cn(r') + S(r')]e^{-(r-r')/\mu} - \frac{e^{-r/\mu}}{\mu^2}\int_\mu^1 \frac{ds}{s}G(\mu,s)\int_a^\infty dr'\, r'\, \tfrac{1}{2}[cn(r') + S(r')] \times \exp\left(\frac{-(r'^2-a^2)^{1/2}}{s}\right) + \frac{1}{\mu}\int_r^\infty dr'\, r' \times \tfrac{1}{2}(cn(r') + S(r'))\,e^{-(r'-r)/\mu}\right]. \quad (2.7)$$

We now have a choice of transform methods: a first order or a second-order differential operator may be applied to suitable definitions of the transform function derived from Eq. (2.7). The author has followed both procedures, and equivalent results are obtained. We choose to present the former method for reasons of clarity and familiarity with plane geometry particle transport analysis.

We define the transform functions

$$\psi(r,\mu) = \frac{1}{\mu}\int_a^r dr'\, r'\, \tfrac{1}{2}[cn(r') + S(r')]\,e^{-(r-r')/\mu} - \frac{e^{-r/\mu}}{\mu^2}\int_\mu^1 \frac{ds}{s}G(\mu,s)\int_a^\infty dr'\, r'\, \tfrac{1}{2}[cn(r') + S(r')] \times \exp\left(\frac{-(r'^2-a^2)^{1/2}}{s}\right), \quad \mu \in (0,1), \quad (2.8)$$

and

$$\psi(r,-\mu)=\frac{1}{\mu}\int_r^\infty dr'\,r'\,\tfrac{1}{2}[cn(r')+S(r')]e^{-(r'-r)/\mu},\qquad \mu\in(0,1). \quad (2.9)$$

Equation (2.7) becomes

$$rn(r)=\int_{-1}^{1}\psi(r,\mu)\,d\mu, \quad (2.10)$$

which, if the appropriate expression for $\psi(r,\mu)$ can be found, provides the inversion transform for the particle density. The definitions in Eqs. (2.8) and (2.9) were selected so that the transform function $\psi(r,\mu)$ would be annihilated by the operator (for the homogeneous case)

$$0=\left[\mu\frac{\partial}{\partial r}+1-\frac{c}{2}\int_{-1}^{1}d\mu'\right]. \quad (2.11)$$

Indeed, upon differentiating Eqs. (2.8) and (2.9) we find

$$\mu\frac{\partial}{\partial r}\psi(r,\mu)+\psi(r,\mu)=\frac{r}{2}[cn(r)+S(r)]$$
$$=\frac{c}{2}\int_{-1}^{1}\psi(r,\mu')d\mu'+\frac{S(r)r}{2}. \quad (2.12)$$

We recognize Eq. (2.12) as the inhomogeneous equation encountered in plane geometry transport. The normal modes to the homogeneous equation are well known, the appropriate completeness properties[1] have been demonstrated, and orthogonality relationships[13] have been derived. In summary,

$$\psi(r,\mu)=\{e^{-r/\nu}\phi(\nu,\mu)\}; \quad (2.13)$$

there are two discrete eigenvalues $\pm\nu_0$, satisfying the dispersion relation

$$1=c\nu_0\tanh^{-1}(1/\nu_0) \quad (2.14)$$

and a set of continuum modes for ν on the real interval $[-1,1]$,

$$\phi(\nu,\mu)=\frac{c\nu}{2}\frac{P}{\nu-\mu}+\lambda(\nu)\,\delta(\nu-\mu),\quad \mu\in[-1,1], \quad (2.15)$$

and

$$\lambda(\nu)=1-c\nu\tanh^{-1}(\nu). \quad (2.16)$$

Here P denotes that integrals over these functions are to be considered in the Cauchy principal value sense and $\delta(\chi)$ is the Dirac delta function. The orthogonality relationships may be written as

$$\int_a^1 d\mu\,W(\mu)\,\phi(\nu,\mu)\,\phi(\nu',\mu)=W(\nu)\,N(\nu)\,\delta(\nu-\nu'), \quad (2.17)$$

where for $\mu\in[-1,1]$:

$$a=-1,$$
$$W(\mu)=\mu,$$
$$N(\nu)=\pm\frac{c}{2}\,\nu_0^2\left[\frac{c}{\nu_0^2-1}-\frac{1}{\nu_0^2}\right],\quad \nu=\pm\nu_0,$$
$$N(\nu)=\left[\lambda^2(\nu)+\left(\frac{c\nu\pi}{2}\right)^2\right],\quad \nu\in[-1,1];$$

for $\mu\in[0,1]$:

$$a=0,$$
$$W(\mu)=(\nu_0-\mu)\gamma(\mu),$$
$$\gamma(\mu)=\frac{c\mu}{2\chi(-\mu)(\nu_0^2-\mu^2)},$$
$$\chi(z)=\exp\left[-\frac{c}{2}\int_0^1\frac{d\mu}{N(\mu)}\left(1+\frac{c\mu^2}{1-\mu^2}\right)\ln(\mu-z)\right]$$
$$W(\nu)N(\nu)=-\left(\frac{c\nu_0}{2}\right)^2\chi(\nu_0),\qquad \nu=\nu_0,$$
$$W(\nu)N(\nu)=(\nu_0-\nu)\gamma(\nu)\left[\lambda^2(\nu)+\left(\frac{\pi c\nu}{2}\right)^2\right],\qquad \nu\in[0,1].$$

A useful technique in finding a particular solution to Eq. (2.12) is given in Appendix A. There it is shown that a particular solution $\psi_p(\mu)$ is given by

$$\psi_p(\chi,\mu)=\int d\nu\,[1/N(\nu)]\,\phi(\nu,\mu)\,\phi_p(\chi,\mu), \quad (2.18)$$

where $\phi_p(\chi,\mu)$ satisfies

$$\mu\frac{\partial\phi_p(\chi,\mu)}{\partial\chi}+\phi_p(\chi,\mu)=S(\chi). \quad (2.19)$$

(We have used the integral notation in the above equation symbolically to represent both discrete eigenfunctions and the entire continuum set, $\nu\in[-1,1]$.)

The final step in the transformation process is to provide boundary conditions to which the solutions of Eq. (2.12) must be subject. This procedure can be shown to be equivalent to that used by Gibbs to determine his expansion coefficients, although his method might be considered more fundamental since he insures the derived form of his transform function is consistent with the defined form. The boundary condition approach appears more simple for this problem, however. From Eqs. (2.8) and (2.9),

$$\text{BC:}\quad \text{(i)}\ \psi(\infty,-\mu)=0,\qquad \mu\in[0,1], \quad (2.20)$$

$$\text{(ii)}\ \psi(a,\mu)=\frac{-e^{-a/\mu}}{\mu^2}\int_\mu^1\frac{ds}{s}\,G(\mu,s)\int_a^\infty dr'\,r'$$
$$\times\tfrac{1}{2}[cn(r')+S(r')]\cdot\exp\left(\frac{-(r'^2-a^2)^{1/2}}{s}\right),$$
$$\mu\in[0,1]. \quad (2.21)$$

The general procedure to be followed is then to obtain a particular solution [Eq. (2.18)] for the problem of interest and use sufficient homogeneous solutions to insure the conditions necessary for completeness[1] in trying to meet the boundry conditions are met. Having determined $\psi(r,\mu)$, we can obtain the particle density from Eq. (2.10).

We note that unlike the homogeneous media problems BC (ii) contains the particle density. This leads to Fredholm integral equations for the expansion coefficients. However, considerable mathematical insight can be gained from the form of the solution derived by this method, certain approximations yield valuable asymptotic results, and the numerical convergence, either a Neumann series method or a discrete ordinates approach, is expected to be rapid. In the following sections we demonstrate the application of the method to several problems of practical interest.

3. SPHERICAL MILNE PROBLEM

A. Transform Solution

We seek the particle density in an infinite, purely moderating medium ($c=1$) containing a central black

absorber of radius "a". This example was selected because of its classical nature and for its practical applications for determining the effect of curvature on the extrapolation distance, a necessary parameter used extensively in reactor physics. The classical assumption that $c = 1$ is not a necessary condition, and the methods employed apply directly for any $c < 1$ case.

As in the planar Milne problem, the discrete eigenfunctions become degenerate since $\nu_0 \to \pm\infty$ as $c \to 1$; the discrete modes then merge into a common value, $\frac{1}{2}$. To meet the boundary conditions requires the addition of the solution

$$\psi(r, \mu) = \tfrac{1}{2}A(r - \mu),$$

The general solution to Eq. (2.12) is then written

$$\psi(r, \mu) = \frac{A_+}{2} + \frac{A_-}{2}(r - \mu) + \int_0^1 d\nu\, \phi(\nu, \mu)\, e^{-r/\nu} A(\nu), \tag{3.1}$$

where

$$A(-\nu) = 0, \qquad \nu \in [0, 1],$$

following the traditional "source condition" of allowing the particle density to diverge, but diverging more slowly than e^r/r as $r \to \infty$.

The application of BC (ii) leads to the equation from which the expansion coefficients may be determined:

$$\frac{A_+}{2} + \frac{A_-}{2}(a - \mu) + \int_0^1 d\nu\, A(\nu)\, e^{-a/\nu}\phi(\nu, \mu) = -\frac{e^{-a/\mu}}{2\mu^2}\int_\mu^1 \frac{ds}{s} G(\mu, s) \int_a^\infty dr'\, r' n(r') \times \exp\left(\frac{-(r'^2 - a^2)^{1/2}}{s}\right). \tag{3.2}$$

We observe that we are expanding a function which contains the unknown density $n(r)$. The general procedure to be followed, when the transformation leads to cases such as this, is to use the inversion integral to obtain the form of the density. This result is used to provide an integral equation from which the coefficients may be determined.

Thus we utilize Eqs. (3.1) and (2.10) to write the density as

$$rn(r) = A_+ + A_- r + \int_0^1 d\nu\, A(\nu)\, e^{-(r-a)/\nu} d\nu, \tag{3.3}$$

where the $A(\nu)$ above is equal to $e^{-a/\nu}$ times the $A(\nu)$ in Eq. (3.2). The substitution of Eq. (3.3) into Eq. (3.2) leads to a valid half-range ($\mu \in [0, 1]$) expansion and is now the equation from which the coefficients may be determined. Due to the involved nature of the integrals encountered, we reserve the details for Appendix B. After these simplifications, Eq. (3.2) becomes

$$\frac{A_+}{2} + \frac{A_-}{2}(a - \mu) + \int_0^1 d\nu\phi(\nu, \mu)A(\nu) = -B(\mu) + \int_0^1 d\nu A(\nu)K(\mu, \nu), \tag{3.4}$$

where

$$B(\mu) = A_+ K(\mu, \infty) + \tfrac{1}{2}A_- e^{-a/\mu} \left(\mu + \frac{a}{\mu}\int_\mu^1 ds\, \frac{I_1(a/\mu s)\sqrt{s^2 - \mu^2}}{\sqrt{s^2 - \mu^2}}\right) \tag{3.5}$$

and

$$K(\mu, \nu) = \frac{e^{-a/\mu}}{\mu^2}\int_\mu^1 \frac{ds}{s} G(\mu, s) \int_0^s \frac{dt}{t} \times e^{-a/t}\, \phi(-\nu, t)\, G(t, s). \tag{3.6}$$

The form of Eq. (3.6) is comparable to that obtained by Sahni[14] using a different technique. Applying the operator

$$\int_0^1 d\mu\, \gamma(\mu),$$

we obtain

$$\frac{A_+}{2} + \frac{A_-}{2}(a - z_0) = \int_0^1 d\nu'\, A(\nu') \int_0^1 d\mu\, \gamma(\mu) K(\mu, \nu') - \int_0^1 d\mu\, \gamma(\mu) B(\mu), \tag{3.7}$$

where z_0 is the well-known planar Milne extrapolation distance. Similarly, applying the operator

$$\int_0^1 d\mu\, \gamma(\mu)\, \phi(\nu', \mu), \qquad \nu' \in [0, 1],$$

yields

$$\frac{A_-\nu}{4} + N(\nu)\gamma(\nu)A(\nu) = \int_0^1 d\nu'\, A(\nu') \int_0^1 d\mu\, \gamma(\mu)\, \phi(\nu, \mu) K(\mu, \nu') - \int_0^1 d\mu\, \gamma(\mu)\, \phi(\nu, \mu) B(\mu). \tag{3.8}$$

The problem has now been reduced to solving Eqs. (3.7) and (3.8) for the expansion coefficients, A_- and $A(\nu)$. (Since the equations are homogeneous in A_+ we arbitrarily normalize this constant to -1.) We may now solve for A_- in terms of $A(\nu)$ in Eq. (3.7) and substitute this result in Eq. (3.8). This procedure leads to an inhomogeneous Fredholm integral equation for $A(\nu)$. Although the analytical representations of the kernel and the inhomogeneous term are complicated, all functions contained therein are tabulated and thus present no special problems for the application of computer methods. We do not pursue the numerical solution for the coefficients in this analysis. However, in previous cases in which Fredholm equations of this type were encountered, rapid convergence of the Newmann-type series was found to be the case.

B. Analysis of Results

In this section we examine the results in several limiting cases of the black sphere radius and compare the form of the solution with previous work.

Having provided a method whereby the coefficients can be determined to a required accuracy, we write the density as

$$n(r) = A_- - \frac{1}{r} + \int_0^1 d\nu\, A(\nu)\, \frac{e^{-(r-a)/\nu}}{r}. \tag{3.9}$$

This form may also be derived using the eigenfunctions developed by Davison[7] for the homogeneous integro-differential equation for the angular density. It is also the form used by Sahni[14] in his analysis. Also, as "r" becomes much greater than "a", the continuum becomes negligible compared to the discrete term. Thus the density becomes $n_D(r) = A_- - (1/r)$,

which can be derived from the diffusion equation for no absorption. Thus, asymptotically our solution is of the correct form. The extrapolation distance λ may be defined in the usual manner.

$$\lambda = n_D(r) \Big/ \frac{dn_D(r)}{dr}\Big|_{r=a/\Sigma},$$

where $n_D(r)$ is the discrete part of the density. Physically, this states that the asymptotic density extrapolates to zero at $r = a - \lambda$. Using the results of Eq. (2.9), we obtain

$$\lambda = a^2 A_- - a. \tag{3.10}$$

In the limit as $a \to \infty$ the black sphere begins to look like a plane to incident particles and we would expect λ to approach z_0, the extrapolation distance for plane geometry. To show this we begin with Eq. (3.7). From Eqs. (3.5) (3.6), $B(\mu)$ and $K(\mu, \nu)$ are obviously of exponential order for large "a" and hence approach zero in the limit. Equation (3.7) reduces to $-\frac{1}{2} + \frac{1}{2}A_-(a - z_0) = 0$ or $A_- = 1/(a - z_0)$.

The extrapolation distance thus becomes

$$\lambda = [a^2/(a - z_0)] - a = az_0/(a - z_0), \tag{3.11}$$

which does indeed approach z_0 as "a" increases without bound.

Another interesting approximation which can be examined analytically is the nonphysical result for λ as $a \to 0$. Classically, this result should result in a value of $\frac{4}{3}$ for λ, although physically this problem for $a = 0$ would not be well posed, since there would be no sink for the particle density produced by the source.

Using series expansions for the transcendental functions in Eqs. (2.5) and (2.6), we can show that as $a \to 0$;

$$K(\mu, \nu) \to -\phi(-\nu, \mu), \tag{3.12}$$

and

$$B(\mu) \to \frac{A_-}{2}\left(\mu - a + \frac{a^2}{2\mu^2}\right) - \tfrac{1}{2}. \tag{3.13}$$

Equation (3.4) in the limit thus becomes

$$-\tfrac{1}{2} + \frac{A_-}{2}(a - \mu) + \int_0^1 d\nu\, A(\nu)\,\phi(\nu, \mu) + \tfrac{1}{2}\int_0^1 d\nu\, A(\nu)\,\frac{\nu}{\nu + \mu} = \frac{A_-}{2}\left(a - \mu + \frac{a^2}{2\mu^2}\right) + \tfrac{1}{2}. \tag{3.14}$$

We have arrived at a half-range expansion which would result in an integral equation for $A(\nu)$, which certainly presents a formidable task if an exact result is to be obtained. However, we note that by defining $A(\nu) = A(-\nu)$ we can write Eq. (3.14) as

$$1 - \int_{-1}^1 d\nu\, A(\nu)\,\phi(\nu, \mu) = A_-\, a^2/4\mu^2. \tag{3.15}$$

It is trivial to show the above is a valid full-range expansion for even functions of μ and $c = 1$; thus full-range orthogonality is applicable. We operate with $\int_{-1}^1 d\mu\, \mu^2$ and make use of the fact that

$$\int_{-1}^1 d\mu\, \mu^2 \phi(\nu, \mu) = 0 \tag{3.16}$$

to obtain $\frac{2}{3} = A_- \frac{1}{4}a^2 2$ or

$$A_- = 4/3a^2. \tag{3.17}$$

The extrapolation distance using this result is

$$\lambda = (4/3a^2)\,a^2 - a \to \tfrac{4}{3} \tag{3.18}$$

as

$$a \to 0.$$

4. THE INHOMOGENEOUS EQUATION

We consider an example of a uniform volumetric source distribution in the moderating medium ($c < 1$) because of its practical application in cell theory in reactor physics and to demonstrate the use of the particular solution derived in Sec. 2. We assume an infinite medium for reasons of simplicity. This restriction is easily removed by applying directly the techniques used in finite slab geometry neutron transport when finding solutions for the transform function $\psi(r, \mu)$. Following the procedure developed in Sec. 2, the transform equation for a uniform source of strength S_ν(#/cm^3 − sec) is

$$\mu\frac{\partial\psi}{\partial r}(r, \mu) + \psi(r, \mu) = \frac{c}{2}\int_{-1}^1 \psi(r, \mu')\,d\mu' + \frac{S_\nu}{2}r, \tag{4.1}$$

with boundary conditions

(i) $rn(r)$ diverge no faster than r at infinity

(ii) $$\psi(a, \mu) = -\frac{e^{-a/\mu}}{\mu^2}\int_\mu^1 G(\mu, s)\,\frac{ds}{s}\int_a^\infty dr'\, r' \tfrac{1}{2}[cn(r') + S_\nu]\cdot \exp\left(\frac{-(r'^2 - a^2)^{1/2}}{s}\right).$$

The particular solution is given by

$$\psi_p(r, \mu) = \int \frac{d\nu}{N(\nu)}\,\nu\,\phi_p(r, \nu)\phi(\nu, \mu), \tag{4.2}$$

where

$$\phi_p(r, \mu) = \tfrac{1}{2}S_\nu(r - \mu)$$

is the solution of

$$\mu\frac{\partial\phi_p(r, \mu)}{\partial r} + \phi_p(r, \mu) = \tfrac{1}{2}S_\nu r.$$

Again the integral in Eq. (4.2) is used symbolically to denote the full set of normal modes.

We can simplify Eq. (4.2) by noting the following expansions,

$$\frac{1}{1 - c} = \int \frac{d\nu}{N(\nu)}\,\nu\phi(\nu, \mu), \tag{4.3}$$

$$\frac{\mu}{1 - c} = \int \frac{d\nu}{N(\nu)}\,\nu^2\phi(\nu, \mu), \tag{4.4}$$

are applicable, reducing the particular solution to

$$\psi_p(r, \mu) = \tfrac{1}{2}S_\nu(r - \mu)/(1 - c). \tag{4.5}$$

We now write the general solution as

$$\psi(r, \mu) = A_+\, e^{-r/\nu_0}\phi(\nu_0, \mu) + \int_0^1 d\nu\, A(\nu)\, e^{-(r-a)/\nu}\phi(\nu, \mu) + \tfrac{1}{2}S_\nu(r - \mu)/(1 - c), \tag{4.6}$$

with

$$rn(r) = A_+ e^{-r/\nu_0} + \int_0^1 d\nu\, A(\nu) e^{-(r-a)/\nu} \phi(\nu, \mu) + S_\nu/(1-c). \quad (4.7)$$

The application of BC (ii) yields the following valid half-range expansion,

$$A_+ e^{-a/\nu_0} \phi(\nu_0, \mu) + \int_0^1 d\nu\, A(\nu)\phi(\nu, \mu) = -B(\mu) + \int_0^1 d\nu\, A(\nu) K(\mu, \nu), \quad (4.8)$$

where

$$B(\mu) = A_+ e^{-a/\nu_0} K(\mu, \nu_0) + \tfrac{1}{2} S_\nu (a - \mu) + \frac{S_\nu}{2}\left(\frac{3-c}{1-c}\right)\frac{e^{-a/\mu}}{\mu^2} \times \left(\mu + \frac{a}{\mu}\int_\mu^1 ds\, s \frac{I_1((a/\mu s)\sqrt{s^2-\mu^2})}{\sqrt{s^2-\mu^2}}\right). \quad (4.9)$$

The similarity with the results of the Milne problem, Eq. (3.4), are to be noted. Again, the form of the solution is amenable to approximation, and numerical techniques are applicable following the application of the orthogonality operators. Other source problems of interest, e.g., the Green's function and the $1/r$ source found frequently in astrophysics, follow with similar ease, requiring only a knowledge of planar particle transport techniques to solve the transform equation. The extension to finite media and critical problems ($c > 1$) follows with only slight modification of the results presented herein.

In summary, the transform procedure developed in this paper extended the work of Mitsis and Gibbs to include a nonhomogeneous media with distributed sources. The results, as in most present-day transport theory analyses, appear as solutions to a set of coupled Fredholm integral equations for the expansion coefficients of a set of "normal modes." These equations lend themselves to various approximations which provide considerable mathematical insight into the nature of the particle density, and, hopefully, to rapid numerical solution. In general, the transform technique provides a framework wherein a large class of particular problems may be solved in a straightforward manner.

ACKNOWLEDGMENT

The author is grateful for the financial support received from a University of Maryland Initiation Grant making possible the major portion of this research.

APPENDIX A: PARTICULAR SOLUTION

We verify here the procedure leading to the particular solution; Eqs. (2.18) and (2.19) satisfy

$$\mu \frac{\partial}{\partial r}\psi_p(r, \mu) + \psi_p(r, \mu) - \frac{c}{2}\int_{-1}^1 \psi_p(r, \mu')\, d\mu' = S(r'). \quad (A1)$$

We insert Eq. (2.18) into the rhs of Eq. (A1) to give

$$\mu \int \frac{d\nu}{N(\nu)}\phi(\nu, \mu)\frac{\partial \phi_p(r, \nu)}{\partial r} + \int \frac{d\nu}{N(\nu)}\phi(\nu, \mu)\phi_p(r, \nu) + \int \frac{d\nu}{N(\nu)}\phi_p(r, \nu)\left(-\frac{c}{2}\int_{-1}^1 \phi(\nu, \mu')\, d\mu'\right) = \text{rhs}. \quad (A2)$$

Equation (2.19) is now used in the first term of Eq. (A2), yielding

$$\mu \int \frac{d\nu}{N(\nu)}\phi(\nu, \mu)\frac{S(r)}{\nu} + \int \frac{d\nu}{N(\nu)}\phi_p(\chi, \nu)\phi(\nu, \mu)\left(1 - \frac{\mu}{\nu}\right) - \frac{c}{2}\int_{-1}^1 \phi(\nu, \mu')\, d\mu' = \text{rhs}. \quad (A3)$$

The integrals in the first term can be shown formally, or more rigorously by a complex contour integrals, to be

$$\frac{1}{\mu} = \int \frac{d\nu}{N(\nu)\nu}\phi(\nu, \mu), \quad (A4)$$

while the expression in wavy brackets in the second term is zero because of the properties of the eigenfunctions $\phi(\nu, \mu)$.

Thus, rhs = $S(r)$ = lhs.

APPENDIX B: FREDHOLM KERNEL DERIVATION

We wish to simplify integrals of the form

$$I = -\frac{e^{-a/\mu}}{\mu^2}\int_\mu^1 \frac{ds}{s} G(\mu, s)\int_a^R dr'\, r' \frac{c}{2} n(r') \times \exp\left(\frac{-(r'^2 - a^2)^{1/2}}{s}\right). \quad (B1)$$

We use the identity given in Eq. (2.5) to obtain

$$I = -\frac{e^{-a/\mu}}{\mu^2}\int_\mu^1 \frac{ds}{s} G(\mu, s)\int_a^R dr'\, r' \frac{c}{2} n(r') \times \int_0^s \frac{dt}{t^2} e^{-r'/t} G(t, s). \quad (B2)$$

Upon interchanging the order of integration, assuming a form of $n(r'), r'\, n(r') = e^{-r'/\nu}$, and using Eq. (2.15), we may perform the integration over r' to yield

$$I = -\frac{e^{-a/\mu}}{\mu^2}\int_\mu^1 \frac{ds}{s} G(\mu, s)\int_0^s \frac{dt}{t} G(t, s)\phi(-\nu, t) \times \left\{\exp\left[-a\left(\frac{1}{t} - \frac{1}{\nu}\right)\right] - \exp\left[-R\left(\frac{1}{t} - \frac{1}{\nu}\right)\right]\right\}. \quad (B3)$$

The definition of $K(\mu, \nu)$ follows straightforwardly on letting $R \to \infty$. For the special case of an infinite medium the useful integral[15],

$$\int_a^\infty dr' \exp[-(r'^2 - a^2)^{1/2}/s] = (a\pi/2)\,[H_1(a/s) - Y_1(a/s) - 2/\pi] \quad (B4)$$

may be used to reduce $K(\mu, \infty)$ to a more tractable form,

$$K(\mu, \infty) = \frac{ae^{-a/\mu}}{4\pi\mu}\left\{H_1\left(\frac{a}{\mu}\right) - Y_1\left(\frac{a}{\mu}\right) - \frac{2}{\pi} + a\int_\mu^1 \frac{ds}{s} \times \frac{I_1((a/\mu s)\sqrt{s^2-\mu^2})}{\sqrt{s^2-\mu^2}} \cdot \left[H_1\left(\frac{a}{s}\right) - Y_1\left(\frac{a}{s}\right) - \frac{2}{\pi}\right]\right\}. \quad (B5)$$

[1] K. M. Case, Ann. Phys. (N.Y.) 9, 1 (1960).
[2] G. J. Mitsis, "Transport Solutions to the Monoenergetic Critical Problems," Report No. ANL-6787, Argonne National Laboratory, Argonne, Illinois, 1963.
[3] M. R. Mendelson, "Solutions of the One-Speed Neutron Transport Equation in Plane Geometry," Ph.D. thesis (unpublished) (University of Michigan, 1964).

4 G. R. Bond, "The Computational Merits of the Singular Eigenfunction Expansion Method," M.S. thesis (North Carolina State University, 1968).
5 E. H. Bareiss and I. K. Abu-Shumays, "On the Structure of the Isotropic Transport Operator in Three Independent Space Variables," Report No. ANL-7328, Argonne National Laboratory, Argonne, Illinois, 1967.
6 B. Davison, Proc. Phys. Soc. (London) **64**, 881 (1951).
7 B. Davison, "Angular Distribution Due to an Isotropic Point Source and Spherically Symmetric Eigensolutions of the Transport Equation," Report No. MT-112, National Research Council of Canada, Chalk River, Ontario, 1945.
8 R. C. Erdmann and C. E. Siewert, J. Math. Phys. **9**, 81 (1968).
9 A. Leonard and T. W. Mullikan, Proc. Natl. Acad. Sci. (U.S.) **52**, 683 (1964).
10 Alan G. Gibbs, J. Math. Phys. **10**, 5, 875 (1969).
11 B. Davison and J. R. Sykes, *Neutron Transport Theory* (Oxford U.P., London, 1958).
12 I. M. Ryzhik and I. S. Gradstein, *Tables of Series, Products and Integrals* (Springer-Verlag, Berlin, 1957).
13 I. Kuscer *et al.*, Ann. Phys. (N.Y.) **30**, 411 (1964).
14 D. C. Sahni, J. Nucl. Energy **20**, 915 (1966).
15 M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Applied Mathematics Series 55 (National Bureau of Standards, Department of Commerce, Washington, D.C., 1964), Eq. 12.1.8.

Decomposition of the Principal Series of Unitary Irreducible Representations of $SU(2,2)$ Restricted to the Subgroup $SU(1,1) \otimes SU(2)$*

Donald D. Cobb†‡ and Edward R. McCliment

Department of Physics and Astronomy, The University of Iowa, Iowa City, Iowa 52240

(Received 24 May 1971)

An explicit form is given for the unitary irreducible representations of the principal nondegenerate series of the group $SU(2,2)$. These representations are then restricted to the subgroup $SU(1,1) \otimes SU(2)$ and are found to be equivalent to the regular representation of $SU(1,1) \otimes SU(2)$.

1. INTRODUCTION

The examination of the assumption that the relativistic scattering operator is a scalar under the transformations which comprise the Poincaré group has given new insights into possible phenomenological descriptions of high energy scattering processes. It enables the expansion of the scattering amplitude in unitary representation functions, or D functions, of its little groups provided that the amplitude is a square integrable function over the appropriate little group manifold. In this manner one obtains a generalization of the usual partial wave analysis from Poincaré symmetry of the scattering amplitude.[1] Toller *et al.* have demonstrated that such an expansion in the crossed channel provides a natural framework for the Regge pole model of high energy scattering phenomenology.[2-4]

The strong forces between hadrons satisfy so-called internal symmetries in addition to the Poincaré space-time symmetry. In particular, the charge independence of these strong forces is believed to be an exact symmetry expressed by the invariance of the scattering operator under the rotations of the isotopic spin group. We propose that a possible way that remnants of broken symmetries which combine Poincaré and internal symmetries are retained by the amplitude is that it continues to have an expansion in unitary representation functions of the little groups of the primordial symmetry group. In this work we shall attempt to find the remnants of a symmetry which combines the internal isotopic spin symmetry with Poincaré symmetry. In accordance with the above hypothesis, we shall assume that the scattering amplitude can be expanded in representation functions of a new "little group" which we take to be $SU(2,2)$.

$SU(2,2)$ is a semisimple, 15-parameter Lie group. The unitary irreducible representations of $SU(2,2)$ are labeled by the eigenvalues of three Casimir operators.[5-7] $SU(2,2)$ contains the direct product subgroup $SU(1,1) \otimes SU(2)$. We obtain a Regge-like expansion with correlated isotopic spin by interpreting $SU(1,1)$ to be the fixed-t Poincaré little group and $SU(2)$ to be the isotopic spin group. Thus the transformations in $SU(1,1) \otimes SU(2)$ are physical space–time and internal symmetry transformations. By restricting the $SU(2,2)$ transformations to include only those in the $SU(1,1) \otimes SU(2)$ subgroup, an irreducible representation of $SU(2,2)$ becomes equivalent to a reducible representation of $SU(1,1) \otimes SU(2)$. The physical significance of assuming the $SU(2,2)$ expansion of the amplitude is found by decomposition of the representations of $SU(2,2)$ into irreducible representations of its subgroup $SU(1,1) \otimes SU(2)$. In this work we shall consider the decomposition of the principal nondegenerate series of representations of $SU(2,2)$. The representations of the principal nondegenerate series are labeled $[\chi_1, \chi_2, M]$, where χ_1 and χ_2 are real and M is integer or half-integer. Mackey's subgroup theorem provides a means of decomposing the representations of the group G into representations of its subgroups provided the representations of G can be written as induced representations.[8,9] In Sec. 2, the elements of $SU(2,2)$ are parametrized in a manner suitable for inducing the principal nondegenerate series $[\chi_1, \chi_2, M]$. In Sec. 3, we state Mackey's theorem and carry out the decomposition of the principal nondegenerate series into irreducible representations of $SU(1,1) \otimes SU(2)$.

2. THE INDUCING SUBGROUP OF $SU(2,2)$ AND ITS RIGHT COSETS

The transformation $g \in SU(2,2)$ can be represented by 4×4 complex matrices which satisfy:

$$\text{(a) unimodular,} \quad \|g\| = 1$$
$$\text{(b) pseudo-unitary,} \quad g^\dagger \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} g = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \tag{2.1}$$

where I is the 2×2 unit matrix. The Iwasawa decomposition of a noncompact Lie group G is given by

$$G = NAK, \tag{2.2}$$

where N is a nilpotent subgroup, A is an Abelian sub-

4 G. R. Bond, "The Computational Merits of the Singular Eigenfunction Expansion Method," M.S. thesis (North Carolina State University, 1968).
5 E. H. Bareiss and I. K. Abu-Shumays, "On the Structure of the Isotropic Transport Operator in Three Independent Space Variables," Report No. ANL-7328, Argonne National Laboratory, Argonne, Illinois, 1967.
6 B. Davison, Proc. Phys. Soc. (London) **64**, 881 (1951).
7 B. Davison, "Angular Distribution Due to an Isotropic Point Source and Spherically Symmetric Eigensolutions of the Transport Equation," Report No. MT-112, National Research Council of Canada, Chalk River, Ontario, 1945.
8 R. C. Erdmann and C. E. Siewert, J. Math. Phys. **9**, 81 (1968).
9 A. Leonard and T. W. Mullikan, Proc. Natl. Acad. Sci. (U.S.) **52**, 683 (1964).
10 Alan G. Gibbs, J. Math. Phys. **10**, 5, 875 (1969).
11 B. Davison and J. R. Sykes, *Neutron Transport Theory* (Oxford U.P., London, 1958).
12 I. M. Ryzhik and I. S. Gradstein, *Tables of Series, Products and Integrals* (Springer-Verlag, Berlin, 1957).
13 I. Kuscer *et al.*, Ann. Phys. (N.Y.) **30**, 411 (1964).
14 D. C. Sahni, J. Nucl. Energy **20**, 915 (1966).
15 M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Applied Mathematics Series 55 (National Bureau of Standards, Department of Commerce, Washington, D.C., 1964), Eq. 12.1.8.

Decomposition of the Principal Series of Unitary Irreducible Representations of $SU(2,2)$ Restricted to the Subgroup $SU(1,1) \otimes SU(2)$*

Donald D. Cobb†‡ and Edward R. McCliment

Department of Physics and Astronomy, The University of Iowa, Iowa City, Iowa 52240

(Received 24 May 1971)

An explicit form is given for the unitary irreducible representations of the principal nondegenerate series of the group $SU(2,2)$. These representations are then restricted to the subgroup $SU(1,1) \otimes SU(2)$ and are found to be equivalent to the regular representation of $SU(1,1) \otimes SU(2)$.

1. INTRODUCTION

The examination of the assumption that the relativistic scattering operator is a scalar under the transformations which comprise the Poincaré group has given new insights into possible phenomenological descriptions of high energy scattering processes. It enables the expansion of the scattering amplitude in unitary representation functions, or D functions, of its little groups provided that the amplitude is a square integrable function over the appropriate little group manifold. In this manner one obtains a generalization of the usual partial wave analysis from Poincaré symmetry of the scattering amplitude.[1] Toller *et al.* have demonstrated that such an expansion in the crossed channel provides a natural framework for the Regge pole model of high energy scattering phenomenology.[2-4]

The strong forces between hadrons satisfy so-called internal symmetries in addition to the Poincaré space-time symmetry. In particular, the charge independence of these strong forces is believed to be an exact symmetry expressed by the invariance of the scattering operator under the rotations of the isotopic spin group. We propose that a possible way that remnants of broken symmetries which combine Poincaré and internal symmetries are retained by the amplitude is that it continues to have an expansion in unitary representation functions of the little groups of the primordial symmetry group. In this work we shall attempt to find the remnants of a symmetry which combines the internal isotopic spin symmetry with Poincaré symmetry. In accordance with the above hypothesis, we shall assume that the scattering amplitude can be expanded in representation functions of a new "little group" which we take to be $SU(2,2)$.

$SU(2,2)$ is a semisimple, 15-parameter Lie group. The unitary irreducible representations of $SU(2,2)$ are labeled by the eigenvalues of three Casimir operators.[5-7] $SU(2,2)$ contains the direct product subgroup $SU(1,1) \otimes SU(2)$. We obtain a Regge-like expansion with correlated isotopic spin by interpreting $SU(1,1)$ to be the fixed-t Poincaré little group and $SU(2)$ to be the isotopic spin group. Thus the transformations in $SU(1,1) \otimes SU(2)$ are physical space–time and internal symmetry transformations. By restricting the $SU(2,2)$ transformations to include only those in the $SU(1,1) \otimes SU(2)$ subgroup, an irreducible representation of $SU(2,2)$ becomes equivalent to a reducible representation of $SU(1,1) \otimes SU(2)$. The physical significance of assuming the $SU(2,2)$ expansion of the amplitude is found by decomposition of the representations of $SU(2,2)$ into irreducible representations of its subgroup $SU(1,1) \otimes SU(2)$. In this work we shall consider the decomposition of the principal nondegenerate series of representations of $SU(2,2)$. The representations of the principal nondegenerate series are labeled $[\chi_1, \chi_2, M]$, where χ_1 and χ_2 are real and M is integer or half-integer. Mackey's subgroup theorem provides a means of decomposing the representations of the group G into representations of its subgroups provided the representations of G can be written as induced representations.[8,9] In Sec. 2, the elements of $SU(2,2)$ are parametrized in a manner suitable for inducing the principal nondegenerate series $[\chi_1, \chi_2, M]$. In Sec. 3, we state Mackey's theorem and carry out the decomposition of the principal nondegenerate series into irreducible representations of $SU(1,1) \otimes SU(2)$.

2. THE INDUCING SUBGROUP OF $SU(2,2)$ AND ITS RIGHT COSETS

The transformation $g \in SU(2,2)$ can be represented by 4×4 complex matrices which satisfy:

$$\text{(a) unimodular,} \quad \|g\| = 1$$
$$\text{(b) pseudo-unitary,} \quad g^\dagger \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} g = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \tag{2.1}$$

where I is the 2×2 unit matrix. The Iwasawa decomposition of a noncompact Lie group G is given by

$$G = NAK, \tag{2.2}$$

where N is a nilpotent subgroup, A is an Abelian sub-

group, and K is the maximal compact subgroup.[10] We shall find it convenient to express the Iwasawa decomposition of $SU(2,2)$ in a basis in which the matrices of A are diagonal. We make the unitary transformation

$$g \to m^{\dagger} g m, \tag{2.3}$$

where

$$m = \frac{1}{\sqrt{2}} \begin{pmatrix} I & iI \\ iI & I \end{pmatrix}.$$

In this basis Eq. (2.1) becomes[11]

$$g^{\dagger} \begin{pmatrix} 0 & iI \\ -iI & 0 \end{pmatrix} g = \begin{pmatrix} 0 & iI \\ -iI & 0 \end{pmatrix}, \quad g \in SU(2,2). \tag{2.4}$$

Specifically for $SU(2,2)$, A is a two parameter Abelian subgroup, the nilpotent subgroup N has six parameters, and the maximal compact subgroup K is $SU(2) \otimes SU(2) \otimes U(1)$ which has eight parameters. In the basis of Eq. (2.4) in which A is diagonal, we have

$$A = \begin{pmatrix} e^{\alpha} \begin{pmatrix} e^{\xi} & 0 \\ 0 & e \end{pmatrix} & 0 \\ 0 & e^{-\alpha} \begin{pmatrix} e^{-\xi} & 0 \\ 0 & e^{\xi} \end{pmatrix} \end{pmatrix} \tag{2.5}$$

with α and ξ real. The nilpotent subgroup N has the form

$$N = \begin{pmatrix} \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} & H(n) \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}^{-1\dagger} \\ 0 & \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}^{-1\dagger} \end{pmatrix}, \tag{2.6}$$

where β is a complex number and $H(n)$ is 2×2 Hermitian matrix defined as $H(n) = n_0 I + \mathbf{n} \cdot \boldsymbol{\sigma}$ with n_0 and $\mathbf{n}$ real. The compact subgroup K takes the form

$$K = \begin{pmatrix} k_+ & ik_- \\ -ik_- & k_+ \end{pmatrix}, \tag{2.7}$$

where

$$k_{\pm} = \tfrac{1}{2}(u_1 e^{-i\psi/2} \pm u_2 e^{i\psi/2}),$$

and $u_{1,2}$ are two independently parametrized 2×2 unitary, unimodular matrices:

$$u_i = \begin{pmatrix} \cos(\theta_i/2) e^{-i(\mu_i + \nu_i)/2} & -\sin(\theta_i/2) e^{-(\mu_i - \nu_i)/2} \\ \sin(\theta_i/2) e^{i(\mu_i - \nu_i)/2} & \cos(\theta_i/2) e^{i(\mu_i + \nu_i)/2} \end{pmatrix}, \quad i = 1, 2. \tag{2.8}$$

The matrices of Eqs. (2.5), (2.6), and (2.7) were obtained by making the Iwasawa decomposition of the associated Lie algebra which is isomorphic to the Dirac algebra, exponentiating, and then making the transformation of Eq. (2.3).

The inducing subgroup H is defined to be[8]

$$H = NAC(A), \tag{2.9}$$

where $C(A)$ is the centralizer of the maximal compact subgroup K with respect to A, i.e., the subgroup of K which commutes with A. In the $SU(2,2)$ case, it is simply a phase,

$$C(A) = \begin{pmatrix} e^{-i\varphi/2} & & & \\ & e^{i\varphi/2} & & \\ & & e^{-i\varphi/2} & \\ & & & e^{i\varphi/2} \end{pmatrix}. \tag{2.10}$$

Combining Eqs. (2.5), (2.6), and (2.10) according to the definition Eq. (2.9), we see that H is the set of transformations

$$H = \begin{pmatrix} e^{\alpha}\Lambda & H(n)(e^{\alpha}\Lambda)^{-1\dagger} \\ 0 & (e^{\alpha}\Lambda)^{-1\dagger} \end{pmatrix}, \tag{2.11}$$

where α is real and Λ is the set of elements of the form[12]

$$\Lambda = \begin{pmatrix} q^{-1} & \beta \\ 0 & q \end{pmatrix}, \tag{2.12}$$

with q and β complex numbers.

Induced representation theory makes use of a right coset decomposition

$$SU(2,2) = \bigcup_c H g_c \tag{2.13}$$

with respect to the inducing subgroup H. A set of right coset representatives $\{g_c\}$ must be a realization of the coset manifold $SU(2,2)/H$ so that each point in the manifold corresponds to a representative. The manifold is covered up to sets of coset measure zero by two disjoint sets of coset representatives $\{g_c^{(+)}\}$ and $\{g_c^{(-)}\}$ of the form

$$g_c^{(\pm)} = \begin{pmatrix} Z^{(\pm)} & 0 \\ (Z^{(\pm)})^{-1\dagger} & H(x)(Z^{(\pm)})^{-1\dagger} \end{pmatrix}, \tag{2.14}$$

in which the 2×2 matrices $H(x) = x_0 I + \mathbf{x} \cdot \boldsymbol{\sigma}$ with x_0, $\mathbf{x}$ real.[13] The matrices $Z^{(\pm)}$ are defined by

$$Z^{(\pm)} = \begin{pmatrix} 1 & 0 \\ z & \pm 1 \end{pmatrix} \tag{2.15}$$

with z a complex number. The entire set of coset representatives $\{g_c^{(+)}\}$ is a subgroup of $SU(2,2)$ but the subset $\{g_c^{(-)}\}$ cannot be reached continuously from the identity by transformations solely within the coset representative subgroup.[14] The union $\bigcup_c H g_c^{(+)}$ is not a subgroup so that this coset decomposition is not valid if the center of $SU(2,2)$ is removed.[15]

The principal nondegenerate series of representations of $SU(2,2)$ labeled by $[\chi_1, \chi_2, M]$ are induced on the representations of the inducing subgroup H given by Eq. (2.11). In general, given a noncompact group G with right coset decomposition $G = \bigcup H g_c$ with respect to its inducing subgroup H, the induced representations are on a Hilbert space $\mathcal{H}(U)$ of functions over the group with the property that

$$\mathcal{H}(U) = \{f(g) \,|\, f(hg) = L(h) f(g) \ \forall h \in H, g \in G\} \tag{2.16}$$

in which $L(h)$ is a representation of the subgroup H. Since every $g = h g_c$ and $h'h = h''$, it is sufficient to consider only functions over the set of coset representatives $\{g_c\}$. The unitary induced representations $U^{(D)}(g)$ are given by

$$U^{[\Omega]}(g)f(g_c) = \sqrt{\kappa(g,g_c)}f(g_cg) = \sqrt{\kappa(g,g_c)}L(h')f(g_c'), \tag{2.17}$$

where the primed elements refer to those that have been obtained from the decomposition of the element $g' = g_cg$ into a right coset element $g' = h'g_c'$. In particular, $SU(2,2)$ has the right coset decomposition given in Eq. (2.11) and Eq. (2.14). The function $\kappa(g,g_c)$ is the Jacobian of the transformation of the coset measure $d\mu(g_cg_0) = \kappa(g_c,g_0)d\mu(g_c)$.[16] The representations (one-dimensional) of the inducing subgroup H are given by

$$h \to L(h) = (e^{\alpha})^{i\chi_1}|q|^{i\chi_2}(q/|q|)^{2M} \tag{2.18}$$

for each $h \in H$.[17] χ_1 and χ_2 are real and M is integer or half-integer. The principal nondegenerate series is induced on $L(h)$,

$$U^{[\chi_1\chi_2 M]}_{(g)}f(\bar{g}_c) = (e^{\alpha'})^{i\chi_1}|q'|^{i\chi_2-2}(q'/|q'|)^{2M}f(g_c'), \tag{2.19}$$

where the primes refer to the decomposition of the element $\bar{g}_cg = h'g_c'$. The norm on $\mathfrak{H}(U^{[\chi_1\chi_2 M]})$ is

$$\|f\|^2 = \int_{SU(2,2)/H} d\mu(g_c)|f(g_c)|^2 \tag{2.20}$$

with coset measure $d\mu(g_c) = d(\mathrm{Re}Z)d(\mathrm{Im}Z)dx_0dx_1dx_2dx_3$. The coset transformation $\bar{g}_cg = h'g_c'$ is obtained by matrix multiplication of elements with the proper form. In particular, we must find the parameters α' and q' of h', the 4-vector $(x_0'\mathbf{x}')$ in $H(x')$, and the complex parameter z' in $Z'^{(\pm)}$ of $g_c'^{(\pm)}$. In terms of the corresponding 4×4 matrices $\bar{g}_chg_c^{(+)} = h'g_c'$ is

$$\begin{pmatrix} \bar{Z} & 0 \\ \bar{Z}^{-1\dagger}H(\bar{x}) & \bar{Z}^{-1\dagger} \end{pmatrix}\begin{pmatrix} e^{\alpha}\Lambda\, H(n)(e^{\alpha}\Lambda)^{-1\dagger} & \\ 0 \qquad (e^{\alpha}\Lambda)^{-1\dagger} & \end{pmatrix}\begin{pmatrix} Z & 0 \\ Z^{-1\dagger}H(x) & Z^{-1\dagger} \end{pmatrix}$$

$$= \begin{pmatrix} e^{\alpha'}\Lambda'\, H(n')(e^{\alpha'}\Lambda')^{-1\dagger} & \\ 0 \qquad (e^{\alpha'}\Lambda')^{-1\dagger} & \end{pmatrix}\begin{pmatrix} Z'^{(\pm)} & 0 \\ (Z'^{(\pm)})^{-1\dagger}H(x') & (Z'^{(\pm)})^{-1\dagger} \end{pmatrix}, \tag{2.21}$$

where $\bar{Z} = \bar{Z}^{(+)}$ and $Z = Z^{(+)}$. The cases of $\bar{Z}^{(-)}$, etc., follow trivially. Carrying out the matrix multiplication, we obtain sets of 2×2 matrix equations which are soluble for the primed variables,

$$(e^{\alpha'}\Lambda'Z'^{(\pm)})^{-1\dagger}H(x') = \bar{Z}^{-1\dagger}[I + H(\bar{x})H(n)](e^{\alpha}\Lambda Z)^{-1\dagger}H(x) + \bar{Z}^{-1\dagger}H(\bar{x})(e^{\alpha}\Lambda Z) \tag{2.22}$$

and

$$(e^{\alpha'}\Lambda'Z'^{(\pm)})^{-1\dagger} = \bar{Z}^{-1\dagger}[I + H(\bar{x})H(n)](e^{\alpha}\Lambda Z)^{-1\dagger}. \tag{2.23}$$

Substitution of Eq. (2.23) into Eq. (2.22) gives a 2×2 matrix equation for $H(x')$,

$$H(x') = H(x) + e^{2\alpha}(\Lambda Z)^{\dagger}[I + H(\bar{x})H(n)]^{-1}H(\bar{x})(\Lambda Z). \tag{2.24}$$

A solution to the set of equations in Eq. (2.24) will exist if the determinant $\Delta(\bar{x},n) \equiv \|I + H(\bar{x})H(n)\|$ does not vanish.[18]

It is easy to show that the matrix $H(y) = [I + H(\bar{x})H(n)]^{-1}H(\bar{x})$ is Hermitian. ΛZ in Eq. (2.24) is a general element of the right cosets of $SL(2,C)$ defined by Naimark in Ref. 12, i.e., almost all 2×2 complex matrices $b \in SL(2,C)$, $\|b\| = 1$, can be written as $b = \Lambda Z$. [If $Z = Z^{(-)}$, the transformations contain improper $SL(2,C)$ elements $b = \Lambda Z^{(-)}$ with $\|b\| = -1$.] Then with $b \equiv \Lambda Z$, Eq. (2.24) contains the transformation $H(y') = b^{\dagger}H(y)b$ of the Hermitian matrix $H(y)$.

It is easy to verify that $H(y')$ is also Hermitian and one can find the equations for y_{μ}' $\mu = 0,1,2,3$. The $SL(2,C)$ transformations of $H(y)$ leave $\|H(y)\| = y_0^2 - \mathbf{y}^2$ invariant and correspond to Lorentz transformations of the 4-vector $y \to y'$. The solution for the 4-vector x_{μ}' is given by

$$x_{\mu}' = x_{\mu} + e^{2\alpha}y_{\mu}', \qquad \mu = 0,1,2,3, \tag{2.25}$$

with y_{μ}' the 4-vector obtained from

$$y_{\mu} = (1/\Delta)(\bar{x}_{\mu} - n_{\mu}\bar{x}_{\nu}^2) \tag{2.26}$$

by the Lorentz transformation corresponding to ΛZ. The quantity $\Delta(\bar{x},n) = 1 + 2(\bar{\mathbf{x}}\cdot\bar{\mathbf{n}} + \bar{x}_0n_0) + \bar{x}_{\mu}^2n_{\mu}^2$.[19]

In order to find α', q' and z', we rewrite Eq. (2.23)

$$e^{\alpha'}\Lambda'Z'^{(\pm)} = \bar{Z}[I + H(\bar{x})H(n)]^{-1\dagger}(e^{\alpha}\Lambda Z). \tag{2.27}$$

Consider the determinants of the left and right sides of Eq. (2.27), $\|LS\| = (\pm)e^{2\alpha'}$ and $\|RS\| = e^{2\alpha}/\Delta(\bar{x},n)$. A solution to Eq. (2.27) can exist if $\|LS\| = \|RS\|$, that is,

$$(\pm)e^{2\alpha'} = e^{2\alpha}/\Delta(\bar{x},n) \tag{2.28}$$

and Eq. (2.28) is the solution for α'.

The choice of sign in Eq. (2.28) depends on the sign of $\Delta(\bar{x},n)$. The plus sign is chosen if $\Delta(\bar{x},n) > 0$ and the minus if $\Delta(\bar{x},n) < 0$. Thus whether $g = hg_c^{(\pm)}$ in Eq. (2.21) moves $\bar{g}_c^{(\pm)}$ to $g_c'^{(+)}$ or $g_c'^{(-)}$ is determined by the *sign* of $\Delta(\bar{x},n)$,

$$\begin{aligned} &\bar{g}_c^{(\pm)}hg_c^{(\pm)} = h'g_c'^{(+)} \quad \text{if } \Delta(\bar{x},n) > 0, \\ &\bar{g}_c^{(\pm)}hg_c^{(\pm)} = h'_cg'^{(-)} \quad \text{if } \Delta(\bar{x},n) < 0. \end{aligned} \tag{2.29}$$

The sign of $\Delta(\bar{x},n)$ will always determine in which piece of the coset space g' will lie. This result is not surprising because $e^{2\alpha}\Delta(\bar{x},n)$ is the determinant of the 2×2 submatrix g_{22}' of $g' = h'g_c'^{(\pm)}$.[15]

In order to find q' and z', we note that the LS and RS of Eq. (2.27) are 2×2 complex matrices with arbitrary real ($\neq 0$) determinants. These matrices can be written as right coset elements by a trivial extension of the $SL(2,C)$ coset decomposition in Ref. 12. Consider an arbitrary 2×2 complex matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

We require $\|A\|$ to be real, but nonzero so that A contains seven parameters. If $a_{22} \neq 0$, A can be decomposed in the following manner:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} s & \zeta \\ 0 & t \end{pmatrix}\begin{pmatrix} 1 & 0 \\ w & 1 \end{pmatrix}, \tag{2.30}$$

where $\|A\| = st$. Application of this decomposition to the left-hand side of Eq. (2.27) and comparing gives, after some algebra,

$$(\pm)q' = \sqrt{|\Delta(\bar{x},n)|}\,(b_{12}F_1 - b_{22}F_2),$$
$$(\pm)z' = (b_{11}F_1 - b_{21}F_2)/(b_{12}F_1 - b_{22}F_2), \tag{2.31}$$

where $b = \Lambda Z$ and

$$F_1 = \bar{z}[1 + (\bar{x}_0 - \bar{x}_3)(n_0 - n_3)(\bar{x}_1 - i\bar{x}_2)(n_1 + in_2)] - (\bar{x}_0 + \bar{x}_3)(n_1 + in_2) - (\bar{x}_1 + i\bar{x}_2)(n_0 - n_3), \tag{2.32}$$

$$F_2 = \bar{z}[(\bar{x}_1 - i\bar{x}_2)(n_0 + n_3) + (\bar{x}_0 - \bar{x}_3)(n_1 - in_2)] - [1 + (\bar{x}_0 + \bar{x}_3)(n_0 + n_3) + (\bar{x}_1 + i\bar{x}_2)(n_1 - in_2)] \tag{2.33}$$

The choice of sign in Eq. (2.31) is the same as in Eq. (2.28). Equations (2.25), (2.29), and (2.31) are the transformation equations necessary in Eq. (2.19) to complete the definition of $[\chi_1, \chi_2, M]$.

3. THE CONTENT OF THE REPRESENTATIONS OF $SU(1,1) \otimes SU(2)$ IN $[\chi_1, \chi_2, M]$

The following summary of Mackey's subgroup theorem contains an outline of the procedure that must be used in its application. We are given an induced representation $U^{[L]}(g)$ of a group G on a Hilbert space $\mathcal{H}(U^{[L]})$ defined by Eq. (2.16). We wish to find the content of the representations of a subgroup $\hat{G} \subset G$. Consider the double coset decomposition

$$G = \bigcup_D H g_D \hat{G}, \tag{3.1}$$

where H is the inducing subgroup for $U^{[L]}(g)$. The double cosets are nonoverlapping. The subspace $\mathcal{H}_D(U^{[L]})$ consisting of the functions over one double coset,

$$\{f(\bar{g}) \,|\, \bar{g} \in H g_D \hat{G}\} \tag{3.2}$$

forms an invariant subspace of $\mathcal{H}(U^{[L]})$ with respect to all right transformations $\hat{g} \in \hat{G}$. The functions of $\mathcal{H}_D$ therefore are a basis for a (reducible) representation of $\hat{G}$. Since the double cosets do not overlap, $\mathcal{H}(U^{[L]})$ decomposes into[20]

$$\mathcal{H}(U^{[L]}) \simeq \int^{\oplus} d\mu(D)\mathcal{H}_D(U^{[L]}). \tag{3.3}$$

Specifically, Mackey's subgroup theorem states that the representations of $\hat{G}$ contained in $\mathcal{H}$ are equivalent to representations induced by a subgroup $H_D \subset \hat{G}$ given by

$$H_D = g_D^{-1} H g_D \cap \hat{G} \tag{3.4}$$

The representations of the new inducing subgroup H_D are obtained from those of H by

$$L_D(h_D) \equiv L(g_D h_D g_D^{-1}) \quad \text{for } h_D \in H_D. \tag{3.5}$$

Since the functions of $\mathcal{H}_D$ are a subset of $\mathcal{H}$, they have the property

$$f(g_D\hat{g}) = f(g_D h_D \hat{g}_c) = f(g_D h_D g_D^{-1} g_D \hat{g}_c) = L_D(h_D) f(g_D \hat{g}_c). \tag{3.6}$$

Thus, for any one double coset $g_D \hat{G}$, we may restrict ourselves to functions over elements $g_D \hat{g}_c$. Let us define $F(\hat{g}_c) = f(g_D \hat{g}_c)$. Then the representation of $\hat{G}$ induced by H_D is given by

$$U^{[L_D]}(\hat{g})F(\hat{g}_c) = F(\hat{g}_c\hat{g}) = L_D(h'_D)F(\hat{g}'_c), \tag{3.7}$$

where $\hat{g}_c\hat{g} = h'_D\hat{g}'_c$. The definition $F(\hat{g}_c) = f(g_D\hat{g}_c)$ and Eq. (3.7) establish the equivalence with $\mathcal{H}_D$. The norm on the set of functions, $\{F\}$, is given by

$$\|F\|^2 = \int_{\hat{G}/H_D} d\mu(\hat{g}) \| F(\hat{g}) \|^2. \tag{3.8}$$

Note that $U^{[L_D]}$, in general, is a reducible representation of $\hat{G}$ which still must be decomposed into unitary irreducible representations of $\hat{G}$. However, Mackey has shown that $U^{[L_D]}$ is contained in the regular representation of $\hat{G}$. The decomposition of the *regular representation* into irreducible representations is known for some noncompact groups.

The above description of Mackey's subgroup theorem implies that it is necessary to have a double coset decomposition of $SU(2,2)$ with respect to its inducing subgroup H and $SU(1,1) \otimes SU(2)$:

$$SU(2,2) = \bigcup_D H g_D [SU(1,1) \otimes SU(2)]. \tag{3.9}$$

To this end, we shall parametrize $SU(1,1) \otimes SU(2)$ in a manner similar to Eq. (2.2) for $SU(2,2)$. In the basis in which the invariant matrix is as in Eq. (2.4), we find that the Iwasawa decomposition of $SU(1,1) \otimes SU(2)$ is given in terms of the matrices

$$\hat{N} = \begin{pmatrix} I & \hat{n}I \\ 0 & I \end{pmatrix},$$
$$\hat{A} = \begin{pmatrix} e^{\hat{\alpha}}I & 0 \\ 0 & e^{-\hat{\alpha}}I \end{pmatrix} \tag{3.10}$$

and

$$\hat{K} = \begin{pmatrix} \hat{u}\cos(\hat{\psi}/2) & \hat{u}\sin(\hat{\psi}/2) \\ -\hat{u}\sin(\hat{\psi}/2) & \hat{u}\cos(\hat{\psi}/2) \end{pmatrix},$$

with $\hat{n}$ and $\hat{\alpha}$ real. The matrix $\hat{u} \in SU(2)$ is parametrized as

$$\hat{u} = \begin{pmatrix} \cos(\hat{\theta}/2)e^{-i(\hat{\mu}+\hat{\nu})/2} & -\sin(\hat{\theta}/2)e^{-i(\hat{\mu}-\hat{\nu})/2} \\ \sin(\hat{\theta}/2)e^{+i(\hat{\mu}-\hat{\nu})/2} & \cos(\hat{\theta}/2)e^{+i(\hat{\mu}+\hat{\nu})/2} \end{pmatrix}. \tag{3.11}$$

Then $SU(1,1) \otimes SU(2) = \hat{N}\hat{A}\hat{K}$ and $\hat{A}, \hat{N}$, and $\hat{K}$ are subgroups of A, N, and K in Eqs. (2.2), (2.5)–(2.7).

A right coset decomposition $SU(1,1) \otimes SU(2) = \bigcup \hat{H}\hat{g}_c$ can be obtained following the procedure of Sec. 2.[c] One finds

$$\hat{H} = \begin{pmatrix} e^{\hat{\alpha}}\hat{u} & \hat{n}e^{-\hat{\alpha}}\hat{u} \\ 0 & e^{-\hat{\alpha}}\hat{u} \end{pmatrix} \tag{3.12}$$

and

$$\hat{g}_c = \begin{pmatrix} I & 0 \\ \hat{x}I & I \end{pmatrix}, \tag{3.13}$$

where $\hat{x}$ is a real number. Almost all $\hat{g} \in SU(1,1) \otimes SU(2) = \hat{N}\hat{A}\hat{K}$ can be written $\hat{g} = \hat{h}\hat{g}_c$, $\hat{h} \in \hat{H}$, $\hat{g}_c \in \{\hat{g}_c\}$, except those transformations with subdeterminant $\|\hat{g}_{22}\| = 0$.

The unitary irreducible representations of $SU(1,1) \otimes SU(2)$ can be written as Kronecker products of the representations $D^{[\lambda]}(\hat{v})$ of $SU(1,1)$ and $D^{[I]}(\hat{u})$ of $SU(2)$, where $\hat{v} \in SU(1,1)$, $\hat{u} \in SU(2)$. Then the representation $[\lambda, I]$ is given by the product $D^{[\lambda]}(\hat{v}) \otimes D^{[I]}(\hat{u})$.[21,22] The diagonal labels j_3, j'_3 and i_3, i'_3 of the matrix elements $D^{[\lambda]}_{j_3 j'_3}(\hat{v}) \times D^{[I]}_{i_3, i'_3}(\hat{u})$ are constrained by[23]

$$i_3 + j_3 = \text{integer}, \quad i'_3 + j'_3 = \text{integer}. \tag{3.14}$$

To find the double coset representatives g_D, we first note that the g_D must themselves be right coset elements of $SU(2,2)$ with respect to H so that we may write

$$g_D = \begin{pmatrix} Z_D & 0 \\ Z_D^{-1\dagger} H(x_D) & Z_D^{-1\dagger} \end{pmatrix}, \tag{3.15}$$

where[24]

$$Z_D = \begin{pmatrix} 1 & 0 \\ Z_D & 1 \end{pmatrix}. \tag{3.16}$$

Next, consider the action of an arbitrary $\hat{g} \in SU(1,1) \otimes SU(2)$ on a double coset representative $g_D\hat{g} = g \in SU(2,2)$. Since $\hat{g}$ can be written as $\hat{h}\hat{g}_c$ and g as hg_c, the transformation of g_D is also given by

$$g_D\hat{h}\hat{g}_c = hg_c \tag{3.17}$$

In general, the set of all transformations in $SU(1,1) \otimes SU(2)$ will map a particular representative g_D onto a set of $SU(2,2)$ right coset representatives $g_D \xrightarrow{\{\hat{g}\}} \{g_c^{(\pm)}\}_D$. If one can find the set $\{g_D\}$ which is mapped uniquely onto $\{g_c^{(\pm)}\}$, this will be the required set of double coset representatives for Eq. (3.9).

In terms of the appropriate transformation matrices, Eq. (3.17) becomes

$$\begin{pmatrix} Z_D & 0 \\ Z_D^{-1\dagger} H(x_D) & Z_D^{-1\dagger} \end{pmatrix} \begin{pmatrix} e^{\hat{\alpha}\hat{u}} & \hat{n}e^{-\hat{\alpha}\hat{u}} \\ 0 & e^{-\hat{\alpha}\hat{u}} \end{pmatrix} \begin{pmatrix} I & 0 \\ \hat{x}I & I \end{pmatrix}$$
$$= \begin{pmatrix} e^{\alpha\Lambda} & H(n)(e^{\alpha\Lambda})^{-1\dagger} \\ 0 & (e^{\alpha\Lambda})^{-1\dagger} \end{pmatrix} \begin{pmatrix} Z^{(\pm)} & 0 \\ (Z^{(\pm)})^{-1\dagger} H(x) & (Z^{(\pm)})^{-1\dagger} \end{pmatrix}. \tag{3.18}$$

These equations are a special case of Eq. (2.21) for the transformation of the cosets of $SU(2,2)$. It is found that by fixing the parameters in g_D to be $Z_D = 1$, $x_{D0} = x_{D3} = \frac{1}{2}$, and $x_{D1} = x_{D2} = 0$ the resulting double coset representative is mapped by the transformations in $SU(1,1) \otimes SU(2)$ onto almost all $SU(2,2)$ right coset representatives $\{g_c\}$. Thus the only double coset representative required for the subgroup theorem is

$$g_D = \begin{pmatrix} Z_D & 0 \\ \frac{1}{2}(I + \sigma_3) & Z_D^{-1\dagger} \end{pmatrix}, \quad Z_D = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \tag{3.19}$$

and almost all $g \in SU(2,2)$ can be decomposed as elements of the double coset $Hg_DSU(1,1) \otimes SU(2)$. With this choice of g_D, the mapping of the coset parameters of $\hat{g}_c$ under the transformation given in Eq. (3.18) is

$$\begin{aligned} x_0 &= \hat{x}_0 \pm |\hat{r}/2|, \\ x_3 &= \pm |\hat{r}/2| \cos\hat{\theta}, \\ x_1 + ix_2 &= \mp |\hat{r}/2| \sin\theta \; e^{-i\hat{v}}, \end{aligned} \tag{3.20}$$

$$\text{and } z = \pm e^{-i\hat{v}} \frac{\sin(\hat{\theta}/2)(|\Delta(\hat{n})| e^{i\hat{\mu}}) \pm \cos(\hat{\theta}/2)}{\cos(\hat{\theta}/2)(|\Delta(\hat{n})| e^{i\hat{\mu}}) \mp \sin(\hat{\theta}/2)}, \tag{3.21}$$

where $\Delta(\hat{n}) = 1 + \hat{n}$ and $r = \sqrt{\hat{x}_1^2 + \hat{x}_2^2 + \hat{x}_3^2}$. The choice of signs in Eqs. (3.20) and (3.21) corresponds to the sign of $\Delta(\hat{n})$. The upper (lower) signs are chosen if $\Delta(\hat{n}) > 0$ ($\Delta(\hat{n}) < 0$). The sign of $\Delta(\hat{n})$ determines whether g_D is mapped into $\{g_c^{(+)}\}$ or $\{g_c^{(-)}\}$.

With $\hat{G} = SU(1,1) \otimes SU(2)$ in Eq. (3.4), we note that the new inducing subgroup H_D is the set of elements satisfying

$$H_D = \{\hat{g} | \hat{g} \in SU(1,1) \otimes SU(2), \; g_D\hat{g} = hg_D\}. \tag{3.22}$$

Thus H_D contains the elements $SU(1,1) \otimes SU(2)$ which send g_D into itself. From the transformation $g_D\hat{g} = hg_c$, Eq. (3.17), it is not difficult to show that the only element of $SU(1,1) \otimes SU(2)$ which sends g_D into g_D is the identity element e and $H_D = \{e\}$.

Finally consider the reducible representation $U(\hat{g})^{[\chi_1,\chi_2,M]}$, where $\hat{g} \in SU(1,1) \otimes SU(2)$. This representation, by the subgroup theorem, is equivalent to a reducible representation $U^{[1]}(\hat{g})$ of $SU(1,1) \otimes SU(2)$, where [1] denotes the fact that this representation is induced by the representation of $\{e\}$. $U^{[1]}$ acts in a Hilbert space of functions $F \in \mathcal{K}_D$ with the property $F(h_D\hat{g}) = F(e\hat{g}) = F(\hat{g})$, $\hat{g} \in SU(1,1) \otimes SU(2)$. Thus F is a function over all of $SU(1,1) \otimes SU(2)$ since $[SU(1,1) \otimes SU(2)]/H_D \simeq SU(1,1) \otimes SU(2)$. The norm on $\mathcal{K}_D$ is

$$\|F_D\|^2 = \int_{\{\hat{g}\}} d\mu(\hat{g}) |F(\hat{g})|^2 \tag{3.23}$$

where $d\mu(\hat{g})$ is a measure in the parameter space of $SU(1,1) \otimes SU(2)$.

Therefore, $\mathcal{K}_D$ is just the space of all square-integrable functions over $SU(1,1) \otimes SU(2)$. A unitary representation which acts in the Hilbert space of all square-integrable functions over a group is the regular representation of that group. Therefore, $U^{[1]}$ is identified as the *regular representation* of $SU(1,1) \otimes SU(2)$.

The decomposition of the regular representation of $SU(2)$ into unitary irreducible representations is well known. The regular representation of $SU(1,1)$ can be decomposed via the $SU(1,1)$ Plancherel formula.[22] Then the content of $[\lambda, I]$ in $[\chi_1, \chi_2, M]$ is

$$U^{[\chi_1,\chi_2,M]}(\hat{v},\hat{u}) \simeq \int^{\oplus} d\mu(\lambda) \sum^{\oplus} (2\iota + 1) D^{[\lambda]}(\hat{v}) \otimes D^{[I]}(\hat{u}), \tag{3.24}$$

where $\hat{v} \in SU(1,1), \hat{u} \in SU(2)$. $\int^{\oplus} d\mu(\lambda)$ is a direct integral over the unitary irreducible representations of $SU(1,1)$ appearing in the Plancherel measure $d\mu(\lambda)$ and $\sum^{\oplus}$ is a direct sum of $SU(2)$ representations. Note that Eq. (3.24) implies that the multiplicity of a representation $[\lambda, I]$ in any $[\chi_1, \chi_2, M]$ is just its multiplicity in the regular representation of $SU(1,1) \otimes SU(2)$. In general, this multiplicity is infinite.

4. CONCLUSION

The content of the unitary irreducible representations of $SU(1,1) \otimes SU(2)$ in the representations of the principal nondegenerate series of $SU(2,2)$ given in Eq. (3.24) is the main result of this work. As a secondary result the transformation of the right cosets of $SU(2,2)$ obtained in Sec. 2 should be of use for further study of $SU(2,2)$ involving its induced representations. Since the entire regular representation of $SU(1,1) \otimes SU(2)$ is contained in each member of the

principal nondegenerate series of $SU(2,2)$ we conclude that the physical content of the specific assumption that a scattering amplitude can be expanded in the set of representation matrices belonging to the principal nondegenerate series of $SU(2,2)$ representations is uninteresting. From the discussion of Ref. 22 we point out, however, that this may be attributed in part to the manner in which we have decided to build $SU(1,1) \otimes SU(2)$ into $SU(2)$. If the full $SU(1,1) \otimes SU(2)$ is imbedded in $SU(2,2)$, it is conceivable that more interesting results may be forthcoming. This matter is presently under investigation.

ACKNOWLEDGMENTS

The authors are pleased to acknowledge several useful conversations concerning the theory of group representations with Dr. Granville Smith, Dr. E. Thieleker, and especially Dr. William Klink.

---

* Portions of this work submitted in fulfillment of the degree Doctor of Philosophy in The University of Iowa.

† NASA Predoctoral Trainee.

‡ Current address: E G & G, Inc., Los Alamos, New Mexico 87544.

[1] The unitary representations of the Poincaré group and the generalized partial wave analysis of two-body scattering amplitudes are discussed by J. Boyce, R. Delbourge, A. Salam, and J. Strathdee in "Partial Wave Analysis I", IAEA Preprint IC/67/9 (Trieste, 1967, unpublished).

[2] L. Sertorio and M. Toller, Nuovo Cimento **33**, 413 (1964); M. Toller, *ibid.* **37**, 631 (1965); **53A**, 671 (1968). See also C. E. Jones *et al.*, "Generalized $O(2,1)$ Expansion for Asymptotically Growing Amplitudes" [to be published in Ann. Phys. (N.Y.)].

[3] F. Hadjioannou, Nuovo Cimento **44**, 185 (1966).

[4] A review of the Regge pole model is given by G. Hite, Rev. Mod. Phys. **41**, 669 (1969).

[5] Y. Murai, Progr. Theoret. Phys. (Kyoto) **9**, 147 (1953); A. Esteve and P. Sona, Nuovo Cimento **32**, 473 (1964); M. I. Graev, Tr. Mosk. Mat. Obs. **7**, 335 (1958) [Am. Math. Soc. Transl. **66**, (1968)].

[6] F. Halbwachs, A. Kihlberg, and V. Müller, Commun. Math. Phys. **3**, 194 (1966).

[7] T. Yao, J. Math Phys. **8**, 1931 (1967).

[8] W. Klink, *Lectures in Theoretical Physics* (U. of Colorado Press, Boulder, Colorado, 1964), Vol. IID.

[9] G. Mackey, *The Theory of Group Representations* (U. of Chicago Press, Chicago, 1955).

[10] R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).

[11] W. Klink, Phys. Rev. **181**, 2027 (1969). See also N. W. Macfadyen, J. Math. Phys. **12**, 1436 (1971).

[12] M. Naimark, *Linear Representations of the Lorentz Group* (MacMillan, New York, 1964). In Chapter III of this reference the principal series of $SL(2,C)$ representations are written as induced representations induced by one-dimensional representations of the subgroup $\{\Lambda\}$.

[13] I is the 2×2 unit matrix and σ is the set of Pauli matrices.

[14] The set of right coset representatives $\{g_c^{(\pm)}\}$ forms a group, of which the set $\{g_c^{(+)}\}$ is a subgroup. The set $\{g_c^{(-)}\}$ is obtained from $\{g_c^{(+)}\}$ by multiplying by the discrete element $\begin{pmatrix} \sigma_3 & 0 \\ 0 & \sigma_3 \end{pmatrix}$ of $SU(2,2)$. The situation is analogous to adding parity to the proper Lorentz group.

[15] If we write the general $SU(2,2)$ element g in terms of a 2×2 submatrices g_{ij}, $g = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$, the determinant $\|g_{22}\|$ lies in the range $-\infty < \|g_{22}\| < \infty$. For all elements of the form $g = hg_c^{(+)}$, $\|g_{22}\| > 0$, meaning another set is necessary. Then the sign of $\|g_{22}\|$ determines whether $g \in \bigcup_c Hg_c^{(+)}$ or $g \in \bigcup_c Hg_c^{(-)}$. See also Eq. (2.21) and the following.

[16] The transformation of the coset measure $\kappa(g_c^{(+)}, g)$ has been found by Esteve and Sona in Ref. 5 for the case $g_c^{(+)} g = hg_c'^{(+)}$. They find $\kappa(g_c^{(+)}) = |q|^{-4}$, where q is a complex parameter of the matrix Λ [Eq. (2.12)].

[17] It is a property of induced representations that the elements of the nilpotent subgroup map into the identity representation.

[18] None of the coset transformations in Eq. (2.21) exist if $\Delta(\vec{x}, n) = 0$. Writing g' as in Ref. 15, these singular transformations correspond to $SU(2,2)$ elements with $\|g'_{22}\| = 0$. Such elements cannot be written as $h' g_c'^{(+)}$. They are contained in an exceptional set of cosets of measure zero.

[19] The choice of metric is such that $n_\mu^2 = n_0^2 - \mathbf{n}^2$, etc.

[20] The notation "$\simeq$" means "is equivalent to" and $\int^{\oplus} d\mu(D)$ denotes a direct integral of subspaces labeled by double cosets. The precise definition of the measure $d\mu(D)$ is given by Mackey, Ref. 9.

[21] Kronecker products of representations are defined in M. Hamermesh, *Group Theory and Its Applications to Physical Problems* (Addison-Wesley, Reading, Mass., 1962).

[22] The unitary representation functions $D^{\lambda}_{j_3 j_3'}$ of $SU(1,1)$ are defined in A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967). There the representations $[\lambda]$ contained in the regular representation of $SU(1,1)$ are labeled by $\lambda \equiv l = \frac{1}{2} + i\,\mathrm{Im}l$ for the principal series and $\lambda \equiv k$, $k = \frac{1}{2}, 1, \frac{3}{2}, \cdots$ for the discrete series.

[23] This restriction is imposed by the manner in which we have imbedded $SU(1,1) \otimes SU(2)$ in $SU(2,2)$. By constructing the 4×4 matrix representation of $SU(1,1) \otimes SU(2)$ which is the open product of the 2×2 matrix representation of $SU(1,1)$ and $SU(2)$, respectively, the elements $\begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$ and $\begin{pmatrix} -1 & \\ & -1 \end{pmatrix} \otimes \begin{pmatrix} -1 & \\ & -1 \end{pmatrix}$ both map into the identity of the 4×4 matrix representation of $SU(1,1) \otimes SU(2)$. $SU(1,1) \otimes SU(2)$ is actually a covering of the group $\hat{G} = \hat{N}\hat{A}\hat{K}$ of Eq. (3.10). In a physical model, the restrictions of Eq. (3.14) essentially amount to the Gell-Mann–Nishijima rule for nonstrange particles.

[24] A general element of the double coset is $hg_D\hat{g}$ where $h \in H$, $\hat{g} \in SU(1,1) \otimes SU(2)$ for fixed g_D. If the identity element of $SU(1,1) \otimes SU(2)$ is chosen, then it is clear that g_D is a right coset element.

Necessary Conditions for N-Representability of Reduced Density Matrices

A. J. Coleman

Department of Mathematics, Queen's University, Kingston, Ontario, Canada.

(Received 17 May 1971)

In this work a large new class of extreme points of the set of N-representable 2-matrices is obtained. They correspond to the exact ground state of an operator $B(g)$ which generalizes the BCS Hamiltonian. Further, three hitherto unpublished necessary conditions for N-representability are obtained.

1. INTRODUCTION

The problem of describing the limitations on a pth-order reduced density matrix, and its related Green's functions, which derive from the fact that it is formed by contraction from an N-fermion (or boson) system, was called the *N-representability problem* by the present author[1] in his paper SFDM-I. Giving it a name merely sharpened and made precise a problem which had been noticed by several previous authors.[2] The significance of the problem arises from the fact that most of the information accessible to physicists about the state of a system of identical particles is contained in the 2-matrix. Thus a "practicable" solution of the N-representability problem for the 2-matrix would almost eliminate the N-particle wavefunction from quantum mechanics.

It was shown in SFDM-I that necessary and sufficient conditions for ensemble N-representability of the 1-

principal nondegenerate series of $SU(2,2)$ we conclude that the physical content of the specific assumption that a scattering amplitude can be expanded in the set of representation matrices belonging to the principal nondegenerate series of $SU(2,2)$ representations is uninteresting. From the discussion of Ref. 22 we point out, however, that this may be attributed in part to the manner in which we have decided to build $SU(1,1) \otimes SU(2)$ into $SU(2)$. If the full $SU(1,1) \otimes SU(2)$ is imbedded in $SU(2,2)$, it is conceivable that more interesting results may be forthcoming. This matter is presently under investigation.

ACKNOWLEDGMENTS

The authors are pleased to acknowledge several useful conversations concerning the theory of group representations with Dr. Granville Smith, Dr. E. Thieleker, and especially Dr. William Klink.

---

* Portions of this work submitted in fulfillment of the degree Doctor of Philosophy in The University of Iowa.

† NASA Predoctoral Trainee.

‡ Current address: E G & G, Inc., Los Alamos, New Mexico 87544.

[1] The unitary representations of the Poincaré group and the generalized partial wave analysis of two-body scattering amplitudes are discussed by J. Boyce, R. Delbourge, A. Salam, and J. Strathdee in "Partial Wave Analysis I", IAEA Preprint IC/67/9 (Trieste, 1967, unpublished).

[2] L. Sertorio and M. Toller, Nuovo Cimento **33**, 413 (1964); M. Toller, *ibid.* **37**, 631 (1965); **53A**, 671 (1968). See also C. E. Jones *et al.*, "Generalized $O(2,1)$ Expansion for Asymptotically Growing Amplitudes" [to be published in Ann. Phys. (N.Y.)].

[3] F. Hadjioannou, Nuovo Cimento **44**, 185 (1966).

[4] A review of the Regge pole model is given by G. Hite, Rev. Mod. Phys. **41**, 669 (1969).

[5] Y. Murai, Progr. Theoret. Phys. (Kyoto) **9**, 147 (1953); A. Esteve and P. Sona, Nuovo Cimento **32**, 473 (1964); M. I. Graev, Tr. Mosk. Mat. Obs. **7**, 335 (1958) [Am. Math. Soc. Transl. **66**, (1968)].

[6] F. Halbwachs, A. Kihlberg, and V. Müller, Commun. Math. Phys. **3**, 194 (1966).

[7] T. Yao, J. Math Phys. **8**, 1931 (1967).

[8] W. Klink, *Lectures in Theoretical Physics* (U. of Colorado Press, Boulder, Colorado, 1964), Vol. IID.

[9] G. Mackey, *The Theory of Group Representations* (U. of Chicago Press, Chicago, 1955).

[10] R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).

[11] W. Klink, Phys. Rev. **181**, 2027 (1969). See also N. W. Macfadyen, J. Math. Phys. **12**, 1436 (1971).

[12] M. Naimark, *Linear Representations of the Lorentz Group* (MacMillan, New York, 1964). In Chapter III of this reference the principal series of $SL(2,C)$ representations are written as induced representations induced by one-dimensional representations of the subgroup $\{\Lambda\}$.

[13] I is the 2×2 unit matrix and σ is the set of Pauli matrices.

[14] The set of right coset representatives $\{g_c^{(\pm)}\}$ forms a group, of which the set $\{g_c^{(+)}\}$ is a subgroup. The set $\{g_c^{(-)}\}$ is obtained from $\{g_c^{(+)}\}$ by multiplying by the discrete element $\begin{pmatrix} \sigma_3 & 0 \\ 0 & \sigma_3 \end{pmatrix}$ of $SU(2,2)$. The situation is analogous to adding parity to the proper Lorentz group.

[15] If we write the general $SU(2,2)$ element g in terms of a 2×2 submatrices g_{ij}, $g = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$, the determinant $\|g_{22}\|$ lies in the range $-\infty < \|g_{22}\| < \infty$. For all elements of the form $g = hg_c^{(+)}$, $\|g_{22}\| > 0$, meaning another set is necessary. Then the sign of $\|g_{22}\|$ determines whether $g \in \bigcup_c Hg_c^{(+)}$ or $g \in \bigcup_c Hg_c^{(-)}$. See also Eq. (2.21) and the following.

[16] The transformation of the coset measure $\kappa(g_c^{(+)}, g)$ has been found by Esteve and Sona in Ref. 5 for the case $g_c^{(+)} g = hg_c'^{(+)}$. They find $\kappa(g_c^{(+)}) = |q|^{-4}$, where q is a complex parameter of the matrix Λ [Eq. (2.12)].

[17] It is a property of induced representations that the elements of the nilpotent subgroup map into the identity representation.

[18] None of the coset transformations in Eq. (2.21) exist if $\Delta(\vec{x}, n) = 0$. Writing g' as in Ref. 15, these singular transformations correspond to $SU(2,2)$ elements with $\|g'_{22}\| = 0$. Such elements cannot be written as $h' g_c'^{(+)}$. They are contained in an exceptional set of cosets of measure zero.

[19] The choice of metric is such that $n_\mu^2 = n_0^2 - \mathbf{n}^2$, etc.

[20] The notation "$\simeq$" means "is equivalent to" and $\int^{\oplus} d\mu(D)$ denotes a direct integral of subspaces labeled by double cosets. The precise definition of the measure $d\mu(D)$ is given by Mackey, Ref. 9.

[21] Kronecker products of representations are defined in M. Hamermesh, *Group Theory and Its Applications to Physical Problems* (Addison-Wesley, Reading, Mass., 1962).

[22] The unitary representation functions $D^{\lambda}_{j_3 j_3'}$ of $SU(1,1)$ are defined in A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967). There the representations $[\lambda]$ contained in the regular representation of $SU(1,1)$ are labeled by $\lambda \equiv l = \frac{1}{2} + i\,\mathrm{Im}l$ for the principal series and $\lambda \equiv k$, $k = \frac{1}{2}, 1, \frac{3}{2}, \cdots$ for the discrete series.

[23] This restriction is imposed by the manner in which we have imbedded $SU(1,1) \otimes SU(2)$ in $SU(2,2)$. By constructing the 4×4 matrix representation of $SU(1,1) \otimes SU(2)$ which is the open product of the 2×2 matrix representation of $SU(1,1)$ and $SU(2)$, respectively, the elements $\begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$ and $\begin{pmatrix} -1 & \\ & -1 \end{pmatrix} \otimes \begin{pmatrix} -1 & \\ & -1 \end{pmatrix}$ both map into the identity of the 4×4 matrix representation of $SU(1,1) \otimes SU(2)$. $SU(1,1) \otimes SU(2)$ is actually a covering of the group $\hat{G} = \hat{N}\hat{A}\hat{K}$ of Eq. (3.10). In a physical model, the restrictions of Eq. (3.14) essentially amount to the Gell-Mann–Nishijima rule for nonstrange particles.

[24] A general element of the double coset is $hg_D\hat{g}$ where $h \in H$, $\hat{g} \in SU(1,1) \otimes SU(2)$ for fixed g_D. If the identity element of $SU(1,1) \otimes SU(2)$ is chosen, then it is clear that g_D is a right coset element.

Necessary Conditions for N-Representability of Reduced Density Matrices

A. J. Coleman

Department of Mathematics, Queen's University, Kingston, Ontario, Canada.

(Received 17 May 1971)

In this work a large new class of extreme points of the set of N-representable 2-matrices is obtained. They correspond to the exact ground state of an operator $B(g)$ which generalizes the BCS Hamiltonian. Further, three hitherto unpublished necessary conditions for N-representability are obtained.

1. INTRODUCTION

The problem of describing the limitations on a pth-order reduced density matrix, and its related Green's functions, which derive from the fact that it is formed by contraction from an N-fermion (or boson) system, was called the *N-representability problem* by the present author[1] in his paper SFDM-I. Giving it a name merely sharpened and made precise a problem which had been noticed by several previous authors.[2] The significance of the problem arises from the fact that most of the information accessible to physicists about the state of a system of identical particles is contained in the 2-matrix. Thus a "practicable" solution of the N-representability problem for the 2-matrix would almost eliminate the N-particle wavefunction from quantum mechanics.

It was shown in SFDM-I that necessary and sufficient conditions for ensemble N-representability of the 1-

matrix D^1 can be expressed in terms of the eigenvalues of D^1—simply that the eigenvalues of ND^1 are at most unity. Alternatively, if I denotes the identity operator in one-particle space, this condition requires that the operator $\mathrm{I} - \mathrm{ND}^1$ be positive semidefinite. Apparently, the N-representability of the 2-matrix D^2 is much more complicated so that necessary and sufficient conditions involve not only limitations on the eigenvalues of D^2 but also complex interrelations among its eigenfunctions. The set of ensemble N-representable 2-matrices, which we denote by $\mathcal{P}_N^2$, is convex. Its closure, in the topology derived from the scalar product $\langle A|B\rangle = \mathrm{tr}(A^\dagger B)$, is compact, and hence the set is determined by its extreme points. The main result of the present paper is a theorem asserting that, for an *arbitrary* antisymmetric 2-particle function g, $D^2(g^N)$ is extreme in $\mathcal{P}_N^2$. Here N is even and g^N is the normalized antisymmetrized geminal power (AGP) function of N-particles derived from g and defined in SFDM-I. A particular case of this result, namely when $D^1(g)$ has all its eigenvalues equal, was proved in SFDM-I (Theorem 9.4). However, lifting the limitation on g constitutes a very great and rather surprising generalization of the previous result, and its proof requires quite different ideas.

In the course of proving the above theorem we show that a certain operator $B(g)$ belongs to the dual of the convex cone determined by $\mathcal{P}_N^2$, and obtain the exact solution for the ground state of $B(g)$. Since $B(g)$ involves two-particle interaction of a more general nature than does the solvable case of the pairing Hamiltonian, we expect that $B(g)$ will serve as a source of model Hamiltonians pertinent to the study of a wide variety of correlation phenomena in many-particle systems. Both Hartree–Fock and BCS systems arise as particular cases, depending on the choice of g.

The paper concludes with the proof of three necessary conditions for N-representability which were announced[3] some years ago but which have not hitherto been published. The first asserts that the rank of an N-representable p-matrix cannot be less than the corresponding value for a Hartree–Fock system, namely the binomial coefficient $\binom{N}{p}$. The other two involve the concept *strong orthogonality* defined in the next section. They illustrate rather dramatically that N-representability of a 2-matrix forces complex interrelations among its eigenfunctions—i.e., among its *natural geminals* or *nags*. If $\{g_i\}, i = 1, 2, \ldots, k$, are nags of an N-particle fermion system in state Ψ and if the g_i are mutually strongly orthogonal, then the sum of the corresponding eigenvalues of $D^2(\Psi)$ is at most $(N-1)^{-1}$. That this result remains true for *all* values of k seemed quite surprising to the author when it was first discovered. Finally, if a nag and its conag are strongly orthogonal, the corresponding eigenvalue is not greater than the Hartree Fock value, namely $\binom{N}{2}^{-1}$.

A survey of the state of the N-representability problem as of the summer of 1967 was given by the author in a report,[4] which is still in print.

2. NOTATION AND TERMINOLOGY

Recall that if Ψ is an antisymmetric wavefunction the p-*operator* $D^p(\Psi)$ sends $2p$ particles $(1\ 2\ \ldots\ p;\ 1'\ 2'\ \ldots\ p')$ into the p-*matrix*

$$D^p(1\ 2\ \ldots\ p;\ 1'\ 2'\ \ldots\ p') = \int_{p+1\cdots N} \Psi(1\ 2\cdots N)\overline{\Psi}(1'\ \ldots\ p'\ p+1\cdots N) = \sum \lambda_i^p \alpha_i^p(1\ \ldots\ p)\overline{\alpha}_i^p(1'\ \ldots\ p'), \tag{2.1}$$

where α_i^p are eigenfunctions and λ_i^p are the corresponding eigenvalues. Note that in this context superscripts are not construed as powers or indices but simply indicate the number of particles occurring in the corresponding functions. The cases $p = 1$ and $p = 2$ are so important as to deserve special terminology: α_i^1 is a *natural orbital* or *norb*; α_i^2 is a *natural geminal* or *nag*. We use the terms *orbital* and *geminal* to denote arbitrary functions of one and two particles, respectively. Physicists and chemists should be warned that what is here, for the sake of brevity, called an orbital they would normally refer to as a *spin-orbital*.

The linear space, over the complex numbers, spanned by $\{\alpha_i^p\}$ will be called the p-*range of* Ψ and denoted by $R^p(\Psi)$. Thus $R^1(\Psi)$ is the familiar space spanned by a one-particle basis set. If Ψ and φ are two antisymmetric (or symmetric) functions, not necessarily of the same number of particles, we say that Ψ and φ are *strongly orthogonal* if $R^1(\Psi)$ is orthogonal to $R^1(\varphi)$, in other words, if all the norbs of Ψ are orthogonal to all the norbs of φ *or*, again, if the ranges of $D^1(\Psi)$ and $D^1(\varphi)$ are orthogonal.

Let A_N denote the antisymmetrizer on N particles, so that if f is an arbitrary function of N particles, $A_N f$ is antisymmetric in N variables. Of course, $A_N f$ may be the zero function. If Ψ^p and Ψ^q, where $p + q = N$, are antisymmetric functions of p and q variables respectively, then

$$\Psi^p \wedge \Psi^q\colon (1\ 2\ \ldots\ N) \to A_N \Psi^p(1\ \ldots\ p)\Psi^q(p+1\ \ldots\ N). \tag{2.2}$$

The function $\Psi^p \wedge \Psi^q$, the "wedge" or "Grassmann" product of Ψ^p and Ψ^q, is an antisymmetric function of N variables. The set of antisymmetric functions equipped with the Grassmann product and ordinary addition of functions forms an algebra or ring. If $\Psi = f \wedge g$, we shall say that f and g are *Grassmann factors* of Ψ. When the context precludes ambiguity, we may refer to f and g simply as *factors* of Ψ and say that f *divides* Ψ. If f is an *orbital*, there is a remarkable theorem, probably due to Grassmann, that f divides Ψ if and only if $f \wedge \Psi = 0$. However, if f is a function of two or more variables, there seems to be no simple criterion for characterizing whether or not f is a factor of Ψ.

If C is a convex subcone of a real linear space V, then the *polar* cone $\bar{C}$ is defined as the set of real valued linear functionals f which are nonnegative on C. That is, $\bar{C} = \{f \mid x \in C \Rightarrow f(x) \geq 0\}$. When V is provided with a nonsingular scalar product, the functional f can be identified with a vector y, say, such that $f(x) = \langle y|x\rangle$ for all x. In this case we define $\bar{C} = \{y \mid x \in C \Rightarrow \langle y|x\rangle \geq 0\}$. The set $\mathcal{P}_N^2$ consisting of the convex closure of 2-operators which are representable by means of normalized N-particle functions, defines a cone obtained by multiplication of the elements of $\mathcal{P}_N^2$ by the nonnegative reals. For the circumambient

vector space V, following Kummer,[5] we take the real space S^2 of bounded symmetric two-particle operators. As the scalar product of $A, B \in S^2$, we take $\mathrm{tr}(AB)$. Then by $\bar{\mathcal{P}}^2_N$ we denote the set $\{B \mid B \in S^2, D^2 \in \mathcal{P}^2_N \Rightarrow \mathrm{tr}(BD^2) \geq 0\}$. For $D^2 \in \mathcal{P}^2_N$ we say that B *exposes* D^2 if $B \in \bar{\mathcal{P}}^2_N$ and $\mathrm{tr}(BD^2) = 0$, and the only other elements of $\mathcal{P}^2_N$ which satisfy this equation are scalar multiples of D^2.

We adopt the common usage of denoting the boundary of a set $\mathcal{P}$ by $\partial\mathcal{P}$. The n-fold tensor product of a linear space H is denoted by H^n; the antisymmetric subspace of H^n by $H^{n\wedge}$. If A^p and B^q are operators on $H^{p\wedge}$ and $H^{q\wedge}$ respectively, then, for $p + q = N$, $A^p \wedge B^q$ is the operator on $H^{N\wedge}$ defined by $A_N A^p \otimes B^q A_N$, where $\otimes$ denotes the usual tensor product of operators.

3. THE OPERATORS $F(g)$ AND $B(g)$

A closed convex cone C is the polar of its own polar $\bar{C}$. Thus, if we knew $\bar{\mathcal{P}}^2_N$, then, in principle, $\mathcal{P}^2_N$ could be characterized as the set $\{D^2 \mid \mathrm{tr}(D^2B) \geq 0, B \in \bar{\mathcal{P}}^2_N\}$. Indeed, since for an element D^2 on the boundary $\partial\mathcal{P}^2_N$ of $\mathcal{P}^2_N$ there is one or more $B \in \bar{\mathcal{P}}^2_N$ such that $\mathrm{tr}(BD^2) = 0$, it is sufficient to know those $B \in \bar{\mathcal{P}}^2_N$ which are exposed by some $D^2 \in \partial\mathcal{P}^2_N$. Note that if $D^2 \in \mathcal{P}^2_N$, $B \in \bar{\mathcal{P}}^2_N$, and $\mathrm{tr}(D^2B) = 0$, then we can conclude that $D^2 \in \partial\mathcal{P}^2_N$ and $B \in \partial\bar{\mathcal{P}}^2_N$.

Thus in order to characterize $\mathcal{P}^2_N$ by means of a set of linear inequalities it is sufficient to use those determined by $B \in \partial\bar{\mathcal{P}}^2_N$. Even these are more than are generally needed since a set of extreme points of $\bar{\mathcal{P}}^2_N$ is enough. For example, $\mathcal{P}^1_N$ is completely characterized as consisting of those D^1 which have unit trace, $\mathrm{tr}D^1 = 1$ and satisfy the two conditions

$$D^1 \geq 0, \quad I - ND^1 \geq 0. \tag{3.1}$$

The first of these asserts that D^1 is a positive semidefinite operator and the second that the eigenvalues of D^1 are not greater than N^{-1}. Conditions (3.1) follow from the fact that if P_φ is the projector onto an arbitrary orbital,

$$P_\varphi \quad \text{and} \quad I - NP_\varphi \tag{3.2}$$

are extreme in $\bar{\mathcal{P}}^2_N$. Since the operators (3.2) exhaust the extreme points of $\bar{\mathcal{P}}^1_N$, (3.1) are sufficient to characterize $\mathcal{P}^1_N$.

The conditions (3.1) are equivalent to results obtained in SFDM-I and constitute a neat and satisfying solution of the ensemble N-representability problem for the 1-matrix. However, to the author's knowledge, a satisfactory solution of the *pure* N-representability problem even for the 1-matrix has still not been obtained. Since any two 1-matrices are unitarily equivalent if they have the same set (with multiplicity) of eigenvalues, and since N-representability is invariant under unitary transformations of the orbital basis, we know that the solution can be formulated as a condition on the eigenvalues of D^1. Unfortunately, for $p > 1$ the N-representability of D^p involves not only conditions on the eigenvalues but also on the interrelations of eigenfunctions. Evidently, this is the reason that for $p > 1$ the N-representability problem is some orders of magnitude more difficult than for $p = 1$.

Since an N-representable D^2 is a positive semidefinite operator on $H^{2\wedge}$ it is clear that the projector P_g onto the geminal g belongs to $\bar{\mathcal{P}}^2_N$ for arbitrary g. It follows from the spectral theorem that P_g is extreme in $\bar{\mathcal{P}}^2_N$. In the next theorem we prove that two more classes of operators $F(g)$ and $B(g)$ each depending on an arbitrary geminal belong to $\bar{\mathcal{P}}^2_N$. Two geminals g_1 and g_2 are equivalent under unitary transformations of the orbital basis if and only if $D^1(g_1)$ and $D^1(g_2)$ have the same eigenvalues with the same multiplicities. Recall that for a geminal these multiplicities are even. Thus, the specification of a monotonely nonincreasing sequence of positive reals, $\lambda_1 = \lambda_2$, $\lambda_2 = \lambda_4, \ldots \lambda_{2s-1} = \lambda_{2s}$, $\sum \lambda_i = 1$, determines a unitary equivalence class of operators $F(g)$ and $B(g)$. For each $r = 2s$ we shall obtain a family, of dimension $s - 1$, of unitary equivalence classes of elements of $\bar{\mathcal{P}}^2_N$. Each class has dimension $r^2 - r - 1$ if $R^1(g)$ is finite, and infinite otherwise.

For g an arbitrary normalized geminal, define

$$F(g) = I^2 - 2ND^1(g) \wedge I^1 + \tfrac{1}{2}N(N-1)P_g, \tag{3.3}$$

$$B(g) = I^2 - (N-2)D^1(g) \wedge I^1 - (N-1)P_g, \tag{3.4}$$

where I^p is the identity on $H^{p\wedge}$ and $P_g = D^2(g)$ is the projector onto the geminal g. Thus $F(g)$ and $B(g)$ are Hermitian operators on $H^{2\wedge}$ and therefore belong to the real space S^2 of all two-particle Hermitian operators which we take as the circumambient space for $\mathcal{P}^2_N$ and $\bar{\mathcal{P}}^2_N$.

Theorem 3.1: The operators $F(g)$ and $B(g)$ belong to $\bar{\mathcal{P}}^2_N$, the polar of the set $\mathcal{P}^2_N$ of N-representable 2-operators.

Proof: (a) We first prove that $F(g) \in \bar{\mathcal{P}}^2_N$. Denote the antisymmetrizer on $N + 2$ particles $1, 2, \ldots, N + 2$ by A_{N+2} and let $\chi(1\ 2 \cdots N+2) = A_{N+2}g(1\ 2)\Psi(3\ 4 \cdots N+2)$ where Ψ is an arbitrary N-particle wavefunction. Then by Theorem 6.2 of SFDM-I, which follows from Sasaki's formula for $p = 2$,

$$\begin{aligned}\binom{N+2}{2}\langle\chi|\chi\rangle &= \binom{N+2}{2}\langle A_{N+2}g\Psi|A_{N+2}g\Psi\rangle \\ &= 1 - 2N\,\mathrm{tr}[D^1(g)D^1(\Psi)] \\ &\quad + \binom{N}{2}\mathrm{tr}[D^2(g)D^2(\Psi)] \\ &= \mathrm{tr}\left[\left(1 - 2ND^1(g) + \binom{N}{2}P_g\right)D^2(\Psi)\right] \\ &= \mathrm{tr}[F(g)D^2(\Psi)]. \end{aligned} \tag{3.5}$$

Since $\langle\chi|\chi\rangle \geq 0$, this implies that $F(g) \in \bar{\mathcal{P}}^2_N$. Note that (3.5) vanishes if and only if $\chi = 0$.

(b) To show that $B(g) \in \bar{\mathcal{P}}^2_N$, recall that if P is an orthogonal projector on a Hilbert space and χ is any vector, then $\|P\chi\| \leq \|\chi\|$ with equality if and only if $P\chi = \chi$. We apply this choosing $P = A_{N+2}$ and $\chi = A_{N+1}g\Psi$ where A_{N+1} acts on the particles numbered 2 to $N + 2$. By applying Theorem 6.2 of SFDM-I twice, with p equal to 1 and 2, we find that

$$2\left[I - 2Na + \binom{N}{2}b\right] \leq (N+2)[1 - Na], \tag{3.6}$$

where

$$a = \mathrm{tr}[D^1(g)D^1(\Psi)], \quad b = \mathrm{tr}[P_g D^2(\Psi)]. \tag{3.7}$$

Note that $a \geq 0, b \geq 0$. It follows from (3.6) that

$$1 - (N-2)a - (N-1)b \geq 0$$

that is

$$\mathrm{tr}[B(g)D^2(\Psi)] \geq 0 \tag{3.8}$$

for arbitrary antisymmetric N-particle functions Ψ. Thus $B(g) \in \bar{\mathcal{O}}_N^2$.

Notice that equality obtains in (3.8) if and only if

$$A_{N+2}g\Psi = A_{N+1}g\Psi. \tag{3.9}$$

In the course of the preceding proof of Theorem 3.1, our curiosity is aroused to discover what pairs g, Ψ imply that

$$\mathrm{tr}[F(g)D^2(\Psi)] = 0 \quad \text{or} \quad \mathrm{tr}[B(g)D^2(\Psi)] = 0.$$

Either of these equations can obtain only for associated pairs of elements of the surfaces $\partial\bar{\mathcal{O}}_N^2$ and $\partial\mathcal{O}_N^2$, respectivly. The answer to this question for $B(g)$ when the rank $r_1(g)$ of g is not less than N is provided by the main theorem of this paper. The case $r_1(g) < N$ is not fully understood. The question for $F(g)$ can be reformulated in terms of the hole or Q matrix[6] taking the form, when does the Q matrix have g as an eigenfunction with zero eigenvalue?

Before we leave Theorem 3.1 let us comment on the proof of part (b). Our method of proof was rather simple minded. Indeed, the key inequality (3.8) already occurred in conclusion (iv) of Theorem 3 of the author's 1962 Uppsala preprint.[7] At that epoch the author was not thinking about convex sets—the relevance of which first became apparent later in one of those celebrated midnight discussions at Sanibel presided over by Per Olov Löwdin—so that he did not draw the important conclusions formulated in Theorem (3.1). There is another, more sophisticated, method of proving part (b) of the theorem. We describe it here since it may possibly lead to greater insight.

For $g(1\ 2)$ and $\Psi(3 \ldots N+2)$ antisymmetric, the function $\chi = g\Psi$, when acted upon by the symmetric group $\mathcal{S}_{N+2}$ permuting variables, generates a linear function space V of dimension δ say, which is a representation space for $\mathcal{S}_{N+2}$. As such V is the direct sum of carrier spaces of irreducible representations of $\mathcal{S}_{N+2}$. One easily concludes, by the theory of induced representations or otherwise, that the only possible irreducibles occurring in V are denoted in the common notation by $[1^{N+2}]$, $[2, 1^N]$, and $[2^2\ 1^{N-2}]$. The carrier spaces V_i, $1 \leq i \leq 3$, of these irreducibles are of dimension $1, N+1$, and $\frac{1}{2}(N+2)(N-1)$, respectively. The sum of these three dimensions is $\binom{N+2}{2}$ which is the number of linearly independent functions which can be generated by $\mathcal{S}_{N+2}$ acting on $\chi = g\Psi$ when g and Ψ are the most general antisymmetric 2 and N particle functions. Thus

$$\chi = \chi_1 + \chi_2 + \chi_3 \quad \text{where } \chi_i \in V_i.$$

Part (a) of Theorem 3.1 was proved simply by noting that $\| \chi_1 \|^2 \geq 0$, where χ_1 is the completely antisymmetrized part $A_{N+2}\chi$ of χ. Similarly, it is possible to prove part (b) of Theorem 3.1 by expressing the fact that $\| \chi_2 \|^2 \geq 0$, since it can be shown that

$$(N+2)\| \chi_2 \|^2 = 2[1 - (N-2)a - (N-1)b], \tag{3.10}$$

where a and b are defined in (3.7). It can also be shown that

$$(N+1)\| \chi_3 \|^2 = (N-1)[1 + 2a + b]. \tag{3.11}$$

Since $a \geq 0$ and $b \geq 0$ it follows from (3.11) that $\chi_3 \neq 0$ so that χ always has a nonzero component in V_3. Thus there are four possibilities for the dimension δ of V according as χ_1 and χ_2 do or do not vanish. All four possibilities can in fact occur as is shown by the examples exhibited in Table I, where [123] for example

TABLE I.

g	Ψ	χ_1	χ_2	δ
[12]	[345]	$\neq 0$	$\neq 0$	$\frac{1}{2}(N+2)(N+1)$
[12]	[123]	0	0	$\frac{1}{2}(N+2)(N-1)$
[12]	[134]	0	$\neq 0$	$\frac{1}{2}N(N+3)$
g	g^N	$\neq 0$	0	$\frac{1}{2}N(N+1)$

denotes a Slater determinant involving three orthonormal orbitals $\varphi_1, \varphi_2, \varphi_3$, g denotes a geminal of rank greater than $N+1$, and g^N is the AGP function generated by g. In the first three lines of Table I the examples correspond to $N = 3$ but are susceptible of trivial generalization corresponding to the dimensions reported for δ.

If g and Ψ are such that *both* (3.5) and (3.8) vanish then by solving two linear equations, we infer that

$$a = N^{-1}, \quad b = 2/N(N-1). \tag{3.12}$$

Since a can be interpreted as a weighted mean of eigenvalues of $D^1(\Psi)$, with weights determined by g, and since $\lambda_1^1(\Psi) \leq N^{-1}$, the first equality in (3.12) occurs if and only if the 1-range $R^1(g)$ of g is contained in the subspace of $R^1(\Psi)$ corresponding to the eigenvalue N^{-1}. Hence, by Corollary 4.3A of SFDM-I, all of the norbs of g are Grassmann factors of Ψ. Thus the rank $r(g)$ of g is $\leq N$. Suppose $r(g) = 2s$ since the rank of a geminal is necessarily even. Then the AGP function g^{2s}, being a function of $2s$ variables and also of rank $2s$, is simply a Slater determinant on the 1-range of g. Since any orbital in $R^1(g)$ is a Grassmann factor of Ψ, so is g^{2s}. Hence $a = N^{-1}$ implies that

$$\Psi = g^{2s} \wedge \varphi, \tag{3.13}$$

where φ is an antisymmetric function of $N - 2s$ variables. Without loss of generality we assume that φ is strongly orthogonal to g. Since g^{2s} is a Slater determinant, $D^2(g^{2s})$ is simply the identity operator on $R^2(g^{2s})$ multiplied by $\binom{N}{2}^{-1}$. With the help of a result due to McWeeny it also follows easily that $b = \binom{N}{2}^{-1}$. The result of McWeeny which we need is the italicized statement on page 361 of his 1960 review article.[8] Since it is very useful and apparently not well known, we repeat it here. Suppose φ_i, $1 \leq i \leq m$, are m mutually strongly orthogonal antisymmetric functions with Grassmann product denoted by $\wedge\ \varphi_i$; then

$$\rho_2(\wedge \varphi_i; 1\ 2, 1'2') = \sum_i \rho_2(\varphi_i; 1\ 2, 1'2') + \sum_{i\neq j} [\rho_1(\varphi_i; 1, 1')\rho_1(\varphi_j; 2, 2') - \rho_1(\varphi_i; 2, 1')\rho_1(\varphi_j; 1, 2')]. \quad (3.14)$$

Here ρ_1 and ρ_2 are the 1- and 2-operators normalized to N and $N(N-1)$, respectively. As a check on this formula take the trace of both sides. The trace of the last expression in (3.14) is zero. Suppose φ_i is a function of N_i particles; then we have, for the lhs, $N(N-1)$ and, for the rhs,

$$\sum N_i(N_i - 1) + \sum_{i\neq j} N_i N_j = \sum N_i^2 + 2\sum_{i<j} N_i N_j - N = N^2 - N$$

which is correct.

Using McWeeny's formula (3.14), we easily derive a converse to the above result. If $\Psi = g^{2s} \wedge \varphi$, where $r(g) = 2s \leq N$ and φ is strongly orthogonal to g, then $a = N^{-1}$ and $b = \binom{N}{2}^{-1}$. We thus have the interesting

Theorem 3.2: If g is a geminal of rank $2s \leq N$ and Ψ is an N-fermion wavefunction, then $\mathrm{tr}[F(g)D^2(\Psi)] = \mathrm{tr}[B(g)D^2(\Psi)] = 0$ if and only if $\Psi = g^{2s} \wedge \varphi$, where φ and g are strongly orthogonal.

4. SOME EXTREME POINTS OF $\mathcal{P}_N^2$

Since a compact convex set is determined by its extreme points, one method of characterizing $\mathcal{P}_N^2$—in other words of solving the N-representability problem for 2-matrices—would be to enumerate all extreme points of $\mathcal{P}_N^2$. It is easy to show that the preimage, under $(N-2)$-contraction, of an extreme point contains a pure state. Erdahl has recently proved that the extreme points of $\mathcal{P}_N^2$ are exposed and has conjectured that their preimages are unique, that is, consist of one pure state. However, in general, the 2-matrix of a pure state is not extreme. It may lie in the boundary but will usually be interior to $\mathcal{P}_N^2$. For $\mathcal{P}_N^1$ we have a fairly complete story. A Slater determinant covers an extreme point of $\mathcal{P}_N^1$. A pure state Ψ with one or more eigenvalues of $D^1(\Psi)$ equal to N^{-1} covers a boundary point of $\mathcal{P}_N^1$. All other Ψ cover interior points.

Theorem 9.4 of SFDM-I asserts that $D^2(g^N)$ is extreme in $\mathcal{P}_N^2$ if g is a geminal of extreme type, that is, if all the eigenvalues of $D^1(g)$ are equal. Witten extended this result in an unpublished work by showing that if g is of extreme type and S is a Slater determinant in s variables strongly orthogonal to g, then $D^2(S \wedge g^{N-s})$ is extreme in $\mathcal{P}_N^2$. Of course, $N-s$ must be even.

In the present section we obtain a major extension of Witten's result by showing that the conclusion remains valid if we allow the antisymmetric geminal g to be arbitrary except that its rank be greater than $N-1$.

Theorem 4.1: If N is even and g is an arbitrary fermion geminal of rank $\geq N$, then an element D^2 of $\mathcal{P}_N^2$ satisfies $\mathrm{tr}[D^2B(g)] = 0$ if and only if D^2 is covered by a pure state Ψ proportional to the AGP function g^N. For all such g, $D^2(g^N)$ is extreme in $\mathcal{P}_N^2$.

Proof: (a) If D^2 is not pure N-representable, assume it is of the form

$$D^2 = \sum_i w^i D^2(\Psi_i), \quad (4.1)$$

where $w^i > 0$, $\sum w^i = 1$, and $D^2(\Psi_i)$ is N-representable by the pure state Ψ_i. We assume that

$$\mathrm{tr}[B(g)D^2] = \sum w^i\, \mathrm{tr}[B(g)D^2(\Psi_i)] = 0. \quad (4.2)$$

But $B(g) \in \tilde{\mathcal{P}}_N^2$; therefore,

$$\mathrm{tr}[B(g)D^2(\Psi_i)] \geq 0$$

and (3.2) is possible only if

$$\mathrm{tr}[B(g)D^2(\Psi_i)] = 0 \quad (4.3)$$

for all Ψ_i.

(b) We are thus led to consider a pure state Ψ such that

$$\mathrm{tr}[B(g)D^2(\Psi)] = 0, \quad (4.4)$$

and we shall show that requiring (4.4) is sufficient to force Ψ to be proportional to g^N. Therefore, each Ψ_i is proportional to g^N, and all the $D^2(\Psi_i)$ in (4.1) are equal to $D^2(g^N)$ and is thus covered by g^N, as was to be proved.

(c) The proof of the theorem has therefore been reduced to showing that (4.4) holds if and only if Ψ is proportional to g^N. However, in the course of the proof of Theorem 3.1 we noted that (4.4) holds if and only if (3.9) is true. But, by Theorem 7.7 of SFDM-I, if $\Psi = g^N$, then (3.9) holds and therefore so does (4.4). Thus the theorem has been proved in one direction, and it follows by the argument in (b) that $D^2(g^N)$ is extreme in $\mathcal{P}_N^2$.

(d) To prove the converse for arbitrary even N is rather more difficult. However, the essence of the proof is completely evident in the particular case $N = 4$, which we therefore set out in detail.

We assume (3.9) in the equivalent form

$$A_5\Psi(1234)g(56) = A_6\Psi(1234)g(56), \quad (4.5)$$

where A_5 acts on the variables 1, 2, 3, 4, 5 and A_6 on all six variables. Define

$$\chi = 5A_5\Psi(1234)g(56) = \Psi(1234)g(56) - \Psi(5234)g(16) - \Psi(1534)g(26) - \Psi(1254)g(36) - \Psi(1235)g(46).$$

If (4.5) is true, then χ is antisymmetric in particles 4 and 6 so that

$$\chi + (46)\chi = 0, \quad (4.6)$$

where by $(46)\chi$ we denote the effect of permuting particles 4 and 6 in χ. We identify particles 3 and 6 in (4.6), multiply the resulting equation by $g(12)$, recall that g and Ψ are antisymmetric, and obtain

$$\Psi(1234)g(12)g(53) - \Psi(1253)g(12)g(34) = \Psi(5234)g(12)g(13) - \Psi(5134)g(12)g(23). \quad (4.7)$$

Symmetrize both sides of Eq. (4.7) in the particles 1, 2, and 3.

The circular permution (123) acting on the first term on the rhs of (4.7) turns it into

$$\Psi(5314)g(23)g(21) = \Psi(5134)g(23)g(12),$$

which is the last term in (4.7) except for sign. Hence the symmetrized rhs of (4.7) is zero.

Since $\Psi(1234)$ is antisymmetric in 1, 2, 3, symmetrizing the first term on the lhs of (4.7) gives rise to $\Psi(1234)A_3 g(12)g(53)$ or $\Psi(1234)g^4(1253)$. The same argument applied to the second term on the lhs of (4.7) leads us to the equation

$$\Psi(1234)g^4(1253) = \Psi(1253)g^4(1234) \tag{4.8}$$

By assumption, $r(g) \geq 4$ so that $g^4 \neq 0$. Therefore,

$$\Psi(1234)/g^4(1234) = \Psi(1253)/g^4(1253). \tag{4.9}$$

Each side of (4.9) is symmetric in its variables, but the rhs does not depend on particle 4, and therefore neither does the lhs, which must therefore be a constant, so that Ψ is proportional to g^4.

For arbitrary even N, the preceding argument needs only minor modifications. We define χ by

$$\chi = (N+1)A_{N+1}\Psi(1\ 2 \cdots N)g(N+1, N+2),$$

where A_{N+1} acts on particles $1, 2, \ldots, N+1$. We observe that (3.9) implies $\chi + (N, N+2)\chi = 0$; in this equation we identify particles $N-1$ and $N+2$, and multiply the resulting equation by $g(12)g(34)\cdots g(N-3, N-2)$. When we symmetrize on $1, 2, \ldots, N-1$, the terms cancel in pairs [like the rhs of (4.7)] except for two terms which give the equation $\Psi(1 \ldots N)g^N(1 \ldots N-1, N+1) = \Psi(1 \ldots N-1, N+1)g^N(1 \ldots N)$. If $r(g) < N$, then $g^N = 0$, and this equation would put no restriction on Ψ. However, under the hypothesis of our theorem $r(g) \geq N$, and so, as before, we conclude that Ψ is proportional to g^N.

In order to extend Theorem 4.1, we need the following result, of which, though it is fairly widely known, there does not seem to be a simple published proof.

Theorem 4.2: If φ is a wavefunction of $N-1$ fermions such that $D^p(\varphi)$ is extreme in $\mathcal{P}^p_{N-1}$, $p < N$, and the orbital f is strongly orthogonal to φ, then $D^p(f \wedge \varphi)$ is extreme in $\mathcal{P}^p_N$.

Proof: Suppose, on the contrary, that

$$D^p(f \wedge \varphi) = \sum w^i D^p(\Psi_i), \tag{4.10}$$

where $w^i > 0$, $\sum w^i = 1$, and Ψ_i are N-fermion wavefunctions. By Corollary 4.3A of SFDM-I, $D^1(f \wedge \varphi)$ has f as a norb of eigenvalue N^{-1}, which is the maximum possible. This will occur only if f is a norb with eigenvalue N^{-1} of each of the Ψ_i. Hence, for each i, $\Psi_i = f \wedge \varphi_i$, where φ_i is an $N-1$ fermion function strongly orthogonal to f.

From the definition of the p-matrix it is easily seen that $D^p(f \wedge \varphi)$ is the sum of a term proportional to $D^p(\varphi)$ and a term, obtained by antisymmetrizing $D^1(f)D^{p-1}(\varphi)$, which is expandable in terms of p-functions, all of which have f as a Grassmann factor and are therefore orthogonal to $R^p(\varphi)$. Applying this remark to both sides of (4.10), we conclude that

$$D^p(\varphi) = \sum w^i D^p(\varphi_i).$$

Since $D^p(\varphi)$ was assumed to be extreme in $\mathcal{P}^p_{N-1}$ it follows that, for all i, $D^p(\varphi_i) = D^p(\varphi)$ and therefore $D^{p-1}(\varphi_i) = D^{p-1}(\varphi)$. But f, $D^p(\varphi_i)$, and $D^{p-1}(\varphi_i)$ completely determine $D^p(\Psi_i)$, which is therefore equal to $D^p(f \wedge \varphi)$ for all i. Hence $D^p(f \wedge \varphi)$ is extreme in $\mathcal{P}^p_N$, as was to be proved.

The set $\mathcal{P}^1_1$ of 1-representable 1-operators is identical to the set of positive 1-operators. By the spectral theorem the extreme points of $\mathcal{P}^1_1$ consist of all projectors onto orbitals. By Theorem 4.2 it follows that if f_1 and f_2 are orthogonal orbitals, $D^1(f_1 \wedge f_2)$ is extreme in $\mathcal{P}^1_2$. Proceeding by induction, we have a new proof of the following well-known fundamental theorem.

Theorem 4.3: If f_i, $1 \leq i \leq N$, are orthonormal orbitals then the Slater determinant $f_1 \wedge f_2 \wedge \cdots \wedge f_N$ covers an extreme point of $\mathcal{P}^1_N$.

It is also a direct consequence of Theorems 4.1 and 4.2 that for N even and $r(g) \geq N$, if f is an orbital which is strongly orthogonal to g, then $D^2(f \wedge g^N)$ is extreme in $\mathcal{P}^2_{N+1}$. Proceeding by induction, we thus obtain the following generalization of Theorem 4.1.

Theorem 4.4: If S is a Slater determinant in σ particles which is strongly orthogonal to the geminal g of rank greater than or equal to 2μ, then $D^2(S \wedge g^{2\mu})$ is extreme in $\mathcal{P}^2_N$ for $N = \sigma + 2\mu$.

The wave-function occurring in Theorem 4.4 can be designated as a GAGP—a generalized geminal power.

5. THE p-RANK OF Ψ

Since the eigenvalues of an N-representable D^1 are not greater than N^{-1} and its trace is unity, there must be at least N natural orbitals. That is, the rank r of an N-representable 1-matrix satisfies the inequality

$$r \geq N. \tag{5.1}$$

However, for $p > 1$ we have little precise knowledge about the rank r_p of $D^p(\Psi)$. We proved in SFDM-I that $\lambda^p_i \leq (N-p+1)^{-1}$. In particular, if $N > 3$, $\lambda^2_i < (N-1)^{-1}$; thus

$$r_2 \geq N. \tag{5.2}$$

Further, since Ψ can be expressed in terms of r orbitals from which only $\binom{r}{2}$ linearly independent geminals can be formed, it follows that

$$r_2 \leq \tfrac{1}{2}r(r-1). \tag{5.3}$$

In general, we expect equality in (5.3), but it is easy to give examples of the inequality. For example, if the φ_i are orthogonal for $1 \leq i \leq 6$, $\Psi = [123] + [456]$ has $\frac{1}{2}r(r-1) = 15$ but $r_2 = 6$. Ruskai[9] has considered the case of infinite rank.

Whereas (5.3) gives an upper bound on r_2, the following theorem gives a lower bound which depends on N.

Theorem 5.1: The rank r_p of an N-representable p-matrix is not less than the binomial coefficient $\binom{N}{p}$.

Proof: Since the rank of a sum of positive operators is greater than or equal to the rank of any summand, it suffices to prove the theorem for a pure N-representable p-matrix.

Suppose $D^p = D^p(\Psi)$; then, by (3.15) of SFDM-I,

$$\Psi = \sum_i c^i \alpha_i^p \alpha_i^q, \tag{5.4}$$

where α_i^p are the natural p-states of Ψ and eigenfunctions of D^p and r_p is equal to the number of α_i^p for which $c^i \neq 0$. Recall that $N = p + q$ and that $\{\alpha_i^p\}$ and $\{\alpha_i^q\}$ are sets of orthonormal functions of p and q particles, respectively. On the other hand, we know from SFDM-I that Ψ can be expressed in terms of the natural orbitals φ_i, so that there exists an antisymmetric tensor $a_{i_1 i_2 \cdots i_N}$ of complex numbers such that

$$\Psi(1\ 2 \cdots N) = \sum a_{i_1 i_2 \cdots i_N} \varphi_{i_1}(1)\varphi_{i_2}(2) \cdots \varphi_{i_N}(N). \tag{5.5}$$

If K is a set of q distinct natural numbers, then we define a function f_K of p particles as follows. Let $\{i_{p+1}, i_{p+2}, \ldots, i_N\} = K$ with $i_{p+1} < i_{p+2} < \cdots < i_N$ then

$$f_K(12 \cdots p) = \sum_{i_1 \cdots i_p} a_{i_1 i_2 \cdots i_N} \varphi_{i_1}(1) \cdots \varphi_{i_p}(p). \tag{5.6}$$

Let δ be the dimension of the complex linear space spanned by the f_K of which there are $\binom{r}{p}$, where r is the 1-rank of Ψ. We shall prove that $\delta \geq \binom{N}{p}$ and that $\delta = r_p$.

If $\Psi \neq 0$, at least one of the coefficients in (5.5) is different from zero. Suppose $a_{1\,2 \cdots N} \neq 0$; let L be a subset of p elements of $\{1, 2, \ldots, N\} = I_N$. Denote by Φ_L the Slater determinant:

$$\Phi_L(1\ 2 \cdots p) = |\varphi_{i_1}(1)\varphi_{i_2}(2) \ldots \varphi_{i_p}(p)|,$$

where $i_1 < i_2 < \cdots < i_p$ and $L = \{i_1, i_2, \ldots, i_p\}$. If K and L are complementary sets in I_N, then Φ_L occurs in the expansion (5.6) with coefficient $a_{1\,2 \cdots N}$ or $-a_{1\,2 \cdots N}$.

For $L \subset I_N$ and $K \subset I_N$, Φ_L cannot occur in the expansion (5.6) if $K \cap L \neq \emptyset$, for in this case the antisymmetry of the coefficients forces them to be zero. The family of all Φ_L constitute an orthonormal set in terms of which all f_K can be expanded. The subfamily F of all f_K for which $K \subset I_N$ contains $\binom{N}{p}$ functions. For fixed $K \subset I_N$, and with $L = I_N - K$, Φ_L occurs in the expansion of f_K, but it does *not* occur in the expansion of any other function in F. It follows that F is a linearly independent set. Thus $\delta \geq \binom{N}{p}$.

We now show that $r_p = \delta$. Multiplying (5.4) and (5.5) by α_j^{q*} and integrating, we find that α_j^p is a linear combination of f_K. Thus $r_p \leq \delta$. Again, multiplying (5.4) and (5.5) by Φ_K^* $(p+1 \ldots N)$ and integrating, we find f_K is a linear combination of the α_i^p. Thus $r_p \geq \delta$. It follows that $r_p = \delta \geq \binom{N}{p}$. QED

6. TWO IMPLICATIONS OF STRONG ORTHOGONALITY

Recall that two antisymmetric (or symmetric) functions are said to be strongly orthogonal when their respective 1-ranges are orthogonal. To postulate strong orthogonality frequently makes a good approximation in chemical systems. For example, two hydroxyl radicals at opposite ends of a huge molecule are almost exactly strongly orthogonal. To approximate a wavefunction by means of an AGP function is certainly as good or better than Hartree–Fock. However, Kutzelnigg[10] has provided numerical evidence for his contention that the wavefunction of a four-electron atom can generally be better approximated by an antisymmetrized product of strongly orthogonal geminals (APSG) than by an AGP function.

The following two results concerning strongly orthogonal nags were announced at Sanibel in 1965 but have not been published hitherto. Suppose that $p = 2$ in (5.4); then, since we refer to α_i^2 as a natural geminal, or nag, we may conveniently, analogously to factor and cofactor, refer to α_i^{N-2} as the *conag* of α_i^2.

Theorem 6.1: If g_i, a natural geminal of the N-fermion function Ψ, is strongly orthogonal to its conag G_i, then λ_i^2 the corresponding eigenvalue of $D^2(\Psi)$ satisfies the inequality

$$\lambda_i^2 \leq 2/N(N-1). \tag{6.1}$$

Equality obtains in (6.1) if and only if Ψ is proportional to $g_i \wedge G_i$.

Proof: The proof follows familiar lines using Sasaki's formula in (6.3) of SFDM-I. Since $\Psi = \sum C_j g_j G_j$ and $\lambda_i^2 = |C_i|^2$ we have

$$\begin{aligned} \lambda_i^2 &= |\langle g_i G_i | \Psi \rangle|^2 \\ &= |\langle A_N g_i G_i | \Psi \rangle|^2 \\ &\leq |\langle A_N g_i G_i | A_N g_i G_i \rangle|, \end{aligned} \tag{6.2}$$

$$\therefore \quad \lambda_i^2 \leq |\langle g_i G_i | A_n g_i G_i \rangle| \tag{6.3}$$

$$\leq [2/N(N-1)][1 - 2N\langle g_i G_i | (13) g_i G_i \rangle + \binom{N}{2} \langle g_i G_i | (13)(24) g_i G_i \rangle]. \tag{6.4}$$

The inequality (6.2) results immediately from the Schwarz inequality since, as always, we assume that Ψ is normalized. The fact that A_N is idempotent and Hermitian implies (6.3) and (6.4) follows from Sasaki's formula. Recall that, in (6.4), (13) and (24) indicated permutations of the particles and, since g_i and G_i are assumed to be strongly orthogonal, the last two terms in the square brackets in (6.4) vanish. Thus

$$\lambda_i^2 \leq 2/N(N-1).$$

Equality obtains in (6.2) if and only if Ψ is proportional to $A_N g_i G_i$, that is, to $g_i \wedge G_i$, as required.

We know that any eigenvalue of D^2 is less than $(N-1)^{-1}$. Intuition led some physicists to believe that it is necessarily less than $\binom{N}{2}^{-1}$ as in the circumstances of Theorem 6.1. There is some numerical evidence that in real systems with few electrons

the value $(\frac{N}{2})^{-1}$ which occurs for the Hartree–Fock model is seldom greatly exceeded. However the possibility that λ_i^2 can be of order N^{-1} rather than just N^{-2} is what allows a system to go superconducting. Thus the value of these bounds is of crucial physical significance.

We know that $(N-1)^{-1}$ is the best possible upper bound for the eigenvalues of an N-representable D^2 if N is even. The following theorem asserts the surprising fact that the *sum* of any number of such eigenvalues has the same upper bound if the eigenvalues correspond to a set of mutually strongly orthogonal nags.

Theorem 6.2: If k is any natural number, if $\{g_i\}$, $1 \le i \le k$, is a set of k mutually strongly orthogonal natural geminals of an N fermion function Ψ, and if λ_i^2 is the eigenvalue of $D^2(\Psi)$ corresponding to g_i, then

$$\sum_{i=1}^{k} \lambda_i^2 \le (N-1)^{-1}. \tag{6.5}$$

Proof: If $\Psi = \sum C_i g_i G_i$, let $\chi = \sum_1^k C_i g_i G_i$. Since the g_i occurring in χ are mutually strongly orthogonal and $|C_i|^2 = \lambda_i^2$,

$$\langle \chi | (13) \chi \rangle = \sum_i^k \lambda_i^2 \operatorname{tr}[D^1(g_i) D^1(G_i)] \ge 0. \tag{6.6}$$

Let $\lambda = \sum_1^k \lambda_i^2$; then

$$\begin{aligned} \lambda^2 &= |\langle \chi | \Psi \rangle|^2 \\ &= |\langle A_N \chi | \Psi|^2 \\ &\le |\langle A_N \chi | A_N \chi \rangle|. \end{aligned}$$

But

$$\begin{aligned} \langle A_N \chi | A_N \chi \rangle &= \langle A_N \chi | A_N A_{N-1} \chi \rangle \\ &\le \langle \chi | A_{N-1} \chi \rangle. \end{aligned} \tag{6.7}$$

Thus

$$\begin{aligned} \lambda^2 &\le \langle \chi | A_{N-1} \chi \rangle \\ &\le [1/(N-1)][\langle \chi | \chi \rangle - \langle \chi | (13) \chi \rangle] \\ &\le [1/(N-1)]\lambda. \end{aligned} \tag{6.8}$$

Since $\lambda > 0$, it follows that $\lambda \le (N-1)^{-1}$, as we wished to prove.

In the preceding argument we again used the Schwarz inequality. The step (6.7) follows from the fact that A_N is a projector and projection does not increase the length of a vector. Equality holds in (6.7) if and only if $A_N \chi = A_{N-1} \chi$. By (6.6), if equality obtains in (6.8), it occurs in (6.5) only if g_i and G_i are strongly orthogonal for all i, $1 \le i \le k$. But then, by Theorem 6.1, $\lambda_i^2 \le (\frac{N}{2})^{-1}$ so that equality can obtain in (6.5) only if $2k \ge N$.

Equality does obtain in (6.5) with $2k = N$ if and only if Ψ is proportional to $g_1 \wedge g_2 \wedge \cdots \wedge g_k$, where the g_i are mutually strongly orthogonal. This follows from Theorem 6.1 from which we can conclude that in the assumed circumstances each of the k geminals g_i is a Grassmann factor of Ψ. For such Ψ, each of the g_i is a nag of Ψ. (Contrast this with the fact that g is a nag of g^N only if g is of extreme type.) The conag of g_1 is $g_2 \wedge g_3 \wedge \cdots \wedge g_k$ and so on. Further $\lambda_1^2 = \lambda_2^2 = \cdots = \lambda_k^2 = (\frac{N}{2})^{-1}$. A single Slater determinant provides the simplest instance of this general result if we choose as the g_i the Slater geminals [12], [34], ..., [$N-1$ N].

7. DISCUSSION

The main result above, Theorem 4.1, assures us of the existence of a new class of extreme points of $\mathcal{O}_N^2$. Though this class is of high dimension, it does not exhaust the set of extreme points. The small subset determined by those g which have rank N is exactly the set of D^2 covered by single Slater determinants. The convex closure of this set is the so-called Slater hull. The essential limitation of the Hartree–Fock approximation, which has dominated the applications of quantum mechanics to physics and chemistry for the past forty years, is that it assumes D^2 lies in the Slater hull.

The proponents of the Green's function technic believe that they can go beyond the HF approximation. As was pointed out by Pruski and the present author,[11] they are able to hold this belief only by gratuitously—and usually unconsciously—assuming that the N-representability problem can be neglected when the hierarchy of equations determining the GF's is truncated. Thus the GF technic as currently practiced rests on faith rather than knowledge! The single exception to this stricture known to the author is the Gorkov ansatz[11] in superconductivity theory. But this *ansatz* is equivalent to the basic assumption of the BCS theory that the wavefunction of a superconductor is an AGP.

The wavefunction appearing in Theorem 4.4 is an AGP supplemented by a Slater determinant and can be briefly described as a GAGP—a generalized antisymmetrized geminal power. In view of the success of the AGP function in superconductivity and nuclear theory, it now appears that it may be well worthwhile to devote as much energy as has been expended on the independent particle model, systematically exploring the possibilities of the GAGP approximation, moving beyond the Slater hull into the convex closure of the extreme points of Theorem 4.4.

In the course of proving Theorem 4.1, it appeared that the exact ground state of the two-particle operator $B(g)$ is g^N.

Perhaps the essential point of this theorem is the assertion that the ground state of the N-particle Hamiltonian formed from $B(g)$ is nondegenerate. When g is of extreme type, $D^1(g)$ is the identity operator on $R^1(g)$. By a special choice of norbs, $B(g)$ then reduces to the Hamiltonian of the simple BCS theory. This occurs if, in the familiar physics notation, we choose $(k, \uparrow)$ and $(-k, \downarrow)$, for a small interval near the Fermi momentum, as norbs. Apart from an additive constant, and a multiplicative constant which measures the strength of the interaction between two electrons, $B(g)$ then reduces to P_g which is the BCS Hamiltonian. This suggests that $B(g)$ might advantageously be studied as a model Hamiltonian for other than superconducting systems. Indeed, it should be possible to solve Eq. (3.3) of[12] SFDM-II, which give the 1-particle occupation numbers $n_\sigma = N\lambda_\sigma^1$, for the eigenvalues λ_σ of $D^1(g)$. If so, then *any* 1-matrix for which the occupation numbers less than unity are evenly degenerate can be covered exactly by a GAGP function. Since for the HF approxi-

mation n_σ is either 1 or 0, it follows that GAGP is extraordinarily more versatile and that $B(g)$ could be used as a model Hamiltonian for a very wide variety of fermion systems.

Regarded as an element of $\bar{\mathcal{P}}^2_N$, $B(g)$ determines a hyperplane which touches $\mathcal{P}^2_N$ at one point and on one side of which $\mathcal{P}^2_N$ must lie. The equation of the envelope of all these hyperplanes can be obtained explicitly, as the author plans to show in a future paper. This envelope bounds a convex cone containing $\mathcal{P}^2_N$. Similarly, the family of operators $F(g)$ and the family of P_g determine two other cones. Thus $\mathcal{P}^2_N$ is interior to the convex set C bounded by these three cones. It is an important unsolved problem to determine the difference set $C - \mathcal{P}^2_N$.

The theorems of Sec. 6 illustrate vividly the thesis of Sec. 1 that N-representability—certainly *pure* N-representability—forces complex interconnections among the eigenvalues and eigenfunctions of $D^2(\Psi)$. Unfortunately Theorem 5.1 rather discourages our hope that 2-matrix technics will open major simplifications for the numerical treatment of electron–electron correlation, implying as it does that the description of D^2 necessarily requires at least $\binom{N}{2}$ geminals. In the current state of the numeric arts, the description of even one geminal is an act which cannot be enterprized lightly or wantonly!

ACKNOWLEDGMENTS

I am grateful to Dr. R. M. Erdahl for frequent helpful discussions bearing on the contents of this paper, to the Mathematical Institute of Oxford University for hospitality during the preparation of the final draft, and to the National Research Council of Canada for financial support.

[1] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963). We refer to this paper as SFDM-I ("The Structure of Fermi Density Matrices I").
[2] R. U. Ayres, Phys. Rev. **111**, 1453 (1958); C. A. Coulson, Rev. Mod. Phys. **32**, 175 (1960); R. H. Tredgold, Phys. Rev. **105**, 1421 (1957).
[3] Notes prepared by J. Veverka of lectures delivered by the author at Sanibel Island, Fla., in January 1965 on "Density Matrix Methods in Quantum Physics."
[4] Queen's Papers on Pure and Applied Mathematics, No. 11 (1968), edited by A. J. Coleman and R. M. Erdahl. This report also contains a valuable survey by Darwin Smith.
[5] H. Kummer, J. Math. Phys. **8**, 2063 (1967).
[6] C. Garrod and J. K. Percus, J. Math. Phys. **5**, 1765, (1964); Eq. (5.8) defines the Q matrix. In second quantization notation it has components $\langle\Psi|a_i a_j a_l^\dagger a_k^\dagger\Psi\rangle$ whereas the 2-matrix has components $\langle\Psi|a_i^\dagger a_j^\dagger a_l a_k\Psi\rangle$.
[7] A. J. Coleman, Technical Report No. 80 (Quantum Chemistry Group, Uppsala University, 1962).
[8] R. McWeeny, Rev. Mod. Phys. **32**, 335 (1960).
[9] B. Ruskai, Phys. Rev. **183**, 23 (1969).
[10] W. Kutzelnigg, Th. Chim. Acta **3**, 241 (1965).
[11] A. J. Coleman and S. Pruski, Can. J. Phys. **43**, 2142 (1965).
[12] A. J. Coleman, J. Math. Phys. **6**, 1425 (1965), SFDM-II.

A Statistical Mechanical Approach to the Problem of a Fluid in an External Field

Kenneth Millard*
Department of Physics, Case Western Reserve University, Cleveland, Ohio
(Received 4 May 1971)

A rigorous, equilibrium statistical mechanical treatment of a fluid in a weak external field is given. The technique involves a cell division which leads to upper and lower bounds for the free energy density. Under a suitable double limiting procedure these limits coalesce, yielding a free energy consisting of a field-free term plus a field-dependent term. The cell division allows a direct physical definition of the local pressure $p(\mathbf{s})$ and the local density $\rho(\mathbf{s})$. This treatment provides a rigorous derivation of the thermodynamics of a fluid in a weak external field and, in particular, the hydrostatic equation $\text{grad}p = -\rho\ \text{grad}\phi$.

I. INTRODUCTION

The problem of a fluid in an external field is often treated in thermodynamics.[1] In such an approach, a cell division of the system is made such that the external field is nearly constant over any given cell, but may vary over the entire system. Each cell is taken to be macroscopic in size, with well-defined thermodynamic parameters. The thermodynamic approach to this problem yields the hydrostatic equation[1] or, equivalently, the variation of the local chemical potential with external field. The hydrostatic equation is used in a variety of applications ranging from a derivation of the barometer formula[1] for an ideal gas to the variation of pressure with height for a fluid in a gravitational field, as used in the experimental determination of critical exponents.[2]

Since the problem of a fluid in an external field is fundamental, it is of interest to obtain a rigorous statistical mechanical approach. This is the objective of the present paper. The treatment here parallels the thermodynamic approach in that a cell division is made and the external field is defined such that in the limit, as the number of cells becomes infinite, the external field is constant over any cell. This is accomplished by a double limiting process.[3] The first limit is the thermodynamic limit of the entire system, and entails simply increasing the number of cells. This limit is followed by the thermodynamic limit of each cell. The second limit insures that each cell is of macroscopic size, with well-defined thermodynamic parameters.

Our approach is to obtain upper and lower bounds on the Helmholtz free energy density for the system in an external field. The bounds consist of the Helmholtz free energy density for each cell in the absence of an external field plus terms involving the interaction of each cell with the external field. Following the above-mentioned double limiting procedure, the upper and lower bounds coalesce determining the Helmholtz free energy density for the entire system. The pressure, density, and chemical potential of each cell are taken to be the local parameters.

mation n_σ is either 1 or 0, it follows that GAGP is extraordinarily more versatile and that $B(g)$ could be used as a model Hamiltonian for a very wide variety of fermion systems.

Regarded as an element of $\bar{\mathcal{P}}^2_N$, $B(g)$ determines a hyperplane which touches $\mathcal{P}^2_N$ at one point and on one side of which $\mathcal{P}^2_N$ must lie. The equation of the envelope of all these hyperplanes can be obtained explicitly, as the author plans to show in a future paper. This envelope bounds a convex cone containing $\mathcal{P}^2_N$. Similarly, the family of operators $F(g)$ and the family of P_g determine two other cones. Thus $\mathcal{P}^2_N$ is interior to the convex set C bounded by these three cones. It is an important unsolved problem to determine the difference set $C - \mathcal{P}^2_N$.

The theorems of Sec. 6 illustrate vividly the thesis of Sec. 1 that N-representability—certainly *pure* N-representability—forces complex interconnections among the eigenvalues and eigenfunctions of $D^2(\Psi)$. Unfortunately Theorem 5.1 rather discourages our hope that 2-matrix technics will open major simplifications for the numerical treatment of electron–electron correlation, implying as it does that the description of D^2 necessarily requires at least $\binom{N}{2}$ geminals. In the current state of the numeric arts, the description of even one geminal is an act which cannot be enterprized lightly or wantonly!

ACKNOWLEDGMENTS

I am grateful to Dr. R. M. Erdahl for frequent helpful discussions bearing on the contents of this paper, to the Mathematical Institute of Oxford University for hospitality during the preparation of the final draft, and to the National Research Council of Canada for financial support.

[1] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963). We refer to this paper as SFDM-I ("The Structure of Fermi Density Matrices I").
[2] R. U. Ayres, Phys. Rev. **111**, 1453 (1958); C. A. Coulson, Rev. Mod. Phys. **32**, 175 (1960); R. H. Tredgold, Phys. Rev. **105**, 1421 (1957).
[3] Notes prepared by J. Veverka of lectures delivered by the author at Sanibel Island, Fla., in January 1965 on "Density Matrix Methods in Quantum Physics."
[4] Queen's Papers on Pure and Applied Mathematics, No. 11 (1968), edited by A. J. Coleman and R. M. Erdahl. This report also contains a valuable survey by Darwin Smith.
[5] H. Kummer, J. Math. Phys. **8**, 2063 (1967).
[6] C. Garrod and J. K. Percus, J. Math. Phys. **5**, 1765, (1964); Eq. (5.8) defines the Q matrix. In second quantization notation it has components $\langle\Psi|a_i a_j a_l^\dagger a_k^\dagger\Psi\rangle$ whereas the 2-matrix has components $\langle\Psi|a_i^\dagger a_j^\dagger a_l a_k\Psi\rangle$.
[7] A. J. Coleman, Technical Report No. 80 (Quantum Chemistry Group, Uppsala University, 1962).
[8] R. McWeeny, Rev. Mod. Phys. **32**, 335 (1960).
[9] B. Ruskai, Phys. Rev. **183**, 23 (1969).
[10] W. Kutzelnigg, Th. Chim. Acta **3**, 241 (1965).
[11] A. J. Coleman and S. Pruski, Can. J. Phys. **43**, 2142 (1965).
[12] A. J. Coleman, J. Math. Phys. **6**, 1425 (1965), SFDM-II.

A Statistical Mechanical Approach to the Problem of a Fluid in an External Field

Kenneth Millard*
Department of Physics, Case Western Reserve University, Cleveland, Ohio
(Received 4 May 1971)

A rigorous, equilibrium statistical mechanical treatment of a fluid in a weak external field is given. The technique involves a cell division which leads to upper and lower bounds for the free energy density. Under a suitable double limiting procedure these limits coalesce, yielding a free energy consisting of a field-free term plus a field-dependent term. The cell division allows a direct physical definition of the local pressure $p(\mathbf{s})$ and the local density $\rho(\mathbf{s})$. This treatment provides a rigorous derivation of the thermodynamics of a fluid in a weak external field and, in particular, the hydrostatic equation $\text{grad}p = -\rho\ \text{grad}\phi$.

I. INTRODUCTION

The problem of a fluid in an external field is often treated in thermodynamics.[1] In such an approach, a cell division of the system is made such that the external field is nearly constant over any given cell, but may vary over the entire system. Each cell is taken to be macroscopic in size, with well-defined thermodynamic parameters. The thermodynamic approach to this problem yields the hydrostatic equation[1] or, equivalently, the variation of the local chemical potential with external field. The hydrostatic equation is used in a variety of applications ranging from a derivation of the barometer formula[1] for an ideal gas to the variation of pressure with height for a fluid in a gravitational field, as used in the experimental determination of critical exponents.[2]

Since the problem of a fluid in an external field is fundamental, it is of interest to obtain a rigorous statistical mechanical approach. This is the objective of the present paper. The treatment here parallels the thermodynamic approach in that a cell division is made and the external field is defined such that in the limit, as the number of cells becomes infinite, the external field is constant over any cell. This is accomplished by a double limiting process.[3] The first limit is the thermodynamic limit of the entire system, and entails simply increasing the number of cells. This limit is followed by the thermodynamic limit of each cell. The second limit insures that each cell is of macroscopic size, with well-defined thermodynamic parameters.

Our approach is to obtain upper and lower bounds on the Helmholtz free energy density for the system in an external field. The bounds consist of the Helmholtz free energy density for each cell in the absence of an external field plus terms involving the interaction of each cell with the external field. Following the above-mentioned double limiting procedure, the upper and lower bounds coalesce determining the Helmholtz free energy density for the entire system. The pressure, density, and chemical potential of each cell are taken to be the local parameters.

For simplicity, we consider a fluid contained in a cubic volume V, of edge length L. We take the fluid to be in the presence of an external field, $w(\mathbf{r}, V)$. The precise form of the external field is described below. We assume that the interaction between any two particles satisfies the conditions[4]

$$u(r) = +\infty \quad \text{for } r < r_0, \tag{1a}$$

$$|u(r)| < Dr^{-3-\epsilon} \quad \text{for } r \geq r_0. \tag{1b}$$

The pair interaction then consists of a hard core of radius r_0 plus an interaction which falls off at least as $r^{-3-\epsilon}$. D and ϵ are finite positive constants. The condition (1) is imposed so that the analysis can be made rigorous. In particular, the hard core places an upper bound on the number of particles that can occupy a specified volume. Condition (1b) allows the neglect, in the thermodynamic limit of the cells, of interactions between particles restricted to disjoint cells.

The configurational portion of the Hamiltonian for a system of N particles is assumed to be the sum of the pair interactions plus the interactions with the external field,

$$H = U + W, \tag{2}$$

where

$$U = \tfrac{1}{2}\sum_{i\neq j}^{N} u(r_{ij}), \tag{3a}$$

$$W = \sum_{i=1}^{N} w(\mathbf{r}_i, V). \tag{3b}$$

The canonical partition function for the fluid in an external field is defined by

$$Q(\beta, N, \Omega) = \frac{\Lambda^{-3N}}{N!}\int_\Omega d\mathbf{r}_1 \cdots \int_\Omega d\mathbf{r}_N e^{-\beta U} e^{-\beta W}, \tag{4}$$

where the temperature is given by $(\beta k)^{-1}$, Ω is the domain of integration with which V is associated, and $\Lambda = h/(2\pi mkT)^{1/2}$ is the thermal wavelength. The Helmholtz free energy density in the thermodynamic limit is defined by

$$f(\beta, \bar{\rho}) = -\beta^{-1}\lim_{V\to\infty} V^{-1}\ln Q(\beta, N, \Omega), \tag{5a}$$

where

$$\bar{\rho} = N/V. \tag{5b}$$

The thermodynamic limit, $V \to \infty$, is taken such that the density $\bar{\rho}$ is held fixed.

At this point it is convenient to introduce the domain ω, which we call the "intensive" domain associated with Ω. Each point $\mathbf{s}$ in ω is obtained from a point $\mathbf{r}$ in Ω by the transformation

$$\mathbf{s} = L^{-1}\mathbf{r}. \tag{6}$$

Since Ω is taken to be a cube of edge length L, this prescription for the "intensive" domain implies that ω is a cube of unit edge length. The "intensive" domain ω is then clearly independent of the volume V.

To assure the existence of $f(\beta, \bar{\rho})$, we place the following condition on $w(\mathbf{r}, V)$:

$$w(\mathbf{r}, V) = \phi(L^{-1}\mathbf{r}), \tag{7}$$

where it is assumed that $\phi(\mathbf{s})$, $\mathbf{s}$ contained in ω, is bounded and continuous. $\phi(\mathbf{s})$ can be thought of as the external field expressed as a function of the "intensive" position $\mathbf{s}$. In a very rough sense, $\int_\Omega w(\mathbf{r}, V)d\mathbf{r}$ is a measure of the external field interaction. Equation (7) and the assumed boundedness and continuity of $\phi(\mathbf{s})$ guarantee that

$$V^{-1}\int_\Omega w(\mathbf{r}, V)d\mathbf{r} = \int_\omega \phi(\mathbf{s})d\mathbf{s} \tag{8}$$

exists and is independent of V. This condition can be viewed as similar in nature to the condition[5] that the integral over all space of a Kac potential exists and is independent of the inverse range parameter.

The entire system is divided into M disjoint cells, each cell being a cube of edge length l and volume v. We then observe that

$$L = M^{1/3}l, \tag{9a}$$

$$V = Mv. \tag{9b}$$

In terms of the "intensive" domain ω, this cell division appears as a division of the unit cube into M cells, each cell of volume M^{-1}.

The above-mentioned double limiting procedure then corresponds to first taking the limit as $M \to \infty$, followed by the second limit as $v \to \infty$. Under the first limit we note that the condition (7) implies that the variation of the external field over any cell approaches zero as $M \to \infty$. This is easily seen by viewing the first limit in terms of the cell division of ω. The external field over any cell is given by $\phi(\mathbf{s})$. As $M \to \infty$, the volume of each cell in the "intensive" domain approaches zero. The external field over any cell then approaches a constant. Such behavior is consistent with the thermodynamic approach to the problem of a fluid in a weak external field.

The local thermodynamic parameters are most conveniently expressed as functions of the "intensive" position $\mathbf{s}$. The major results proved below are

$$f(\beta, \bar{\rho}) = \int_\omega [\phi\rho + f^0(\beta, \rho)]d\mathbf{s}, \tag{10a}$$

where

$$\bar{\rho} = \int_\omega \rho\, d\mathbf{s} \tag{10b}$$

and

$$\text{grad}p = -\rho\ \text{grad}\phi\,. \tag{10c}$$

The term $f^0(\beta, \rho)$ is the Helmholtz free energy density for the fluid in the absence of an external field. $p(\mathbf{s})$ and $\rho(\mathbf{s})$ are the local pressure and density, respectively, where

$$p = \rho\frac{\partial f^0(\beta, \rho)}{\partial\rho} - f^0(\beta, \rho). \tag{11}$$

Equation (10a) expresses the Helmholtz free energy as an average of a local function consisting of the local Helmholtz free energy density in the absence of an external field plus a term involving a product of the external field and the local density. Equation (10b) simply states that $\bar{\rho}$ is equal to the average of the local density. Equation (10c) is the hydrostatic equation expressed in terms of the local parameters. Under restricted conditions it is also shown that

$$\bar{p} = \int_\omega p d\mathbf{s}, \tag{12}$$

where $\bar{p}$ represents the bulk pressure defined by

$$\bar{p} = \bar{\rho}\frac{\partial f(\beta,\bar{\rho})}{\partial\bar{\rho}} - f(\beta,\bar{\rho}). \tag{13}$$

Equation (12) leads to the interpretation that $\bar{p}$ is the average of the local pressure.

II. PROOF OF THE MAJOR RESULTS

Let $\{n_\alpha\}$ be a particular configuration of cell occupations such that

$$\sum_{\alpha=1}^{M} n_\alpha = N. \tag{14}$$

Let Ω_α, $\alpha = 1, \ldots, M$, be the domain of cell α. The partition function can then be written[6] as

$$Q(\beta,N,\Omega) = \sum_{\{n_\alpha\}}{}'\left(\prod_{\gamma=1}^{M}\frac{\Lambda^{-3n_\gamma}}{n_\gamma!}\int_{\Omega_\gamma} d\mathbf{r}_1^{(\gamma)}\cdots\int_{\Omega_\gamma} d\mathbf{r}_{n_\gamma}^{(\gamma)}\right)e^{-\beta W}e^{-\beta U}, \tag{15}$$

where $\mathbf{r}_k^{(\gamma)}$ is the position of the kth particle in cell γ. The summation is over all cell occupations such that Eq. (14) is satisfied. We now establish upper and lower bounds on the partition function. Under the double limiting procedure these bounds coalesce, giving an exact expression for $f(\beta,\bar{\rho})$. This procedure provides a rigorous proof of the familiar "maximum term" method.

A. Upper Bound for $f(\beta,\bar{\rho},\Omega)$

We first establish an upper bound for $f(\beta,\bar{\rho},\Omega)$ or, equivalently, a lower bound for $Q(\beta,N,\Omega)$. Since each term in the summation in Eq. (15) is nonnegative, a lower bound for the partition function is obtained by choosing any single term, i.e., a set $\{n_\alpha\}$, such that Eq. (14) is satisfied:

$$Q(\beta,N,\Omega) \geq \left[\prod_{\gamma=1}^{M}\frac{\Lambda^{-3n_\gamma}}{n_\gamma!}\int_{\Omega_\gamma} d\mathbf{r}_1^{(\gamma)}\cdots\int_{\Omega_\gamma} d\mathbf{r}_{n_\gamma}^{(\gamma)}\right]e^{-\beta W}e^{-\beta U}. \tag{16}$$

Define

$$w_\gamma(V) = \max_{\mathbf{r}\in\Omega_\gamma} w(\mathbf{r},V), \tag{17}$$

where the maximum is over all $\mathbf{r}$ contained in cell γ. Then, clearly

$$W \leq \sum_{\alpha=1}^{M} n_\alpha w_\alpha \tag{18}$$

for each occupation configuration $\{n_\alpha\}$. The bound in Eq. (16) can then be weakened as follows:

$$Q(\beta,N,\Omega) \geq \left(\prod_{\gamma=1}^{M}\frac{\Lambda^{-3n_\gamma}}{n_\gamma!}e^{-\beta n_\gamma w_\gamma}\int_{\Omega_\gamma} d\mathbf{r}_1^{(\gamma)}\cdots\int_{\Omega_\gamma} d\mathbf{r}_{n_\gamma}^{(\gamma)}\right)e^{-\beta U}. \tag{19}$$

We now write each integral over a cell as the sum of an integral over the cell free volume[7] and an integral over a corridor. The domain Ω_α' of the cell free volume is defined such that $\Omega_\alpha' \in \Omega_\alpha$ and the boundary of Ω_α' is everywhere a distance $t/2$ from the boundary of Ω_α. For the case where Ω_α is a cube of edge length l, Ω_α' is a concentric cube of edge length $l' = l - t$. Using this definition of free volume and the fact that the integrand in Eq. (19) is nonnegative, we obtain

$$Q(\beta,N,\Omega) \geq \left[\prod_{\gamma=1}^{M}\frac{\Lambda^{-3n_\gamma}}{n_\gamma!}e^{-\beta n_\gamma w_\gamma}\int_{\Omega_\gamma'} d\mathbf{r}_1^{(\gamma)}\cdots\int_{\Omega_\gamma'} d\mathbf{r}_{n_\gamma}^{(\gamma)}\right]e^{-\beta U}. \tag{20}$$

Since the integrals in Eq. (20) are over the free volumes of the cells, we are assured that

$$|\mathbf{r}_k^{(\gamma)} - \mathbf{r}_j^{(\delta)}| \geq t \quad \text{for } \gamma \neq \delta. \tag{21}$$

U can now be written as

$$U = \sum_{\alpha=1}^{M} U_\alpha + U'. \tag{22}$$

U_α contains all intracell interactions for cell α, and U' is all intercell interactions. The inequality (20) can then be weakened in the following way:

$$\begin{aligned} Q(\beta,N,\Omega) &\geq \exp(-\beta U'_{\max})\prod_{\gamma=1}^{M}\left(\frac{\Lambda^{-3n_\gamma}}{n_\gamma!}e^{-\beta n_\gamma w_\gamma}\right. \\ &\quad\left.\times\int_{\Omega_\gamma'} d\mathbf{r}_1^{(\gamma)}\cdots\int_{\Omega_\gamma'} d\mathbf{r}_{n_\gamma}^{(\gamma)}e^{-\beta U_\gamma}\right) \\ &= \exp(-\beta U'_{\max})\prod_{\gamma=1}^{M}[e^{-\beta n_\gamma w_\gamma}Q^0(\beta,n_\gamma,\Omega_\gamma')]. \end{aligned} \tag{23}$$

$U'_{\max}$ represents an upper bound on U' and Q^0 is the canonical partition function for a fluid in the absence of an external field. Taking the logarithm of both sides of the inequality (23) and dividing by βV, we obtain

$$f(\beta,\bar{\rho},\Omega) \leq M^{-1}\sum_{\alpha=1}^{M}[w_\alpha\rho_\alpha + f^0(\beta,\rho_\alpha,\Omega_\alpha')] + V^{-1}U'_{\max}, \tag{24a}$$

where $f^0(\beta,\rho,\Omega)$ is the Helmholtz free energy density for a fluid in the absence of an external field and where

$$\rho_\alpha = n_\alpha/v, \qquad \bar{\rho} = N/V. \tag{24b}$$

The condition (14) requires

$$M^{-1}\sum_{\alpha=1}^{M}\rho_\alpha = \bar{\rho}. \tag{25}$$

Since $f^0(\beta,\rho_\alpha,\Omega_\alpha')$ converges uniformly in the thermodynamic limit,[7] we are guaranteed that there exists an $\epsilon(v)$, independent of ρ_α, such that

$$|f^0(\beta,\rho_\alpha,\Omega_\alpha') - f^0(\beta,\rho_\alpha)| < \epsilon(v), \tag{26a}$$

where

$$\lim_{v\to\infty}\epsilon(v) = 0. \tag{26b}$$

Inequality (24a) can then be written as

$$f(\beta,\bar{\rho},\Omega) \leq M^{-1}\sum_{\alpha=1}^{M}[w_\alpha\rho_\alpha + f^0(\beta,\rho_\alpha)] + \epsilon(v) + V^{-1}U'_{\max}. \tag{27}$$

An upper bound on U' for an interaction satisfying Eq. (1) is given by[8]

$$U' \leq AN n_{\max} t^{-3-\epsilon} \tag{28}$$

for $t > 2r_0$, where A is a finite positive constant and $n_{\max}$ is the maximum occupation of a cell free volume. Since we have assumed a hard core of radius r_0, it follows[9] that

$$n_{\max} < (l' + 2r_0)^3\rho_c, \qquad l' \geq 2r_0, \tag{29}$$

where ρ_c is the close-packing density. Combining Eqs. (28) and (29), we can take $U'_{\max}$ to be

$$U'_{\max} = AN\rho_c(l' + 2r_0)^3 t^{-3-\epsilon}. \tag{30}$$

B. Lower Bound for $f(\beta,\bar{\rho},\Omega)$

We now determine a lower bound for $f(\beta,\bar{\rho},\Omega)$. To accomplish this, we bound $Q(\beta,N,\Omega)$ above by choosing

the maximum term in the summation appearing in Eq. (15). Since each term in this summation is positive and since there are at most

$$g(N, M) = [(N + M - 1)!]/[N!(M - 1)!] \tag{31}$$

terms in the sum, we find

$$Q(\beta, N, \Omega) \leq g(N, M) \max_{\{n_\alpha\}} \times \left[\left(\prod_{\gamma=1}^{M} \frac{\Lambda^{-3n_\gamma}}{n_\gamma!} \int_{\Omega_\gamma} d\mathbf{r}_{n_\gamma}^{(\gamma)} \cdots \int_{\Omega_\gamma} d\mathbf{r}_{n_\gamma}^{(\gamma)}\right) e^{-\beta W} e^{-\beta U}\right]. \tag{32}$$

The maximum is taken with respect to all cell occupations $\{n_\alpha\}$, satisfying Eq. (14). Define

$$\tilde{w}_\gamma(V) = \max_{\mathbf{r}\in\Omega_\gamma} w(\mathbf{r}, V) - \min_{\mathbf{r}\in\Omega_\gamma} w(\mathbf{r}, V). \tag{33}$$

Then, using Eqs. (3b), (17) and (33), we observe that, for a given set $\{n_\alpha\}$,

$$W \geq \sum_{\alpha=1}^{M} n_\alpha [\min_{\mathbf{r}\in\Omega_\alpha} w(\mathbf{r}, V)] \geq \sum_{\alpha=1}^{M} n_\alpha w_\alpha - n_{\max} \tilde{W}, \tag{34a}$$

where

$$\tilde{W} = \sum_{\alpha=1}^{M} \tilde{w}_\alpha. \tag{34b}$$

The bound given by Eq. (32) can then be weakened as

$$Q(\beta, N, \Omega) \leq g(N, M) \exp(\beta n_{\max} \tilde{W}) \max_{\{n_\alpha\}} \times \left[\left(\prod_{\gamma=1}^{M} \frac{\Lambda^{-3n_\gamma}}{n_\gamma!} e^{-\beta n_\gamma w_\gamma} \int_{\Omega_\gamma} d\mathbf{r}_1^{(\gamma)} \cdots \int_{\Omega_\gamma} d\mathbf{r}_{n_\gamma}^{(\gamma)}\right) e^{-\beta U}\right]. \tag{35}$$

U can be written as the sum of two terms as in Eq. (22). Note, however, that no corridor has been constructed at this point. The inequality (35) can then be written as

$$Q(\beta, N, \Omega) \leq g(N, M) \exp(\beta n_{\max} \tilde{W}) \exp(-\beta U'_{\min}) \max_{\{n_\alpha\}} \times \left(\prod_{\gamma=1}^{M} [e^{-\beta n_\gamma w_\gamma} Q^0(\beta, n_\gamma, \Omega_\gamma)]\right). \tag{36}$$

$U'_{\min}$ represents a lower bound on the intercell interactions U'. In writing Eq. (36) we have anticipated the fact [see Eq. (39)] that the $U'_{\min}$ used in this calculation is independent of any particular occupation $\{n_\alpha\}$. Taking the logarithm of both sides of Eq. (36) and dividing by βV, we obtain

$$f(\beta, \bar{\rho}, \Omega) \geq - \max_{\{\rho_\alpha\}} \left(- M^{-1} \sum_{\gamma=1}^{M} [w_\gamma \rho_\gamma + f^0(\beta, \rho_\gamma, \Omega_\gamma)]\right) - (\beta V)^{-1} \ln g(N, M) - V^{-1} n_{\max} \tilde{W} + V^{-1} U'_{\min}, \tag{37}$$

subject to Eq. (25). In obtaining Eq. (37), the logarithm and max were commuted. This is valid since the logarithm is a monotonic function. Using the uniform convergence properties of f^0, i.e., Eq. (26), we find

$$f(\beta, \bar{\rho}, \Omega) \geq \min_{\{\rho_\alpha\}} \left(M^{-1} \sum_{\gamma=1}^{M} [w_\gamma \rho_\gamma + f^0(\beta, \rho_\gamma)]\right) - \epsilon(v) - (\beta V)^{-1} \ln g(N, M) - V^{-1} n_{\max} \tilde{W} + V^{-1} U'_{\min}. \tag{38}$$

A fictitious corridor for each cell is now constructed so as to coincide with the corridor constructed in obtaining the upper bound for $f(\beta, \bar{\rho}, \Omega)$. For such a construction and an interaction satisfying Eq. (1), an expression for $U'_{\min}$ can be written[10] as

$$U'_{\min} = - AN\rho_c (l' + 2r_0)^3 t^{-3-\epsilon} - \rho_c \Phi' M[(l' + t + 2r_0 M^{-1/3})^3 - (l' - 2r_0)^3]. \tag{39}$$

Φ' is a lower bound on the interaction of any particle with all its neighbors and A is the same constant appearing in Eq. (28). For interactions satisfying Eq. (1), Φ' is a finite constant.[11]

C. The Free Energy

We operate on both sides of Eqs. (27) and (38) with the double limiting procedure $\lim_{v\to\infty} \lim_{M\to\infty}$. The second limit, $v \to \infty$, can be constructed[7] such that

$$t/v^{1/3} \to 0, \qquad v/t^{3+\epsilon} \to 0 \quad \text{as } v \to \infty. \tag{40}$$

Using Eqs. (30), (39), and (40), we observe that

$$\lim_{v\to\infty} \lim_{M\to\infty} V^{-1} U'_{\max} = 0, \tag{41a}$$

$$\lim_{v\to\infty} \lim_{M\to\infty} V^{-1} U'_{\min} = 0. \tag{41b}$$

Further, by Eqs. (26b) and (31) we find

$$\lim_{v\to\infty} \lim_{M\to\infty} \epsilon(v) = 0, \tag{42a}$$

$$\lim_{v\to\infty} \lim_{M\to\infty} V^{-1} \ln g(N, M) = 0. \tag{42b}$$

Finally, using Eq. (29), we observe that

$$|V^{-1} n_{\max} \tilde{W}| \leq M^{-1} \rho_c \tilde{W}. \tag{43}$$

But by Eqs. (33) and (34), the limit as $M \to \infty$ of $M^{-1}\tilde{W}$ is just the difference of the upper and lower Riemann integrals of $\phi(\mathbf{s})$, which must vanish. Therefore,

$$\lim_{M\to\infty} V^{-1} n_{\max} \tilde{W} = 0. \tag{44}$$

We thus conclude that

$$f(\beta, \bar{\rho}) = \lim_{v\to\infty} \lim_{M\to\infty} \min_{\{\rho_\alpha\}} \left(M^{-1} \sum_{\gamma=1}^{M} [w_\gamma \rho_\gamma + f^0(\beta, \rho_\gamma)]\right), \tag{45}$$

subject to Eq. (25). In writing this expression we have used the fact that Eq. (27) was valid for any choice of $\{\rho_\alpha\}$ satisfying Eq. (25).

Since $f^0(\beta, \rho)$ is a convex downward function,[7] we are assured that an extremum of the bracket term in Eq. (45) is an absolute minimum. Extremizing the bracket term in Eq. (45), using a Lagrange multiplier to satisfy the constraint Eq. (25), we obtain the condition

$$w_\gamma + \frac{\partial f^0}{\partial \rho_\gamma} = \lambda, \tag{46}$$

where λ is the Lagrange multiplier. This set of equations determine $\{\tilde{\rho}_\gamma\}$ which minimize the bracket term in Eq. (45). Equation (45) can be written as

$$f(\beta, \bar{\rho}) = \lim_{v\to\infty} \lim_{M\to\infty} \left(M^{-1} \sum_{\gamma=1}^{M} [w_\gamma \tilde{\rho}_\gamma + f^0(\beta, \tilde{\rho}_\gamma)]\right), \tag{47}$$

where $\{\tilde{\rho}_\gamma\}$ are determined by Eqs. (25) and (46).

The second term on the left-hand side of Eq. (46) is just the chemical potential for a fluid of density ρ_γ in the absence of an external field, i.e., we define

$$\mu_\gamma = \frac{\partial f^0(\beta, \rho_\gamma)}{\partial \rho_\gamma}. \tag{48}$$

Equation (46) is then simply

$$w_\gamma + \mu_\gamma = \lambda. \tag{49}$$

If this equation is written for two adjacent cells, say cells γ and $\gamma + 1$, we obtain the equality

$$\mu_{\gamma+1} - \mu_\gamma = -(w_{\gamma+1} - w_\gamma). \tag{50}$$

We observe that, as $M \to \infty$, the right-hand side of Eq. (47) becomes a Riemann integral, i.e., in terms of the "intensive volume" the summation in Eq. (47) appears as a Riemann sum over divisions of volume M^{-1}. Equation (47) can then be written as (suppressing the tilde notation)

$$f(\beta, \bar{\rho}) = \int_\omega [\phi\rho + f^0(\beta, \rho)]\, d\mathbf{s}, \tag{51}$$

where $\rho(\mathbf{s})$ is determined by Eqs. (25) and (46). Equation (25) also goes over to a Riemann integral,

$$\bar{\rho} = \int_\omega \rho d\mathbf{s}. \tag{52}$$

D. Hydrostatic Equation and Local Pressure

If both sides of Eq. (50) are divided by $M^{-1/3}$, this expression goes, in the limit $M \to \infty$, to a differential relation between μ and ϕ,

$$\text{grad}\mu = -\text{grad}\phi. \tag{53}$$

In the limit as $M \to \infty$, Eqs. (46) and (49) go over to their appropriate continuum identities. Equations (46), (51), and (52) then determine the Helmholtz free energy density for a fluid in an external field exactly.

Equation (53) is equivalent to the hydrostatic equation,

$$\text{grad}p = -\rho\, \text{grad}\phi, \tag{54}$$

where p is the local pressure associated with a fluid in the absence of an external field with density ρ, i.e.,

$$p = \rho \frac{\partial f^0}{\partial \rho} - f^0 = \rho\mu - f^0. \tag{55}$$

Using the continuum form of Eq. (49) and integrating Eq. (55) over the "intensive volume", we obtain the expression

$$\int_\omega p d\mathbf{s} = \lambda \int_\omega \rho d\mathbf{s} - \int_\omega (\phi\rho + f^0) d\mathbf{s}. \tag{56}$$

Via Eqs. (51) and (52), this equation can be written as

$$\int_\omega p d\mathbf{s} = \lambda\bar{\rho} - f(\beta, \bar{\rho}). \tag{57}$$

From the form of this expression it is quite tempting to associate $\int_\omega p d\mathbf{s}$ with the pressure and λ with the chemical potential of the system in an external field. The association of λ, given by Eq. (49), with the chemical potential is the basis for the usual thermodynamic treatment of this problem.[1] The association of $\int_\omega p d\mathbf{s}$ with the pressure of the system with an external field is reminiscent of the definition of local pressure due to Stillinger and Buff.[12]

The above association can, in fact, be made precise as follows. The pressure $\bar{p}$ and the chemical potential $\bar{\mu}$ for a fluid in an external field can be defined by

$$\bar{p} = \bar{\rho} \frac{\partial f(\beta, \bar{\rho})}{\partial \bar{\rho}} - f(\beta, \bar{\rho}), \tag{58a}$$

$$\bar{\mu} = \frac{\partial f(\beta, \bar{\rho})}{\partial \bar{\rho}}. \tag{58b}$$

If $\phi(\mathbf{s})$ and $f^0(\beta, \rho)$ are such that a partial derivative with respect to $\bar{\rho}$ can be interchanged with the integral over the domain ω appearing in Eqs. (51) and (52), then Eqs. (58) can be used to determine $\bar{p}$ and $\bar{\mu}$ explicitly. Assuming the interchange of the derivative and integral is valid,[13] we obtain from Eqs. (51) and (52)

$$\bar{\mu} = \frac{\partial f}{\partial \bar{\rho}} = \int_\omega \frac{\partial}{\partial \bar{\rho}} [\phi\rho + f^0] d\mathbf{s} = \int_\omega \left[\phi + \frac{\partial f^0}{\partial \rho}\right] \frac{\partial \rho}{\partial \bar{\rho}}\, d\mathbf{s}$$

$$= \lambda \int_\omega \frac{\partial \rho}{\partial \bar{\rho}} d\mathbf{s} = \lambda. \tag{59}$$

Equation (57) then implies that

$$\bar{p} = \int_\omega p d\mathbf{s}. \tag{60}$$

Under this assumption of interchanging the order of differentiation and integration, we thus observe that the association of $\bar{p}$ with $\int_\omega p d\mathbf{s}$ and $\bar{\mu}$ with λ is exact.

ACKNOWLEDGMENTS

The author wishes to thank Professor Harvey S. Leff for helpful discussions and suggestions relating to this work and Dr. Joseph Czika for initial discussions on the subject.

---

* Present address, Department of Physics, University of Missouri-Rolla, Rolla, Missouri 65401.

1 See, for example: E. A. Guggenheim, *Thermodynamics* (North-Holland, Amsterdam, 1967), 5th ed., Chap. 9; D. ter Haar and H. N. S. Wergeland, *Elements of Thermodynamics* (Addison-Wesley, Reading, Mass., 1966), Chap. 9.

2 See, for example, Sec. 2 of P. Heller, Rept. Progr. Phys. 30 (Part II), 731 (1967).

3 We note that the order of these limits can, in fact, be interchanged. The bounding arguments leading to Eqs. (27) and (38) are carried out *before* either limit is taken. Equations (41)–(44) remain valid if the order of the limits are interchanged. Further, the bracket term in Eq. (45) is independent of v. We therefore conclude that $f(\beta, \bar{\rho}) = \lim_{v\to\infty} \lim_{M\to\infty} f(\beta, \bar{\rho}, \Omega) = \lim_{M\to\infty} \lim_{v\to\infty} f(\beta, \bar{\rho}, \Omega)$.

4 These conditions are the same as Eqs. (2. 2) and (2. 3) of O. Penrose, J. Math. Phys. 4, 1312 (1963).

5 See, for example, Eq. (1. 9) of J. L. Lebowitz and O. Penrose, J. Math. Phys. 7, 98 (1966).

6 This identity is analogous to Eq. (4. 2) of Ref. 5.

7 See, M. E. Fisher, Arch. Ratl. Mech. 17, 377 (1964).

8 See Eq. (2. 9) of Ref. 5.

9 See Eq. (4. 10) of Ref. 5.

10 This expression is obtained by combining Eqs. (4. 7)–(4. 10) and the stronger bound appearing in Eq. (4. 11) of Ref. 5.

11 See Ref. 4.

12 See discussion by J. Czika, Ph.D. thesis (Case Western Reserve University, 1971).

13 For conditions relating to this interchange, see T. M. Apostol, *Mathematical Analysis* (Addison-Wesley, Reading, Mass., 1957), pp. 219–20.

Lower Bounds of the Energy Eigenvalues of Systems Containing Identical Particles

Robert H. T. Yeh
Department of Physics and Astronomy, State University of New York, Buffalo, New York 14214
(Received 28 July 1971)

Lower bounds of the energy eigenvalues of systems containing identical particles were obtained by a generalization of the method of Calogero and Marchioro [J. Math. Phys. 10, 562 (1969)].

It is desired to find a lower bound for the ith energy eigenvalue E_N^i, $i = 1, 2, \ldots, E_N^i \geq E_N^j$ if $i > j$), of the following Hamiltonian of N identical particles,

$$H_N = \sum_{i=1}^{N} T_i + \sum_{i,j=1}^{N}{}' V_{ij} + \sum_{i=1}^{N} W_i, \tag{1}$$

where $\sum'$ excludes $i = j$ terms, $T_i(W_i)$ is the kinetic energy (external potential energy) of the ith particle, and V_{ij} $(= V_{ji})$ is one-half of the two-body potential energy between particle i and particle j.

We shall derive a lower bound of E_N^i by the Rayleigh-Ritz method.[1] Let us consider the following Hamiltonian of a comparison system,

$$H_n(h, g, f) = \sum_{i=1}^{n} h_i T_i + \sum_{i,j=1}^{n}{}' g_{ij} V_{ij} + \sum_{i=1}^{n} f_i W_i, \tag{2}$$

where $n \leq N$, and h, g, f's are arbitrary numbers satisfying the following constraints:

$$\sum_{i=1}^{n} h_i = N, \qquad \sum_{i=1}^{n}{}' g_{ij} = N(N-1), \qquad \sum_{i=1}^{n} f_i = N. \tag{3}$$

Let us now choose the orthonormal set of eigenfunctions of H_N, $\{\psi_N^1, \psi_N^2, \ldots, \psi_N^i\}$, as the trial functions of $H_n(h, g, f)$. Although ψ_N's are not eigenfunctions of $H_n(h, g, f)$, because they are completely symmetric or antisymmetric under the exchange of any two coordinates i, j, we have from Eq. (3)

$$(\psi_N^m, H_n(h, g, f)\psi_N^l) = E_N^m \delta_{lm}. \tag{4}$$

Thus by using the Rayleigh-Ritz method, we have

$$E_N^i \geq E_n^i(h, g, f). \tag{5}$$

This result is a direct generalization of Calogero and Marchioro's result[2] for the ground state energy. In particular, Theorem 1 of Ref. 2 can be generalized as follows:

$$E_N^i \geq \frac{N}{N-1} E_{N-1}^i\left(1; \frac{N-1}{N-2}; 1\right) \geq \frac{N}{2} E_2^i(1; N-1; 1), \tag{6}$$

or the ith eigenvalue of H_N is not less than $N/2$ times the ith energy eigenvalue of the system composed of two such particles interacting with the same external potential and among themselves through a two-body interparticle potential, which is $N-1$ times stronger than the original interparticle potential. It is easy to see that Eq. (5) can be generalized to systems containing several different groups of identical particles, provided that, in addition to Eq. (3) for each group of identical particles, we also impose the following constraints:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij} = NM, \tag{7}$$

for the interaction energy $\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij} v_{ij}$, of two groups of identical particles in the comparison Hamiltonian. Those theorems of Ref. 2 which are based upon the separation of H_N into several decoupled subsystems have no simple analogy in the case of higher energy eigenvalues, as the nth $(n \neq 1)$ energy eigenvalue of the total subsystem is in general not equal to the sum of the nth energy eigenvalues of subsystems.

---

[1] S. H. Gould, *Variational Methods for Eigenvalue Problems* (Oxford U.P., London, 1966), 2nd ed., Chap. 4.

[2] F. Calogero and C. Marchioro, J. Math. Phys. 10, 562 (1969).

Upper and Lower Bounds of the Eigenvalues of a Second-Order Linear Self-Adjoint Differential Equation

Robert H. T. Yeh
Department of Physics and Astronomy, State University of New York, Buffalo, New York 14214
(Received 30 July 1971)

Upper and lower bounds of the eigenvalues of a second-order linear self-adjoint differential equation were obtained by a generalization of the methods of Nordtvedt [J. Math. Phys. 8, 1406 (1967)].

It has been shown that[1] the eigenvalue problem of various types of second-order linear, self-adjoint, differential equation

$$L[u] + \lambda\rho u = 0, \qquad \rho > 0, \tag{1}$$

for a domain G with the boundary condition $u = 0$.

(1) can be reformulated into maximum–minimum problems of the quadratic functional– $\int_G uL(u)dv / (\int_G \rho u^2 dv)$.

(2) If eigenfunctions of Eq. (1) are ordered according to increasing eigenvalues, then the nodes of the nth eigenfunctions u_n, $n = 1, 2, \ldots,$ divide the domain G into no more than n subdomains. (In particular, u_1 has no node and u_2 has exactly one node in G).

(3) The nth eigenvalue for a domain G never exceeds the nth eigenvalue for a subdomain of G.

To find an upper bound of the nth eigenvalue λ_n, let us first consider a trial function u_T, which has $(n-1)$ nodes in G. [I.e., u_T divides G into exactly n subdomains G_α. u_T can always be constructed from simple polynomials. E.g., $\pi_i(u - u_i)(y - y_i)$ if u has two

independent variables. The optimum value of the upper bound is obtained by adjusting the position of the nodes x_i, y_i's.] Then we clearly have

$$\lambda_T^{(\alpha)} \equiv -\int_{G_\alpha} u_T L(u_T)dv \Big/ \left(\int_{G_\alpha} \rho u_T{}^2 dv\right) \geq \lambda_1^{(\alpha)}, \quad (2)$$

where $\lambda_1^{(\alpha)}$ is the smallest eigenvalue of Eq. (1) with domain G_α and the boundary condition $u = 0$. If two neighboring subdomains G_α, G_β had different fundamental tones ($\lambda_1^{(\alpha)} \neq \lambda_1^{(\beta)}$), then by using the property (3) stated above, we could raise the fundamental tone of one and lower that of the other by moving the nodal point (x_i, y_i's) which they have in common, so that both subdomains would have the same fundamental tones. Thus, there exists eigenfunctions of Eq. (1) with $(n-1)$ nodes and $\max\{\lambda_1^{(\alpha)}\} \geq \lambda^{(n)}$, where $\lambda^{(n)}$ is the smallest eigenvalue of Eq. (1) which has an eigenfunction with $(n-1)$ nodes. This together with Eq. (2) implies that

$$\bar{\lambda}_n \equiv \max\{\lambda_T^{(\alpha)}\} \geq \lambda^{(n)} \geq \lambda_n, \quad (3)$$

where the last inequality follows from the property (2) stated in the first paragraph. Equation (3) is the desired upper bound of λ_n. It involves the computation of $2n$ integrals of Eq. (2), $\alpha = 1, \ldots, n$, in comparison with the computation of $2n'^2, n' \geq n$, matrix element integrals and the diagonalization of a $n' \times n'$ matrix in the Rayleigh–Ritz procedure.[2] Equation (3) is a direct generalization of Nordtvedt's result[3] for the one-dimensional Sturm–Liouville equation.

In many cases, we can integrate the following expressions:

$$T_r(L^{-1}\rho)^2 = \sum_i \int_G u_i \rho (L^{-1}\rho)^2 u_i dv = \sum_i \lambda_i^{-2}, \quad (4)$$

where L^{-1} is the inverse operator of L. Then, a lower bound of λ_n can be obtained via the following inequality:

$$\lambda_n{}^2 \geq [T_r(L^{-1}\rho)^2 - \lambda_m{}^{-2}]^{-1} \geq [T_r(L^{-1}\rho)^2 - \bar{\lambda}_m{}^{-2}]^{-1}, \quad (5)$$

where $\bar{\lambda}_m$ is an upper bound of λ_m (any $m, m \neq n$), obtained, say, by using Eq. (3) for λ_m.

Equations (3) and (5) also hold for eigenvalue problems [Eq. (1)] with homogeneous boundary condition of the form $\partial u/\partial n + \sigma u = 0$, where $\partial/\partial n$ denotes differentiation in the direction of the outer normal. This is because the nth eigenvalue λ_n' of Eq. (1), with the boundary condition $\partial u/\partial n + \sigma u = 0$, is never larger than the nth eigenvalue λ_n of the corresponding problem with the boundary condition $u = 0$[2]. Thus $\bar{\lambda}_n$ of Eq. (3) is also an upper bound of λ_n'. It is plausible that these results still hold for more general boundary conditions and for differential operators involving higher-order derivatives, as usually a corresponding quadratic functional for this generalized eigenvalue problem can be found.

---

[1] R. Courant and D. Hilbert; *Methods of Mathematical Physics* (Interscience, New York, 1937), 1st English ed., Chap. 6.

[2] S. H. Gould, *Variational Methods for Eigenvalue Problems* (Oxford U.P., London, 1966), 2nd ed., Chap. 4.

[3] K. Nordtvedt, Jr., J. Math. Phys. 8, 1406 (1967).

Reduction of a Class of Nonlinear Integral Equations to a Cauchy System

Harriet Kagiwada and Robert Kalaba
University of Southern California, Los Angeles, California

and

Chung-Chun Yang
Naval Research Laboratory, Washington, D.C.

(Received 29 June 1971)

It is shown that a wide class of nonlinear integral equations can be transformed into a Cauchy system. Then it is shown that a solution of the Cauchy system provides a solution of the original nonlinear integral equation. Such reductions are important because modern computers can solve initial value problems with speed and accuracy. There are intended applications in the theories of multiple scattering, optimal filtering, and lateral inhibition of neural systems. This new approach makes no use of successive approximations or series expansions.

1. INTRODUCTION

In recent years much effort has been devoted to the transformation of linear Fredholm integral equations[1] into Cauchy systems, i.e., differential equations with known initial conditions.[2–4] The importance of this reduction resides in the fact that modern digital computers can integrate large systems of ordinary differential equations subject to known initial conditions with speed and accuracy.[5]

In spite of the theoretical emphasis on linear functional equations, nature resolutely provides us with example after example of nonlinear behavior in which cause and effect are not proportional.[6,7] In particular, nonlinear integral equations of Hammerstein type have been studied much.[8,9] This paper is devoted to the equivalence between the nonlinear integral equation in Eq. (1) below and a certain Cauchy system. The Cauchy system is derived and validated in Sec. 2 and 3, and Sec. 4 provides some computational aspects.

Our method does not involve the usual successive approximations or series expansions.

2. DERIVATION

Consider the class of nonlinear integral equations

$$u(t) = g(t, \lambda) + \lambda \int_0^1 k(t, y, \lambda, u(y))dy, \quad 0 \leq t \leq 1, \quad 0 \leq \lambda \leq \Lambda. \quad (1)$$

To call attention to the dependence of a solution u upon λ, as well as upon t, we shall write

independent variables. The optimum value of the upper bound is obtained by adjusting the position of the nodes x_i, y_i's.] Then we clearly have

$$\lambda_T^{(\alpha)} \equiv -\int_{G_\alpha} u_T L(u_T)dv \Big/ \left(\int_{G_\alpha} \rho u_T{}^2 dv\right) \geq \lambda_1^{(\alpha)}, \quad (2)$$

where $\lambda_1^{(\alpha)}$ is the smallest eigenvalue of Eq. (1) with domain G_α and the boundary condition $u = 0$. If two neighboring subdomains G_α, G_β had different fundamental tones ($\lambda_1^{(\alpha)} \neq \lambda_1^{(\beta)}$), then by using the property (3) stated above, we could raise the fundamental tone of one and lower that of the other by moving the nodal point (x_i, y_i's) which they have in common, so that both subdomains would have the same fundamental tones. Thus, there exists eigenfunctions of Eq. (1) with $(n-1)$ nodes and $\max\{\lambda_1^{(\alpha)}\} \geq \lambda^{(n)}$, where $\lambda^{(n)}$ is the smallest eigenvalue of Eq. (1) which has an eigenfunction with $(n-1)$ nodes. This together with Eq. (2) implies that

$$\bar{\lambda}_n \equiv \max\{\lambda_T^{(\alpha)}\} \geq \lambda^{(n)} \geq \lambda_n, \quad (3)$$

where the last inequality follows from the property (2) stated in the first paragraph. Equation (3) is the desired upper bound of λ_n. It involves the computation of $2n$ integrals of Eq. (2), $\alpha = 1, \ldots, n$, in comparison with the computation of $2n'^2, n' \geq n$, matrix element integrals and the diagonalization of a $n' \times n'$ matrix in the Rayleigh–Ritz procedure.[2] Equation (3) is a direct generalization of Nordtvedt's result[3] for the one-dimensional Sturm–Liouville equation.

In many cases, we can integrate the following expressions:

$$T_r(L^{-1}\rho)^2 = \sum_i \int_G u_i \rho (L^{-1}\rho)^2 u_i dv = \sum_i \lambda_i^{-2}, \quad (4)$$

where L^{-1} is the inverse operator of L. Then, a lower bound of λ_n can be obtained via the following inequality:

$$\lambda_n{}^2 \geq [T_r(L^{-1}\rho)^2 - \lambda_m{}^{-2}]^{-1} \geq [T_r(L^{-1}\rho)^2 - \bar{\lambda}_m{}^{-2}]^{-1}, \quad (5)$$

where $\bar{\lambda}_m$ is an upper bound of λ_m (any $m, m \neq n$), obtained, say, by using Eq. (3) for λ_m.

Equations (3) and (5) also hold for eigenvalue problems [Eq. (1)] with homogeneous boundary condition of the form $\partial u/\partial n + \sigma u = 0$, where $\partial/\partial n$ denotes differentiation in the direction of the outer normal. This is because the nth eigenvalue λ_n' of Eq. (1), with the boundary condition $\partial u/\partial n + \sigma u = 0$, is never larger than the nth eigenvalue λ_n of the corresponding problem with the boundary condition $u = 0$[2]. Thus $\bar{\lambda}_n$ of Eq. (3) is also an upper bound of λ_n'. It is plausible that these results still hold for more general boundary conditions and for differential operators involving higher-order derivatives, as usually a corresponding quadratic functional for this generalized eigenvalue problem can be found.

---

[1] R. Courant and D. Hilbert; *Methods of Mathematical Physics* (Interscience, New York, 1937), 1st English ed., Chap. 6.

[2] S. H. Gould, *Variational Methods for Eigenvalue Problems* (Oxford U.P., London, 1966), 2nd ed., Chap. 4.

[3] K. Nordtvedt, Jr., J. Math. Phys. 8, 1406 (1967).

Reduction of a Class of Nonlinear Integral Equations to a Cauchy System

Harriet Kagiwada and Robert Kalaba
University of Southern California, Los Angeles, California

and

Chung-Chun Yang
Naval Research Laboratory, Washington, D.C.

(Received 29 June 1971)

It is shown that a wide class of nonlinear integral equations can be transformed into a Cauchy system. Then it is shown that a solution of the Cauchy system provides a solution of the original nonlinear integral equation. Such reductions are important because modern computers can solve initial value problems with speed and accuracy. There are intended applications in the theories of multiple scattering, optimal filtering, and lateral inhibition of neural systems. This new approach makes no use of successive approximations or series expansions.

1. INTRODUCTION

In recent years much effort has been devoted to the transformation of linear Fredholm integral equations[1] into Cauchy systems, i.e., differential equations with known initial conditions.[2–4] The importance of this reduction resides in the fact that modern digital computers can integrate large systems of ordinary differential equations subject to known initial conditions with speed and accuracy.[5]

In spite of the theoretical emphasis on linear functional equations, nature resolutely provides us with example after example of nonlinear behavior in which cause and effect are not proportional.[6,7] In particular, nonlinear integral equations of Hammerstein type have been studied much.[8,9] This paper is devoted to the equivalence between the nonlinear integral equation in Eq. (1) below and a certain Cauchy system. The Cauchy system is derived and validated in Sec. 2 and 3, and Sec. 4 provides some computational aspects.

Our method does not involve the usual successive approximations or series expansions.

2. DERIVATION

Consider the class of nonlinear integral equations

$$u(t) = g(t, \lambda) + \lambda \int_0^1 k(t, y, \lambda, u(y))dy, \quad 0 \leq t \leq 1, \quad 0 \leq \lambda \leq \Lambda. \quad (1)$$

To call attention to the dependence of a solution u upon λ, as well as upon t, we shall write

$$u = u(t,\lambda), \quad 0 \le t \le 1, \quad 0 \le \lambda \le \Lambda. \tag{2}$$

Equation (1) becomes

$$u(t,\lambda) = g(t,\lambda) + \lambda\int_0^1 k(t,y,\lambda,u(y,\lambda))\,dy, \quad 0 \le t \le 1, \quad 0 \le \lambda \le \Lambda. \tag{3}$$

The parameter λ may occur naturally, or it may be artificially introduced. Assume that a solution u is differentiable in λ, and differentiate both sides of Eq. (3) to obtain the relation

$$u_\lambda(t,\lambda) = g_\lambda(t,\lambda) + \int_0^1 k(t,y,\lambda,u(y,\lambda))\,dy + \lambda\int_0^1 k_\lambda(t,y,\lambda,u(y,\lambda))\,dy + \lambda\int_0^1 k_u(t,y,\lambda,u(y,\lambda))u_\lambda(y,\lambda)\,dy. \tag{4}$$

Equation (4) is viewed as a linear Fredholm integral equation in which the kernel is k_u, and the first three terms on the right-hand side are forcing terms.

Assume that the linear Fredholm integral equation

$$w(t,\lambda) = F(t,\lambda) + \lambda\int_0^1 k_u(t,y,\lambda,u(y,\lambda))w(y,\lambda)\,dy, \quad 0 \le t \le 1, \tag{5}$$

possesses a unique solution for $0 \le \lambda \le \Lambda$, where u is a solution of Eq. (3) and F is arbitrary. In terms of the resolvent kernel K, the solution of Eq. (5) is represented as

$$w(t,\lambda) = F(t,\lambda) + \lambda\int_0^1 K(t,y,\lambda)F(y,\lambda)dy, \quad 0 \le t \le 1. \tag{6}$$

The resolvent kernel K itself satisfies the integral equation

$$K(t,y,\lambda) = k_u(t,y,\lambda,u(y,\lambda)) + \lambda\int_0^1 k_u(t,y',\lambda,u(y',\lambda))K(y',y,\lambda)dy', \quad 0 \le t,y \le 1, \quad 0 \le \lambda \le \Lambda. \tag{7}$$

Returning to Eq. (4), we find that the solution u_λ may be represented in the form

$$u_\lambda(t,\lambda) = \Psi(t,\lambda) + \lambda\int_0^1 K(t,y',\lambda)\Psi(y',\lambda)dy', \tag{8}$$

where the function Ψ is given by

$$\Psi(t,\lambda) = g_\lambda(t,\lambda) + \int_0^1 k(t,y',\lambda,u(y',\lambda))dy' + \lambda\int_0^1 k_\lambda(t,y',\lambda,u(y',\lambda))dy', \quad 0 \le t \le 1, \quad 0 \le \lambda \le \Lambda. \tag{9}$$

Equations (8) and (9) form a differential equation for the unknown function u, the other unknown function being the resolvent kernel K. From Eq. (3) we see that the initial condition on the function u at $\lambda = 0$ is

$$u(t,0) = g(t,0), \quad 0 \le t \le 1. \tag{10}$$

Next we obtain a differential equation for the resolvent kernel K. Through differentiation, Eq. (7) becomes

$$K_\lambda(t,y,\lambda) = k_{u\lambda}(t,y,\lambda,u(y,\lambda)) + k_{uu}(t,y,\lambda,u(y,\lambda))u_\lambda(y,\lambda) + \int_0^1 k_u(t,y',\lambda,u(y',\lambda))K(y',y,\lambda)dy' + \lambda\int_0^1 k_{u\lambda}(t,y',\lambda,u(y',\lambda))K(y',y,\lambda)dy' + \lambda\int_0^1 k_{uu}(t,y',\lambda,u(y',\lambda))u_\lambda(y',\lambda)K(y',y,\lambda)dy' + \lambda\int_0^1 k_u(t,y',\lambda,u(y',\lambda))K_\lambda(y',y,\lambda)dy'. \tag{11}$$

This is considered to be a linear Fredholm integral equation for the function K_λ. The kernel is as in Eq. (5), and the first five terms on the right-hand side are taken to be inhomogeneous terms. For simplicity we introduce the auxiliary function Q to be the sum of these forcing terms, and, making use of Eq. (8), we see that

$$\begin{aligned} Q(t,y,\lambda) = {} & k_{u\lambda}(t,y,\lambda,u(y,\lambda)) + k_{uu}(t,y,\lambda,u(y,\lambda)) \\ & \times \left(\Psi(y,\lambda) + \lambda\int_0^1 K(y,y',\lambda)\Psi(y',\lambda)dy'\right) \\ & + \int_0^1 k_u(t,y',\lambda,u(y',\lambda))K(y',y,\lambda)dy' \\ & + \lambda\int_0^1 k_{u\lambda}(t,y',\lambda,u(y',\lambda))K(y',y,\lambda)dy' \\ & + \lambda\int_0^1 k_{uu}(t,y',\lambda,u(y',\lambda)) \\ & \times \left(\Psi(y',\lambda) + \lambda\int_0^1 K(y',y'',\lambda)\Psi(y'',\lambda)dy''\right) \\ & \times K(y',y,\lambda)dy'. \end{aligned} \tag{12}$$

It follows that the solution of Eq. (11) may be written in the form

$$K_\lambda(t,y,\lambda) = Q(t,y,\lambda) + \lambda\int_0^1 K(t,y',\lambda)Q(y',y,\lambda)dy', \quad 0 \le t,y \le 1, \quad 0 \le \lambda \le \Lambda. \tag{13}$$

The initial condition on the resolvent kernel K at $\lambda = 0$ is

$$K(t,y,0) = k_w(t,y,0,g(y,0)), \quad 0 \le t,y \le 1, \tag{14}$$

which follows from Eq. (7).

Let us now summarize the Cauchy system for the functions u and K. The initial conditions at $\lambda = 0$ are given in Eqs. (10) and (14). The differential equations are Eqs. (8) and (13), the auxiliary functions Ψ and Q being defined in Eqs. (9) and (12).

Notice that solution of the Cauchy system provides the desired parameter study in λ, and no successive approximations or series expansions are used. A numerical technique is given in Sec. 4.

3. VALIDATION

We shall now show that a solution of the Cauchy system in Eqs. (10), (14), (8), (13), (9), and (12) provides a solution of the nonlinear integral equation (3). We begin by showing that K, as determined by the Cauchy system, satisfies Eq. (7). Introduce the auxiliary function α by means of the definition

$$\alpha(t,y,\lambda) = k_u(t,y',\lambda,u(y',\lambda)) + \lambda\int_0^1 k_u(t,y',\lambda,u(y',\lambda))K(y',y,\lambda)dy', \quad 0 \le t,y \le 1, \quad 0 \le \lambda \le \Lambda. \tag{15}$$

Then by differentiation and use of Eqs. (8) and (13) we see that

$$\begin{aligned} \alpha_\lambda(t,y,\lambda) = {} & k_{u\lambda}(t,y,\lambda,u(y,\lambda)) + k_{uu}(t,y,\lambda,u(y,\lambda))u_\lambda(y,\lambda) \\ & + \int_0^1 k_u(t,y',\lambda,u(y',\lambda))K(y',y,\lambda)dy' \\ & + \lambda\int_0^1 k_{u\lambda}(t,y',\lambda,u(y',\lambda))K(y',y,\lambda)dy' \\ & + \lambda\int_0^1 k_{uu}(t,y',\lambda,u(y',\lambda))u_\lambda(y',\lambda)K(y',y,\lambda)dy' \\ & + \lambda\int_0^1 k_u(t,y',\lambda,u(y',\lambda))K_\lambda(y',y,\lambda)dy', \end{aligned} \tag{16}$$

$$\alpha_\lambda(t,y,\lambda) = k_{u\lambda}(t,y,\lambda,u(y,\lambda)) + k_{uu}(t,y,\lambda,u(y,\lambda)) \times \left(\Psi(y,\lambda) + \lambda\int_0^1 K(y,y',\lambda)\Psi(y',\lambda)dy'\right) + \int_0^1 k_u(t,y',\lambda,u(y',\lambda))K(y',y,\lambda)dy' + \lambda\int_0^1 k_{u\lambda}(t,y',\lambda,u(y',\lambda))K(y',y,\lambda)dy' + \lambda\int_0^1 k_{uu}(t,y',\lambda,u(y',\lambda)) \times \left(\Psi(y',\lambda) + \lambda\int_0^1 K(y',y'',\lambda)\Psi(y'',\lambda)dy''\right) \times K(y',y,\lambda)dy' + \lambda\int_0^1 k_u(t,y',\lambda,u(y',\lambda)) \times \left(Q(y',y,\lambda) + \lambda\int_0^1 K(y',y'',\lambda)Q(y'',y,\lambda)dy''\right)dy'. \quad (17)$$

It follows that

$$\alpha_\lambda(t,y,\lambda) = Q(t,y,\lambda) + \lambda\int_0^1\left(k_u(t,y',\lambda,u(y',\lambda)) + \lambda\int_0^1 k_u(t,y'',\lambda,u(y'',\lambda))K(y'',y',\lambda)dy''\right) \times Q(y',y,\lambda)dy'. \quad (18)$$

Keeping the definition of α in mind [Eq. (15)], we find that the function α satisfies the differential equation

$$\alpha_\lambda(t,y,\lambda) = Q(t,y,\lambda) + \lambda\int_0^1 \alpha(t,y',\lambda)Q(y',y,\lambda)dy', \quad 0 \le t,y \le 1. \quad (19)$$

The initial condition at $\lambda = 0$ is

$$\alpha(t,y,0) = k_u(t,y,0,g(y,0)), \quad 0 \le t,y \le 1. \quad (20)$$

Comparing Eqs. (19) and (20) against Eqs. (13) and (14), and, assuming uniqueness of solution, it is seen that

$$\alpha(t,y,\lambda) = K(t,y,\lambda) \quad (21)$$

or

$$K(t,y,\lambda) = k_u(t,y,\lambda,u(y,\lambda)) + \lambda\int_0^1 k_u(t,y',\lambda,u(y',\lambda))K(y',y,\lambda)dy', \quad 0 \le t,y \le 1, \quad 0 \le \lambda \le \Lambda. \quad (22)$$

Lastly, we introduce the function β to be

$$\beta(t,\lambda) = g(t,\lambda) + \lambda\int_0^1 k(t,y,\lambda,u(y,\lambda))dy, \quad 0 \le 1 \le 1, \quad 0 \le \lambda \le \Lambda. \quad (23)$$

Differentiation with respect to λ shows that

$$\beta_\lambda(t,\lambda) = g_\lambda(t,\lambda) + \int_0^1 k(t,y,\lambda,u(y,\lambda))dy + \lambda\int_0^1 k_\lambda(t,y,\lambda,u(y,\lambda))dy + \lambda\int_0^1 k_u(t,y,\lambda,u(y,\lambda))u_\lambda(y,\lambda)dy. \quad (24)$$

Using Eqs. (9) and (8), we reduce Eq. (24) to the form

$$\beta_\lambda(t,\lambda) = \Psi(t,\lambda) + \lambda\int_0^1 k_u(t,y,\lambda,u(y,\lambda)) \times \left(\Psi(y,\lambda) + \lambda\int_0^1 K(y,y',\lambda)\Psi(y',\lambda)dy'\right)dy \quad (25)$$

or

$$\beta_\lambda(t,\lambda) = \Psi(t,\lambda) + \lambda\int_0^1\left(k_u(t,y,\lambda,u(y,\lambda)) + \lambda\int_0^1 k_u(t,y',\lambda,u(y',\lambda))K(y',y,\lambda)dy'\right) \times \Psi(y,\lambda)dy. \quad (26)$$

Making use of Eq. (22), we observe that

$$\beta_\lambda(t,\lambda) = \Psi(t,\lambda) + \lambda\int_0^1 K(t,y,\lambda)\Psi(y,\lambda)dy, \quad 0 \le t \le 1. \quad (27)$$

Furthermore, from the definition of the function β in Eq. (23) we have

$$\beta(t,0) = g(t,0), \quad 0 \le t \le 1. \quad (28)$$

Keeping Eqs. (8) and (10) in mind, we see that

$$\beta(t,\lambda) = u(t,\lambda), \quad 0 \le t \le 1, \quad 0 \le \lambda \le \Lambda, \quad (29)$$

i.e.,

$$u(t,\lambda) = g(t,\lambda) + \lambda\int_0^1 k(t,y,\lambda,u(y,\lambda))dy, \quad 0 \le t \le 1, \quad 0 \le \lambda \le \Lambda, \quad (30)$$

which completes the demonstration.

4. NUMERICAL ANALYSIS

The Cauchy system may be handled numerically by using the method of lines.[10] Much previous experience indicates the efficacy of such an approach.[2,3] The general idea is to approximate integrals on the interval (0, 1) by means of a quadrature formula of order N,

$$\int_0^1 f(y)dy \cong \sum_{i=1}^N f(r_i)w_i. \quad (31)$$

In that way the differential integral equations are approximated by a system of ordinary differential equations as closely as desired. By introducing

$$u(r_i,\lambda) = u_i(\lambda), \quad (32)$$

$$\Psi(r_i,\lambda) = \Psi_i(\lambda), \quad (33)$$

$$K(r_i,r_j,\lambda) = K_{ij}(\lambda), \quad i,j = 1,2,\ldots,N, \quad 0 \le \lambda \le \Lambda, \quad (34)$$

Eq. (8) becomes, for example,

$$du_i(\lambda)/d\lambda = \Psi_i(\lambda) + \lambda\sum_{m=1}^N K_{im}(\lambda)\Psi_m(\lambda)w_m, \quad (35)$$

with similar equations approximating Eqs. (13), (9), and (10) and the initial condition in Eqs. (10) and (14). In all, there are $N^2 + N$ ordinary differential equations subject to known initial conditions. In 1971 this is reasonable for $N \cong 50$.

Successful calculations for the Ambarzumian integral equation of radiative transfer have been done.[11,12]

5. DISCUSSION

In this paper we have shown the equivalence between a general class of nonlinear integral equations and a Cauchy system. Much remains to be studied analytically, especially with regard to bifurcation phenomena. On the computational side, we plan to undertake sample calculations in radiative transfer,[12] optimal filtering and modulation,[13] and neurophysiology.[7]

---

[1] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. I.

[2] H. Kagiwada and R. Kalaba, J. Assoc. Computing Machinery **17**, 412 (1970).

[3] H. Kagiwada, R. Kalaba, and B. Vereeke, J. Approximation Theory **1**, 355 (1968).

[4] J. Buell, J. Casti, R. Kalaba, and S. Ueno, J. Math. Phys. **11**, 1673 (1970).

[5] G. Bekey and W. Karplus, *Hybrid Computation* (Wiley, New York, 1968).
[6] F. Tricomi, *Integral Equations* (Interscience, New York, 1957).
[7] H. Landahl, "Mathematical Models in the Behavior of the Central Nervous System", in *Mathematical Problems in the Biological Sciences* (American Mathematical Society, Providence, R.I., 1962).
[8] M. Vainberg, *Variational Methods for the Study of Nonlinear Operators* (Holden-Day, San Francisco, 1964).
[9] M. Vainberg and B. Trenogin, *Teoriya Vetvleniya Reshenii Nelineinykh Uravnenii* (Moskva, 1969).
[10] I. Berezin and N. Zhidkov, *Computing Methods* (Addison-Wesley, Palo Alto, 1965).
[11] J. Buell, A. Fymat, and R. Kalaba, "An Initial Value Method for the Ambarzumian Integral Equation", J. Quant. Spectroscopy Radiative Transfer (to be published).
[12] V. Sobolev, *A Treatise on Radiation Transfer* (Van Nostrand, New York, 1963).
[13] H. Van Trees, *Detection, Estimation and Modulation Theory* (Wiley, New York, 1968).

Operator Valued Measures in Quantum Mechanics: Finite and Infinite Processes*

Paul A. Benioff
Chemistry Division, Argonne National Laboratory, Argonne, Illinois 60439
(Received 25 June 1971)

In this work, operator valued measures are used to study finite and infinite sequences of measurements. It is shown that to each such process q^Δ there is uniquely associated a probability operator measure O^{q^Δ} which contains all the statistical properties of the process. In order to make this association for infinite processes, the operator valued equivalent of the Kolmogorov extension theorem is needed. This theorem is given and proved. It is then shown that for each q^Δ and each set E of possible outcome sequences, there are two ways to find the probability that carrying out q^Δ on a system in state ρ gives an outcome sequence ϕ in E. The usual method of repeating q^Δ on ρ over and over again generates a sequence α of outcome sequences ϕ. The probability is obtained as the limit relative frequency that $\alpha(j)$ is in E, for $j = 0, 1, \cdots$. The other, new, method is the repeated measurement of $O_E^{q^\Delta}$ on ρ. The remarkable aspect of this equivalence is that the mathematical procedures of the usual method for determining if $\alpha(j)$ is in E or not 'disappear' into the operators $O_E^{q^\Delta}$ of the new method. This is discussed in some detail and examples are given.

I. INTRODUCTION

Although not yet in wide use, operator valued measures appear to be a useful tool, particularly for the study of processes in Quantum Mechanics. This is exemplified by the work of Davies and Lewis[1] who have given a very general treatment applicable to processes with a finite number of steps. In particular, their description of sequences of measurements allows use of observables with continuous spectra and does not use Von Neumann's projection postulates.[2]

In this paper, the relationship between compound processes and operator valued measures is studied further. The work here is more restricted than that of Davies and Lewis in that the usual formalism, which is restricted to observables with discrete spectra and uses Von Neumann's projection postulate, is employed. It is more general in that the treatment includes processes with an infinite number of steps.

A main result of this work is that to each compound process of a finite or infinite number of steps there is associated in a unique manner a probability operator measure which contains or encodes all the statistical properties of the process. Furthermore, this association is exhaustive in that all processes of the type studied here are included. That is, the operator sequence, the time spacing between measurements, and the time-dependence of the Hamiltonian are arbitrary. The construction of this mapping for both finite and infinite processes occupies Sec. III.

Section II collects together some mathematical properties of positive operator measures which are relevant to this work. The main result which is needed for infinite processes is Theorem 4. This theorem is an extension to positive operator measures of a theorem of Kolmogorov[3,4] for probability measures, which states the existence and uniqueness of a probability measure on an infinite product space, given a consistent set of probabilities on the component finite product spaces. As proved here, Theorem 4 is more general than is needed and may have use in extending the work of Davies and Lewis[1] to include infinite processes.

If one assumes that to each bounded self-adjoint operator on an appropriate Hilbert space there corresponds an observation procedure,[5] then a remarkable property of these operator measures emerges. It is first seen that for each event E in a σ-algebra Σ of subsets of the certain event Ω there are two ways to determine the probability that carrying out the process on a system in state ρ gives an outcome sequence in E. One is the usual statistical method which consists of carrying out the process on the system over and over, infinitely many times and determining the limit relative frequency that the elements of the resulting sequence lie in E. The other, which is new, consists of repeated measurements of the particular operator O_E which the probability operator measure assigns to the event E, on the system in state ρ.

This equivalence as well as a "first order" equivalence, which removes the infinite repetition of carrying out the process on a system in state ρ, is discussed in Sec. IV. There it is shown by means of sample space constructions that the two methods are equivalent.

In Sec. V this equivalence is discussed further. It is seen to provide an interesting link between mathematical procedures and physical operations in the following sense: 1) Each $E \in \Sigma$ corresponds to some mathematical property Q where $E = [\phi \mid Q(\phi)\ \text{true}]$. 2) In the standard statistical method, one generates from one or more outcome sequences, an infinite sequence θ_E of 0's and 1's, whose limit relative frequency gives the desired probability $\mathrm{Tr}\rho O_E$ that carrying out the process on a system in state ρ gives an outcome sequence in E. This generation involves a sequence of mathematical decision procedures about whether $Q(\phi)$ is true or false for different ϕ as well as (in the "first order" case) the construction

[5] G. Bekey and W. Karplus, *Hybrid Computation* (Wiley, New York, 1968).
[6] F. Tricomi, *Integral Equations* (Interscience, New York, 1957).
[7] H. Landahl, "Mathematical Models in the Behavior of the Central Nervous System", in *Mathematical Problems in the Biological Sciences* (American Mathematical Society, Providence, R.I., 1962).
[8] M. Vainberg, *Variational Methods for the Study of Nonlinear Operators* (Holden-Day, San Francisco, 1964).
[9] M. Vainberg and B. Trenogin, *Teoriya Vetvleniya Reshenii Nelineinykh Uravnenii* (Moskva, 1969).
[10] I. Berezin and N. Zhidkov, *Computing Methods* (Addison-Wesley, Palo Alto, 1965).
[11] J. Buell, A. Fymat, and R. Kalaba, "An Initial Value Method for the Ambarzumian Integral Equation", J. Quant. Spectroscopy Radiative Transfer (to be published).
[12] V. Sobolev, *A Treatise on Radiation Transfer* (Van Nostrand, New York, 1963).
[13] H. Van Trees, *Detection, Estimation and Modulation Theory* (Wiley, New York, 1968).

Operator Valued Measures in Quantum Mechanics: Finite and Infinite Processes*

Paul A. Benioff
Chemistry Division, Argonne National Laboratory, Argonne, Illinois 60439
(Received 25 June 1971)

In this work, operator valued measures are used to study finite and infinite sequences of measurements. It is shown that to each such process q^Δ there is uniquely associated a probability operator measure O^{q^Δ} which contains all the statistical properties of the process. In order to make this association for infinite processes, the operator valued equivalent of the Kolmogorov extension theorem is needed. This theorem is given and proved. It is then shown that for each q^Δ and each set E of possible outcome sequences, there are two ways to find the probability that carrying out q^Δ on a system in state ρ gives an outcome sequence ϕ in E. The usual method of repeating q^Δ on ρ over and over again generates a sequence α of outcome sequences ϕ. The probability is obtained as the limit relative frequency that $\alpha(j)$ is in E, for $j = 0, 1, \cdots$. The other, new, method is the repeated measurement of $O_E^{q^\Delta}$ on ρ. The remarkable aspect of this equivalence is that the mathematical procedures of the usual method for determining if $\alpha(j)$ is in E or not 'disappear' into the operators $O_E^{q^\Delta}$ of the new method. This is discussed in some detail and examples are given.

I. INTRODUCTION

Although not yet in wide use, operator valued measures appear to be a useful tool, particularly for the study of processes in Quantum Mechanics. This is exemplified by the work of Davies and Lewis[1] who have given a very general treatment applicable to processes with a finite number of steps. In particular, their description of sequences of measurements allows use of observables with continuous spectra and does not use Von Neumann's projection postulates.[2]

In this paper, the relationship between compound processes and operator valued measures is studied further. The work here is more restricted than that of Davies and Lewis in that the usual formalism, which is restricted to observables with discrete spectra and uses Von Neumann's projection postulate, is employed. It is more general in that the treatment includes processes with an infinite number of steps.

A main result of this work is that to each compound process of a finite or infinite number of steps there is associated in a unique manner a probability operator measure which contains or encodes all the statistical properties of the process. Furthermore, this association is exhaustive in that all processes of the type studied here are included. That is, the operator sequence, the time spacing between measurements, and the time-dependence of the Hamiltonian are arbitrary. The construction of this mapping for both finite and infinite processes occupies Sec. III.

Section II collects together some mathematical properties of positive operator measures which are relevant to this work. The main result which is needed for infinite processes is Theorem 4. This theorem is an extension to positive operator measures of a theorem of Kolmogorov[3,4] for probability measures, which states the existence and uniqueness of a probability measure on an infinite product space, given a consistent set of probabilities on the component finite product spaces. As proved here, Theorem 4 is more general than is needed and may have use in extending the work of Davies and Lewis[1] to include infinite processes.

If one assumes that to each bounded self-adjoint operator on an appropriate Hilbert space there corresponds an observation procedure,[5] then a remarkable property of these operator measures emerges. It is first seen that for each event E in a σ-algebra Σ of subsets of the certain event Ω there are two ways to determine the probability that carrying out the process on a system in state ρ gives an outcome sequence in E. One is the usual statistical method which consists of carrying out the process on the system over and over, infinitely many times and determining the limit relative frequency that the elements of the resulting sequence lie in E. The other, which is new, consists of repeated measurements of the particular operator O_E which the probability operator measure assigns to the event E, on the system in state ρ.

This equivalence as well as a "first order" equivalence, which removes the infinite repetition of carrying out the process on a system in state ρ, is discussed in Sec. IV. There it is shown by means of sample space constructions that the two methods are equivalent.

In Sec. V this equivalence is discussed further. It is seen to provide an interesting link between mathematical procedures and physical operations in the following sense: 1) Each $E \in \Sigma$ corresponds to some mathematical property Q where $E = [\phi \mid Q(\phi)\ \text{true}]$. 2) In the standard statistical method, one generates from one or more outcome sequences, an infinite sequence θ_E of 0's and 1's, whose limit relative frequency gives the desired probability $\mathrm{Tr}\rho O_E$ that carrying out the process on a system in state ρ gives an outcome sequence in E. This generation involves a sequence of mathematical decision procedures about whether $Q(\phi)$ is true or false for different ϕ as well as (in the "first order" case) the construction

of some mappings. 3) By means of the equivalence, this whole *mathematical procedure* of generating θ_E "disappears" into an appropriate operator O_E, in the sense that it is replaced by the purely *physical operation* of measuring O_E on ρ over and over again.

Besides some examples which illustrate this property, Sec. V includes some brief comments on possible extensions of the construction given here.

II. MATHEMATICAL PROPERTIES

Let Ω be a set and Σ a σ-field of subsets of Ω. That is, Σ is a set of subsets of Ω which contains Ω and is closed under countable unions and complementations. Let $B(\mathcal{H})$ be the set of all bounded linear operators on a Hilbert space $\mathcal{H}$. Let $B(\mathcal{H})^+$ be the restriction of $B(\mathcal{H})$ to the positive operators. A is a positive operator if $A \geq 0$; that is $(\psi, A\psi) \geq 0$ for each ψ in $\mathcal{H}$. Each positive operator is self-adjoint[6] and thus each A in $B(\mathcal{H})^+$ is self-adjoint.

A positive operator measure O is defined to be a mapping with domain Σ and range in $B(\mathcal{H})^+$, such that O is finitely additive and strongly countably additive on Σ. That is, for each sequence $\{E_n | n = 0, 1, \cdots\}$ of pairwise disjoint sets in Σ, with $E = \bigcup_n E_n$,

$$O_E = \sum_n O_{E_n}, \tag{1}$$

where the implied convergence is in the strong operator topology.

It is an immediate consequence of this definition that

$$O_\Phi = \mathbf{0}, \tag{2}$$

where Φ is the empty set and $\mathbf{0}$ is the zero element of $B(\mathcal{H})$.

This definition is a specialization of the more general definition of operator valued measures as countably additive set functions with range in $B(X, Y)$, the set of all bounded linear operators from a Banach space X to a Banach space Y. Since $B(X, Y)$ is a Banach space under the uniform operator topology,[7] many, but not all properties of the better known vector valued measures[8,9] can be taken over to operator valued measures. An example of a property which apparently cannot be taken over is that while Pettis' theorem[8] gives the result that, in the operator topologies, weak and strong countable additivity are equivalent, they do not imply uniform countable additivity.[10]

The definition of Eq. (1) can be trivially extended by allowing Σ to be a field (or a ring)[11] and we shall often do so without comment. Theorems 1–3 which follow, hold for rings and σ-rings as well as for fields and σ-fields.

Some properties of O are as follows[11]: if $E, F \in \Sigma$ and $E \subseteq F$, then

$$O_F - O_E = O_{F-E} \tag{3}$$

and

$$O_E \leq O_F. \tag{4}$$

These results follow from Eq. (2), the positivity of O, and $F = E \cup (F - E)$ with $E \cap (F - E) = \Phi$. Also, for each $E \in \Sigma$,

$$0 \leq O_E \leq O_\Omega, \tag{5}$$

and thus O is bounded by O_Ω.

Also, one has

$$O_{E \cap F} \leq O_E \text{ and } O_F, \tag{6}$$

$$O_{E \cup F} \geq O_E \text{ and } O_F, \tag{7}$$

$$O_{E'} = O_\Omega - O_E, \tag{8}$$

$$O_F = O_{F \cap E} + O_{F \cap E'}, \tag{9}$$

where $E' = \Omega - E$ is the complement of E. Note that for any E if $(\psi, O_E\psi) = 0$ for all $\psi \in \mathcal{H}$, then $O_E = \mathbf{0}$.

Let $\sigma(O_E)$ denote the spectrum of O_E. The above shows that $\sigma(O_E) \subseteq [0, \|O_\Omega\|]$ for each E.

A probability operator measure is defined to be a positive operator valued measure for which

$$O_\Omega = 1. \tag{10}$$

Let ρ be any normalized state over $\mathcal{H}$. Clearly by Eqs. (1)–(5) and (10), the scalar set function $P_{O\rho}$ defined for each $E \in \Sigma$ by

$$P_{O\rho}(E) = \mathrm{Tr}\rho O_E \tag{11}$$

is a probability measure. We shall be mainly concerned here with probability operator measures. Let O be a probability operator measure. Then for each $E \in \Sigma$, O_E is a positive contraction operator and the spectrum of O_E lies in $[0, 1]$. Also if $(\psi, O_E\psi) = 1$ for each $\psi \in \mathcal{H}$, then $O_E = 1$.

It is clear from the above definition that spectral measures of self-adjoint operators in $B(\mathcal{H})$ are simple examples of probability operator measures. The range set of a spectral measure is a set of mutually commuting projection operators. The measures studied here generalize spectral measures in that the operators in the range set do not have to commute or be projection operators.

Berberian[11] has given a different definition of positive operator measures which replaces strong countable additivity by finite additivity and strong continuity from below. We show that the two definitions are equivalent.

A positive operator measure on a field Z is *strongly continuous from above (below)* if for all nonincreasing (nondecreasing) sequences of sets $E_0, E_1, \cdots$ in Z such that $\lim_n E_n = E \in Z$, one has $O_E = \mathrm{s(trong)} - \lim_n O_{E_n}$. O is *s-continuous* on Z if it is both s-continuous from above and below.

Theorem 1: If O is a positive operator measure on a field Z of subsets of Ω then O is s-continuous on Z.

Proof: a) O is s-continuous from below.

Let $\{E_n | n = 0, 1, \ldots\}$ be a nondecreasing sequence of sets in Z with $\lim_n E_n = E \in Z$. Then $E = \bigcup_n E_n = E_0 \cup E_1 - E_0 \cup \cdots = A_0 \cup A_1 \cup \cdots$, where $A_n = E_n - E_{n-1}$, $A_0 = E_0$ and the A_n are pairwise disjoint.

Since Z is a field, $A_n \in Z$ and the strong σ-additivity of O gives

$$\left\| \left(O_E - O_{\cup_{j=0}^n E_j}\right)\psi \right\|$$
$$\leq \left\| \left(O_E - \sum_{j=0}^n O_{A_j}\right)\psi \right\| + \left\| \left(\sum_{j=0}^n O_{A_j} - O_{\cup_{j=0}^n E_j}\right)\psi \right\|$$
$$= \left\| \left(O_E - \sum_{j=0}^n O_{A_j}\right)\psi \right\| \to 0 \text{ as } n \to \infty$$
$$\text{as } \sum_{j=0}^n O_{A_j} = O_{\cup_{j=0}^n E_j} = O_{E_n}.$$

b) O is s-continuous from above.

Let $\{E_n \mid n = 0, 1, \cdots\}$ be a nonincreasing sequence of sets in Z with $\lim_n E_n = \cap_n E_n = E \in Z$. By the set identities, $E' = \cup_n E_n'$ for the complements of E and E_n. Since $\{E_n\}$ is nonincreasing, $\{E_n'\}$ is nondecreasing with E_n' and E' both in Z. Equation (8) and part a above give

$$\left\| \left(O_E - O_{\cap_{j=0}^n E_j}\right)\psi \right\| = \left\| \left(O_{E'} - O_{\cup_{j=0}^n E_j'}\right)\psi \right\| \to 0 \text{ as } n \to \infty. \quad \text{QED}$$

Theorem 2: Let O be a finitely additive $B(\mathcal{H})^+$ valued set function on a field Z. If O is either s-continuous from below or s-continuous from above at Φ, then O is strongly σ-additive on Z.

Proof: If O is finitely additive, Eq. (8) holds. Repetition of the proof of part b, of Theorem 1 gives that O is s-continuous from above. Thus it is sufficient to prove the theorem under the condition that O is s-continuous from above at Φ.

Let $\{E_n\}$ be a sequence of pairwise disjoint sets in Z with $\cup_n E_n = E \in Z$ and let $F_n = \cup_{j=0}^n E_n$. Clearly the sequence of sets $\{E - F_n\}$ is nonincreasing with $\lim_n E - F_n = \Phi$ and $E - F_n \in Z$. Since $F_n \subseteq E$, Eqs. (2) and (3) and the finite additivity give with s-continuity from the above at Φ,

$$\left\| (O_E - \sum_{j=0}^n O_{E_j})\psi \right\| = \left\| (O_E - O_{F_n})\,\psi \right\| = \| O_{E-F_n}\psi \| \to 0$$

as $n \to \infty$.

Thus

$$O = \text{s} - \lim_n \sum_{j=0}^n O_{E_j}. \quad \text{QED}$$

From these two theorems, one concludes that Berberian's definition of positive operator measures and that of Eq. (1) are equivalent. Note, too, that by Pettis' theorem,[8] strong continuity is equivalent to weak continuity.

We come now to the extension theorem.

Theorem 3: Let O be a positive operator measure on a field Z. Then O can be extended to a positive operator measure O' on the minimal σ-field Σ over Z such that $O' = O$ on Z. Furthermore, the extension is unique in that if O_1 and O_2 are extensions of O then $O_1 = O_2$.

Berberian[11] has given a proof of this theorem which uses the Hahn extension theorem[8] to extend the family $[(\psi, O_- \psi) \mid \psi \in \mathcal{H}]$ of scalar measures to the family $[(\psi, O_-' \psi) \mid \psi \in \mathcal{H}]$. He then shows that this family defines a positive operator measure on Σ. One can also give a more constructive proof of this theorem by using the transfinite induction process[12] which generates Σ from Z to extend O from Z to Σ.[13]

Here, the importance of the extension theorem for positive operator measures is that it allows one to prove the theorem needed to assign probability operator measures to infinite processes. Let R^n and R^ω denote the Cartesian product of n copies and denumerably many copies, respectively, of the real line and let $\mathfrak{B}(R^n)$ and $\mathfrak{B}(R^\omega)$ be the respective σ-fields of Borel subsets of R^n and R^ω. A subset E of R^ω is a Borel cylinder set if it can be written in the form

$$E = F \times R \times R \times \cdots$$

with $F \in \mathfrak{B}(R^n)$ for some n. Clearly for each cylinder set, there is a least n such that this property holds. F is called a *base* of the set E. Let $\mathfrak{F}$ be the set of all Borel cylinder subsets of R^ω. Clearly $\mathfrak{F}$ is a subfield of $\mathfrak{B}(R^\omega)$.

Let $\{O^n\}$ be a sequence of positive operator measures such that O^n is defined on $\mathfrak{B}(R^n)$ for $n = 1, 2, \cdots$. The sequence is said to be consistent if for each m and n with $m \geq n$ and for each pair of sets E, F with $E \in \mathfrak{B}(R^m)$, $F \in \mathfrak{B}(R^n)$ such that $E = F \times R^{m-n}$, one has $O_E^m = O_F^n$.

Theorem 4: Let $\{O^n\}$ be a consistent sequence of positive operator measures on the sequence $\{\mathfrak{B}(R^n)\}$ of σ-fields. Then $\{O^n\}$ extends to a unique positive operator measure O on the σ-field $\mathfrak{B}(R^\omega)$ such that for each $E \in \mathfrak{F}$ with base $F \in \mathfrak{B}(R^n)$, $O_E = O_F^n$.

Proof: The proof follows that given by Kolmogorov[3,4] for probability measures. Let $\mathfrak{F}$ be the field defined above. Define an operator measure O' on $\mathfrak{F}$ by $O_E' = O_F^n$ for each set $E \in \mathfrak{F}$ with base $F \in \mathfrak{B}(R^n)$. Since the O^n are consistent and σ-additive on the $\mathfrak{B}(R^n)$, O' is well defined and finitely additive on $\mathfrak{F}$. If one can prove that O' is strongly continuous at Φ on $\mathfrak{F}$, then it follows from Theorem 2 that O' is strongly σ-additive on $\mathfrak{F}$. The extension theorem then gives the desired result with $O = O'$ on $\mathfrak{F}$.

For the proof of the strong continuity of O' at Φ, it is sufficient to prove the converse. That is, let $\{E_n\}$ be a nonincreasing sequence of sets in $\mathfrak{F}$ such that for some vector ψ in $\mathcal{H}$ and some $\epsilon > 0$, $\| O_{E_n}' \psi \| > \epsilon$ for each n. We must show that $\lim_n E_n$ is not empty. Furthermore, it is an inessential simplification to assume that $\{E_n\}$ is such that for each n, E_n has a base $F_n \in \mathfrak{B}(R^n)$.

Now by construction, $\| O_{F_n}^n \psi \| = \| O_{E_n}' \psi \| > \epsilon$. For each n, O^n is s-continuous from below (Theorem 1) and regular[11] on $\mathfrak{B}(R^n)$. Thus for each $\delta > 0$, there is some bounded closed set B_n such that

$$B_n \subset F_n \text{ and } \| (O_{F_n}^n - O_{B_n}^n)\psi \| = \| O_{F_n - B_n}^n \psi \| < \frac{\delta}{2^{n+1}}. \quad (3')$$

Let $G_n \in \mathfrak{F}$ be the cylinder set with base B_n. Then $G_n \subset E_n$ and

$$\| (O_{F_n}^n - O_{B_n}^n)\psi \| = \| (O_{E_n}' - O_{G_n}')\psi \| = \| O_{E_n - G_n}' \psi \| < \frac{\delta}{2^{n+1}}$$

Define W_n by $W_n = \cap_{j=0}^n G_n$. Since $E_n - W_n = \cup_{j=0}^n (E_n - G_j) \subseteq \cup_{j=0}^n (E_j - G_j)$, one has

$$\| O'_{E_n - W_n} \psi \| \leq \sum_{j=0}^{n} \| O'_{E_j - G_j} \psi \| < \delta.$$

Since $W_n \subset G_n \subset E_n$, Eq. (4) gives

$$\delta > \| (O'_{E_n} - O'_{W_n}) \psi \| > \| O'_{E_n} \psi \| - \| O'_{W_n} \psi \| \text{ or } \| O_{W_n} \psi \| > \epsilon - \delta > 0$$

if δ is sufficiently small. Thus W_n is not empty and we can choose a point ϕ^n in W_n. Carrying this out for each n yields a sequence $\{\phi^n\}$ such that $\phi^{n+k} \in W_n$ for each $k = 0, 1, \cdots$. Thus ϕ^{n+k} is in G_n and E_n and $\phi_n^{n+k} \in B_n$ where ϕ_n^m is the initial segment of ϕ^m of length n.

Consider the sequence $\phi^m(0)$ for $m = 0, 1, \cdots$. Since B_n is bounded, $\{\phi^m(0)\}$ contains a convergent subsequence $\phi^{m_0(i)}(0)$ with $\lim_i \phi^{m_0(i)}(0) = x_0 \in R$. Similarly, the sequence $\phi^{m_0(i)}(1)$ for $i = 0, 1, \ldots$ contains a convergent subsequence $\phi^{m_{0,1}(i)}(1)$ with $\lim_i \phi^{m_{0,1}(i)}(1) = x_1$. Continuing in this fashion, one generates for each k a sequence $\{\phi^{m_{0,1,\ldots,k(i)} | i = 0, 1, \cdots}\}$ of sequences such that for each $j \leq k$, $\lim_i \phi^{m_{0,1,\ldots k^{(i)}}}(j) = x_j$.

By the diagonal procedure, one chooses a single sequence of sequences $\{\phi^{m_{0,1,\ldots,k(k)} | k = 0, 1, \ldots}\}$ such that $\lim_k \phi^{m_{0,1,\ldots,k(k)}}(j) = x_j$ for each j. From this one can define a sequence ϕ by $\phi(j) = x_j$ for each j. By construction and the fact that B_n is closed, one has $\phi_n \in B_n$ for each n. So $\phi \in G_n \subset E_n$ for each n. As a result, $\phi \in \cap_n E_n$ so $\lim_n E_n$ is not empty. QED

We now give some other, more general properties of operator measures. Let $\theta(\Sigma, B(\mathcal{H}))$ be the set of all operator measures from Σ to $B(\mathcal{H})$. A general $B(\mathcal{H})$ valued operator measure is a set function with domain Σ and range in $B(\mathcal{H})$, the set of *all* linear bounded operators on $\mathcal{H}$, which satisfies Eqs. (1) and (2). $\theta(\Sigma, B(\mathcal{H}))$ is a Banach space under the norm[7]

$$\| O \| = \sup_{E \in \Sigma} \| O_E \| . \tag{12}$$

That is, it is closed under scalar multiplication and addition of measures, it is closed with respect to the norm, Eq. (12), and it is complete. A proof of this follows Dunford and Schwartz (Ref. 7, pp. 160–63).

Let T be any set transformation which is measurable and preserves set operations. A set transformation T is measurable if T takes sets in Σ into sets in Σ. T preserves set operations if $T\Phi = \Phi$, $T\Omega = \Omega$, $T \cup_n E_n = \cup_n TE_n$, and $T(E - F) = TE - TF$ for all subsets E, F, E_n of Ω. In what follows, T will always denote such a transformation and $T^n E$ denotes n applications of T to E.

Each transformation T as defined above induces a linear transformation $\mathcal{T}: \theta(\Sigma, B(\mathcal{H})) \to \theta(\Sigma, B(\mathcal{H}))$ according to

$$\mathcal{T} O_E = O_{TE} \tag{13}$$

for each $O \in \theta(\Sigma, B(\mathcal{H}))$ and each $E \in \Sigma$. To see that Eq. (13) induces a linear transformation let $O' = \alpha O + \beta O''$ with α, β arbitrary complex numbers. Then $\mathcal{T} O'_E = O'_{TE} = \alpha O_{TE} + \beta O''_{TE} = \alpha \mathcal{T} O_E + \beta \mathcal{T} O''_E$ for each $E \in \Sigma$.

Let $\mathcal{T}(\theta)$ denote the set of all linear transformations $\mathcal{T}: \theta(\Sigma, B(\mathcal{H})) \to \theta(\Sigma, B(\mathcal{H}))$. Clearly $\mathcal{T}(\theta)$ is an algebra under the usual definitions of multiplication and addition. It is also a normed algebra under the norm given by

$$\| \mathcal{T} \| = \sup_{\| O \| \leq 1} \| \mathcal{T} O \| . \tag{14}$$

Any $\mathcal{T}$ in $\mathcal{T}(\theta)$ induced by a T through Eq. (13) is a contraction operator.

Proof:

$$\| \mathcal{T} \| = \sup_{\| O \| \leq 1} \| \mathcal{T} O \| = \sup_{\| O \| \leq 1} \sup_{E \in \Sigma} \| O_{TE} \| \leq \sup_{\| O \| \leq 1} \sup_{E \in \Sigma} \| O_E \| \leq 1,$$

where the measureability of T and Eq. (12) have been used.

Define $\overline{\mathcal{T}}_n$ by

$$\overline{\mathcal{T}}_n O = \frac{1}{n} \sum_{j=0}^{n-1} \mathcal{T}^j O \tag{15}$$

for each O in $\theta(\Sigma, B(\mathcal{H}))$ where $(\mathcal{T}^j O)_E = O_{T^j E}$ for each $E \in \Sigma$. Clearly $\overline{\mathcal{T}}_n$ is linear and bounded if $\mathcal{T}$ is. Of special interest are those O for which $\overline{O}^{\mathcal{T}}$ defined by

$$\overline{O}^{\mathcal{T}} = \lim_n \overline{\mathcal{T}}_n O = \overline{\mathcal{T}} O \tag{16}$$

exists.

The relevant ergodic theorem is the following.[14]

Theorem 5: Let $\| \overline{\mathcal{T}}_n \| \leq c$ for each n for some fixed constant c. Then the set of O for which $\overline{\mathcal{T}}_n O$ converges is a subspace θ' of $\theta(\Sigma, B(\mathcal{H}))$ which contains all those O for which the set $\overline{\mathcal{T}}_n O$ is weakly sequentially compact and $\mathcal{T}^n O/n \to 0$.

Also $\overline{\mathcal{T}}$ is a projection operator on θ' which decomposes θ' into two subspaces θ'_1 and θ'_N where any O in $\theta'_1 = \overline{\mathcal{T}} \theta'$ is $\mathcal{T}$ invariant (that is $\mathcal{T} O = O$) and $\theta'_N =$ closure of $(1 - \mathcal{T})\theta'$.

The proof is given in Dunford and Schwartz (Ref. 7, pp. 660–62).

For any $\mathcal{T}$ induced by a T by Eq. (13), $\| \mathcal{T} \| \leq 1$. Thus $\| \overline{\mathcal{T}}_n \| \leq 1$ and $\mathcal{T}^n O/n \to 0$. In this case the weak sequential compactness of $\{\overline{\mathcal{T}}_n O\}$ is a sufficient condition of convergence for $\overline{\mathcal{T}}_n O$. Different convergence conditions for contraction operators are given by Sine.[15]

For any $\mathcal{T}$ induced by a T, $\overline{O}^{\mathcal{T}}$ has the following properties (provided $\lim_n \overline{\mathcal{T}}_n O$ exists):

$$\overline{O}^{\mathcal{T}}_{TE} = \overline{O}^{\mathcal{T}}_E \tag{17}$$

for each E in Σ. This is a result of $\mathcal{T} \overline{\mathcal{T}} O = \overline{\mathcal{T}} O$. For any F in Σ which is T invariant ($TF = F$),

$$\overline{O}^{\mathcal{T}}_F = O_F . \tag{18}$$

Thus O and $\overline{O}^{\mathcal{T}}$ coincide on the sub σ-field of T invariant sets in Σ. Finally one has

Theorem 6: Let $\mathcal{T}$ be induced by some measurable T which preserves set operations. Then if O is a probability operator measure and $\overline{O}^{\mathcal{T}}$ exists, $\overline{O}^{\mathcal{T}}$ is a probability operator measure.

Proof: By Eqs. (12), (15), (16), and the hypothesis of the theorem, one has that for each n and ψ in $\mathcal{H}$, $\mathcal{T}_n O_- \psi$ is a vector measure and $\lim_n \mathcal{T}_n O_E \psi = \overline{O}_E^{\mathcal{T}}$ exists for each E in Σ. By a result of Dunford and Schwartz (Ref. 7, Theorem 6, p. 321), $\overline{O}^{\mathcal{T}}\psi$ is a vector measure and thus $\overline{O}^{\mathcal{T}}$ is an operator measure. That $\overline{O}^{\mathcal{T}}$ is a probability operator measure follows from that property of O and Eqs. (4), (15), (16), and (18).

QED

This completes a brief survey of the properties of positive operator measures which are relevant for this work. Other properties can either be obtained from similar properties for vector valued measures[8,9] or can be obtained directly from the literature.[10,11,16]

III. COMPOUND PROCESSES

A. Finite Processes

We begin with some definitions—a finite process q^Δ of n steps consists of the following things: (1) a sequence Δ of $n-1$ times, (2) a unitary time translation operator $U(t,t')$, and (3) a sequence q of n observables. q^Δ corresponds to the process—measure $q(0)$ on an system at time $t=0$, observe outcome, measure $q(1)$ on the "same" system at Δ_1, observe outcome, . . . , and measure $q(n-1)$ on the "same" system at Δ_{n-1} and observe the outcome.

In this work, q is restricted to be a sequence of discrete observables. That is, for each j, $q(j)$ is required to be in $B(\mathcal{H})_D$, the set of all self-adjoint operators in $B(\mathcal{H})$ with discrete spectra. (The convention is followed here of identifying observables with the self-adjoint operators to which they correspond). "Coarse grained" operators which replace the continuous ones with a finite or infinite number of real line intervals and whose eigenvalues label the intervals are included here.

For each j let

$$q(j) = \sum_{s \in S_j} s P_s^{q_j}, \tag{19}$$

where S_j is the at most countably infinite eigenvalue set of $q(j)$ and $P_s^{q_j}$ is the projection operator corresponding to eigenvalue s of $q(j)$. Let $S^n = S_0 \times \cdots \times S_{n-1}$. S^n is the set of all possible outcome sequences ϕ_n of q^Δ. Let Σ^n be the set of all subsets of S^n. Clearly Σ^n is a σ-field.

As an intuitive basis for the construction, consider a two step process, where

$$q(0) = A_0 = \sum_{s \in S_0} s P_s^0 \quad \text{and} \quad q(1) = A_1 = \sum_{s' \in S_1} s' P_{s'}^1.$$

According to the Von Neumann projection postulate,[2] if one measures A_0 on a system in state ρ at time zero and observes outcome s_0, then state ρ becomes ρ_{s_0} with

$$\rho_{s_0} = \frac{P_{s_0}^0 \rho P_{s_0}^0}{\mathrm{Tr} P_{s_0} \rho}. \tag{20}$$

If one waits until time Δ_1 and measures the observable whose operator is

$$A_1 = \sum_{s' \in S_1} s' P_{s'}^1$$

and observes outcome s_1, the state ρ_{s_0} becomes $\rho_{s_0 s_1}$ with

$$\rho_{s_0 s_1} = \frac{P_{s_1}^1 U(\Delta_1, 0) \rho_{s_0} U^\dagger(\Delta_1, 0) P_{s_1}^1}{\mathrm{Tr}(P_{s_1}^1 U(\Delta_1, 0) \rho_{s_0} U^\dagger(\Delta_1, 0) P_{s_1}^1)}. \tag{21}$$

The denominator of Eq. (20) is the probability of observing outcome s_0 when A_0 is measured on ρ, and the denominator of Eq. (21) is the conditional probability for observing outcome s_1 when A_1 is measured after Δ_1, given that s_0 is observed on state ρ at time 0.

The unconditional probability of observing $\phi_2(0) = s_0$ at time 0 and $\phi_2(1) = s_1$ at a time Δ_1 later is given by

$$P_\rho\{\phi_2\} = \mathrm{Tr}(P_{s_1}^1 U(\Delta_1, 0) P_{s_0}^0 \rho P_{s_0}^0 U^\dagger(\Delta_1, 0) P_{s_1}^1). \tag{22}$$

In general, the unconditional probability that the outcome sequence ϕ_2 lies in some subset E of $S_0 \times S_1$ is just

$$P_\rho(E) = \sum_{\phi_2 \in E} P_\rho\{\phi_2\}. \tag{23}$$

A standard empirical meaning of these probabilities is that if one repeats over and over the process—prepare a system in state ρ, measure A_0, observe outcome, measure A_1 after a wait of Δ_1, observe outcome, discard system—then $P_\rho(E)$ is the limit relative frequency that a two outcome sequence in the infinite sequence of two outcome sequences so obtained will be found to lie in E. Equivalently, one can regard the state $\rho_{s_0 s_1}$ as that assigned to a selective preparation procedure with two filtering steps and the probability of Eq. (22) is the fraction of systems passed by the filter.

The construction process for an n step process q^Δ is as follows: For each eigenvalue sequence ϕ_n of length n define the operator $\beta_{\phi_n}^{q^\Delta}$ by

$$\beta_{\phi_n}^{q^\Delta} = P_{\phi_n(n-1)}^{q(n-1)} U(\Delta_{n-1}, \Delta_{n-2}) P_{\phi_n(n-2)}^{q(n-2)} \cdots P_{\phi_n(1)}^{q(1)} U(\Delta_1, 0) P_{\phi_n(0)}^{q(0)}. \tag{24}$$

For each singleton set $\{\phi_n\}$, define the operator $O_{\{\phi_n\}}$ by

$$O_{\{\phi_n\}} = \beta_{\phi_n}^{q^\Delta \dagger} \beta_{\phi_n}^{q^\Delta} \tag{25}$$

and for each $E \in \Sigma^n$ define O_E by

$$O_E = \sum_{\phi_n \in E} O_{\{\phi_n\}}, \tag{26}$$

where the implied limit is in the strong operator topology. Finally for Φ set $O_\Phi = \mathbf{0}$, the zero operator in $B(\mathcal{H})$.

It is clear that this construction associates to each E in Σ^n for which the limit implied by Eq. (26) exists, a positive self-adjoint operator O_E in $B(\mathcal{H})$. Clearly $O_{\{\phi_n\}}$ is self-adjoint and positive. The positivity follows from $0 \le \| \beta_{\phi_n} \psi \|^2 = (\beta_{\phi_n}\psi, \beta_{\phi_n}\psi) = (\psi, O_{\{\phi_n\}}\psi)$ for each ψ and ϕ_n.

In order to show that O_E exists for each $E \in \Sigma^n$, one needs O_{S^n}. From Eqs. (24)–(26) one has

$$O_{S^n} = \sum_{\phi_n(0) \in S_0} \cdots \sum_{\phi_n(n-1) \in S_{n-1}} O_{\{\phi_n\}} = 1. \tag{27}$$

To get this result the facts that $U^\dagger(\Delta_j, \Delta_{j-1})U(\Delta_j, \Delta_{j-1}) = \mathbf{1}$ and $\Sigma_{s\in S_j} P_s^{q_j} = \mathbf{1}$ for each $j < n$ have been used.

Let $\{E_j\}$ be some nondecreasing sequence of finite subsets of E such that $\lim_j E_j = E$. Consider the sequence $(\psi, O_{E_j}\psi)$ for any $\psi \in \mathcal{K}$ with O_{E_j} given by a finite sum over ϕ_n, Eq. (26). By Eq. (27), $O_{E_j} \leq 1$ for each j and thus $(\psi, O_{E_j}\psi)$ is a nondecreasing sequence bounded from above by (ψ,ψ). So $\lim_j (\psi, O_{E_j}\psi) = (\psi, O_E\psi)$ exists and, for each $\epsilon > 0$, $(\psi, (O_E - O_{E_m})\psi) < \epsilon^2$ for all $m >$ some N.

Now for any m, $0 \leq O_E - O_{E_m} \leq \mathbf{1}$. Thus $\| (O_E - O_{E_m})\psi\|^2 = (\psi, (O_E - O_{E_m})^2\psi) \leq (\psi, (O_E - O_{E_m})\psi) < \epsilon^2$ and $\|(O_E - O_{E_m})\psi\| < \epsilon$. So one has $O_E = s - \lim_j O_{E_j}$.

From the above, one has that O is a mapping with domain Σ^n and range in $B(\mathcal{K})^+$ with $O_\Phi = \mathbf{0}$ and $O_{S^n} = \mathbf{1}$. To see that O is a probability operator measure, let $\{E_m\}$ be a sequence of pairwise disjoint subsets of S^n with $E = \cup_m E_m$. Then Eq. (26) gives

$$O_E = \sum_{\phi_n \in E} O_{\{\phi_n\}} = \sum_m \sum_{\phi_n \in E_m} O_{\{\phi_n\}} = \sum_m O_{E_m} \tag{28}$$

or O is strongly σ-additive. [This follows from the fact that for each vector ψ and each ϕ_n in S^n $(\psi, O_{\{\phi_n\}}\psi) \geq 0$. Since the sum of a convergent series of nonnegative numbers is independent of the arrangement of the terms,

$$\sum_{\phi_n \in E} (\psi, O_{\{\phi_n\}}\psi) = \sum_m \sum_{\phi_n \in E_m} (\psi, O_{\{\phi_n\}}\psi).]$$

Thus, Eqs. (1), (2), and (10) are satisfied and O is a probability operator measure on Σ^n.

The above results show that the construction of Eqs. (24)–(27) associates to each process q^Δ a unique probability operator measure O (or O^{q^Δ}) which depends on the sequences q and Δ and, through the operators $U(\Delta_j, \Delta_{j-1})$, on the Hamiltonian H. For, given q, Δ, and H, the operators $\beta_{\phi_n}^{q^\Delta}$ are uniquely defined by Eq. (24) and thus Eqs. (25)–(27) define O_E uniquely for each $E \in \Sigma^n$.

This construction includes all Hamiltonians no matter what time variation they have over the time interval $\sum_{j=0}^{n-1} \Delta_j$. Thus all external fields are included. Note also that if q is fixed, then for each Δ and each H the construction assigns a unique O in the space $\Phi(\Sigma^n, B(\mathcal{K}))$.

Besides depending on q, Δ, and H, O also depends on the projection axiom.[2] This was used to write down Eq. (24). As is well known,[17] there are measurement procedures for which the axiom is invalid. If one assumes that for each observable there is a procedure which satisfies the axiom, then this problem can be avoided. The treatment of the more general case in which Von Neumann's projection axiom is not valid, will be postponed to future work.

Finally, it should be noted that this construction includes processes which generate finite sequences of independent uncorrelated single measurements. To see this let $\mathcal{K}^n = \otimes_{i=0}^{n-1} \mathcal{K}_1$ be the tensor product of n copies of $\mathcal{K}$ and $B(\mathcal{K}^n)$ the algebra of bounded linear operators on $\mathcal{K}^n$. Let the process q be such that for each $j \leq n-1$, $q(j) = \mathbf{1}_0 \times \cdots \times \mathbf{1}_{j-1} \times A_j \times \mathbf{1}_{j+1} \times \cdots \times \mathbf{1}_{n-1}$, where $A_j \in B(\mathcal{K}_j)_D$. Let the time translation operator $U(0, \ldots, n-1, t, t')$ be such that $U(0, \ldots, n-1, t, t') = \otimes_{j=0}^{n-1} U_j(t, t')$. The use of these definitions in Eqs. (24) and (25) gives the results that $\beta_{\phi_n}^{q^\Delta} = \otimes_{j=0}^{n-1} P_{\phi_n(j)}^{q_j} U_j(\Delta_j, 0)$ and $O_{\{\phi_n\}} = \otimes_{j=0}^{n-1} U_j^\dagger(\Delta_j, 0) P_{\phi_n(j)}^{q_j} U_j(\Delta_j, 0)$. From these results and Eq. (26), it is clear that the probability operator measure O associated with this q^Δ has domain Σ^n and range in $B(\mathcal{K}^n)$.

Now let the state $\rho(0, \ldots, n-1)$ be given by $\rho(0, \ldots, n-1) = \otimes_{j=0}^{n-1} \rho_j$, where ρ_j is a state over $\mathcal{K}_j$. In this case the scalar probability measure $P_{O\rho}$, which describes the sequence of measurements is a product measure as $P_{O\rho}\{\phi_n\} = \mathrm{Tr}\rho O_{\{\phi_n\}} = \Pi_{j=0}^{n-1} \mathrm{Tr}(\rho_j U_j^t(\Delta_j, 0) P_{\phi_n(j)}^{q_j} U_j(\Delta_j, 0))$. Thus in this case carrying out q^Δ on a system in slate ρ is equivalent statistically to a sequence of n-independent uncorrelated measurements in which the jth consists of measuring $U_j^t(\Delta_j, 0) A_j U(\Delta_j, 0)$ on a system in state ρ_j.

B. Infinite Processes

For an infinite process, q now becomes an infinite sequence of observables or self-adjoint operators in $B(\mathcal{K})_D$ and Δ becomes an infinite sequence of time intervals. Again we make the simplifying restriction that for each j, $q(j)$ has only a discrete point spectrum S_j; operators with continuous spectra will be treated in future work.

The process q^Δ here corresponds to the extension of the finite case. That is "measure $q(0)$ on a system at $t = 0$, observe outcome, measure $q(1)$ at Δ_1 on the same system, observe outcome, ..." In general such a process takes an infinite amount of time to carry out, unless, of course, $\Sigma_n \Delta_n$ is finite.

Let $S^\omega = S_0 \times S_1 \times \cdots$ with ϕ denoting an element of S^ω. Clearly S^ω is the set of all possible outcome sequences for q^Δ. Let $\mathfrak{F}$ denote the field of all cylinder subsets of S^ω and Σ^ω the minimal σ-field over $\mathfrak{F}$. Since the eigenvalues of observables are real numbers and each S_j is countable, one has $S^\omega \subset R^\omega$ and $\Sigma^\omega \subset \mathfrak{B}(R^\omega)$. The latter follows from the facts that $\mathfrak{F} \subset \mathfrak{F}^R$, the field of all Borel cylinder subsets of R^ω and $\mathfrak{B}(R^\omega)$ is the minimal σ-field over $\mathfrak{F}^R$.

The probability operator measure O for the process q^Δ is obtained as follows: One first constructs for each n a probability operator measure O^n on Σ^n given by

$$O_E^n = \sum_{\phi_n \in E} O_{\{\phi_n\}} \tag{29}$$

for each $E \in \Sigma^n$. Here $\phi_n \in S^n$, the Cartesian product of the first n elements of S^ω and $O_{\{\phi_n\}}$, is given by Eq. (25), where the operator $\beta_{\phi_n}^{q^\Delta}$ is constructed from the first n steps of the process q^Δ by the prescription of Eq. (24).

In this manner, one generates from q^Δ a sequence $\{O^n\}^{q^\Delta}$ of probability operator measured defined on the sequence $\{\Sigma^n\}$ of σ-fields. One must show that $\{O^n\}^{q^\Delta}$ is consistent. To this end, let $n > m$ and let $E \subset S^n$ and $F \subset S^m$ be such that

$$E = F \times S_m \times \cdots \times S_{n-1}.$$

One must show that $O_E^n = O_F^m$. But this follows from the facts that $U^\dagger U = \mathbf{1}, \sum_{s\in S_j} P_s^{q(j)} = \mathbf{1}$ for each j and $P^2 = P$ as

$$\begin{aligned} O_E^n = &\sum_{\phi_m \in F} \sum_{s_m \in S_m} \cdots \sum_{s_{n-1}\in S_{n-1}} \beta_{\phi_m}^{q^\Delta\dagger} U^\dagger(\Delta_m, \Delta_{m-1}) \\ &\times P_{s_m}^{q(m)} U^\dagger(\Delta_{m+1,}\Delta_m) \cdots U^\dagger(\Delta_{n-1,n-2}) P_{s_{n-1}}^{q(n-1)} \\ &\times P_{s_{n-1}}^{q(n-1)} U(\Delta_{n-1,}\Delta_{n-2}) \cdots P_{s_m}^{q(m)} U(\Delta_{m,}\Delta_{m-1}) \beta_{\phi_m}^{q^\Delta} \\ = &\, O_F^m. \end{aligned}$$

Thus $\{O^n\}^{q^\Delta}$ is consistent.

The construction of O on Σ^ω from the sequence $\{O^n\}^{q^\Delta}$ proceeds as follows: For each n, one constructs a probability operator measure O'^n on $\mathfrak{B}(R^n)$ from O^n on Σ^n according to

$$O_B'^n = O_{E_B}^n, \tag{30}$$

for each n and each Borel set $B \in \mathfrak{B}(R^n)$ and where $E_B = B \cap S^n$. Equivalently $O_B'^n = \Sigma O_{\{\phi_n\}}^n$, where the sum is over all sequences in S^n which lie in B. It is clear from this construction that O'^n is well defined. For if B_1 and B_2 are Borel sets in $\mathfrak{B}(R^n)$ with $B_1 \equiv B_2 \bmod S^n$, then Eq. (30) gives $O_{B_1}'^n = O_{B_2}'^n$. Furthermore, since $\Sigma^n \subset \mathfrak{B}(R^n)$, for each $E \in \Sigma^n$, $O_E'^n = O_E^n$. From this argument it is also clear that $\{O'^n\}$ is a consistent sequence of probability operator measures on $\{\mathfrak{B}(R^n)\}$.

Theorem 4 now gives the result that there exists a unique probability operator measure O' on $\mathfrak{B}(R^\omega)$ such that for each cylinder set $G \in \mathfrak{F}^R$ with base $B \in \mathfrak{B}(R^n)$,

$$O_G' = O_B'^n. \tag{31}$$

Since Σ^ω is a sub σ-field of (R^ω), one defines O as the restriction of O' to Σ^ω. That is, O is given by

$$O_E = O_E' \tag{32}$$

for each $E \in \Sigma^\omega$.

Thus by the above construction, there is associated with each infinite process q^Δ a unique probability operator measure O. The uniqueness follows from the uniqueness of the O'^n Eq. (30), and theorem 4. That O is a probability operator measure follows from Eqs. (30)–(32), $S^\omega \in \Sigma^\omega$, and $\Sigma^\omega \subset (R^\omega)$.

It is of interest to consider the singleton sets $\{\phi\}$. Since for each j, S_j is countable, $\{\phi\} \in \Sigma^\omega$ for each $\phi \in \Sigma^\omega$, so O is defined on each $\{\phi\}$. In fact one can define $O_{\{\phi\}}$ by

$$O_{\{\phi\}} = \text{s-lim}_n O_{\{\phi_n\}} = \text{s-lim}_n \beta_{\phi_n}^{q^\Delta\dagger} \beta_{\phi_n}^{q^\Delta}$$

with $\beta_{\phi_n}^{q^\Delta}$ given by Eq. (24) and ϕ_n the initial segment of ϕ of length n.

Clearly the right-hand limit exists and equals $O_{\{\phi\}}$. To see this, note that $\{\phi_n\}$ is the base for the cylinder set $F_n = \cap_{j=0}^{n-1} E_{\phi(j),j}$, where $E_{\phi(j),j} = [\theta \mid \theta(j) = \phi(j)]$. Since $F_0, F_1 \cdots$ is a nonincreasing sequence of sets with $\lim_n F_n = \{\phi\}$, Theorem 1 gives the desired result. By Eqs. (2) and (4) one has that $\mathbf{0} \le O_{\{\phi\}} \le O_{F_n}$ for each n.

A common type of infinite process is one for which the associated (scalar) probability measure P is nontrivial. That is $P\{\phi\} = 0$ for each $\phi \in S^\omega$. This corresponds here to the situation in which q^Δ and ρ are such that $\text{Tr}\rho O_{\{\phi\}}^{q^\Delta} = 0$ for each $\phi \in S^\omega$.

Examples of this type include infinite sequences of uncorrelated independent single measurements. It appears that these can be included in the construction given here by extending the corresponding construction of the finite case to the (nonseparable) Hilbert space $\mathcal{H}^\omega$, which is the infinite tensor product $\otimes_{i=1}^\infty \mathcal{H}_i$ of copies of $\mathcal{H}$.[18] However we will not go into this here.

As in the finite case, O^{q^Δ} depends on q, Δ, and the Hamiltonian H. In this case, however for processes which last forever, H can have an arbitrary time-dependence for all of calendar time. That is, the time-dependence of H need not be cyclic or repetitive. This is in contrast to the finite case in that if one wishes to consider an infinite number of repetitions of a finite process then, with respect to calendar time, the time variation of H must be repetitive.[19]

IV. CONSEQUENCE OF THE ASSOCIATION OF O^{q^Δ} TO q^Δ

A. Higher-Order Equivalence

A very interesting equivalence property arises from the constructions of the previous section. We shall consider infinite ($\delta = \omega$) and finite ($\delta =$ some n) processes together. Suppose one considers carrying out a δ step process q^Δ on a system in state ρ and wants to know what the probability is that the outcome sequence lies in some set E in Σ^δ. By the previous section this probability is given by $\text{Tr}\rho O_E$.

Now the standard method of determining this probability is to carry out an infinite sequence of repetitions[20] of q^Δ on a system in state ρ. (In this paper, an infinite repetition of measuring an observable A on ρ or carrying out q^Δ on ρ means the infinite repetition of the following: prepare a system in state ρ; measure A and observe outcome, or carry out q^Δ and obtain an outcome sequence; discard system.) The result is an infinite sequence α of sequences in S^δ. That is, $\alpha(j)$ is a sequence in S^δ for each j. Then the limit relative frequency that each of the sequences $\alpha(0), \alpha(1), \cdots$ is in E is given by $\text{Tr}\rho O_E$. That is, generate an infinite sequence $\theta_E\alpha$ of 0's and 1's from α by the prescription $(\theta_E\alpha)(j) = 1$ if $\alpha(j)$ is in E and $(\theta_E\alpha)(j) = 0$ if $\alpha(j)$ is not in E. Then $\bar{M}\theta_E\alpha = \text{Tr}\rho O_E$ where $\bar{M}$ denotes the limit mean. Note that in the simplest case of $\delta = 1$, O^{q^Δ} reduces to the spectral measure for $q(0)$.

The remarkable point is that there is another equivalent method of determining the value of $\text{Tr}\rho O_E$. This is, carry out an infinite repetition of measuring O_E on a system in state ρ. Then the limit mean of the resulting infinite sequence β of outcomes equals $\text{Tr}\rho O_E$ or $\bar{M}\beta = \text{Tr}\rho O_E$.

Before proving that these two methods are equivalent, one should note that the latter method implicitly assumes that to each self-adjoint operator in $B(\mathcal{H})$ there corresponds a procedure for measuring it.[5] This assumption has been criticized on the grounds that it is negated by the existence of superselection

rules[22] and that no one knows how to measure most self-adjoint operators.[23] As regards the first criticism, it is to be noted that the existence or nonexistence of superselection rules is not related to this correspondence assumption. For under the 1-to-1 correspondence between observables and self-adjoint operators, superselection rules are statements about the structure of $B(\mathcal{H})$ and $\mathcal{H}$.

Without denying the second criticism, we would only point out that the correspondence statement is an existence statement only. It by no means implies that there exists an effective procedure for discovering which measurement procedure is associated to any given operator.

It should also be noted that this equivalence is the extension to sequences of well-known properties of operators in Quantum Mechanics. To see this consider the case for $n = 1$. Then O^q is given by $O_E^q = \sum_{s\in E} P_s^{q(0)}$ for each $E \subset S_0$ and is equivalent to the spectral measure for $q(0) = \sum_{s\in S_0} sP_s^{q(0)}$. In this case the first method becomes "measures $q(0)$ on ρ over and over again to give an infinite outcome sequence α in $S_0^\omega = S_0 \times S_0 \times \cdots$ and generate the 0–1 sequence $\theta_E(\alpha)$ by answering—Is $\alpha(j) \in E$?—for each j." The second method is "measure O_E on ρ over and over again generating a 0–1 outcome sequence β'." Here these methods are closely related because both β and $\theta_E(\alpha)$ are 0–1 sequences and all the O_E^q are mutually commuting projections which also commute with $q(0)$.

Another well-known case concerns the sum of two observables A and B. As is well known, there are two ways to measure $\mathrm{Tr}\rho(A + B)$ which follow from the linearity of $B(\mathcal{H})$. One is to measure A over and over again on ρ giving an outcome sequence α_A and then to measure B over and over again on ρ giving α_B. Then $\overline{M}\alpha_A + \overline{M}\alpha_B = \overline{M}(\alpha_A + \alpha_B) = \mathrm{Tr}\rho(A + B)$. The other method is to measure $A + B$ over and over again on ρ which gives an outcome sequence β. Then $\overline{M}\beta = \mathrm{Tr}\rho(A + B)$. Of course if A and B do not commute, the spectrum of $A + B$ need have no simple relation to that of A and of B. Similar arguments apply to $f(A)$ where f is any Borel function of the real numbers.

At the risk of belaboring the obvious and to point out an important but not so obvious aspect of such proofs we prove that the two methods are equivalent. The proof consists in constructing for each method an appropriate sample space, finding the correct random function which represents the possible outcomes of a single measurement, and then showing that the limit mean random function is almost everywhere equal to $\mathrm{Tr}\rho O_E$. An entirely similar proof holds for the better known cases of $A + B$ and $f(A)$ for observables A and B.

The sample space for the standard method of determining $\mathrm{Tr}\rho O_E$ is $((S^\delta)^\omega, (\Sigma^\delta)^\omega, \tilde{P}_{O\rho})$, where $(S^\delta)^\omega$ is the set of all infinite sequences α of sequences in S^δ, $(\Sigma^\delta)^\omega$ is the minimal σ-field over the set of all subsets of $(S^\delta)^\omega$ of the form $F_{El} = [\alpha \mid \alpha(l) \in E]$, where $l = 0, 1, \cdots$ and $E \in \Sigma^\delta$. $\tilde{P}_{O\rho}$ is the product probability measure on $(\Sigma^\delta)^\omega$ defined by

$$\tilde{P}_{O\rho}F_{El} = \mathrm{Tr}\rho O_E \tag{33}$$

for each l and each $E \in \Sigma^\delta$ and O is the probability operator measure associated with the process q^Δ. The measure defined by Eq. (33) on the class of sets of the form F_{El} extends uniquely[24] to a product measure on $(\Sigma^\delta)^\omega$.

Let $T_1 : (S^\delta)^\omega \to (S^\delta)^\omega$ be the one-sided shift operator defined by

$$(T_1\alpha)(j) = \alpha(j + 1) \tag{34}$$

for each j and each $\alpha \in (S^\delta)^\omega$. In a standard fashion, T_1 induces a measurable transformation $T_1' : (\Sigma^\delta)^\omega \to (\Sigma^\delta)^\omega$ defined by

$$T_1'E = [\alpha \mid T_1\alpha \in E]. \tag{35}$$

Clearly T_1' preserves all the set operations.

Let $L((S^\delta)^\omega(\Sigma^\delta)^\omega)$ be the set of all Borel functions $f : (S^\delta)^\omega \to R$. T_1' induces a transformation $\mathcal{T}_1$ on L defined on the characteristic functions by

$$\mathcal{T}_1 I_E = I_{T_1'E} \tag{36}$$

for each $E \in (\Sigma^\delta)^\omega$. Finally for each $f \in L((S^\delta)^\omega, (\Sigma^\delta)^\omega)$ define $\overline{f}_m$ by

$$\overline{f}_m = \frac{1}{m}\sum_{j=0}^{m-1} \mathcal{T}_1^j f \tag{37}$$

and $\overline{f}$ by

$$\overline{f} = \lim_m \overline{f}_m \tag{38}$$

if the limit exists.

For the one-sided shift operator and the product probability measure of Eq. (33), T_1' is $P_{O\rho}$ measure preserving and the pointwise ergodic theorem holds,[25] so $\overline{f}$ exists, $\tilde{P}_{O\rho}$ almost everywhere for each f in L. Furthermore, $\mathcal{T}_1$ is $P_{O\rho}$ indecomposable[25] and thus

$$\overline{f}(-) = \int_{(S^\delta)^\omega} f(\alpha)\tilde{P}_{O\rho}(d\alpha), \tag{39}$$

$\tilde{P}_{O\rho}$ almost everywhere for each f in L.

Now for each E in Σ^δ define f_E by

$$f_E(\alpha) = I_{F_{E,0}}(\alpha), \tag{40}$$

where $F_{E,0} = [\alpha \mid \alpha(0) \in E]$. Clearly $f_E(\alpha) = 1(0)$ if $\alpha(0) \in E$ ($\alpha(0) \notin E$). From Eqs. (34)–(36), $\mathcal{T}_1^j f_E = I_{F_{E,j}}$ and thus $\overline{f}_E(\alpha)$ is the limit relative frequency for finding $\alpha(0), \alpha(1), \ldots$ in E. Equations (33), (39), and (40) give the result that

$$\overline{M}\theta_E(-) = \overline{f}_E(-) = \tilde{P}_{O\rho}F_{E,0} = \mathrm{Tr}\rho O_E, \tag{41}$$

$\tilde{P}_{O\rho}$ almost everywhere for each $E \in \Sigma^\delta$.

For the infinite repetition of the measurement of O_E on a system in state ρ, the relevant sample space is $(R^\omega, \mathcal{B}(R^\omega), \tilde{P}_{O_E\rho})$ where R^ω is the set of all infinite sequences β of real numbers. $\mathcal{B}(R^\omega)$ is the set of all Borel subsets of R^ω. As in the case of $(\Sigma^\delta)^\omega$, it is the minimal σ-field over the set of all $F_{Bl} = [\beta \mid \beta(l) \in B]$ for $l = 0, 1, \cdots$ and B a Borel subset of R. $\tilde{P}_{O_E\rho}$ is the product probability measure on $\mathcal{B}(R^\omega)$ given by

$$\tilde{P}_{O_E\rho}F_{Bl} = P_{O_E\rho}B = \int_B d(\mathrm{Tr}(\rho\mathcal{E}_r^{O_E})) \tag{42}$$

for each l and Borel subset B of R. Here $\mathcal{E}^{O_E}$ is the spectral measure[26] of the operator O_E and $\mathcal{E}_r^{O_E} = \mathcal{E}^{O_E}((-\infty, r])$, where $(-\infty, r]$ is the set of all real numbers $\leq r$. As before $\tilde{P}_{O_E\rho}$ defined as in Eq. (42) from a probability measure on $\mathfrak{B}(R)$ extends uniquely[24] to a product probability measure on $\mathfrak{B}(R^\omega)$.

As before T_1 is the one-sided shift operator on R^ω with $(T_1\beta)(j) = \beta(j+1)$. As in Eqs. (35) and (36), T_1 induces a T_1' on $\mathfrak{B}(R^\omega)$ and a $\mathcal{T}$ on $L(R^\omega, \mathfrak{B}(R^\omega))$, with $\overline{g}_m$ and $\overline{g}$ defined as in Eqs. (37) and (38) for each g in L. Since $P_{O_E\rho}$ is T_1 invariant and a product probability measure on $\mathfrak{B}(R^\omega)$, the pointwise ergodic theorem and the ergodic hypothesis[25] give that $\overline{g}$ exists and

$$\overline{g}(-) = \int_{R^\omega} g(\beta)\tilde{P}_{O_E\rho}(d\beta), \tag{43}$$

$\tilde{P}_{O_E\rho}$ almost everywhere for each g in L.

Since the spectrum of O_E lies in the interval $[0,1]$, the measure $\tilde{P}_{O_E\rho}$ is concentrated on the subset $[0,1]^\omega$ of R^ω, and $\tilde{P}_{O_E\rho}(R^\omega - [0,1]^\omega) = 0$ for all ρ.

Now a measurement of O_E on any system gives a number in $[0,1]$. So the random variable g in L describing the first measurement is given by

$$g(\beta) = \lim_m \sum_{j=0}^{m} \frac{j}{m} I_{F_{[j/m,(j+1)/m),0}}(\beta) \tag{44}$$

for each β. Here I_E is the characteristic function for the set E and $F_{[j/m,j+1/m),0} = [\beta | j/m \leq \beta(0) < (j+1)/m]$. Note that if $0 < \beta(0) < 1$, $g(\beta) = \beta(0)$ and that the value of $g(\beta)$ for $\beta(0) > 1$ or < 0 is irrelevant since such β are always in $R^\omega - [0,1]^\omega$.

Now $\mathcal{T}^k g(\beta)[= \beta(k)$ if $0 < \beta(k) < 1]$ is the random variable describing the kth measurement. The limit mean of the outcome sequence of an infinite repetition of measurements of O_E on a system in state ρ is described in L by $\overline{g}(-)$. Equations (42)–(44) give[27]

$$\begin{aligned}\overline{M}(-) = \overline{g}(-) &= \lim_m \sum_{j=0}^{m} \frac{j}{m} \tilde{P}_{O_E\rho} F_{[j/m,(j+1)/m),0} \\ &= \int_{0-}^{1} r\,d(\mathrm{Tr}(\rho \mathcal{E}_r^{O_E})) = \mathrm{Tr}\rho O_E,\end{aligned} \tag{45}$$

$\tilde{P}_{O_E\rho}$ almost everywhere.

From Eqs. (41) and (45), one has the result that $\overline{M}\theta_E(-) = \mathrm{Tr}\rho O_E, \tilde{P}_{O\rho}$ almost everywhere and $\overline{M}(-) = g(-) = \mathrm{Tr}\rho O_E, \tilde{P}_{O_E\rho}$ almost everywhere.

Now the usual procedure in proofs of this type is to conclude from this that

$$\overline{M}\theta_E(\alpha) = \mathrm{Tr}\rho O_E = M\beta, \tag{46}$$

where α and β are the sequences obtained by actually carrying out the respective processes. We want to stress that, *in common with all comparisons between an empirical limit mean and an expectation value*, this step implies the assumption that a (scalar) measure P assigned to a process is "correct" for the process. That is, "all" properties (with "all" suitably defined)[21,28] of outcome sequences which are true P almost everywhere, are true for the sequence obtained by actually doing the process. For the cases considered here, one assumes that $\tilde{P}_{O\rho}$ is "correct" for an infinite repetition of q^Δ on ρ and $P_{O_E\rho}$ is "correct" for an infinite repetition of measurements of O_E on ρ.

The not so obvious aspect of such proofs is that the assumption as well as the proper definition of "correctness" is highly nontrivial. It is the central concept of a definition of agreement between theory and experiment discussed elsewhere[21,28] and will be discussed more in future work. However, from here on, the usual procedure will be followed of finishing proofs of this type without comment by assuming implicitly that the given (scalar) measure is "correct" for the process being considered.

This equivalence sheds new light on the meaning of O_E. For it shows that by means of the O associated with the process q^Δ, questions about the statistical properties of q^Δ, which are usually formulated in the probability sample space of infinite repetitions of the process, can be reduced to (conceptually) simpler questions about $\mathrm{Tr}\rho O_E$ and the spectrum of O_E for each $E \in \Sigma^\delta$ and each state ρ.

This reduction process appears more powerful for infinite processes than for finite ones. For it allows questions about "higher order" sequences, i.e., infinite sequences of δ step processes to be reduced to questions about single infinite processes each of which is an infinite repetition of measurements of O_E for some E. Thus this reduction appears to be a powerful tool for it shows that considerations of "first order processes" seem to be sufficient.

It might be objected that in order to set up this equivalence one had to consider infinite repetitions of δ step processes which for $\delta = \omega$ are impossible to carry out. However, we now show that the equivalence does not really depend on this by considering a lower, or "first order" equivalence, in which a single carrying out of q^Δ suffices.

B. First Order Equivalence

Let q^Δ be a δ step process and let $T: S^\delta \to S^\delta$ be a Σ^δ measurable transformation on S^δ such that the set transformation T'[Eq. (35)] induced by T preserves set operations. Furthermore, we require that q^Δ and T be such that the limit mean probability operator measure $\overline{O}^{\mathcal{T}}$ exists [Eq. (16)], where $\mathcal{T}$ is induced by T' on $\theta(\Sigma^\delta, B(\mathcal{K}))$ and O is the probability operator measure associated with the process q^Δ by the construction of Sec. III. Also the state ρ must be such that for the probability measure $P_{O\rho}$ constructed from O and ρ according to Eq. (11), the only T' invariant sets are Φ and $S^\delta, P_{O\rho}$ almost surely.

Now suppose one considers carrying out q^Δ on a system in state ρ and wants to know what the limit relative frequency is for finding $\alpha, T\alpha, T^2\alpha$, etc., in some set $E \in \Sigma^\delta$, where α denotes a possible outcome sequence. Clearly one way to find this out is to carry out q^Δ on a system in state ρ *only once*. Then one must go through the purely mathematical construction of generating the sequence $\alpha, T\alpha, T^2\alpha, \cdots$ and then construct the 0–1 sequence $\theta_E^T\alpha$ according to $(\theta_E^T\alpha)(j) = 1(0)$ if $T^j\alpha \in E (T^j\alpha \notin E)$ for $j = 0, 1, \cdots$. Then $\overline{M}\theta_E^T\alpha$ gives the desired limit relative frequency.

Again the interesting aspect of the probability operator measure O is that under the restrictions given, this limit relative frequency can also be obtained by repeating over and over a measurement of $\overline{O}_E^{\mathcal{T}}$ on ρ. As before, this equivalence is shown by means of sample space constructions. For the first process,

let I_E be the characteristic function for the set $E \in \Sigma^\delta$ where I_E is defined on the sample space $(S^\delta, \Sigma^\delta, P_{O\rho})$. Let τ be induced on the I_E by T' according to $\tau I_E = I_{T'E}$ for each $E \in \Sigma^\delta$ and define the limit mean random variable $\bar{I}_E^\tau$ by

$$\bar{I}_E^\tau(-) = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \tau^j I_E(-). \tag{47}$$

Now since $\overline{O}^\tau$ exists, Eqs. (11)–(16) give

$$P_{\bar{O}\rho}^\tau E = \mathrm{Tr}\rho\overline{O}_E^\tau = \lim_n \frac{1}{n} \sum_{j=0}^{n-1} \mathrm{Tr}\rho O_{T'^j E}$$

$$= \lim_n \frac{1}{n} \sum_{j=0}^{n-1} P_{O\rho}\, T'^j E = \overline{P}_{O\rho}^{T'} E$$

exists for each $E \in \Sigma^\delta$. Thus by the pointwise ergodic theorem and ergodic criterion,[25] $\bar{I}_E^\tau(-)$ exists $P_{O\rho}$ almost everywhere. Since Φ and S^δ are $P_{O\rho}$ almost surely the only T' invariant sets, the ergodic hypotheses holds, and thus, by the indecomposability theorem,[25]

$$\bar{I}_E^\tau(-) = \overline{P}_{O\rho}^{T'} E = \mathrm{Tr}\rho\overline{O}_E^\tau, \tag{48}$$

$P_{O\rho}$ almost everywhere. Clearly by Eq. (47), $\bar{M}\theta_E^T\alpha = \bar{I}_E^\tau(\alpha)$ for each α and E.

For the second process of repeating over and over the measurement of $\overline{O}_E^\tau$ on ρ, the sample space construction of Eqs. (42)–(45) can be taken over exactly if one replaces O_E everywhere by $\overline{O}_E^\tau$. Equations (42)–(45) then yield

$$\bar{M}(-) - \bar{g}(-) = \mathrm{Tr}\rho\overline{O}_E^\tau, \tag{49}$$

$P_{\bar{O}_E^\tau \rho}$ almost everywhere.

As in the case of Eq. (46), Eqs. (48) and (49) give the result that

$$\bar{M}\theta_E^T\alpha = \mathrm{Tr}\rho\overline{O}_E^\tau = \bar{M}\beta, \tag{50}$$

where α and β are the outcome sequences which would actually be obtained from the respective two processes.

V. DISCUSSION

There are some other properties of the association of probability operator measures to process q^Δ which are worth noting.

Possibly the most interesting property of these measures is that through the equivalences just discussed, they provide an interesting link between *purely mathematical properties* and *physical operations*. This stems from the circumstance that to each set $E \in \Sigma^\delta$ there corresponds a unique mathematical property Q of elements of S^δ, where $E = [\phi \mid Q(\phi)\ \mathrm{true}]$, i.e., E is the set of all elements ϕ of S^δ for which $Q(\phi)$ is true or which have property Q.

To see this in more detail suppose an observer contemplates carrying out the (finite or infinite) process q^Δ on a system in state ρ and wants to know what the probability is that the resulting outcome sequence has the *mathematical property* Q. Previous discussion shows that the standard way to answer this is to carry out an infinite sequence of repetitions of q^Δ on ρ which gives a sequence α in $(S^\delta)^\omega$. Now one must carry out an infinite sequence of decision procedures about the mathematical property Q. That is, one must determine the truth or falseness of the mathematical propositions $Q(\alpha(0))$, $Q(\alpha(1))$, $\cdots$. This generation of the 0–1 sequence $\theta_E(\alpha)$ is a a *purely mathematical procedure*. The desired probability that the outcome sequence has property Q is then given by $\bar{M}\theta_E(\alpha)$ where $E = [\phi \mid Q(\phi)\ \mathrm{true}]$.

Now the discussion of the previous section has shown that there is an entirely equivalent method of determining the probability that the outcome sequence obtained from doing q^Δ on ρ has property Q. This is, carry out the *physical operation* of repeatedly measuring $O_E^{q^\Delta}$ on ρ and determine $\bar{M}\beta$ where β is the outcome sequence. By the previous section, $\bar{M}\beta = \mathrm{Tr}\rho O_E^{q^\Delta}$ where $E = [\phi \mid Q(\phi)\ \mathrm{true}]$.

Thus one sees that the purely mathematical operation of "decide whether $Q(\alpha(0))$, $Q(\alpha(1))$, $\cdots$ are true or not" in effect, "disappears" into the operator $O_E^{q^\Delta}$. For the physical operation of "repeating" q^Δ on ρ over and over followed by the mathematical operation of generating θ_E for $E = [\phi \mid Q(\phi)\ \mathrm{true}]$ is equivalent to the physical operation of measuring $O_E^{q^\Delta}$ on ρ over and over again. Here equivalence means that $\bar{M}\theta_E\alpha = \bar{M}\beta = \mathrm{Tr}\rho O_E$.

This reduction property appears even more dramatically in the case of the "first order" equivalence. For in this case, the physical operation of repeating q^Δ on ρ over and over is replaced by carrying out q^Δ on ρ only once and then generating, as a mathematical operation, the sequence $\alpha, T\alpha, T^2\alpha, \ldots, T^j\alpha, \ldots$. Here the complete mathematical procedure of generating this sequence and then deciding for each j if $Q(T^j\alpha)$ is true or not "disappears" into the operator $\overline{O}_E^\tau$ where $E = [\phi \mid Q(\phi)\ \mathrm{true}]$. That is, carrying out q^Δ on ρ only once followed by the *mathematical procedure* outlined above is equivalent to the *physical operation* of measuring $\overline{O}_E^\tau$ on ρ over and over infinitely many times.

Again this reduction property is an extension to sequences of properties of operators in quantum mechanics which, although well known, perhaps has not been expressed in the same way as is done here. For the case $n = 1$, let α be the outcome sequence resulting from measuring $q(0) = \sum_{s \in S_0} sP_s^{q(0)}$ over and over again on ρ. Then the mathematical decision procedure whereby one answers—Is $Q(\alpha(j))$ true?—for each j where $E = [s \mid Q(s)\ \mathrm{true}] \subset S_0$ disappears into the physical operation of measuring $O_E = \sum_{s \in E} P_s^{q(0)}$ on ρ over and over again. As noted before in this case, each O_E commutes with $q(0)$.

For the case of two self-adjoint operators A and B, a somewhat different reduction holds. If one wants to measure $\mathrm{Tr}\rho(A + B)$ by measuring A and B each on ρ over and over again giving α_A and α_B, respectively, then there is a mathematical operation which must be performed. That is, one must either add $\bar{M}\alpha_A$ to $\bar{M}\alpha_B$ or α_A to α_B. By the equivalence discussed in the previous section, this addition operation disappears into the physical operation of measuring $A + B$ on ρ over and over again.

This argument also applies to the operations of multiplication by real numbers and computing $f(r)$ where f is any Borel function of the real numbers. One way to measure $f(A)$ on a system in state ρ is to measure

A on ρ over and over again and then compute $f'\alpha$ [where $(f'\alpha)(j) = (f(\alpha(j))$ for each j] from the outcome sequence α. The other way is to measure $f(A)$ on ρ over and over again giving the outcome sequence β. Since $\overline{M}f'\alpha = \mathrm{Tr}\rho f(A) = \overline{M}\beta$ the mathematical operation of computing $f'\alpha$ disappears into the physical operation of measuring $f(A)$ on ρ over and over again.

Two examples will illustrate these properties of probability operator measures associated with compound processes. As a simple example, let Q be the property $\phi(3) = s$ with s some fixed element of S_3. That is, $Q(\phi) \equiv \phi(3) = s$. Set $E = [\phi \mid Q(\phi)\,\mathrm{true}] = [\phi \mid \phi(3) = s]$.

Now the probability that carrying out q^Δ on a system in state ρ will yield an outcome sequence with property Q can be obtained in two ways. The method of repeating q^Δ on ρ over and over yields an infinite sequence $\alpha \in (S^\delta)^\omega$ for which one must carry out the mathematical decision procedures for "$\mathrm{Is}(\alpha(0))(3) = s\,?$," "$\mathrm{Is}(\alpha(1))(3) = s\,?$" etc. to obtain $\theta_E\alpha$. The other method consists of measuring $O_E^{q^\Delta}$ on ρ over and over again to obtain $\beta \in [0,1]^\omega$. From previous considerations one must have $\overline{M}\theta_E\alpha = \overline{M}\beta = \mathrm{Tr}\rho O_E^{q^\Delta}$. The mathematical procedure used to construct $\theta_E\alpha$ from α has "disappeared" into $O^{q^\Delta}_{[\phi \mid \phi(3)=s]}$.

This example transfers to the "first order" equivalence case as follows. Let $T: S^\delta \to S^\delta$ be Σ^δ measurable. As examples of such T let the δ step process q be such that $S_j = S$ independent of j for each $j < \delta$. Then T can be induced by any bijection $V: S \to S$ by $(T\phi)(j) = V(\phi(j))$ for each j. Or if $\delta = \omega$, T can be the one sided shift operator. For the first method one carries out q^Δ on a system in state ρ only once and generates the sequence $\alpha, T\alpha, T^2\alpha, \cdots$ from the outcome sequence α so obtained. Then one obtains $\theta_E^T\alpha$ by carrying out the mathematical decision procedures for "$Is\ \alpha(3) = s\,?$," "$Is\ (T\alpha)(3) = s\,?$," etc. The equivalent physical method consists in measuring $\overline{O}^{\mathcal{T}}_{[\phi \mid \phi(3)=s]}$ on ρ over and over, where $O = O^{q^\Delta}$ and $\mathcal{T}$ is induced on $\theta(\Sigma^\delta_\delta B(\mathfrak{K}))$ by T, Eq. (13). (We assume $\overline{O}^{\mathcal{T}}$ exists.) As before, the mathematical procedure used to generate $\theta_E^T\alpha$ from α has disappeared into $\overline{O}^{\mathcal{T}}_{[\phi \mid \phi(3)=s]}$.

As a second example, let f be some (definable)[21,28] function from S^δ to R and let Q be given by

$$Q(\phi) \equiv f(\phi) = \int_{S^\delta} f(\theta) P_{O\rho}(d\theta).$$

For the "higher order" equivalence case, one carries out q^Δ on ρ over and over and thereby obtains the sequence $\alpha \in (S^\delta)^\omega$. He then generates $\theta_E\alpha$ by answering "does $f(\alpha(j)) = \int_{S^\delta} f(\theta) P_{O\rho}(d\theta)\,?$" for each j where $O = O^{q^\Delta}$. This involves a relatively complex mathematical procedure as one must evaluate the integral. The equivalent physical operation is the measurement of $O^{q^\Delta}_{[\phi \mid Q(\phi)\,\mathrm{true}]}$ on ρ over and over again. For the first order equivalence one carries out q^Δ on ρ only once, which gives an $\alpha \in S^\delta$, and generates $\theta_E^T\alpha$ by answering "does $f(T^j\alpha) = \int_{S^\delta} f(\theta) P_{O\rho}(d\theta)\,?$" for each j. The equivalent physical operation is the measurement of $\overline{O}^{\mathcal{T}}_{[\phi \mid Q(\phi)\,\mathrm{true}]}$ on ρ over and over again.

This example illustrates the possibility that the probability operator measure O^{q^Δ} and the state ρ can occur in the defining relation for a set in Σ^δ. It will be seen in future work that such relations *must* in fact be considered. This property for scalar probability measures was used extensively elsewhere.[28]

These, and many other examples which can be constructed, illustrate some of the properties of this equivalence, or embedding of mathematical procedures into operators in $B(\mathfrak{K})$ by means of the probability operator measures O^{q^Δ} associated with processes q^Δ. It will be seen in future work that this embedding is, in essence, the transferral of properties and procedures which are in the metaphysical world, i.e., outside of quantum mechanics, into the physical world, or into quantum mechanics. This process is similar to the extension process used in mathematical logic where a formal theory *Th* can be extended to a larger theory *Th'*, which can describe some of the metamathematics of *Th*.

Finally some brief comments about the construction of O^{q^Δ} from a δ step process q^Δ are in order. It was tacitly assumed here that the process q^Δ was such that $q(j+1)$ and $\Delta(j+1)$ were independent of the observed outcome of measuring $q(j)k$ for any $k \leq j$. This assumption should be removed to allow decision procedures where $q(j+1)$ and $\Delta(j+1)$ can depend on the outcomes already obtained.

Also the restriction of $q(j)$ to be an operator with discrete eigenvalues should be removed. This is especially true if one wishes to consider processes q whose range set includes any of the operators O_E. The generalization to statements more general than that of Von Neumann's projection axiom should also be made. This appears to be necessary since, as Margenau has observed,[17] there are measurement processes for which the usual projection axiom is false. The methods developed by Davies and Lewis[1] appear to be relevant here.

It was tacitly assumed here that the transition from step j to step $j+1$ is a time translation only and was represented by a unitary operator $U(\Delta_{j+1}, \Delta_j)$. As far as the discussion and constructions given in this paper are concerned, neither of these restrictions are necessary. Δ can include space–time translations, rotations, etc. Also it is sufficient for the derivations and proofs given here that $U(\Delta_{j+1}, \Delta_j)$ be an isometry only. It need not be unitary.

In conclusion, it should be noted that there is a construction of a probability operator measure from the process q^Δ which is dual to the construction given here. To see this one notes that the constructions of this paper begin with

$$O_{\{\phi_n\}} = \beta^{q^{\Delta\dagger}}_{\phi_n} \beta^{q^\Delta}_{\phi_n} \tag{25}$$

for each ϕ_n and n. However there is also a dual construction which begins with

$$O'_{\{\phi_n\}} = \beta^{q^\Delta}_{\phi_n} \beta^{q^{\Delta\dagger}}_{\phi_n}.$$

This construction cannot be considered the same as that beginning with Eq. (25) as there is no reason to assume that $\beta^{q^\Delta}_{\phi_n}$ is a normal operator, or that $O'_{\{\phi_n\}} = O_{\{\phi_n\}}$.

In some respects, this dual construction for finite q^Δ is related to the original construction for the time reversal of the process q^Δ. This can be seen from the structure of $\beta^{q^\Delta}_{\phi_n}$, Eq. (24), and the fact that the

time reversal q^{Δ^r} of the process q^{Δ} is such that $q^r(j) = (q(n-j))^r$, the time reversal of the operator $q(n-j)$, and $\Delta_j^r = \Delta_{n-j}$. But this is a subject for future work.

ACKNOWLEDGMENT

The author wishes to thank Dr. Joe Moyal for many very helpful discussions.

* Based on work performed under the auspices of the U.S. Atomic Energy Commission.

1 E. B. Davies and J. T. Lewis, Commun. Math. Phys. **17**, 239 (1970); E. B. Davies, J. Functional Analysis **6**, 318 (1970).
2 J. Von Neumann, *Mathematical Foundations of Quantum Mechanics*, translated by R. T. Beyer (Princeton U. P., Princeton, N.J., 1955), pp. 335, 347.
3 A. N. Kolmogorov, *Foundations of Probability Theory* (Chelsea, New York, 1956), 2nd English ed., pp. 29–33.
4 M. Loève, *Probability Theory* (Van Nostrand, Princeton, N.J., 1963), 3rd ed., pp. 92–94.
5 See Ref. 2, p. 313.
6 S. K. Berberian, *Introduction to Hilbert Space* (Oxford U.P., New York, 1961), p. 149.
7 N. Dunford and J. T. Schwartz, *Linear Operators, Part I* (Wiley, New York, 1957), Chap. VI, Sec. 1.
8 See Ref. 7, Chap. III and Chap. IV, Sec. 10.
9 N. Dinculeanu, *Vector Measures* (Pergamon, London, 1967).
10 I. Dobrakov, Czech. Math. J. **20**, 511 (1970).
11 S. K. Berberian, *Notes on Spectral Theory* (Van Nostrand, New York, 1966).
12 F. Hausdorff, *Set Theory* (Chelsea, New York, 1962), 2nd ed., Secs. 18 and 32.
13 P. A. Benioff (unpublished); see also G. Fox, Can. J. Math. **20**, 1246 (1968).
14 See Ref. 7, pp. 660–62.
15 R. Sine, Proc. Amer. Math. Soc. **24**, 438 (1970).
16 J. DeAraya, Pac. J. Math. **29**, 1 (1969).
17 H. Margenau, Phil. Sci. **30**, 1 (1963).
18 J. Von Neumann, Compos. Math. **6**, 1 (1937); Collected Works, Vol. III, pp. 323–99; Y. Nakagami Kodai, Math. Sem. Rep. **22**, 341 (1970).
19 H. Ekstein, Phys. Rev. **153**, 1397 (1967); **184**, 1315 (1969).
20 Obviously one cannot actually carry out an infinite number of steps of any procedure. Nevertheless infinite repetitions, etc., are of the essence in the interpretation of probabilities as limit relative frequencies. (This is discussed in more detail in Ref. 21.) Thus we shall not hesitate to speak of infinite processes and even infinite sequences of infinite processes as completed entities. This is similar to the usual interpretation of mathematics in which one considers infinite mathematical entities as completed or at hand.
21 P. A. Benioff, J. Math. Phys. **11**, 2553 (1970).
22 J. L. Park and H. Margenau, "The Logic and Noncommutability of Quantum Mechanical Operators and its Empirical Consequences," Landé Festschrift (to be published).
23 E. P. Wigner, Amer. J. Phys. **31**, 6 (1963).
24 P. Halmos, *Measure Theory* (Van Nostrand, Princeton, N.J., 1962), p. 157.
25 See Ref. 7, pp. 669, 675. See also Ref. 4, pp. 423–35.
26 J. M. Jauch, *Foundations of Quantum Mechanics* (Addison Wesley, Reading, Mass., 1968), Chap. 4: The usual definition of a spectral measure given in the above reference is applied to mixed states in Eq. (42) in an obvious manner.
27 See Ref. 24, Chap. V.
28 P. A. Benioff, J. Math. Phys. **12**, 360 (1971).

Occupation Statistics for Parallel Dumbbells on a 2 x *N* Lattice Space

R. B. McQuistan and S. J. Lichtman*
Department of Physics and Laboratory for Surface Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

and

L. P. Levine
Electrical Department and Laboratory for Surface Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

(Received 29 July 1971)

It is shown that occupation statistics for parallel dumbbells on a $2 \times N$ lattice obeys the central limit theorem. On the basis of this conclusion, a modification of the maximum term method is utilized to enumerate for large N the number of distinguishable arrangements arising when indistinguishable, parallel dumbbells are placed on a $2 \times N$ array. The coverage at which the maximum number of arrangements occurs and the pseudovariance for the distribution are also determined.

I. INTRODUCTION

The statistics which describe the occupation of lattice spaces by correlated particles such as dumbbells is unique in three ways:

(1) The occupation of a lattice site insures the occupation of a nearest neighbor site, thus particle orientation must be considered;

(2) the vacancy of a lattice site cannot serve as the sole criterion for a site to be occupied because of the possible existence of isolated vacancies;

(3) there is no equivalence between vacant sites and occupied sites but rather between vacancies and *pairs* of occupied sites.

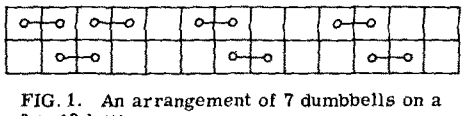

FIG. 1. An arrangement of 7 dumbbells on a 2×13 lattice.

It follows, therefore, that statistical mechanical treatments of physico-chemical systems which may be represented by dumbbells on a lattice space encounter a number of fundamental difficulties. Foremost among these problems is the manner in which the orientational degeneracy of the particles on the lattice space complicates the determination of the grand canonical partition function. To resolve this difficulty, one must be able to calculate the degeneracy of a state represented by a specified number of dumbbells on a lattice space. One is lead to inquire, therefore, into the number of possible ways in which dumbbells may be arranged on a lattice space.[1]

In previous papers we have discussed the statistics[2] and kinetics[3,4] of occupation of one-dimensional arrays of dumbbells and an occupation recursion relation for a $2 \times N$ lattice.[5]

Section II is concerned with demonstrating that the statistics of occupation for parallel dumbbells on a $2 \times N$ lattice (see Fig. 1) conforms to the central

time reversal q^{Δ^r} of the process q^{Δ} is such that $q^r(j) = (q(n-j))^r$, the time reversal of the operator $q(n-j)$, and $\Delta_j^r = \Delta_{n-j}$. But this is a subject for future work.

ACKNOWLEDGMENT

The author wishes to thank Dr. Joe Moyal for many very helpful discussions.

* Based on work performed under the auspices of the U.S. Atomic Energy Commission.

1 E. B. Davies and J. T. Lewis, Commun. Math. Phys. **17**, 239 (1970); E. B. Davies, J. Functional Analysis **6**, 318 (1970).
2 J. Von Neumann, *Mathematical Foundations of Quantum Mechanics*, translated by R. T. Beyer (Princeton U. P., Princeton, N.J., 1955), pp. 335, 347.
3 A. N. Kolmogorov, *Foundations of Probability Theory* (Chelsea, New York, 1956), 2nd English ed., pp. 29–33.
4 M. Loève, *Probability Theory* (Van Nostrand, Princeton, N.J., 1963), 3rd ed., pp. 92–94.
5 See Ref. 2, p. 313.
6 S. K. Berberian, *Introduction to Hilbert Space* (Oxford U.P., New York, 1961), p. 149.
7 N. Dunford and J. T. Schwartz, *Linear Operators, Part I* (Wiley, New York, 1957), Chap. VI, Sec. 1.
8 See Ref. 7, Chap. III and Chap. IV, Sec. 10.
9 N. Dinculeanu, *Vector Measures* (Pergamon, London, 1967).
10 I. Dobrakov, Czech. Math. J. **20**, 511 (1970).
11 S. K. Berberian, *Notes on Spectral Theory* (Van Nostrand, New York, 1966).
12 F. Hausdorff, *Set Theory* (Chelsea, New York, 1962), 2nd ed., Secs. 18 and 32.
13 P. A. Benioff (unpublished); see also G. Fox, Can. J. Math. **20**, 1246 (1968).
14 See Ref. 7, pp. 660–62.
15 R. Sine, Proc. Amer. Math. Soc. **24**, 438 (1970).
16 J. DeAraya, Pac. J. Math. **29**, 1 (1969).
17 H. Margenau, Phil. Sci. **30**, 1 (1963).
18 J. Von Neumann, Compos. Math. **6**, 1 (1937); Collected Works, Vol. III, pp. 323–99; Y. Nakagami Kodai, Math. Sem. Rep. **22**, 341 (1970).
19 H. Ekstein, Phys. Rev. **153**, 1397 (1967); **184**, 1315 (1969).
20 Obviously one cannot actually carry out an infinite number of steps of any procedure. Nevertheless infinite repetitions, etc., are of the essence in the interpretation of probabilities as limit relative frequencies. (This is discussed in more detail in Ref. 21.) Thus we shall not hesitate to speak of infinite processes and even infinite sequences of infinite processes as completed entities. This is similar to the usual interpretation of mathematics in which one considers infinite mathematical entities as completed or at hand.
21 P. A. Benioff, J. Math. Phys. **11**, 2553 (1970).
22 J. L. Park and H. Margenau, "The Logic and Noncommutability of Quantum Mechanical Operators and its Empirical Consequences," Landé Festschrift (to be published).
23 E. P. Wigner, Amer. J. Phys. **31**, 6 (1963).
24 P. Halmos, *Measure Theory* (Van Nostrand, Princeton, N.J., 1962), p. 157.
25 See Ref. 7, pp. 669, 675. See also Ref. 4, pp. 423–35.
26 J. M. Jauch, *Foundations of Quantum Mechanics* (Addison Wesley, Reading, Mass., 1968), Chap. 4: The usual definition of a spectral measure given in the above reference is applied to mixed states in Eq. (42) in an obvious manner.
27 See Ref. 24, Chap. V.
28 P. A. Benioff, J. Math. Phys. **12**, 360 (1971).

Occupation Statistics for Parallel Dumbbells on a 2 x *N* Lattice Space

R. B. McQuistan and S. J. Lichtman*
Department of Physics and Laboratory for Surface Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

and

L. P. Levine
Electrical Department and Laboratory for Surface Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

(Received 29 July 1971)

It is shown that occupation statistics for parallel dumbbells on a $2 \times N$ lattice obeys the central limit theorem. On the basis of this conclusion, a modification of the maximum term method is utilized to enumerate for large N the number of distinguishable arrangements arising when indistinguishable, parallel dumbbells are placed on a $2 \times N$ array. The coverage at which the maximum number of arrangements occurs and the pseudovariance for the distribution are also determined.

I. INTRODUCTION

The statistics which describe the occupation of lattice spaces by correlated particles such as dumbbells is unique in three ways:

(1) The occupation of a lattice site insures the occupation of a nearest neighbor site, thus particle orientation must be considered;

(2) the vacancy of a lattice site cannot serve as the sole criterion for a site to be occupied because of the possible existence of isolated vacancies;

(3) there is no equivalence between vacant sites and occupied sites but rather between vacancies and *pairs* of occupied sites.

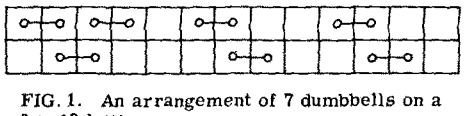

FIG. 1. An arrangement of 7 dumbbells on a 2×13 lattice.

It follows, therefore, that statistical mechanical treatments of physico-chemical systems which may be represented by dumbbells on a lattice space encounter a number of fundamental difficulties. Foremost among these problems is the manner in which the orientational degeneracy of the particles on the lattice space complicates the determination of the grand canonical partition function. To resolve this difficulty, one must be able to calculate the degeneracy of a state represented by a specified number of dumbbells on a lattice space. One is lead to inquire, therefore, into the number of possible ways in which dumbbells may be arranged on a lattice space.[1]

In previous papers we have discussed the statistics[2] and kinetics[3,4] of occupation of one-dimensional arrays of dumbbells and an occupation recursion relation for a $2 \times N$ lattice.[5]

Section II is concerned with demonstrating that the statistics of occupation for parallel dumbbells on a $2 \times N$ lattice (see Fig. 1) conforms to the central

limit theorem. Expressions are developed for $A_p[q, 2 \times N]$, the number of ways q parallel, indistinguishable dumbbells can be arranged on a $2 \times N$ lattice. It follows from such expressions, that for large N the occupation statistics of dumbbells can be described by a Gaussian distribution characterized by a first moment and a pseudo variance. This latter subject is discussed in Sec. III.

II. CENTRAL LIMIT THEOREM FOR PARALLEL DUMBBELLS ON A $2 \times N$ LATTICE

In this Section we will show that for large values of N, all significant values of $A_p[q, 2 \times N]$ occur at values of q which differ only negligibly from the product of N and the first moment of A_p. Specifically, we will define $\theta \equiv q/N, 0 \le \theta \le 1$ to be the fraction occupied lattice sites and we will demonstrate that the normalized fractions of all arrangements occurring in the range between θ and $\theta + d\theta$ are insignificant for all values of θ except where $\theta = \mu'(1)$, the first moment of $A_p[N\theta, 2 \times N]$.

We begin by defining Δ_N, the normalization and $\mu_N(m)$, the mth noncentral moment of $A_p[q, 2 \times N]$,

$$\Delta_N \equiv \sum_{q=0}^{\infty} A_p[q, 2 \times N] \tag{1}$$

and

$$\mu_N(m) \equiv \frac{1}{\Delta_N} \sum_{q=0}^{\infty} q^m A_p[q, 2 \times N]. \tag{2}$$

Clearly $\mu_N(0) = 1$ and $\mu_N(m)$ may be considered to be the mth noncentral moment of the nonnegative, normalized function

$$A_p[q, 2 \times N]/\Delta_N, \text{where } 0 \le A_p[q, 2 \times N]/\Delta_N \le 1. \tag{3}$$

In Appendix A we show that a recursion relation for $A_p[q, 2 \times N]$ may be written in the form

$$A_p[q, 2 \times N] = \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} C_{ij} A_p[(q-i), 2 \times (N-j)], \tag{4}$$

where the C_{ij} are constants independent of q and N; and $A_p[q, 2 \times N] \equiv 0$ if $q > N$. We may thus write $\mu_N(m)$ as

$$\mu_N(m) = \frac{1}{\Delta_N} \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} C_{ij} q^m A_p[(q-i), 2 \times (N-j)]. \tag{5}$$

It will prove convenient to change variables at this point: $Z \equiv q - i$. Equation (5) then becomes

$$\begin{aligned}
\mu_N(m) &= \frac{1}{\Delta_N} \sum_{Z=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} C_{ij} (Z+i)^m A_p[Z, 2 \times (N-j)] \\
&= \frac{1}{\Delta_N} \sum_{Z=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \sum_{k=0}^{m} C_{ij} \binom{m}{k} i^{m-k} Z^k A_p \\
&\quad \times [Z, 2 \times (N-j)] \\
&= \frac{1}{\Delta_N} \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \sum_{k=0}^{m} C_{ij} \binom{m}{k} i^{m-k} \Delta_{N-j} \\
&\quad \times \sum_{Z=0}^{\infty} \frac{Z^k A_p[Z, 2 \times (N-j)]}{\Delta_{N-j}} \\
&= \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} C_{ij} \binom{m}{k} i^{m-k} \frac{\Delta_{N-j}}{\Delta_N} \mu_{N-j}(k)
\end{aligned} \tag{6}$$

We next examine the quantity Δ_{N-j}/Δ_N. It is shown in Appendix B that

$$\Delta_N = K \mu_g^{2N+2} \tag{7}$$

for large N, where K and μ_g, the golden proportion, are constants. Inserting this result in Eq. (6) yields

$$\mu_N(m) = \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} C_{ij} \binom{m}{k} i^{m-k} \mu_g^{-2j} \mu_{N-j}(k). \tag{8}$$

We now write A_p as a function of θ, the fraction of occupied lattice sites, i.e.,

$$A_p[q, 2 \times N]/\Delta_N = A_p[N\theta, 2 \times N]/\Delta_N. \tag{9}$$

As N increases, this ratio must become independent N because A_p and $\Delta_N = K\mu_g^{2N+2}$ are monotonically increasing and the ratio is bounded [see Eq. (3)]. Thus we may write

$$\lim_{N \to \infty} \frac{A_p[N\theta, 2 \times N]}{\Delta_N} = f(\theta). \tag{10}$$

Equation 2 then becomes

$$\begin{aligned}
\mu_N(m) &= \sum_{q=0}^{\infty} q^m \frac{A_p[q, 2 \times N]}{\Delta_N} \\
&= N^m \sum_{q=0}^{\infty} \frac{\theta^m A_p[N\theta, 2 \times N]}{\Delta_N} \\
&\simeq N^m \sum_{\theta=0}^{1} \theta^m f(\theta) = N^m \mu'(m),
\end{aligned} \tag{11}$$

where $\mu'(m)$ is independent of N.

Utilizing Eq. (11) in Eq. (8) yields

$$\begin{aligned}
\mu'(m) N^m &= \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \sum_{k=0}^{m} C_{ij} \binom{m}{k} i^{m-k} \mu_g^{-2j} \mu'(k) [N-j]^k \\
&= \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \sum_{k=0}^{m} \sum_{l=0}^{k} C_{ij} \binom{m}{k} i^{m-k} \mu_g^{-2j} \mu'(k) \binom{k}{l} (-j)^l N^{k-l}.
\end{aligned} \tag{12}$$

We note here that when $l = 0$ and $k = m$, we obtain

$$\sum_{i=0}^{\infty} \sum_{j=1}^{\infty} C_{ij} \mu_g^{-2j} \mu'(m) N^m \equiv \mu'(m) N^m \tag{13}$$

or

$$\sum_{i=0}^{\infty} \sum_{j=1}^{\infty} C_{ij} \mu_g^{-2j} = 1. \tag{14}$$

The only dependence of the right-hand side of Eq. (12) on N is contained in the factor N^{k-l}; therefore, all the coefficients of N^{k-l} must sum to zero for $k - l \ne m$;

$$0 = \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \sum_{k=0}^{m} \sum_{l=0}^{k} C_{ij} \binom{m}{k} i^{m-k} \mu_g^{-2j} \mu'(k) \binom{k}{l} (-j)^l \tag{15}$$

for all values of $k - l \ne m$.

To determine the properties of μ' let us examine the coefficient of N^{m-1}, i.e., $k - l = m - 1$. This means that either $k = m$ and $l = 1$ or that $k = m - 1$ and $l = 0$:

$$\begin{aligned}
0 &= -\sum_{i=0}^{\infty} \sum_{j=1}^{\infty} C_{ij} \mu_g^{-2j} \mu'(m) mj \\
&\quad + \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} C_{ij} \mu_g^{-2j} \mu'(m-1) mi
\end{aligned}$$

or

$$0 = \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} C_{ij} \mu_g^{-2j} [-j\mu'(m) + i\mu'(m-1)], \tag{16}$$

so that

$$\mu'(m) = \mu'(m-1)\left\{\frac{\sum_{i=0}^{\infty}\sum_{j=1}^{\infty} C_{ij}\mu_g^{-2ji}}{\sum_{i=0}^{\infty}\sum_{j=1}^{\infty} C_{ij}\mu_g^{-2ij}}\right\} \tag{17}$$

$$= k\,\mu'(m-1),$$

where

$$k \equiv \frac{\sum_{i=0}^{\infty}\sum_{j=1}^{\infty} C_{ij}\mu_g^{-2ji}}{\sum_{i=0}^{\infty}\sum_{j=1}^{\infty} C_{ij}\mu^{-2ij}} \text{ is}$$

a constant.

Thus

$$\mu'(m) = k^m \tag{18}$$

because

$$\mu'(0) = 1.$$

If we use the C_{ij} determined in Appendix A, we see that

$$\mu'(1) = k = (1 - 1/\sqrt{5}). \tag{19}$$

For a function whose noncentral moments obey Eq. 18 we have

$$f(\theta) = 1 \quad \text{for} \quad \theta = \mu'(1),$$
$$f(\theta) = 0 \quad \text{for} \quad \theta \neq \mu'(1), \tag{20}$$

and so, for large N the only significant terms in the sum

$$\sum_{\theta=0}^{1} A_p[N\theta, 2\times N] \tag{21}$$

occur when θ does not deviate significantly from $\mu'(1)$. Thus the statistics obeys the central limit theorem.

III. ENUMERATION OF DISTINGUISHABLE ARRANGEMENTS OF PARALLEL DUMBBELLS ON A $2 \times N$ LATTICE

It has been shown[2] that $A[q, 1\times N]$, the number of distinguishable arrangements possible when q indistinguishable dumbbells are placed on a one-dimensional array of N compartments, is given by (see Fig. 2)

$$A[q, 1\times N] = \binom{N-q}{q} \tag{22}$$

when the occupation of a compartment precludes further occupation (a Fermi-Dirac constraint).

Equation (22) arises from the following reasoning. In this situation, there are $N-2q$ vacancies and q dumbbells or a total of $N-2q+q = N-q$ objects. $A[q, 1\times N]$ is then the number of arrangements of $N-q$ objects of which $N-2q$ are vacancies and q are dumbbells.

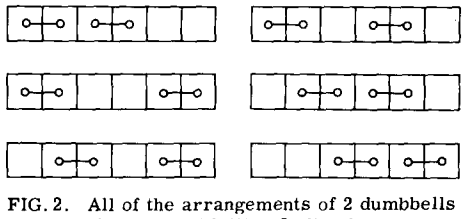

FIG. 2. All of the arrangements of 2 dumbbells on a one dimensional lattice 6 sites long.

Thus if parallel dumbbells are placed on a $2\times N$ array (see Fig. 1) in such a way that there are n_1 dumbbells in the top row and n_2 in the bottom, then $a(n_1, n_2, 2\times N)$ the number of arrangements possible is

$$a(n_1, n_2, 2\times N) = \binom{N-n_1}{n_1}\binom{N-n_2}{n_2} \tag{23}$$

To determine $A_p[q, 2\times N]$, Eq. (23) must be summed over all possible values of n_1 and n_2 subject to the constraint that

$$n_1 + n_2 = q,$$

i.e.,

$$A_p[q, 2\times N] = \sum_{n_1, n_2} \binom{N-n_1}{n_1}\binom{N-n_2}{n_2}. \tag{24}$$

Two cases must be considered in evaluating this sum

$$\text{(a)} \quad 0 \le q \le [N/2], \tag{25}$$

$$\text{(b)} \quad [N/2] \le q \le 2\,[N/2], \tag{26}$$

where $[N/2]$ is the maximum integer contained in $N/2$.

Here (a) represents the case in which all the dumbbells can be placed on one row and (b) the situation when they cannot.

For (a), Eq. (24) becomes

$$A_p[q, 2\times N] = \sum_{n_1=0}^{q} \binom{N-n_1}{n_1}\binom{N-q+n_1}{q-n_1} \tag{27}$$

and, for (b), Eq. (24) is

$$A_p[q, 2\times N] = \sum_{n_1=q-[n/2]}^{[N/2]} \binom{N-n_1}{n_1}\binom{N-q+n_1}{q-n_1}. \tag{28}$$

However, we may add to Eq. (28)

$$\sum_{n_1=0}^{q-[N/2]-1} \binom{N-n_1}{n_1}\binom{N-q+n_1}{q-n_1} + \sum_{n_1=[N/2]+1}^{q} \binom{N-n_1}{n_1}\binom{N-q+n_1}{q-n_1} \tag{29}$$

because both these sums are zero for $[N/2] \le q \le 2\,[N/2]$. Thus for both cases we may write

$$A_p[q, 2\times N] = \sum_{n_1=0}^{q} \binom{N-n_1}{n_1}\binom{N-q+n_1}{q-n_1}$$

$$= \sum_{n_1=0}^{q} a_{n_1}, \tag{30}$$

where

$$a_{n_1} \equiv \binom{N-n_1}{n_1}\binom{N-q+n_1}{q-n_1}. \tag{31}$$

To evaluate Eq. (30) we determine the maximum term and perform a Taylor expansion of $\ln a_{n_1}$ about the maximum, keeping the first three terms only:

$$\ln a_{n_1} = \ln(N-n_1)! - \ln n_1! - \ln(N-2n_1)! + \ln(N-q+n_1)! - \ln(q-n_1)! - \ln(N-2q+2n_1)! \tag{32}$$

Utilizing the Stirling approximation, we may write Eq. (32) as

$$\begin{aligned}\ln a_{n_1} = &- \ln 2\pi + (N - n_1 + \tfrac{1}{2})\ln(N - n_1) \\ &- (n_1 + \tfrac{1}{2})\ln(n_1) - (N - 2n_1 + \tfrac{1}{2})\ln(N - 2n_1) \\ &+ (N - q + n_1 + \tfrac{1}{2})\ln(N - q + n_1) \\ &- (q - n_1 + \tfrac{1}{2})\ln(q - n_1) - (N - 2q + 2n_1 + \tfrac{1}{2}) \\ &\times \ln(N - 2q + 2n_1). \end{aligned} \tag{33}$$

Then

$$\begin{aligned}\frac{\partial \ln a_{n_1}}{\partial n_1} = &\ln\left(\frac{(q - n_1)(N - 2n_1)^2(N - q + n_1)}{n_1(N - 2q + 2n_1)^2(N - n_1)}\right) \\ &+ \tfrac{1}{2}[-(N - n_1)^{-1} - (n_1)^{-1} + 2(N - 2n_1)^{-1} \\ &+ (N - q + n_1)^{-1} + (q - n_1)^{-1} \\ &- 2(N - 2q + 2n_1)^{-1}, \end{aligned} \tag{34}$$

which vanishes for

$$n_1^* = q/2,$$

i.e., the maximum number of arrangements occurs when the dumbbells are evenly distributed between the top and bottom rows. Thus $a_{n_1}^*$, the maximum value of a_{n_1}, is given by

$$a_{n_1}^* = \binom{N - q/2}{q/2}^2 . \tag{35}$$

The third term in the Taylor expansion of a_{n_1} is then

$$\left(\frac{1}{\partial!}\frac{\partial^2 \ln a_{n_1}}{\partial n_1^2}\right)_{n_1^*} = -\frac{1}{2!}\left(\frac{2^3N^2q(N - q)(2N - q) - 2^4N(N - q)^3 - 2^2q^2(2N - q)^2}{q^2(N - q)^2(2N - q)^2}\right) \equiv -\frac{\sigma_1^{-2}}{2!} \tag{36}$$

Thus, from Eqs. (35) and (36),

$$\ln a_{n_1} = \ln a_{n_1}^* - \frac{[n_1 - (q/2)]^2}{2!\sigma_1^2} \tag{37}$$

or

$$a_{n_1} = a_{n_1}^* \exp\left(-\frac{[n_1 - (q/2)]^2}{2!\sigma_1^2}\right). \tag{38}$$

Substituting Eq. (38) in Eq. (30) yields

$$A_p[q, 2 \times N] = a_{n_1}^* \sum_{n_1=0}^{q} \exp\left(\frac{-[n_1 - (q/2)]^2}{2!\sigma_1^2}\right) \tag{39}$$

Let $2n_1 - q \equiv 2m$, then Eq. (39) becomes

$$\begin{aligned}A_p[q, 2 \times N] &= a_{n_1}^* \sum_{m=-q/2}^{q/2} \exp\left[\frac{-m^2}{2!\sigma_1^2}\right] \\ &\cong a_{n_1}^* \int_{-\infty}^{\infty} \exp\left[\frac{-\chi^2}{2!\sigma_1^2}\right] d\chi \\ &= a_{n_1}^* \sqrt{2\pi}\,\sigma_1, \end{aligned} \tag{40}$$

where

$$\sigma_1^2 \equiv \left(\frac{q^2(N - q)^2(2N - q)^2}{8N^2q(N - q)(2N - q) - 16N(N - q)^3 - 4q^2(2N - q)^2}\right). \tag{41}$$

For $0 \ll q \ll N$, Eq. (41) reduces to

$$\sigma_1^2 = \left(\frac{q(N - q)(2N - q)}{8N^2}\right). \tag{42}$$

Figure 3 shows A_p(calc)/A_p(true) for various values of N, where A_p(true) is determined from Eq. (24) and A_p(calc) from Eq. (40). Equation (42) has been utilized to calculate σ_1^2.

For the purposes of some calculations it may be sufficient to represent $A_p[q, 2 \times N]$ as a Gaussian distribution characterized by a mean value and a pseudo variance. To do so we must first determine the value of q which maximizes $A_p[q, 2 \times N]$:

$$\begin{aligned}\ln A_p[q, 2 \times N] &= \ln a_{n_1}^* + \ln\sigma_1 + \tfrac{1}{2}\ln(2\pi) \\ &= 2\ln[N - q/2]! - 2\ln[q/2]! - 2\ln[N - q]! \\ &\quad + \ln\sigma_1 + \tfrac{1}{2}\ln(2\pi). \end{aligned} \tag{43}$$

Using the Stirling approximation for $0 \ll q \ll N$, Eq. (43) may be written as

$$\begin{aligned}\ln A_p[q, 2 \times N] = &\, 2(N - q/2 + \tfrac{1}{2})\ln(N - q/2) \\ &- 2(q/2 + \tfrac{1}{2})\ln(q/2) - 2(N - q + \tfrac{1}{2})\ln[N - q) \\ &+ \ln\sigma_1 + \tfrac{1}{2}\ln(2\pi) \end{aligned} \tag{44}$$

Because $a_{n_1}^*$ varies much more rapidly with q than

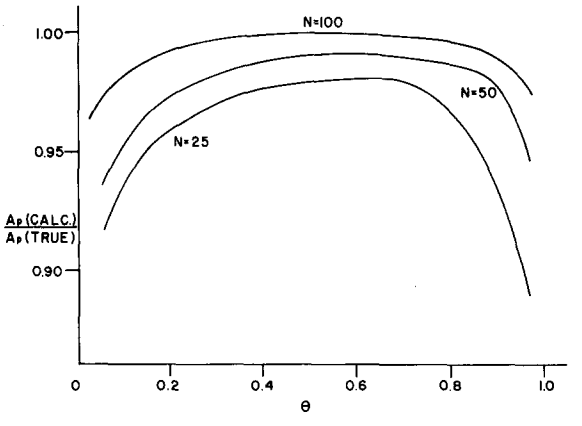

FIG. 3. The ratio A_p(calc)/A_p(true) as a function of θ for various values of N, where A_p(calc) is determined from Eq. (40) and (42) and A_p(true) is determined from Eq. (30).

does σ_1 we shall neglect the derivatives of σ_1 with respect to q. Thus

$$\frac{\partial \ln A_p[q, 2 \times N]}{\partial q} = -\ln(2N - q) - \ln(q) + 2\ln(N - q) + \ln 4 = \ln\left(\frac{4(N-q)^2}{q(2N-q)}\right). \quad (45)$$

This will vanish when

$$\theta^* \equiv (q/N) = \{1 - 1/\sqrt{5}\} \cong 0.552. \quad (46)$$

Thus the most probable value of $A_p[q, 2 \times N]$, i.e., the maximum value of the number of arrangements, occurs when the coverage is approximately 55%, in agreement with Eq. (19). In other words, as one would expect, the first moment is the same as the most probable value or maximum value of $A_p[q, 2 \times N]$.

With this result we may for large N perform a Taylor expansion of $\ln A_p[q, 2 \times N]$ about its maximum value, i.e., about $q^* = N(1 - 1/\sqrt{5})$,

$$\ln A_p[q, 2 \times N] = \ln A_p^* + \frac{1}{2!}\left(\frac{\partial^2 \ln A_p}{\partial q^2}\right)_{q=q^*} (q - q^*)^2 \quad (47)$$

or (see Fig. 4)

$$A_p[q, 2 \times N] = \frac{\mu_g{}^{2N+2}}{5\sigma\sqrt{2\pi}} \exp\left[\frac{-(q - q^*)^2}{2!\sigma^2}\right], \quad (48)$$

where from the evaluation of the derivative of Eq. (45) at q^* we see that

$$\sigma^2 \equiv \frac{2N}{(\sqrt{5})^3} \text{ (a pseudovariance)}. \quad (49)$$

We note here also that the standard deviation for simple particles on the same $2 \times N$ lattice space would be $\frac{1}{2}(2N)^{1/2}$, i.e., the standard deviation for dumbbells is approximately 60% of that for simple particles. The ratio of the standard deviation to the most probable value for dumbbells is approximately $\frac{1}{2}$ of that for simple particles. Thus the Gaussian distribution for dumbbells, as indicated by Eq. (48), is sharper than that of the Gaussian distribution for simple particles.

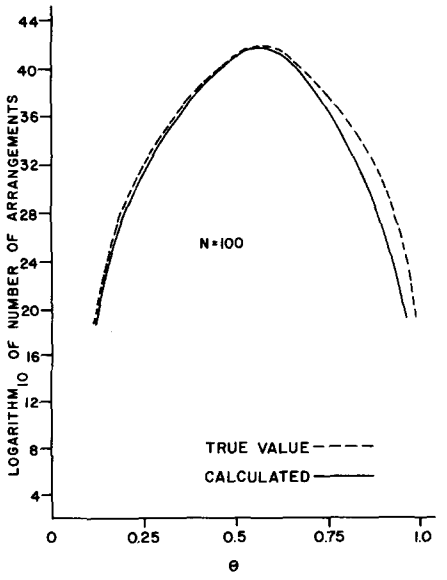

FIG. 4. A comparison of the number of arrangements vs. θ as calculated by Eq. (30) (true) and Eq. (48) (calculated) for $N = 100$.

APPENDIX A

In this appendix we seek to determine a recursion relation for $A_p[q, 2 \times N]$. For this purpose we define two configurations of lattice sites $L_I(N)$ and $L_{II}(N)$, (see Fig. A1). $L_I(N)$ is an array of lattice sites arranged in two adjacent aligned rows of N sites each. $L_{II}(N)$ is an array composed of sites arranged in two adjacent aligned rows; one row of N sites and the other row of $N + 1$ sites. Let $A_p[q, 2 \times N]$ be the number of arrangements of q parallel indistinguishable dumbbells on $L_I(N)$ and $B[q, 2 \times N]$ be the number of arrangements of q indistinguishable parallel dumbells on $L_{II}(N)$.

Theorem I:

$$B[q, 2 \times N] = A_p[q, 2 \times N] + B[q - 1, 2 \times (N - 1)]. \quad (A1)$$

Proof: Let $b[q, 2 \times N]$ be the set of all parallel arrangements of q indistinguishable dumbbells on a L_{II} lattice, $c[q, 2 \times N]$, the subset of $b[q, 2 \times N]$ in which the extra compartment is vacant, and $d[q, 2 \times N]$ in which the extra compartment is filled. Clearly $c[q, 2 \times N] \cap d[q, 2 \times N] = \phi$, a null set. In addition every member of $b[q, 2 \times N]$ will be found in either $c[q, 2 \times N]$ or $d[q, 2 \times N]$, i.e., $c[q, 2 \times N] \cup d[q, 2 \times N]$. We conclude that $\#b[q, 2 \times N]$, the number of members of the set $b[q, 2 \times N]$ is given by

$$\#b[q, 2 \times N] = \#c[q, 2 \times N] + \#d[a, 2 \times N] \equiv B[q, 2 \times N]. \quad (A2)$$

The extra compartment of the $L_{II}(N)$ array is unoccupied in the set $c[q, 2 \times N]$ so that $\#c[q, 2 \times N] \equiv A_p[q, 2 \times N]$. If the extra compartment is occupied, then the adjacent compartment in the same row is also occupied. Hence all other possible arrangements must involve the remaining $q - 1$ dumbbells on the remainder of the array which is a $L_{II}(N - 1)$ array. The number of elements in $d[q, 2 \times N]$ is therefore $B[q - 1, 2 \times (N - 1)]$. Theorem I follows from Eq. (A2).

Theorem II:

$$A_p[q, 2 \times N] = A_p[q, 2 \times (N - 1)] + 2B[q - 1, 2 \times (N - 2)] + A_p[q - 2, 2 \times (N - 2)]. \quad (A3)$$

Proof: Let $a[q, 2 \times N]$ be the set of all possible arrangements on a $L_I(N)$ array and let $e_1[q, 2 \times N]$,

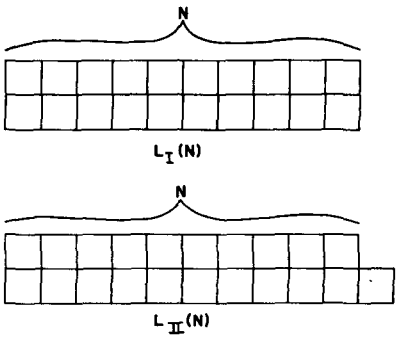

FIG. A1. An $L_I(N)$ array and an $L_{II}(N)$ array.

$e_2[q, 2 \times N, \ldots, e_4[q, 2 \times N]$ be the subsets of $a[q, 2 \times N]$ in which the end two lattice sites are occupied in a manner shown in Fig. A2. Since every member $e_k[q, 2 \times N]$ differs from any and every member of $e_j[q, 2 \times N], k \neq j$, we state that $e_j[q, 2 \times N] \cap e_k[q, 2 \times N] = \phi, k \neq j$. In addition, these four configurations are clearly complete in that they completely describe the occupation of the two end lattice sites

$$\bigcup_{k=1}^{4} e_k[q, 2 \times N] = a[q, 2 \times N]. \tag{A4}$$

We conclude that

$$\#a[q, 2 \times N] = \#e_1[q, 2 \times N] + \#e_2[q, 2 \times N] + \#e_3[q, 2 \times N] + \#e_4[q, 2 \times N] \equiv A_p[q, 2 \times N]. \tag{A5}$$

The set $e_1[q, 2 \times N]$ contains only those arrangements in which the end sites are unoccupied. All q dumbbells are then arranged on a $L_I(N-1)$ array. Both sets $e_2[q, 2 \times N]$ and $e_3[q, 2 \times N]$ have one site empty and one filled. This implies that one adjacent compartment in the same row is also occupied. In each of these sets, the remaining $q-1$ dumbbells are arranged on a $L_{II}(N-2)$ array. In the set $e_4[q, 2 \times N]$, two dumbbells occupy the end two sites. Then the remaining $q-2$ dumbbells are arranged on a $L_I(N-2)$ array. Application of Eq. (A5) yields Theorem II.

If Eq. (A3) is solved for $B[q-1, 2 \times (N-2)]$,

$$2B[q-1, 2 \times (N-2)] = A_p[q, 2 \times N] - A_p[q, 2 \times (N-1)] - A_p[q-2, 2 \times (N-2)], \tag{A6}$$

so that by reindexing q and N we may also write

$$2B[q, 2 \times N] = A_p[q+1, 2 \times (N+2)] - A_p[q+1, 2 \times (N+1)] - A_p[q-1, 2 \times N] \tag{A7}$$

and these values for the B can be substituted in Eq. (A1). If the result is then reindexed in q and N, we obtain

$$A_p[q, 2 \times N] = A_p[q, 2 \times (N-1)] + A_p[q-1, 2 \times (N-1)] + A_p[q-1, 2 \times (N-2)] + A_p[q-2, 2 \times (N-2)] - A_p[q-3, 2 \times (N-3)]. \tag{A8}$$

Thus

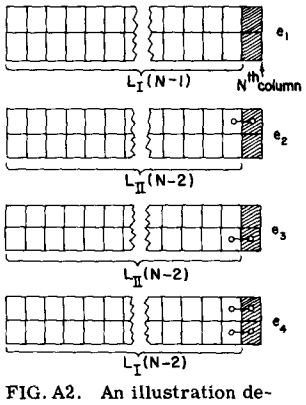

FIG. A2. An illustration defining the subsets e_1, e_2, e_3, and e_4 in terms of the occupation of the Nth column.

$$C_{00} = 0, \quad C_{11} = 1, \quad C_{20} = 0, \quad C_{30} = 0,$$
$$C_{01} = 1, \quad C_{12} = 1, \quad C_{21} = 0, \quad C_{31} = 0,$$
$$C_{22} = 1, \quad C_{32} = 0,$$
$$C_{33} = -1.$$

APPENDIX B

In this appendix we show that

$$\Delta_N = K\mu_g^{2N+2}, \tag{B1}$$

where K and μ_g are constants.

To accomplish this we imagine that the $2 \times N$ lattice is rearranged into a $1 \times 2N$ lattice. $A[q, 1 \times 2N]$, the number of arrangements of a q dumbbells on the $1 \times 2N$ lattice, minus $X(q, 1 \times 2N)$, the number of arrangements in which a dumbbell occupies the N and $N+1$ lattice sites is equal to $A_p[q, 2 \times N]$, i.e.,

$$A_p[q, 2 \times N] = A[q, 1 \times 2N] - X(q, 1 \times 2N). \tag{B2}$$

As N becomes large, the sum of Eq. (B2) over all possible values of q may be written as

$$\sum_{q=0}^{2(N/2)} A_p[q, 2 \times N] \approx \sum_{q=0}^{[N]} A[q, 1 \times N] \tag{B3}$$

because the sum of X over q will be insignificant, i.e., the sum of all arrangements in which the N and $N+1$ lattice sites are occupied will be negligible when compared to the sum of all possible arrangements for all possible q. Then

$$\lim_{N \to \infty} \sum_q A_p[q, 2 \times N] \approx \sum_q A[q, 1 \times 2N]. \tag{B4}$$

We have shown previously[5] that

$$A[q, 1 \times 2N] = \binom{2N-q}{q}. \tag{B5}$$

Thus

$$\sum_q A[q, 1 \times 2N] = \sum_{q=0}^{[N]} \binom{2N-q}{q} = f_{2N+1}, \tag{B6}$$

where f_R is the Rth Fibonacci number. Because

$$f_{2N+1} = \frac{1}{\sqrt{5}} \left\{ \left(\frac{1+\sqrt{5}}{2} \right)^{2N+2} - \left(\frac{1-\sqrt{5}}{2} \right)^{2N+2} \right\}, \tag{B7}$$

we may write for large N,

$$\sum_{q=0}^{[N]} \binom{2N-q}{q} = K\mu_g^{2N+2}, \tag{B8}$$

where $K \equiv 1/\sqrt{5}$ and μ_g, the golden proportion is given by $\mu_g \equiv \frac{1}{2}(1+\sqrt{5})$.

Thus for large N,

$$\Delta_N \equiv \sum_{q=0}^{[N]} A_p[q, 2 \times N] = K'\mu_g^{2N+2}. \tag{B9}$$

We can also determine μ_g in terms of the C_{ij} in our recursion relation [Eqs. (3) and (A8)]

$$K'\mu_g^{2N+2} = \sum_{q=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} C_{ij} A_p[q-i, 2 \times (N-j)] = \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} C_{ij} K'\mu_g^{2(N-j)+2}$$

$$= K'\mu_g^{2N+2} \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} C_{ij} \mu_g^{-2j}$$

$$\pm K'\mu_g^{2N+2} \sum_{j=1}^{\infty} \sum_{i=0}^{\infty} C_{ij} \chi^j \qquad \text{(B10)}$$

or

$$1 = \sum_{j=0}^{\infty} C_j \chi^j, \qquad \text{(B11)}$$

where $\chi \equiv \mu_g^{-2}$ and the χ are the smallest positive root of Eq. (B11) in which $C_0 \equiv -1$.

---

* Present Address: University of Wisconsin-Madison, Madison, Wisconsin 53706

1 S. G. Brush, Rev. Mod. Phys. **39**, 883 (1967).
2 D. Lichtman and R. B. McQuistan, J. Math. Phys. **8**, 2441 (1967).
3 R. B. McQuistan and D. Lichtman, J. Math. Phys. **9**, 1630 (1968).
4 R. B. McQuistan, J. Math. Phys. **10**, 2205 (1969).
5 R. B. McQuistan and S. J. Lichtman, J. Math. Phys. **11**, 3095 (1970).

Perturbation Calculation of an Algebraic Realization of Spin Waves

C. J. Liu

Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

(Received 17 June 1971)

Via the results of a previous paper on an algebraic realization of spin-wave theory [J. Math. Phys. **12**, 2144 (1971)], the Hamiltonian of the Heisenberg model of ferromagnetism is regrouped into a main part and a perturbed part. By introducing a generalized Bethe–Salpeter equation it is shown that the main part can be diagonalized by the single-particle states, two-particle states, and two-particle bound states. To study the scatterings of the perturbed part, an equation for the amplitude of the two-particle scattering states is derived. At low temperatures this equation is solved by power series expansion. The modification of thermodynamic free energy and spontaneous magnetization due to perturbation is equal to the first Born approximation multiplied by a function $Q(S)$.

1. INTRODUCTION

In a recent paper, the spin wave theory has been studied on the basis of an algebraic realization of spin operators.[1] The Hamiltonian under consideration is the Heisenberg model of ferromagnetism,

$$\mathcal{H} = -H \sum_l S_l^{(z)} - \sum_{l,m} J_{lm} \mathbf{S}_l \cdot \mathbf{S}_m, \qquad (1.1)$$

where $\mathbf{S}_l$ is the spin operator of the atom at the lth site, satisfying the algebraic relations

$$[S_l^{(+)}, S_m^{(-)}] = 2\delta_{lm} S_l^{(z)}, [S_l^{(z)}, S_m^{(\pm)}] = \pm \delta_{lm} S_l^{(\pm)}. \qquad (1.2)$$

The spin operators can be realized in the form

$$\mathbf{S}_l = \Phi_l^\dagger \sigma \Phi_l, \qquad (1.3)$$

with

$$\Phi_l \equiv \begin{pmatrix} \beta_l + \xi_\beta \\ b_l + \xi_b \end{pmatrix}, \qquad (1.4)$$

where σ is the Pauli matrix. Both ξ_β and ξ_b are arbitrary c-numbers while β_l and b_l are boson operators that commute with each other. In particular, we choose $\xi_\beta^2 = 2S$ and $\xi_b = 0$. The Fourier transform for β_l and b_l is

$$\beta_k = (1/\sqrt{N}) \sum_l e^{il\cdot\mathbf{k}} \beta_l,$$
$$b_k = (1/\sqrt{N}) \sum_l e^{il\cdot\mathbf{k}} b_l. \qquad (1.5)$$

We note (cf. Ref. 1) that the β particles are "spurions" which carry only spin quantum number without excitation energy. The dynamic properties of spin wave are essentially due to the b particles which are the observed spin waves, i.e., magnons. Therefore, it suffices to write down the effective Hamiltonian in place of Eq. (1.1):

$$\mathcal{H} = E_0 + \sum_{\mathbf{k}} \{H + 2S[J(0) - J(\mathbf{k})]\} b_k^\dagger b_k - (1/2N) \times \sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4} [J(\mathbf{k}_1 - \mathbf{k}_3) + J(\mathbf{k}_1 - \mathbf{k}_4)] b_{k_1}^\dagger b_{k_2}^\dagger b_{k_3} b_{k_4} \times \delta(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3 - \mathbf{k}_4), \qquad (1.6)$$

where

$$E_0 = -NS^2 J(0) - NSH \qquad (1.7)$$

and

$$J(k) = J \sum_1 e^{i\mathbf{k}\cdot\mathbf{l}}. \qquad (1.8)$$

The Hamiltonian of Eq. (1.6) is obtainable from that of Holstein and Primakoff by putting[2]

$$(1 - b_l^\dagger b_l / 2S)^{1/2} = 1 \qquad (1.9)$$

and keeping the interaction term

$$(a_l^\dagger a_l)(a_m^\dagger a_m). \qquad (1.10)$$

Since Eq. (1.10) yields the two-particle bound states, it plays an essential role. In the limiting case of long wavelengths, calculations show that the two-particle bound states are no less important than the single-particle states.[1] Recently, bound states have been observed in far infrared measurements.[3] Elaborations are focused on anisotropic effect.[4,5] However, our approach is easily extended to include such an effect.

The existence of the bound states make Born approximation inapplicable. For this reason we regroup the Hamiltonian of Eq. (1.6) in the following way:

$$\mathcal{H} = \mathcal{H}_I + \mathcal{H}_{II}, \qquad (1.11)$$

where

$$\mathcal{H}_I = E_0 + \sum_{\mathbf{k}} \{H + 2S[J(0) - J(\mathbf{k})]\} b_k^\dagger b_k - [J(0)/N] \times \sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4} b_{k_1}^\dagger b_{k_2}^\dagger b_{k_3} b_{k_4} \delta(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3 - \mathbf{k}_4) \qquad (1.12)$$

and

$$= K'\mu_g^{2N+2} \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} C_{ij} \mu_g^{-2j}$$

$$\pm K'\mu_g^{2N+2} \sum_{j=1}^{\infty} \sum_{i=0}^{\infty} C_{ij} \chi^j \qquad \text{(B10)}$$

or

$$1 = \sum_{j=0}^{\infty} C_j \chi^j, \qquad \text{(B11)}$$

where $\chi \equiv \mu_g^{-2}$ and the χ are the smallest positive root of Eq. (B11) in which $C_0 \equiv -1$.

---

* Present Address: University of Wisconsin-Madison, Madison, Wisconsin 53706

1 S. G. Brush, Rev. Mod. Phys. **39**, 883 (1967).
2 D. Lichtman and R. B. McQuistan, J. Math. Phys. **8**, 2441 (1967).
3 R. B. McQuistan and D. Lichtman, J. Math. Phys. **9**, 1630 (1968).
4 R. B. McQuistan, J. Math. Phys. **10**, 2205 (1969).
5 R. B. McQuistan and S. J. Lichtman, J. Math. Phys. **11**, 3095 (1970).

Perturbation Calculation of an Algebraic Realization of Spin Waves

C. J. Liu

Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

(Received 17 June 1971)

Via the results of a previous paper on an algebraic realization of spin-wave theory [J. Math. Phys. **12**, 2144 (1971)], the Hamiltonian of the Heisenberg model of ferromagnetism is regrouped into a main part and a perturbed part. By introducing a generalized Bethe–Salpeter equation it is shown that the main part can be diagonalized by the single-particle states, two-particle states, and two-particle bound states. To study the scatterings of the perturbed part, an equation for the amplitude of the two-particle scattering states is derived. At low temperatures this equation is solved by power series expansion. The modification of thermodynamic free energy and spontaneous magnetization due to perturbation is equal to the first Born approximation multiplied by a function $Q(S)$.

1. INTRODUCTION

In a recent paper, the spin wave theory has been studied on the basis of an algebraic realization of spin operators.[1] The Hamiltonian under consideration is the Heisenberg model of ferromagnetism,

$$\mathcal{H} = -H \sum_l S_l^{(z)} - \sum_{l,m} J_{lm} \mathbf{S}_l \cdot \mathbf{S}_m, \qquad (1.1)$$

where $\mathbf{S}_l$ is the spin operator of the atom at the lth site, satisfying the algebraic relations

$$[S_l^{(+)}, S_m^{(-)}] = 2\delta_{lm} S_l^{(z)}, [S_l^{(z)}, S_m^{(\pm)}] = \pm \delta_{lm} S_l^{(\pm)}. \qquad (1.2)$$

The spin operators can be realized in the form

$$\mathbf{S}_l = \Phi_l^\dagger \sigma \Phi_l, \qquad (1.3)$$

with

$$\Phi_l \equiv \begin{pmatrix} \beta_l + \xi_\beta \\ b_l + \xi_b \end{pmatrix}, \qquad (1.4)$$

where σ is the Pauli matrix. Both ξ_β and ξ_b are arbitrary c-numbers while β_l and b_l are boson operators that commute with each other. In particular, we choose $\xi_\beta^2 = 2S$ and $\xi_b = 0$. The Fourier transform for β_l and b_l is

$$\beta_k = (1/\sqrt{N}) \sum_l e^{il\cdot\mathbf{k}} \beta_l,$$
$$b_k = (1/\sqrt{N}) \sum_l e^{il\cdot\mathbf{k}} b_l. \qquad (1.5)$$

We note (cf. Ref. 1) that the β particles are "spurions" which carry only spin quantum number without excitation energy. The dynamic properties of spin wave are essentially due to the b particles which are the observed spin waves, i.e., magnons. Therefore, it suffices to write down the effective Hamiltonian in place of Eq. (1.1):

$$\mathcal{H} = E_0 + \sum_{\mathbf{k}} \{H + 2S[J(0) - J(\mathbf{k})]\} b_k^\dagger b_k - (1/2N) \times \sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4} [J(\mathbf{k}_1 - \mathbf{k}_3) + J(\mathbf{k}_1 - \mathbf{k}_4)] b_{k_1}^\dagger b_{k_2}^\dagger b_{k_3} b_{k_4} \times \delta(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3 - \mathbf{k}_4), \qquad (1.6)$$

where

$$E_0 = -NS^2 J(0) - NSH \qquad (1.7)$$

and

$$J(k) = J \sum_1 e^{i\mathbf{k}\cdot\mathbf{l}}. \qquad (1.8)$$

The Hamiltonian of Eq. (1.6) is obtainable from that of Holstein and Primakoff by putting[2]

$$(1 - b_l^\dagger b_l / 2S)^{1/2} = 1 \qquad (1.9)$$

and keeping the interaction term

$$(a_l^\dagger a_l)(a_m^\dagger a_m). \qquad (1.10)$$

Since Eq. (1.10) yields the two-particle bound states, it plays an essential role. In the limiting case of long wavelengths, calculations show that the two-particle bound states are no less important than the single-particle states.[1] Recently, bound states have been observed in far infrared measurements.[3] Elaborations are focused on anisotropic effect.[4,5] However, our approach is easily extended to include such an effect.

The existence of the bound states make Born approximation inapplicable. For this reason we regroup the Hamiltonian of Eq. (1.6) in the following way:

$$\mathcal{H} = \mathcal{H}_I + \mathcal{H}_{II}, \qquad (1.11)$$

where

$$\mathcal{H}_I = E_0 + \sum_{\mathbf{k}} \{H + 2S[J(0) - J(\mathbf{k})]\} b_k^\dagger b_k - [J(0)/N] \times \sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4} b_{k_1}^\dagger b_{k_2}^\dagger b_{k_3} b_{k_4} \delta(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3 - \mathbf{k}_4) \qquad (1.12)$$

and

$$\mathcal{H}_{II} = (1/2N) \sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4} [2J(0) - J(\mathbf{k}_1 - \mathbf{k}_2) - J(\mathbf{k}_1 - \mathbf{k}_4)] b^+_{k_1} b^+_{k_2} b_{k_3} b_{k_4} \delta(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3 - \mathbf{k}_4). \quad (1.13)$$

By introducing a generalized Bethe–Salpeter equation we shall show that the Hamiltonian given in Eq. (1.12) can be diagonalized by single-particle states, two-particle states, and two-particle bound states. Accordingly, we can treat $\mathcal{H}_I$ as the main part and $\mathcal{H}_{II}$ as the perturbed part.

Finally, the modification of thermodynamic free energy and spontaneous magnetization is calculated. To accomplish this, an equation for the amplitude of two-particle scattering states is derived. The inhomogeneous part gives the first Born approximation. At low temperature the equation is solved by power expansion. The result shows that the correction is equal to the first Born approximation multiplied by a function $Q(S)$.

2. THE BETHE–SALPETER EQUATION

Consider $\mathcal{H}_I$ as the main part. The equation of motion is given by

$$\left(H + 2S[J(0) - J(\mathbf{k})] + \frac{1}{i}\frac{\partial}{\partial t}\right) b_k = 2J(0)/N \sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3} b^+_{k_1} b_{k_2} b_{k_3} \delta(\mathbf{k} + \mathbf{k}_1 - \mathbf{k}_2 - \mathbf{k}_3) \quad (2.1)$$

where

$$J(0) - J(\mathbf{k}) = \sum_{\mathbf{l}} \sum_{n} [(-)^{n+1}/(2n)!](\mathbf{k}\cdot\mathbf{l})^{2n}. \quad (2.2)$$

Define the Fourier transform

$$\psi(\mathbf{x}, t) = (1/\sqrt{N}) \sum_{\mathbf{k}} b_k(t) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (2.3)$$

therefore,

$$[\psi(\mathbf{x}, t), \psi^+(\mathbf{x}', t)] = \delta(\mathbf{x} - \mathbf{x}')$$

and

$$[\psi(\mathbf{x}, t), \psi(\mathbf{x}', t)] = 0. \quad (2.4)$$

The equation of motion, in terms of $\psi(\mathbf{x}, t)$, is

$$\left(H - \Lambda(\nabla) + \frac{1}{i}\frac{\partial}{\partial t}\right)\psi(x) = 2J(0)\psi^+(x)\psi(x)\psi(x), \quad (2.5)$$

where

$$\Lambda(\nabla) \equiv 2S \sum_{\mathbf{l}} \sum_{n} [1/(2n)!](\mathbf{l}\cdot\nabla)^{2n}. \quad (2.6)$$

We note that the operator $\Lambda(\nabla)$ introduced here is to generalize the field equation.[6]

The following calculation is parallel to the method used in Ref. 1. Define first the Bethe–Salpeter wave function:

$$\chi_{\mathbf{k}E\alpha}(x_1, x_2) = \langle 0 | T(\psi(x_1)\psi(x_2)) | \mathbf{k}, E, \alpha \rangle, \quad (2.7)$$

where $T(\psi(x_1)(\psi(x_2))$ is the time-ordered product of $\psi(x_1)$ and $\psi(x_2)$. Using the equation of motion and the commutation relations of ψ, one finds

$$\left(H - \Lambda(\nabla_1) + \frac{1}{i}\frac{\partial}{\partial t_1}\right)\left(H - \Lambda(\nabla_2) + \frac{1}{i}\frac{\partial}{\partial t_2}\right)\chi_{\mathbf{k}E\alpha}(x_1, x_2) = i2J(0)\delta(t_1 - t_2)\delta(\mathbf{x}_1 - \mathbf{x}_2)\chi_{\mathbf{k}E\alpha}(x_1, x_2). \quad (2.8)$$

Now, we introduce the propagator function

$$S_F(x_1 - x_2) = \langle 0 | T(\psi(x_1)\psi^+(x_2)) | 0 \rangle. \quad (2.9)$$

S_F satisfies the equation

$$\left(H - \Lambda(\nabla_1) + \frac{1}{i}\frac{\partial}{\partial t_1}\right) S_F(x_1 - x_2) = -i\delta(t_1 - t_2)\delta(\mathbf{x}_1 - \mathbf{x}_2). \quad (2.10)$$

Using $S_F(x)$ to integrate Eq. (2.8), we obtain the Bethe–Salpeter integral equation

$$\chi_{\mathbf{k}E\alpha}(x_1, x_2) = \chi^\circ_{\mathbf{k}E\alpha}(x_1, x_2) - i2J(0) \times \sum_{\mathbf{y}} \int dt_y S_F(x_1 - y)\chi_{\mathbf{k}E\alpha}(y, y) S_F(x_2 - y) \quad (2.11)$$

where the inhomogeneous term $\chi^\circ_{\mathbf{k}E\alpha}(x_1, x_2)$ satisfies the source-free equation.

By taking $x_1 = x_2$, Eq. (2.11) can be written

$$\chi_{\mathbf{k}E\alpha}(x, x) = \chi^\circ_{\mathbf{k}E\alpha}(x, x) - \sum_{\mathbf{y}} \int dt_y V(x - y)\chi_{\mathbf{k}E\alpha}(y, y), \quad (2.12)$$

where the kernel $V(x)$ is defined by

$$V(x) \equiv i2J(0)S_F(x)S_F(x). \quad (2.13)$$

Since

$$S_F(x) = \frac{i}{N} \sum_{\mathbf{k}} \int \frac{dE}{2\pi} \frac{1}{E - E_k + i_\epsilon} e^{i(\mathbf{k}\cdot\mathbf{x} - Et)} \quad (2.14)$$

with

$$E_k \equiv H + 2S[J(0) - J(\mathbf{k})], \quad (2.15)$$

we obtain

$$V(x) = \frac{1}{N} \sum_{k} \int \frac{dE}{2\pi} V(\mathbf{k}, E) e^{i(\mathbf{k}\cdot\mathbf{x} - Et)}, \quad (2.16)$$

where

$$V(\mathbf{k}, E) \equiv \frac{4J(0)}{N} \sum_{\mathbf{q}} \frac{1}{E - (E_{\mathbf{k}/2+\mathbf{q}} + E_{\mathbf{k}/2-\mathbf{q}})}. \quad (2.17)$$

Define the two-particle states by

$$\chi_{\mathbf{k}E\alpha}(x, x) = \chi_\alpha(\mathbf{p}, \mathbf{q}) \exp\{i[(\mathbf{p} + \mathbf{q})\cdot\mathbf{x} - (E_p + E_q)t]\} \quad (2.18)$$

with

$$\mathbf{p} + \mathbf{q} = \mathbf{k}. \quad (2.19)$$

Subsititution of Eq. (2.18) into Eq. (2.11) gives

$$\chi_\alpha(\mathbf{p}, \mathbf{q}) = [1 + V(\mathbf{p} + \mathbf{q}, E_p + E_q)]^{-1}, \quad (2.20)$$

where

$$V(\mathbf{p} + \mathbf{q}, E_p + E_q) = \frac{J(0)}{NSJ} \sum_{\mathbf{k}'} \times \frac{1}{\sum \cos[\frac{1}{2}(\mathbf{p} + \mathbf{q})\cdot l]\{\cos(\mathbf{k}'\cdot\mathbf{l}) - \cos[\frac{1}{2}(\mathbf{p} - \mathbf{q})\cdot\mathbf{l}]\}}. \quad (2.21)$$

The left-hand side of Eq. (2.21) depends on $(\mathbf{p} + \mathbf{q})$ while the right-hand side depends on $(p_i - q_i)^{2n}$ $i = x, y, z$ and $n > 0$. Hence, there are only two possibilities:

$$\text{(i) } \mathbf{p} = ip \quad \text{and} \quad \mathbf{q} = jq \quad i \neq j, \ i, j = z, y, z; \qquad \text{(ii) } \mathbf{p} = \mathbf{q}. \quad (2.22)$$

Case (i) leads to an inconsistency. The reason is

parallel to the arguments presented in Ref. 1. On the other hand, case (ii) yields

$$\chi_{kE\alpha}(x, x) = \chi_\alpha(\mathbf{k}) \exp[i(\mathbf{k}\cdot\mathbf{x} - \epsilon_k t)] \tag{2.23}$$

with

$$\epsilon_k = 2E_{k/2}. \tag{2.24}$$

The representation of $\chi_\alpha(\mathbf{k})$ is given in the Appendix A.[7]

We now define the two-particle bound state

$$\chi_{B,\mathbf{k}\omega\alpha}(x, x) = \chi_{B,\alpha}(\mathbf{k}) \exp[i(\mathbf{k}\cdot\mathbf{x} - \omega_k t)]. \tag{2.25}$$

Substitution of Eq. (2.25) into the homogeneous Bethe–Salpeter equation yields

$$V(k, \omega_k) = -1. \tag{2.26}$$

Define

$$M = 2H + 4S[J(0) - J(\tfrac{1}{2}\mathbf{k})] - \omega_k; \tag{2.27}$$

then Eq. (2.26) turns out to be

$$\frac{4J(0)}{N} \sum_{\mathbf{q}} \frac{1}{M + 8SJ \sum_{\mathbf{l}} \cos(\tfrac{1}{2}\mathbf{k}\cdot\mathbf{l}) \sin^2(\tfrac{1}{2}\mathbf{q}\cdot\mathbf{l})} = 1. \tag{2.28}$$

At low temperature, $ka \ll 1$, the integrand is independent of the total momentum $\mathbf{k}$. We note that Eqs. (2.27) and (2.28) agree with the results obtained in Ref. 1 if only the leading terms are considered. M can be interpreted as the binding energy of the bound state.

By restricting ourselves only to two-particle processes, the Hamiltonian of Eq. (1.12) takes the form

$$\mathcal{H}_1 = E_0 + \sum_{\mathbf{k}} E_k b_k^\dagger b_k + \sum_{\mathbf{k}} \epsilon_k c_k^\dagger c_k + \sum_{\mathbf{k}} \omega_k B_k^\dagger B_k, \tag{2.29}$$

where c_k and $c_k^\dagger$ are the annihilation and creation operators of the *two-particle states* while B_k and $B_k^\dagger$ are the annihilation and creation operators of the *two-particle bound states*. The present calculations can be generalized to include the anisotropic effects. These are presented in the Appendix B.

When we evaluate the thermodynamic free energy from the Hamiltonian of Eq. (2.29), we have to bear in mind that at low temperature the number of excitations, i.e., spin waves, are very small compared with N.[1] The ground states are characterized by the "condensation" of β particles while the observed spin waves are b particles. Therefore, we neglect the effect of binding energy in the calculation of thermodynamic free energy.

In the case of simple cubic lattices with nearest-neighbor exchange interactions, the free energy can be expressed as an expression:

$$\begin{aligned} F = E_0 - N\beta^{-}\{&[Z_{5/2}(\beta H) + 2^{5/2} Z_{5/2}(2\beta H)]\theta^{3/2} \\ &+ \tfrac{3}{4}\pi[Z_{7/2}(\beta H) + 2^{3/2} Z_{7/2}(2\beta H)]\theta^{5/2} \\ &\times + \tfrac{33}{32}\pi^2[Z_{9/2}(\beta H) + 2^{1/2} Z_{9/2}(2\beta H)]\theta^{7/2} + \cdots\} \end{aligned} \tag{2.30}$$

with

$$z_n(x) \equiv \sum_{r=1} \frac{1}{r^n} e^{-rx} \tag{2.31}$$

and

$$\theta \equiv [\tfrac{4}{3}\pi\beta S J(0)]^{-1}. \tag{2.32}$$

The spontaneous magnetization is therefore given by

$$\begin{aligned} \mathfrak{M} = N\{&S - (1 + 2^{7/2})\zeta(\tfrac{3}{2})\theta^{3/2} - \tfrac{3}{4}\pi(1 + 2^{5/2})\zeta(\tfrac{5}{2})\theta^{5/2} \\ &- \tfrac{33}{32}\pi^2(1 + 2^{3/2})\zeta(\tfrac{7}{2})\theta^{7/2} + \cdots\}, \end{aligned} \tag{2.33}$$

where $\zeta(n) = Z_n(0)$.

3. THE SCATTERINGS

To investigate two-particle scattering states, it suffices to consider $\mathcal{H}_{II}$ as the interaction part and $\mathcal{H}_0$ as the noninteraction part, which is given by

$$\mathcal{H}_0 = \sum_{\mathbf{k}} E_k b_k^\dagger b_k. \tag{3.1}$$

In the following, we use the conventional method of perturbation. The two-particle states satisfy the familiar equation

$$|\psi(\mathbf{p}, \mathbf{q})\rangle = \left(1 - \frac{1}{\mathcal{H}_0 + \mathcal{H}_{II} - E_p - E_q + i\epsilon} \mathcal{H}_{II}\right) \phi(\mathbf{p}, \mathbf{q}), \tag{3.2}$$

where $\epsilon \to 0^+$ and $|\phi(\mathbf{p}, \mathbf{q})\rangle = b_p^\dagger b_q^\dagger |0\rangle$.

Now, we write $|\psi(\mathbf{p}, \mathbf{q})\rangle$ in the form

$$|\psi(\mathbf{p}, \mathbf{q})\rangle = \sum_{\mathbf{k}_1, \mathbf{k}_2} \psi(\mathbf{p}, \mathbf{q}; \mathbf{k}_1, \mathbf{k}_2) b_{k_1}^\dagger b_{k_2}^\dagger |0\rangle \tag{3.3}$$

with

$$\psi(\mathbf{p}, \mathbf{q}; \mathbf{k}_1, \mathbf{k}_2) = \psi(\mathbf{p}, \mathbf{q}; \mathbf{k}_2, \mathbf{k}_1). \tag{3.4}$$

Hence Eq. (3.2) becomes

$$\{\mathcal{H}_0 + \mathcal{H}_{II} - E_p - E_q + i\epsilon\} |\psi(\mathbf{p}, \mathbf{q})\rangle = i\epsilon |\phi(\mathbf{p}, \mathbf{q})\rangle. \tag{3.5}$$

By substituting Eq. (3.3) into Eq. (3.5), one finds

$$\begin{aligned} &\psi(\mathbf{p}, \mathbf{q}; \mathbf{k}_1, \mathbf{k}_2) \\ &\quad = [\delta(\mathbf{p} - \mathbf{k}_1)\delta(\mathbf{q} - \mathbf{k}_2) + \delta(\mathbf{p} - \mathbf{k}_2)\delta(\mathbf{q} - \mathbf{k}_1)] \\ &\qquad - \frac{2}{N} \sum_{\mathbf{k}} \frac{J(0) - J(\mathbf{k})}{E_{k_1} + E_{k_2} - E_p - E_q + i\epsilon} \\ &\qquad \times \psi(\mathbf{p}, \mathbf{q}; \mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}). \end{aligned} \tag{3.6}$$

Equation (3.6) may be solved by iteration. But it is necessary since we are interested only in finding the correction of the thermodynamic free energy which is given by[8]

$$\Delta F = (1/2N) \sum_{\mathbf{p},\mathbf{q}} \langle n_p \rangle \langle n_q \rangle A(\mathbf{p}, \mathbf{q}), \tag{3.7}$$

where

$$\langle n_p \rangle = 1/\{\exp[(\beta E_p) - 1]\} \tag{3.8}$$

and

$$A(\mathbf{p}, \mathbf{q}) \equiv N\langle \phi(\mathbf{p}, \mathbf{q}) | \mathcal{H}_{II} | \psi(\mathbf{p}, \mathbf{q}) \rangle. \tag{3.9}$$

Define

$$f(\mathbf{p}, \mathbf{q}; \mathbf{k}_1, \mathbf{k}_2) \equiv \sum_{\mathbf{k}} [J(0) - J(\mathbf{k})] \psi(\mathbf{p}, \mathbf{q}; \mathbf{k}_1 + \mathbf{k}, \mathbf{k}_2 - \mathbf{k}). \tag{3.10}$$

Therefore, Eq. (3.6) becomes

$$\begin{aligned} f(\mathbf{p}, \mathbf{q}; \mathbf{p}, \mathbf{q}) = \tfrac{1}{2}[J(0) - J(\mathbf{p} - \mathbf{q})] - \frac{2}{N} \sum_{\mathbf{k}} \\ \times \frac{J(0) - J(\mathbf{k})}{E_{p+k} + E_{q-k} - E_p - E_q + i\epsilon} f(\mathbf{p}, \mathbf{q}; \mathbf{p} + \mathbf{k}, \mathbf{q} - \mathbf{k}). \end{aligned} \tag{3.11}$$

It follows from Eqs. (3.3), (3.5), and (3.9) that

$$A(\mathbf{p},\mathbf{q}) = 4f(\mathbf{p},\mathbf{q};\mathbf{p},\mathbf{q}), \tag{3.12}$$

i.e.,

$$A(\mathbf{p},\mathbf{q}) = 2[J(0) - J(\mathbf{p}-\mathbf{q})] + (1/NS)\sum_{\mathbf{k}} K(\mathbf{p},\mathbf{q};\mathbf{k})A(\mathbf{p}+\mathbf{k},\mathbf{q}-\mathbf{k}), \tag{3.13}$$

where

$$K(p,q;k) \equiv \frac{J(0) - J(\mathbf{k})}{J(\mathbf{p}) + J(\mathbf{q}) - J(\mathbf{p}+\mathbf{k}) - J(\mathbf{q}-\mathbf{k})} \tag{3.14}$$

The inhomogeneous part of Eq. (3.13) gives the first Born approximation. Since at low temperature the number of excitations are very small in compared with N, i.e., $\sum_p \langle n_p\rangle \ll N$, therefore the leading term contributing to ΔF comes from

$$2J\sum_{\mathbf{l}} -\frac{1}{4!}\left(\frac{4\cdot 3}{2}\,(\mathbf{p}\cdot\mathbf{l})^2(\mathbf{q}\cdot\mathbf{l})^2\right). \tag{3.15}$$

Consequently,

$$\Delta F = -(3\pi/4S)N\beta^{-1}[Z_{5/2}(\beta H)]^2\theta^4. \tag{3.16}$$

The modification of spontaneous magnetization is, accordingly,

$$\Delta\mathfrak{M} = -(3\pi/2S)N\zeta(\tfrac{3}{2})\xi(\tfrac{5}{2})\theta^4. \tag{3.17}$$

Now let us take into account the corrections due to the homogeneous part of Eq. (3.13). We note first that in the case of $k_1 a \ll 1$ and $k_2 a \ll 1$ the kernel of Eq. (3.14) is approximated by

$$K(\mathbf{p},\mathbf{q};\mathbf{k}) \simeq \tfrac{1}{2}\left(1 - \frac{Ja\sum_{i=1}^{3}(p_i - q_i)\sin(ak_i)}{J(0) - J(\mathbf{k})}\right) \tag{3.18}$$

Write the invariant form of $A(\mathbf{p},\mathbf{q})$ as

$$A(\mathbf{p},\mathbf{q}) = Q_1 a^2 \sum_{i=1}^{3}(p_i - q_i)^2 + Q_2 a^2 \mathbf{p}\cdot\mathbf{q} + Q_3 a^4 \sum_{i=1}^{3}(p_i - q_i)^4 + Q_4 a^4 \sum_{i=j}^{3}(p_i - q_i)^2 \times (p_j - q_j)^2 + Q_5 a^4(\mathbf{p}\cdot\mathbf{q})^2 + O(a^6). \tag{3.19}$$

By substituting Eqs. (3.18) and (3.19) into Eq. (3.13), one can easily show that

$$Q_2 = Q_4 = Q_5 = 0, \tag{3.20}$$

$$Q_1 = \frac{2S}{2S - 1 + 2R} \quad \text{and} \quad Q_3 = \frac{2S}{2S - 1 + 4R}\left(-\frac{2}{4!}\right), \tag{3.21}$$

where

$$R \equiv \frac{1}{3(2\pi)^3}\int_{-\pi}^{\pi}\!\!\int dx\,dy\,dz\,\frac{x\sin x + y\sin y + z\sin z}{3 - \cos x - \cos y - \cos z}. \tag{3.22}$$

Therefore, the correction to the thermodynamic free energy and spontaneous magnetization of Eqs. (3.16) and (3.17) are modified by the factor $Q(S)$ which is given by

$$Q(S) = 2S/(2S - 1 + 4R). \tag{3.23}$$

ACKNOWLEDGMENTS

The author is indebted to Professor Y. Chow and Professor H. Umezawa for valuable discussions.

APPENDIX A: REPRESENTATION OF $\chi_\alpha(\mathbf{k})$

When $\mathbf{p} = \mathbf{q}$, Eqs. (2.19), (2.20), and (2.21) yield

$$\chi_\alpha(\mathbf{k})^{-1} = 1 - \frac{J(0)}{SJ}\left(G^{(0)} + \sum_{\mathbf{l}}[1 - \cos(\tfrac{1}{2}\mathbf{k}\cdot\mathbf{l})]G_l^{(1)} + \sum_{\mathbf{l}_1,\mathbf{l}_2}[1 - \cos(\tfrac{1}{2}\mathbf{k}\cdot\mathbf{l}_1)][1 - \cos(\tfrac{1}{2}\mathbf{k}\cdot\mathbf{l}_2)] \times G_{l_1 l_2}^{(2)} + \cdots\right), \tag{A1}$$

where

$$G_{l_1\cdots l_n}^{(n)} \equiv \frac{1}{N}\sum_{\mathbf{q}}\frac{[1\cos(\mathbf{q}\cdot\mathbf{l}_1)]\cdots[1 - \cos(\mathbf{q}\cdot\mathbf{l}_n)]}{\left(\sum_{\mathbf{l}}[1 - \cos(\mathbf{q}\cdot\mathbf{l})]\right)^{n+1}} \tag{A2}$$

The function defined by Eq. (A2) satisfies the sum rules

$$\sum_{\mathbf{l}_1,\ldots,\mathbf{l}_n} G_{l_1\cdots l_n}^{(n)} = \sum_{\mathbf{l}_1,\ldots,\mathbf{l}_{n-1}} G_{l_1\cdots l_{n-1}}^{(n-1)} = \cdots = G^{(0)} \tag{A3}$$

and the inequalities

$$G_{l_1\cdots l_n}^{(n)} \le G_{l_1\cdots l_{n-1}}^{(n-1)} \le \cdots \le G^{(0)}. \tag{A4}$$

At low temperature, $ka \ll 1$, $\chi_\alpha(\mathbf{k})$ can be power expanded by ka.

APPENDIX B: EFFECTS OF ANISOTROPY

Consider the anisotropic Hamiltonian

$$\mathfrak{H} = -H\sum_{\mathbf{l}} S_l^{(z)} - \sum_{\mathbf{l},\mathbf{m}} J_{lm}\mathbf{S}_l\cdot\mathbf{S}_m - D\sum_{\mathbf{l}}(S_l^{(z)})^2, \tag{B1}$$

where $D \ge 0$.

The main part has the form

$$\mathfrak{H}_1 = E_0 - NDS^2 + \sum_{\mathbf{k}} E_k b_k^+ b_k - \frac{D + J(0)}{N} \times \sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4} b_{k_1}^+ b_{k_2}^+ b_{k_3} b_{k_4}\delta(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3 - \mathbf{k}_4), \tag{B2}$$

where

$$E_k = H + D(2S - 1) + 2S[J(0) - J(\mathbf{k})]. \tag{B3}$$

If we follow the same arguments as in Sec. 2, it is straightforward to show that the two-particle state has energy $\epsilon_k = 2E_{k/2}$. On the other hand, the energy of the two-particle bound state is $\omega_k = 2E_{k/2} - M$.

Consider, instead of the anisotropic Hamiltonian of Eq. (B1),

$$\mathfrak{H} = -H\sum_{\mathbf{l}} S_l^{(z)} - \sum_{\mathbf{l},\mathbf{m}} J_{lm}[S^{(z)}S_m^{(z)} + \sigma S_l^{(+)}S_m^{(-)}], \tag{B4}$$

where $\sigma \ge 0$.

In this case one finds

$$\mathfrak{H}_1 = E_0 + \sum E_k b_k^+ b_k - \frac{J(0)}{N} \times \sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4} b_{k_1}^+ b_{k_2}^+ b_{k_3} b_{k_4}\delta(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3 - \mathbf{k}_4), \tag{B5}$$

where

$$E_k = H + 2S[J(0) - \sigma J(\mathbf{k})]. \tag{B6}$$

The two-particle state has energy $\epsilon_k = 2E_{k/2}$ following a similar argument as before. The energy of the two-particle bound state is

$$\omega_k = 2E_{k/2} - M', \tag{B7}$$

with M' satisfying

$$\frac{4J(0)}{N}\sum_{\mathbf{q}}\frac{1}{M' + 8SJ\sigma\sum_{\mathbf{l}}\cos(\tfrac{1}{2}\mathbf{k}\cdot\mathbf{l})\sin^2(\tfrac{1}{2}\mathbf{q}\cdot\mathbf{l})} = 1. \tag{B8}$$

1 C. J. Liu and Yutze Chow, J. Math. Phys. **12**, 2174 (1971).
2 T. Holstein and H. Primakoff, Phys. Rev. 58, 1098 (1940); F. Keffer, in *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1966), Vol. 18.
3 J. B. Torrance, Jr. and M. Tinkham, Phys. Rev. **187**, 595 (1969), and the papers cited there.
4 J. B. Torrance, Jr. and M. Tinkham, Phys. Rev. **187**, 587 (1969).
5 R. Silberglitt and J. B. Torrance, Jr., Phys. Rev. B **2**, 772 (1970).
6 Y. Takahashi and H. Umezawa, Progr. Theoret. Phys. (Kyoto) **9**, 14 (1953); Y. Takahashi and H. Umezawa, Nucl. Phys. **51**, 193 (1964). See also H. Umezawa, *Quantum Field Theory* (North-Holland, Amsterdam, 1956); Y. Takahashi, *An Introduction to Field Quantization* (Pergamon, New York, 1969).
7 When $ka \to \pi$ $V(k, \epsilon_k) \to \infty$, therefore, the two-particle states do not exist at the zone corner. However, we are interested only at low temperature, i.e., $ka \ll 1$.
8 See, e.g., A. I. Akhiezer, V. G. Bar'yakhtar, and S. V. Peletminskii, *Spin Waves* (North-Holland, Amsterdam, 1968).

Nonlinear Initial-Boundary-Value Problem for Convection, Diffusion, Ionization, and Recombination Processes*

H. E. Wilhelm
Colorado State University, Fort Collins, Colorado

(Received 28 June 1971)

The time-dependent convection–diffusion equation with ionization and recombination reactions is reduced by means of a nonlinear transformation to a differential equation, in which the nonlinear term represents a small perturbation. The general procedure of solution for the corresponding nonlinear initial-boundary-value problem is then established by means of the method of successive approximations. Uniqueness and convergence of the analytical solution are discussed. As applications, the temporal change of an initial distribution of electrons and ions is discussed for a finite box system and an infinitely extended system, respectively.

PROBLEM

In the evaluation of experiments concerned with the measurement of ionization and recombination coefficients, one has in general to consider the simultaneous concentration changes of the reacting particles due to diffusion and convection. The corresponding initial-boundary-value problem is of general interest in plasma kinetics[1–3] and radiation chemistry.[4–6] This problem has been treated analytically in the steady-state case by Wilhelm[7] and numerically in the time-dependent case (without a first-order reaction) by Gray and Kerr[8] and Reinhardt.[9] A comprehensive discussion of a variety of nonlinear initial-boundary-value problems of second order has been given recently by Montroll[10] with the motivation of clarifying the underlying mathematical principles for their analysis. In the following, the nonlinear initial-boundary-value problem for convection and diffusion, ionization, and recombination is subject to a nonlinear transformation and subsequently solved by the method of successive approximations.

Mathematically, the initial-boundary-value problem for convection (flow field $\mathbf{v}$) and collective diffusion (diffusion coefficient D)[11] of the electrons and ions in a partially ionized plasma in presence of ionization (ionization coefficient ϵ)[12] and recombination (recombination coefficient α)[13] reactions is described by

$$\frac{\partial n}{\partial t} + \mathbf{v}\cdot\nabla n = D\nabla^2 n + \epsilon n - \alpha n^2, \qquad \mathbf{r} \neq \mathbf{s},\ t \neq 0, \tag{1}$$

where

$$n(\mathbf{r}, t)_{t=0} = n_0(\mathbf{r}), \qquad \mathbf{r} \neq \mathbf{s}, \tag{2}$$

and

$$n(\mathbf{r}, t)_{\mathbf{r}=\mathbf{s}} = \phi(\mathbf{s}), \qquad t \neq 0, \tag{3}$$

are the initial and boundary conditions ($\mathbf{s}$ = position vector of boundary). The electron-ion density $n(\mathbf{r}, t)$ is assumed to be small compared to the density of the homogeneous background gas (α, D, ϵ = const).

NONLINEAR TRANSFORMATION

In absence of diffusion ($D = 0$) and convection ($\mathbf{v} = \mathbf{0}$) a given particle density $n(0)$ would change with time as a result of ionization and recombination according to $n(t) = n(0)e^{\epsilon t}/[1 - (\alpha/\epsilon)(1 - e^{\epsilon t})n(0)]$. In presence of diffusion ($D \neq 0$) and convection ($\mathbf{v} \neq 0$), therefore, a trial transformation is attempted in the form

$$n(\mathbf{r}, t) = \frac{u(\mathbf{r}, t)e^{\epsilon t}}{[1 - (\alpha/\epsilon)(1 - e^{\epsilon t})u(\mathbf{r}, t)]} \tag{4}$$

with

$$\frac{\partial n}{\partial t} = \left(\frac{\partial u}{\partial t} + \epsilon u - \alpha u^2\right)\frac{e^{\epsilon t}}{[1 - (\alpha/\epsilon)(1 - e^{\epsilon t})u]^2}, \tag{5}$$

$$\nabla n = \frac{\nabla u e^{\epsilon t}}{[1 - (\alpha/\epsilon)(1 - e^{\epsilon t})u]^2}, \tag{6}$$

$$\nabla^2 n = \frac{\nabla^2 u e^{\epsilon t}}{[1 - (\alpha/\epsilon)(1 - e^{\epsilon t})u]^2} + \frac{2(\alpha/\epsilon)(1 - e^{\epsilon t})e^{\epsilon t}(\nabla u)^2}{[1 - (\alpha/\epsilon)(1 - e^{\epsilon t})u]^3}. \tag{7}$$

By substitution of Eqs. (4)–(7), the initial-boundary-value problem defined by Eqs. (1)–(3) is reduced to

$$\frac{\partial u}{\partial t} + \mathbf{v}\cdot\nabla u = D\nabla^2 u - \sigma, \qquad \mathbf{r} \neq \mathbf{s},\ t \neq 0, \tag{8}$$

where

$$u(\mathbf{r}, t)_{t=0} = n_0(\mathbf{r}), \qquad \mathbf{r} \neq \mathbf{s}, \tag{9}$$

$$u(\mathbf{r}, t)_{\mathbf{r}=\mathbf{s}} = \frac{\phi(\mathbf{s})e^{-\epsilon t}}{[1 - (\alpha/\epsilon)(1 - e^{-\epsilon t})\phi(\mathbf{s})]}, \qquad t \neq 0, \tag{10}$$

The two-particle state has energy $\epsilon_k = 2E_{k/2}$ following a similar argument as before. The energy of the two-particle bound state is

$$\omega_k = 2E_{k/2} - M', \tag{B7}$$

with M' satisfying

$$\frac{4J(0)}{N}\sum_{\mathbf{q}}\frac{1}{M' + 8SJ\sigma\sum_{\mathbf{l}}\cos(\tfrac{1}{2}\mathbf{k}\cdot\mathbf{l})\sin^2(\tfrac{1}{2}\mathbf{q}\cdot\mathbf{l})} = 1. \tag{B8}$$

1 C. J. Liu and Yutze Chow, J. Math. Phys. **12**, 2174 (1971).
2 T. Holstein and H. Primakoff, Phys. Rev. 58, 1098 (1940); F. Keffer, in *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1966), Vol. 18.
3 J. B. Torrance, Jr. and M. Tinkham, Phys. Rev. **187**, 595 (1969), and the papers cited there.
4 J. B. Torrance, Jr. and M. Tinkham, Phys. Rev. **187**, 587 (1969).
5 R. Silberglitt and J. B. Torrance, Jr., Phys. Rev. B **2**, 772 (1970).
6 Y. Takahashi and H. Umezawa, Progr. Theoret. Phys. (Kyoto) **9**, 14 (1953); Y. Takahashi and H. Umezawa, Nucl. Phys. **51**, 193 (1964). See also H. Umezawa, *Quantum Field Theory* (North-Holland, Amsterdam, 1956); Y. Takahashi, *An Introduction to Field Quantization* (Pergamon, New York, 1969).
7 When $ka \to \pi$ $V(k, \epsilon_k) \to \infty$, therefore, the two-particle states do not exist at the zone corner. However, we are interested only at low temperature, i.e., $ka \ll 1$.
8 See, e.g., A. I. Akhiezer, V. G. Bar'yakhtar, and S. V. Peletminskii, *Spin Waves* (North-Holland, Amsterdam, 1968).

Nonlinear Initial-Boundary-Value Problem for Convection, Diffusion, Ionization, and Recombination Processes*

H. E. Wilhelm
Colorado State University, Fort Collins, Colorado

(Received 28 June 1971)

The time-dependent convection–diffusion equation with ionization and recombination reactions is reduced by means of a nonlinear transformation to a differential equation, in which the nonlinear term represents a small perturbation. The general procedure of solution for the corresponding nonlinear initial-boundary-value problem is then established by means of the method of successive approximations. Uniqueness and convergence of the analytical solution are discussed. As applications, the temporal change of an initial distribution of electrons and ions is discussed for a finite box system and an infinitely extended system, respectively.

PROBLEM

In the evaluation of experiments concerned with the measurement of ionization and recombination coefficients, one has in general to consider the simultaneous concentration changes of the reacting particles due to diffusion and convection. The corresponding initial-boundary-value problem is of general interest in plasma kinetics[1–3] and radiation chemistry.[4–6] This problem has been treated analytically in the steady-state case by Wilhelm[7] and numerically in the time-dependent case (without a first-order reaction) by Gray and Kerr[8] and Reinhardt.[9] A comprehensive discussion of a variety of nonlinear initial-boundary-value problems of second order has been given recently by Montroll[10] with the motivation of clarifying the underlying mathematical principles for their analysis. In the following, the nonlinear initial-boundary-value problem for convection and diffusion, ionization, and recombination is subject to a nonlinear transformation and subsequently solved by the method of successive approximations.

Mathematically, the initial-boundary-value problem for convection (flow field $\mathbf{v}$) and collective diffusion (diffusion coefficient D)[11] of the electrons and ions in a partially ionized plasma in presence of ionization (ionization coefficient ϵ)[12] and recombination (recombination coefficient α)[13] reactions is described by

$$\frac{\partial n}{\partial t} + \mathbf{v}\cdot\nabla n = D\nabla^2 n + \epsilon n - \alpha n^2, \qquad \mathbf{r} \neq \mathbf{s},\ t \neq 0, \tag{1}$$

where

$$n(\mathbf{r}, t)_{t=0} = n_0(\mathbf{r}), \qquad \mathbf{r} \neq \mathbf{s}, \tag{2}$$

and

$$n(\mathbf{r}, t)_{\mathbf{r}=\mathbf{s}} = \phi(\mathbf{s}), \qquad t \neq 0, \tag{3}$$

are the initial and boundary conditions ($\mathbf{s}$ = position vector of boundary). The electron-ion density $n(\mathbf{r}, t)$ is assumed to be small compared to the density of the homogeneous background gas (α, D, ϵ = const).

NONLINEAR TRANSFORMATION

In absence of diffusion ($D = 0$) and convection ($\mathbf{v} = \mathbf{0}$) a given particle density $n(0)$ would change with time as a result of ionization and recombination according to $n(t) = n(0)e^{\epsilon t}/[1 - (\alpha/\epsilon)(1 - e^{\epsilon t})n(0)]$. In presence of diffusion ($D \neq 0$) and convection ($\mathbf{v} \neq 0$), therefore, a trial transformation is attempted in the form

$$n(\mathbf{r}, t) = \frac{u(\mathbf{r}, t)e^{\epsilon t}}{[1 - (\alpha/\epsilon)(1 - e^{\epsilon t})u(\mathbf{r}, t)]} \tag{4}$$

with

$$\frac{\partial n}{\partial t} = \left(\frac{\partial u}{\partial t} + \epsilon u - \alpha u^2\right)\frac{e^{\epsilon t}}{[1 - (\alpha/\epsilon)(1 - e^{\epsilon t})u]^2}, \tag{5}$$

$$\nabla n = \frac{\nabla u e^{\epsilon t}}{[1 - (\alpha/\epsilon)(1 - e^{\epsilon t})u]^2}, \tag{6}$$

$$\nabla^2 n = \frac{\nabla^2 u e^{\epsilon t}}{[1 - (\alpha/\epsilon)(1 - e^{\epsilon t})u]^2} + \frac{2(\alpha/\epsilon)(1 - e^{\epsilon t})e^{\epsilon t}(\nabla u)^2}{[1 - (\alpha/\epsilon)(1 - e^{\epsilon t})u]^3}. \tag{7}$$

By substitution of Eqs. (4)–(7), the initial-boundary-value problem defined by Eqs. (1)–(3) is reduced to

$$\frac{\partial u}{\partial t} + \mathbf{v}\cdot\nabla u = D\nabla^2 u - \sigma, \qquad \mathbf{r} \neq \mathbf{s},\ t \neq 0, \tag{8}$$

where

$$u(\mathbf{r}, t)_{t=0} = n_0(\mathbf{r}), \qquad \mathbf{r} \neq \mathbf{s}, \tag{9}$$

$$u(\mathbf{r}, t)_{\mathbf{r}=\mathbf{s}} = \frac{\phi(\mathbf{s})e^{-\epsilon t}}{[1 - (\alpha/\epsilon)(1 - e^{-\epsilon t})\phi(\mathbf{s})]}, \qquad t \neq 0, \tag{10}$$

and

$$\sigma \equiv -\frac{2(\alpha/\epsilon)D(1-e^{\epsilon t})(\nabla u)^2}{[1-(\alpha/\epsilon)(1-e^{\epsilon t})u]}, \quad \mathbf{r}\neq\mathbf{s},\ t\neq 0. \tag{11}$$

The nonlinear transformation is generally valid, since Eq. (4) provides a unique interrelation between $n(\mathbf{r},t)\geq 0$ and $u(\mathbf{r},t)\geq 0$, which is free from singularities for any point $\{\mathbf{r},t\}$.

A comparison of Eqs. (8) and (1) indicates that the nonlinear term $\sigma(\mathbf{r},t)$ is negligibly small in Eq. (8) for any time $0\leq t\leq\infty$ if $\delta(\mathbf{r},t)=\sigma/(\alpha u^2 e^{\epsilon t})$ is small compared to one, i.e., if for any interior point $\mathbf{r}\neq\mathbf{s}$,

$$\delta(\mathbf{r},t)=\frac{2(D/\epsilon)(1-e^{-\epsilon t})(\nabla \ln u)^2}{[1-(\alpha/\epsilon)(1-e^{\epsilon t})u]}\ll 1. \tag{12}$$

It is seen that $\delta(r,t)\ll 1$ for intermediate times $0<t<\infty$, if $u(\mathbf{r},t)$ is an analytically well-behaved function in the interior region excluding the boundaries $\mathbf{r}=\mathbf{s}$ where $u(\mathbf{r},t)$ may be discontinuous.[14] In particular,

$$\lim_{t\to 0}\delta(\mathbf{r},t)=0, \qquad \lim_{t\to\infty}\delta(\mathbf{r},t)=0. \tag{13}$$

SUCCESSIVE APPROXIMATION

The transformed initial-boundary-value problem, Eqs. (8)–(10), can now be solved analytically by means of the method of successive approximations. In this approach, the small nonlinear term $\sigma(\mathbf{r},t)$ is treated as a perturbation.

νth Approximation, $u(\mathbf{r},t)=u_\nu(\mathbf{r},t)$. The initial-boundary-value problem defined by Eqs. (8)–(10) reduces by evaluating the perturbation $\sigma(\mathbf{r},t)$ in the $(\nu-1)$th approximation to

$$\begin{aligned}
&\frac{\partial u_\nu}{\partial t}+\mathbf{v}\cdot\nabla u_\nu = D\nabla^2 u_\nu-\sigma_\nu,\\
&\sigma_\nu=-\frac{2(\alpha/\epsilon)D(1-e^{\epsilon t})(\nabla u_{\nu-1})^2}{[1-(\alpha/\epsilon)(1-e^{\epsilon t})u_{\nu-1}]},\\
&u_\nu(\mathbf{r},t)_{t=0}=n_0(\mathbf{r}),\\
&u_\nu(\mathbf{r},t)_{\mathbf{r}=\mathbf{s}}=\frac{\phi(\mathbf{s})e^{-\epsilon t}}{[1-(\alpha/\epsilon)(1-e^{-\epsilon t})\phi(\mathbf{s})]}.
\end{aligned} \tag{14}$$

This is a linear, parabolic problem with a source term $\sigma_\nu[u_{\nu-1}(\mathbf{r},t),t]$ known from the $(\nu-1)$th approximation for $\nu\geq 1$, while in the *zeroth approximation*, $\nu=0$,

$$\sigma_0\equiv 0. \tag{15}$$

In any approximation, $\nu\geq 0$, the original function $n(\mathbf{r},t)$ is given in terms of $u_\nu(\mathbf{r},t)$ by Eq. (4). Thus, the reduction of the nonlinear initial-boundary-value problem to a linear one is established, i.e., its analytical solution.[15]

APPLICATIONS

As an illustration, the method of solution is applied to initial-boundary-value problems for a three-dimensional box and an infinite spherical system, respectively. Convective flow fields are assumed to be absent ($\mathbf{v}=0$), and only the zeroth and first approximations will be evaluated. Higher approximations are not required, since generally the zeroth approximation is within the experimental uncertainties.

A. Finite System

The temporal development of an initial ionization distribution $n_0(x,y,z)$ in a box with side lengths a,b,c and homogeneous particle densities at the walls $[n(\mathbf{s},t)=\phi_0]$ is described by the (transformed) initial-boundary-value problem

$$\begin{aligned}
&\frac{\partial u}{\partial t}=D\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right)-\sigma,\\
&\sigma \equiv -2\frac{\alpha}{\epsilon}D(1-e^{\epsilon t})\left(\frac{(\partial u/\partial x)^2+(\partial u/\partial y)^2+(\partial u/\partial z)^2}{1-(\alpha/\epsilon)(1-e^{\epsilon t})u}\right),
\end{aligned} \tag{16}$$

where

$$u(x,y,z,t)_{t=0}=n_0(x,y,z), \tag{17}$$

$$\left.\begin{aligned}
&u(x,y,z,t)_{x=0,a}=\mu(t)\\
&u(x,y,z,t)_{y=0,b}=\mu(t)\\
&u(x,y,z,t)_{z=0,c}=\mu(t)
\end{aligned}\right\}\quad \mu(t)\equiv\frac{\phi_0 e^{-\epsilon t}}{1-(\alpha/\epsilon)(1-e^{-\epsilon t})\phi_0}, \tag{18}$$

and $0\leq x\leq a$, $0\leq y\leq b$, $0\leq z\leq c$. By means of the linear substitution

$$u(x,y,z,t)=\mu(t)+w(x,y,z,t), \tag{19}$$

the initial-boundary-value problem defined in Eqs. (16)–(18) is reduced to

$$\frac{\partial w}{\partial t}=D\nabla^2 w-\left(\sigma+\frac{d\mu}{dt}\right), \tag{20}$$

where

$$w(x,y,z,t)_{t=0}=n_0(x,y,z)-\phi_0 \tag{21}$$

and

$$\begin{aligned}
&w(x,y,z,t)_{x=0,a}=0, \quad w(x,y,z,t)_{y=0,b}=0,\\
&\qquad\qquad w(x,y,z,t)_{z=0,c}=0.
\end{aligned} \tag{22}$$

From Eqs. (20)–(22) and (19), one obtains for $\sigma=\sigma_0=0$ the solution in the *zeroth approximation*, $u=u_0$,

$$\begin{aligned}
u_0=\mu(t)&+\int_0^a\int_0^b\int_0^c[n_0(\xi,\eta,\zeta)-\phi_0]\\
&\times G(x,y,z\,|\,\xi,\eta,\zeta,t)d\xi d\eta d\zeta\\
&\times\int_0^t\left(\frac{d\mu}{dt}\right)_{t=\tau}d\tau\int_0^a\int_0^b\int_0^c G(x,y,z\,|\,\xi,\eta,\zeta,t-\tau)d\xi d\eta d\zeta,
\end{aligned} \tag{23}$$

where $G(x,y,z\,|\,\xi,\eta,\zeta,t)$ is the source function of Eq. (20) [boundary conditions according to Eq. (22)],

$$\begin{aligned}
&G(x,y,z\,|\,\xi,\eta,\zeta,t)\\
&=\frac{2^3}{abc}\sum_{\lambda=1}^{\infty}\sum_{\mu=1}^{\infty}\sum_{\nu=1}^{\infty}\exp\left\{-\left[\left(\frac{\lambda\pi}{a}\right)^2+\left(\frac{\mu\pi}{b}\right)^2+\left(\frac{\nu\pi}{c}\right)^2\right]\right\}Dt\,\sin\frac{\lambda\pi}{a}x\,\sin\frac{\mu\pi}{b}y\,\sin\frac{\nu\pi}{c}z\,\sin\frac{\lambda\pi}{a}\xi\,\sin\frac{\mu\pi}{b}\eta\,\sin\frac{\nu\pi}{c}\zeta.
\end{aligned} \tag{24}$$

Similarly, there results from Eqs. (20)–(22) and (19), by setting $\sigma = \sigma_1 \neq 0$, the solution in the *first approximation*, $u = u_1$,

$$u_1 = u_0 - \int_0^t d\tau \int_0^a\int_0^b\int_0^c \sigma_1(\xi,\eta,\zeta,\tau) \times G(x,y,z\,|\,\xi,\eta,\zeta,t-\tau)d\xi d\eta d\zeta, \quad (25)$$

where

$$\sigma_1 = -2\frac{\alpha}{\epsilon}D(1-e^{\epsilon t})\,\frac{(\partial u_0/\partial x)^2 + (\partial u_0/\partial y)^2 + (\partial u_0/\partial z)^2}{1-(\alpha/\epsilon)(1-e^{\epsilon t})u_0} \quad (26)$$

is given in terms of the solution of the zeroth approximation and $G(x,y,z\,|\,\xi,\eta,\zeta,t)$ by Eq. (24).

According to Eqs. (4), (21), and (24), the particle density $n_0(\mathbf{r},t)$ is in the *zeroth approximation*,

$$n(\mathbf{r},t) = \frac{u_0(\mathbf{r},t)e^{\epsilon t}}{1-(\alpha/\epsilon)(1-e^{\epsilon t})u_0(\mathbf{r},t)}, \quad (27)$$

where

$$u_0(\mathbf{r},t) = \sum_{\lambda=1}^{\infty}\sum_{\mu=1}^{\infty}\sum_{\nu=1}^{\infty} A_{\lambda\mu\nu} \times \exp(-\gamma_{\lambda\mu\nu}t)\sin\frac{\lambda\pi}{a}x\sin\frac{\mu\pi}{b}y\sin\frac{\nu\pi}{c}z + \sum_{\lambda=1}^{\infty}\sum_{\mu=1}^{\infty}\sum_{\nu=1}^{\infty} B_{\lambda\mu\nu}\gamma_{\lambda\mu\nu}S_{\lambda\mu\nu}(t)\sin\frac{\lambda\pi}{a}x\sin\frac{\mu\pi}{b}y\sin\frac{\nu\pi}{c}z, \quad (28)$$

$$A_{\lambda\mu\nu} \equiv \frac{2^3}{abc}\int_0^a\int_0^b\int_0^c n_0(\xi,\eta,\zeta)\sin\frac{\lambda\pi}{a}\xi\sin\frac{\mu\pi}{b}\eta\sin\frac{\nu\pi}{c}\zeta\,d\xi d\eta d\zeta, \quad (29)$$

$$B_{\lambda\mu\nu} \equiv \left(\frac{2}{\pi}\right)^3\frac{[1-(-1)^{\lambda}]}{\lambda}\frac{[1-(-1)^{\mu}]}{\mu}\frac{[1-(-1)^{\nu}]}{\nu}, \quad (30)$$

$$S_{\lambda\mu\nu}(t) \equiv \phi_0\exp(-\gamma_{\lambda\mu\nu}t)\int_0^t\left(1-\frac{\alpha}{\epsilon}(1-e^{-\epsilon\tau})\phi_0\right)^{-1} \times \exp[(\gamma_{\lambda\mu\nu}-\epsilon)\tau]d\tau \quad (31)$$

and

$$\gamma_{\lambda\mu\nu} \equiv [(\lambda\pi/a)^2 + (\mu\pi/b)^2 + (\nu\pi/c)^2]D. \quad (32)$$

It is noted that the $B_{\lambda\mu\nu}$ are the Fourier coefficients of the function which is $f(\mathbf{r}) = 1$ within[14] $0 < x < a$, $0 < y < b$, $0 < z < c$:

$$\sum_{\lambda=1}^{\infty}\sum_{\mu=1}^{\infty}\sum_{\nu=1}^{\infty} B_{\lambda\mu\nu}\sin\frac{\lambda\pi}{a}x\sin\frac{\mu\pi}{b}y\sin\frac{\nu\pi}{c}z = 1, \quad \mathbf{r} \neq \mathbf{s}. \quad (33)$$

The characteristics of the solution given by Eqs. (27) and (28) depend in general on whether (i) the boundary conditions correspond to an equilibrium or nonequilibrium state and (ii) the ionization coefficient ϵ is smaller, equal, or larger than the first eigenvalue

$$\gamma_{111} = [(\pi/a)^2 + (\pi/b)^2 + (\pi/c)^2]D. \quad (34)$$

The corresponding steady-state solutions $n_\infty(\mathbf{r})$ are obtained from the respective solutions $n(\mathbf{r},t)$ by the operation

$$\lim_{t\to\infty} n(\mathbf{r},t) = n_\infty(\mathbf{r}). \quad (35)$$

It should be noted that this $\lim n_\infty(\mathbf{r})$ is not necessarily applicable at the boundaries $\mathbf{r} = \mathbf{s}$, since in general, for mathematical reasons,

$$\lim_{\mathbf{r}\to\mathbf{s}}\left(\lim_{t\to\infty} n(\mathbf{r},t)\right) \neq \lim_{t\to\infty}\left(\lim_{\mathbf{r}\to\mathbf{s}} n(\mathbf{r},t)\right). \quad (36)$$

1. Equilibrium Boundary Condition, $\phi_0 = \epsilon/\alpha$

If $\phi_0 = \epsilon/\alpha$, then $\mu(t) = \phi_0$ and $S_{\lambda\mu\nu}(t) = \phi_0[1 - \exp(-\gamma_{\lambda\mu\nu}t)]/\gamma_{\lambda\mu\nu}$ by Eq. (31). In this case, Eqs. (27) and (28) give as steady-state solution in the *zeroth approximation*,

$$n_\infty(\mathbf{r}) = \epsilon/\alpha, \quad \epsilon \gtrless \gamma_{111}. \quad (37)$$

Hence, an equilibrium density $\phi_0 = \epsilon/\alpha$ as a boundary value leads to a single, homogeneous steady-state distribution $n_\infty(\mathbf{r}) = \phi_0$, $0 \le x \le a$, $0 \le y \le b$, $0 \le z \le c$, for any $\epsilon \lessgtr \gamma_{111}$.

2. Nonequilibrium Boundary Conditions, $0 < \phi_0 \neq \epsilon/\alpha$

If $0 < \phi_0 \neq \epsilon/\alpha$, then the integral $S_{\lambda\mu\nu}(t)$ is nontrivial, and has by Eq. (31) the asymptotic properties

$$\lim_{t\to\infty} S_{\lambda\mu\nu}(t) = 0, \quad \epsilon \lessgtr \gamma_{\lambda\mu\nu}, \quad (38)$$

$$\lim_{t\to\infty}[e^{\epsilon t}S_{\lambda\mu\nu}(t)] = \frac{\phi_0}{(\gamma_{\lambda\mu\nu}-\epsilon)[1-(\alpha/\epsilon)\phi_0]}, \quad \epsilon < \gamma_{\lambda\mu\nu}, \quad (39)$$

$$\lim_{t\to\infty}[e^{\epsilon t}S_{\lambda\mu\nu}(t)] = \infty, \quad \epsilon \ge \gamma_{\lambda\mu\nu}. \quad (40)$$

Accordingly, Eqs. (27) and (28) give as steady-state solutions in the *zeroth approximation*,

$$n_\infty(\mathbf{r}) = \frac{\phi_0}{[1-(\alpha/\epsilon)\phi_0]}\sum_{\lambda=1}^{\infty}\sum_{\mu=1}^{\infty}\sum_{\nu=1}^{\infty}B_{\lambda\mu\nu}\frac{\gamma_{\lambda\mu\nu}}{(\gamma_{\lambda\mu\nu}-\epsilon)} \times \sin\frac{\lambda\pi}{a}x\sin\frac{\mu\pi}{b}y\sin\frac{\nu\pi}{c}z \times \left(1+\frac{(\alpha/\epsilon)\phi_0}{[1-(\alpha/\epsilon)\phi_0]}\sum_{\lambda=1}^{\infty}\sum_{\mu=1}^{\infty}\sum_{\nu=1}^{\infty}B_{\lambda\mu\nu}\frac{\gamma_{\lambda\mu\nu}}{(\gamma_{\lambda\mu\nu}-\epsilon)} \times \sin\frac{\lambda\pi}{a}x\sin\frac{\mu\pi}{b}y\sin\frac{\nu\pi}{c}z\right)^{-1}, \quad \epsilon < \gamma_{111}, \quad (41)$$

$$n_\infty(\mathbf{r}) = \epsilon/\alpha, \quad 0 < x < a,\ 0 < y < b,\ 0 < z < c, \quad \epsilon \ge \gamma_{111};\ n_\infty(\mathbf{s}) = \lim_{\mathbf{r}\to\mathbf{s}} n(\mathbf{r},t) = \phi_0. \quad (42)$$

Hence, a nonequilibrium density $\phi_0 \neq \epsilon/\alpha$ as a boundary value gives rise to a steady-state distributions $n_\infty(\mathbf{r})$, which are inhomogeneous within $0 \le x \le a$, $0 \le y \le b$, $0 \le z \le c$ and different in the cases $\epsilon < \gamma_{111}$ and $\epsilon \ge \gamma_{111}$.

The steady-state solutions in Eq. (41) and (42) satisfy the boundary condition $n_\infty(\mathbf{s}) = \phi_0$ as can be shown by means of Eq. (33); e.g., the relevant expression in Eq. (41) becomes

$$\lim_{\mathbf{r}\to\mathbf{s}}\sum_{\lambda=1}^{\infty}\sum_{\mu=1}^{\infty}\sum_{\nu=1}^{\infty}B_{\lambda\mu\nu}\frac{\gamma_{\lambda\mu\nu}}{\gamma_{\lambda\mu\nu}-\epsilon}\sin\frac{\lambda\pi}{a}x\sin\frac{\mu\pi}{b}y\sin\frac{\nu\pi}{c}z = \lim_{\mathbf{r}\to\mathbf{s}}\left(1+\epsilon\sum_{\lambda=1}^{\infty}\sum_{\mu=1}^{\infty}\sum_{\nu=1}^{\infty}\frac{B_{\lambda\mu\nu}}{\gamma_{\lambda\mu\nu}-\epsilon} \times \sin\frac{\lambda\pi}{a}x\sin\frac{\mu\pi}{b}y\sin\frac{\nu\pi}{c}z\right) = 1.$$

3. Nonequilibrium Boundary Condition, $\phi_0 = 0$

If $\phi_0 = 0$, then $\mu(t) = 0$ and $S_{\lambda\mu\nu}(t) = 0$ by Eq. (31). In this case, Eqs. (27) and (28) give as steady-state solutions in the *zeroth approximation*,

$$n_\infty(\mathbf{r}) = 0, \quad \epsilon < \gamma_{111}, \quad (43)$$

$$n_\infty(\mathbf{r}) = \left(A_{111} \sin\frac{\pi}{a}x \sin\frac{\pi}{b}y \sin\frac{\pi}{c}z\right) \times \left(1 + \frac{\alpha}{\epsilon}A_{111} \sin\frac{\pi}{a}x \sin\frac{\pi}{b}y \sin\frac{\pi}{c}z\right)^{-1}, \quad \epsilon = \gamma_{111}, \tag{44}$$

$$n_\infty(\mathbf{r}) = \epsilon/\alpha \quad (0 < x < a,\ 0 < y < b,\ 0 < z < c), \quad \epsilon > \gamma_{111};\ n_\infty(\mathbf{s}) = \lim_{\mathbf{r}\to\mathbf{s}} n(\mathbf{r}, t) = 0. \tag{45}$$

Hence, the special nonequilibrium boundary value $\phi_0 = 0$ leads to steady-state distributions $n_\infty(\mathbf{r})$, which are different in all possible cases $\epsilon < \gamma_{111}$, $\epsilon = \gamma_{111}$, and $\epsilon > \gamma_{111}$.

In an analogous way, the solution in the first approximation [Eq. (4) with Eq. (25)] may be discussed.

B. Infinite System

The temporal development of an initial ionization distribution $n_0(r)$ of the form of a Gaussian

$$n_0(\mathbf{r}) = [N_0/(\pi b^2)^{3/2}]e^{-r^2/b^2} \tag{46}$$

(N_0, b = constant parameters) in an infinite medium is described by the (transformed) initial-boundary-value problem

$$\frac{\partial u}{\partial t} = D\left(\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r}\right) - \sigma,$$

$$\sigma \equiv -2\frac{\alpha}{\epsilon}D(1 - e^{\epsilon t})\left(\frac{(\partial u/\partial r)^2}{1 - (\alpha/\epsilon)(1 - e^{\epsilon t})u}\right), \tag{47}$$

where

$$u(r, t)_{t=0} = n_0(r), \tag{48}$$

$$\lim_{r\to\infty} u(r, t) = 0, \tag{49}$$

and $0 \le r \le \infty$, $r \equiv \sqrt{x^2 + y^2 + z^2}$. From Eqs. (46)–(49), one obtains for $\sigma = \sigma_0 = 0$ the solution in the *zeroth approximation*, $u = u_0$,

$$u_0(r, t) = \int_{-\infty}^{+\infty}\!\!\int\!\!\int n_0(\xi, \eta, \zeta)\tilde{G}(x, y, z \mid \xi, \eta, \zeta, t)d\xi d\eta d\zeta \tag{50}$$

where $\tilde{G}(x, y, z \mid \xi, \eta, \zeta, t)$ is the source function of Eq. (47) [boundary conditions according to Eq. (49)],

$$\tilde{G}(x, y, z \mid \xi, \eta, \zeta, t) = \frac{\exp\{-[(x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2]/4Dt\}}{(4\pi Dt)^{3/2}}. \tag{51}$$

Similarly, there results from Eqs. (47)–(49), by setting $\sigma = \sigma_1 \neq 0$, the solution in the *first approximation*, $u = u_1$,

$$u_1 = u_0 - \int_0^t d\tau \int_0^a\int_0^b\int_0^c \sigma_1(\xi, \eta, \zeta, \tau) \times \tilde{G}(x, y, z \mid \xi, \eta, \zeta, t - \tau)d\xi d\eta d\zeta, \tag{52}$$

where

$$\sigma_1 = -2\frac{\alpha}{\epsilon}D(1 - e^{\epsilon t})\left(\frac{(\partial u_0/\partial r)^2}{1 - (\alpha/\epsilon)(1 - e^{\epsilon t})u_0}\right) \tag{53}$$

is given in terms of the solution u_0 of the zeroth approximation and $\tilde{G}(x, y, z \mid \xi, \eta, \zeta, t)$ by Eq. (51).

According to Eqs. (4), (50), and (51), the particle density is in the *zeroth approximations:*

$$n(r, t) = \frac{u_0(r, t)e^{\epsilon t}}{1 - (\alpha/\epsilon)(1 - e^{\epsilon t})u_0(r, t)}, \tag{54}$$

where

$$u_0(r, t) = \frac{N_0}{[\pi(4Dt + b^2)]^{3/2}}e^{-r^2/(4Dt+b^2)}. \tag{55}$$

Equations (54) and (55) indicate that only a single, homogeneous steady-state solution exists,

$$n_\infty(r) = \epsilon/\alpha, \tag{56}$$

which corresponds to the state of thermodynamic equilibrium. This result is generally valid for systems of infinite extension.

In an analogous way, the solution in the first approximation [Eq. (4) with Eq. (52)] may be discussed.

UNIQUENESS AND CONVERGENCE

The nonlinear initial-boundary-value problem describing the temporal change of electron–ion distributions by collective convection, diffusion, ionization, and recombination [Eqs. (1)–(3)] has been reduced by means of a nonlinear transformation [Eq. (4)] to an initial-boundary-value problem, in which the nonlinear term is small. It has been shown that the solution of the transformed initital-boundary-value problem [Eqs. (8)–(10)] can be obtained by means of the method of successive approximations from linear initial-boundary-value problems with a known source term [Eqs. (13)–(15)].

From the physical nature of the initial-boundary-value problem [Eqs. (1)–(3)], one can infer that it has a unique solution. In order to prove mathematically the uniqueness, the function $n(\mathbf{r}, t)$ is expanded in Sturm–Liouville characteristic functions. This leads to an infinite system of coupled, nonlinear integral equations for the time-dependent expansion coefficients, for which only a single solution can be shown to exist.[16]

In applications, the number of successive approximations to be evaluated is determined by the accuracy required. The sequence of successive approximations,

$$\{u_\nu(\mathbf{r}, t)\} = u_0(\mathbf{r}, t), u_1(\mathbf{r}, t), \ldots, u_\nu(\mathbf{r}, t), \ldots,$$

converges towards the solution $u(\mathbf{r}, t)$ if a $N(\epsilon \mid \mathbf{r}, t)$ exists to every small $\epsilon > 0$ such that the condition

$$|u_\nu(\mathbf{r}, t) - u_\mu(\mathbf{r}, t)| < \epsilon, \quad \text{for } \nu, \mu > N(\epsilon \mid \mathbf{r}, t)$$

is satisfied.[17] If G designates the Green's function of the homogeneous problem, the λth approximation is of the general form

$$u_\lambda(\mathbf{r}, t) = u_0(\mathbf{r}, t) - \int_0^t d\tau \iiint_{\mathfrak{R}} \sigma_\lambda(\xi, \eta, \zeta, \tau) \times G(x, y, z \mid \xi, \eta, \zeta, t - \tau)d\xi d\eta d\zeta.$$

The Green's function G can be assumed to be bounded in the space $\mathfrak{R}$ for $0 \le t < \infty$. The convergence criterion reduces then, for every small $\epsilon' > 0$, to

$$|\sigma_\nu(\mathbf{r}, t) - \sigma_\mu(\mathbf{r}, t)| < \epsilon', \quad \text{for } \nu, \mu > N(\epsilon' \mid \mathbf{r}, t).$$

It follows that the successive approximations con-

verge since $0 \leq |\sigma_\lambda(\mathbf{r}, t)| < \epsilon'$ if $0 \leq \delta(\mathbf{r}, t) \ll 1$ for $0 \leq t < \infty$ as shown in Eq. (12).

Initial-boundary-value problems with boundary conditions of the type considered (first kind) have solutions, which are discontinuous across the boundaries $\mathbf{r} = \mathbf{s}$ for all times $0 < t < \infty$, since the Fourier series representations of $u(\mathbf{r}, t)$ [and, hence, also $n(\mathbf{r}, t)$] are discontinuous functions of $\mathbf{r}$ as $\mathbf{r}$ approaches $\mathbf{s}$ if $\phi_0 \neq 0$.[14] Furthermore, in the limit $t = \infty$, the solutions for $u(\mathbf{r}, t)$ are discontinuous functions of $\mathbf{r}$ as $\mathbf{r}$ approaches $\mathbf{s}$ in the case $\epsilon > \gamma_{111}$ [and hence also $n(\mathbf{r}, t)$ if $\epsilon > \gamma_{111}$ and $\phi_0 \neq \epsilon/\alpha$], since the mathematical result of the successive limiting processes $t \to \infty$ and $\mathbf{r} \to \mathbf{s}$ depends in general on their sequence [Eq. (36)]. Such discontinuities do not imply any mathematical contradictions,[14] quite apart from the fact that the infinite time point $t = \infty$ is physically not realizable.[18]

---

* Supported in part by the U.S. Office of Naval Research.

1 E. W. MacDaniel, *Collision Phenomena in Ionized Gases* (Wiley, New York, 1964).
2 H. W. Drawin, Z. Phys. **225**, 470 (1969); 483 (1969).
3 J. Rajchl, Bull. Astron. Inst. Czech. **16**, 282 (1965); **20**, 363 (1969); **21**, 262 (1970).
4 A. Kupperman, "Diffusion Kinetics in Radiation Chemistry" in *Actions chimiques et biologiques des radiations, 5me serie*, edited by M. Haissinsky, (Masson, Paris, 1961).
5 H. A. Schwarz, J. Phys. Chem. **73**, 1928 (1969).
6 R. Schiller, "The Early Stages of Radiation-Produced Charge Carriers," in *Current Topics in Radiation Research*, edited by M. Ebert and A. Howard (North-Holland, Amsterdam, 1970), Vol. 6.
7 H. E. Wilhelm, J. Math. Phys. **11**, 824 (1970).
8 E. P. Gray and D. E. Kerr, Ann. Phys. **17**, 276 (1962).
9 W. A. Reinhardt, NASA TN D-2126 (1964).
10 E. W. Montroll, "Lectures on Nonlinear Rate Equations, Especially Those with Quadratic Nonlinearities," in *Lectures in Theoretical Physics*, edited by A. O. Barut and W. E. Britton (Gordon & Breach, New York, 1968).
11 W. Schottky, Physik Z. **25**, 324 (1924).
12 G. Elwert, Z. Naturforsch. **7a**, 432 (1952).
13 L. P. Pitaevskii, Soviet Phys.—JETP **15**, 919 (1962). I. S. Veselovskii, Soviet Phys.—JETP **25**, 687 (1967).
14 B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1956), pp. 268–72.
15 For the solution of linear boundary-value-problems with non-uniform, time-dependent, and mixed boundary conditions, see: A. Datzeff, *Sur le probleme de la propagation de la chaleur dans les corps solides* (Ganthier-Villars, Paris, 1963); L. V. Kantorovich and V. I. Krylov, *Approximate Methods of Higher Analysis* (Interscience, New York, 1964); H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids* (Claredon, Oxford, 1968); C. H. Muntz, Math. Z. **38**, 323 (1934); T. S. Chow, Z. Angew. Math. Phys. **8**, 478 (1957); E. F. Afanasev, Differential Equations **1**, 509 (1965); Y. F. Rutner and L. Skryabina, Differential Equations **2**, 570 (1966); R. B. Kelman, J. Math. Phys. **7**, 333 (1967); G. A. Grinberg, Prikl. Mat. Mekh. **31**, 193 (1967).
16 M. R. Siddiqi, Proc. Cambridge Phil. Soc. **31**, 195 (1935).
17 S. G. Mikhlin, *Integral Equations* (Pergamon, New York, 1957).
18 B. M. Budak, A. A. Samarskii, and A. N. Tikhonov, *Mathematical Physics* (Macmillan, New York, 1964).

Timelike Asymptotic Series of Almost Local Fields

A. H. Völkel
Institut für Theoretische Physik, Freie Universität Berlin, 1 Berlin 33, Arnimallee 3, Germany
(Received 12 November 1970)

Timelike asymptotic series for many-particle matrix elements of products of almost local fields are derived which generalize and extend the Araki–Haag series of quasilocal operators. An interpretation of the asymptotic leading terms in the form of contributions from disconnected intermediate particle states is given. A discussion of the dependence of the asymptotic leading terms on the smearing in the space variables is presented.

I. INTRODUCTION

Within the framework of quantum field theory, the connection between interacting fields and scattering matrix elements is established by the weak asymptotic condition via the *LSZ* reduction technique.[1]

Let $|0\rangle$ denote the vacuum state and

$$|M_i a_i \gamma_i\rangle = a_{\text{in}}^{\text{out}}(M_i, a_i, \gamma_i)^* |0\rangle = \int \frac{d^3p}{2\omega_i(\mathbf{p})} \psi_{M_i a_i}(\mathbf{p}) a_{\text{in}}^{\text{out}}(M_i, \mathbf{p}, \gamma_i)^* |0\rangle, \quad (1)$$

a discrete one-particle wave packet state with mass M_i and internal (symmetry) quantum numbers γ_i.[2] Here a_{ex} are the asymptotic creation and destruction operators satisfying the canonical (anti) commutation relations

$$[a_{\text{ex}}(M_i, \mathbf{p}, \gamma), a_{\text{ex}}(M_j, \mathbf{k}, \epsilon)^*]_{\mp} = 2\omega_i(\mathbf{p})\delta(\mathbf{p}-\mathbf{k})\delta_{ij}\delta_{\gamma\epsilon}, \quad (2)$$
$$\omega_i(\mathbf{p}) = : + \sqrt{(M_i^2 + \mathbf{p}^2)^{1/2}}.$$

$\{\psi_{M_i a_i}(\mathbf{p})\}$ is a set of S_3 functions complete and orthonormal in L_2 with respect to the measure $\omega_i(p)^{-1} d^3p$:

$$\int \frac{d^3p}{2\omega_i(p)} \psi_{M_i a}(\mathbf{p})^* \psi_{M_i b}(\mathbf{p}) = \delta_{ab},$$
$$\sum_a \psi_{Ma}(\mathbf{p}) \psi_{Ma}(\mathbf{k})^* = 2\omega_M(\mathbf{p})\delta(\mathbf{p}-\mathbf{k}). \quad (3)$$

Finally let D_{ex} denote the domain in the Hilbert space $\mathfrak{h}$, which is given by finite linear combinations of many particle states with nonoverlapping wavefunctions in velocity space. A set of wavefunctions $\{\psi_1(\mathbf{p}_1), \ldots, \psi_r(\mathbf{p}_r)\}$ is said to be nonoverlapping in velocity space, if for all p_i from the support of $\psi_i(\mathbf{p}_i)$, we have pairwise

$$\omega_j(\mathbf{p}_j)\mathbf{p}_i \neq \omega_i(\mathbf{p}_i)\mathbf{p}_j, \quad i \neq j, i, j = 1, \ldots, r. \quad (4)$$

With these notations the weak asymptotic conditions reads[1]: If $A(f^{x^0}) = \int d^4y f(y^0 - x^0, \mathbf{y}) A(y)$ is an almost local field,[3] with $\langle 0|A(f)|0\rangle = 0$, if one of the state vectors Ψ, Φ is from D_{ex} and if the point $x = (x^0, x)$ moves to infinity in a timelike direction, then

$$\langle \Psi | A(f^{x^0}) | \Phi \rangle \xrightarrow[x^0 \to \pm\infty]{} \sum_i \{\langle 0|A(f^{x^0})|i\rangle \langle \Psi | a_{\text{in}}^{\text{out}}(i) | \Phi\rangle + \langle i|A(f^{x^0})|0\rangle \langle \Psi | a_{\text{in}}^{\text{out}}(i)^* | \Phi\rangle\}. \quad (5)$$

Here the sum runs over all discrete one-particle states and f is an element from S_4.

Araki and Haag[4] recognized the expression on the right-hand side of (5) to be the leading term (if $\langle 0|A(f)|0\rangle = 0$) of an asymptotic series which explicitely reads: If $A(f)$ is an almost local field and $\Psi_{\text{ex}}, \Phi_{\text{ex}}$ are both from D_{ex} then we have

$$\begin{aligned}\langle \Psi_{\text{in}}^{\text{out}} | A(f^{x^0}) | \Phi_{\text{in}}^{\text{out}} \rangle &= \langle 0|A(f^{x^0})|0\rangle \langle \Psi_{\text{in}}^{\text{out}} \| \Phi_{\text{in}}^{\text{out}} \rangle \\ &+ \sum_i \{\langle i|A(f^{x^0})0\rangle \langle \Psi_{\text{in}}^{\text{out}} | a_{\text{in}}^{\text{out}}(i)^* | \Phi_{\text{in}}^{\text{out}} \rangle \\ &+ \langle 0|A(f^{x^0})|i\rangle \langle \Psi_{\text{in}}^{\text{out}} | a_{\text{in}}^{\text{out}}(i) | \Phi_{\text{in}}^{\text{out}} \rangle\} \\ &+ \sum_{i,j} \langle i|A(f^{x^0})|j\rangle^T \langle \Psi_{\text{in}}^{\text{out}} | a_{\text{in}}^{\text{out}}(i)^* a_{\text{in}}^{\text{out}}(j) | \Phi_{\text{in}}^{\text{out}} \rangle \\ &+ R(\Psi, \Phi, f^{x^0}), \end{aligned} \quad (6)$$

with the remainder R being bounded in x^0 by

$$|R(\Psi, \Phi, f^{x^0})| < B(\Psi, \Phi, f)(x^0)^{-N} \quad (7)$$

for $x^0 \lesssim 0$ and any arbitrary positive integer N. The distribution of indices (6) is out for $x^0 > 0$ and in for $x^0 < 0$.

The first term in (6) is a constant by translational invariance and often put equal to zero. Since the matrix elements of $A(f^x)$ between vacuum and one-particle states are smooth solution of the Klein–Gordon equation, the curly bracket in (6) behaves asymptotically like $(x^0)^{-3/2}$.

Finally by means of the stationary phase method, one can show[4] that the double sum in (6) behaves like $(x^0)^{-3}$. In other words by Eq. (6), we have reduced the many-particle matrix elements of a field operator to the three possible types of vacuum and one-particle matrix elements with well-known asymptotic behavior plus a remainder vanishing faster than any inverse polynomial.

Of course the smearing of the field $A(f)$ in the space variables $\mathbf{x}$ can be dropped in (6). According to Borchers[5] almost local fields[6] together with all their derivatives are bounded operator-valued functions in $\mathbf{x}$ after smearing in the time variable x^0. If we drop the smearing in space, then (6) holds uniformly in $\mathbf{x}$.

Furthermore for self-adjoint operators $A_i(t)$, which destroy the vacuum state, it was shown in Ref. 4 that

$$\lim_{t \to \pm\infty} t^{3n} \left\langle \Psi_{\text{ex}} \middle| \prod_{i=1}^{n} A_i(t) \middle| \Phi_{\text{ex}} \right\rangle$$

exists for Ψ_{ex} or $\Phi_{\text{ex}} \in D_{\text{ex}}$ and is essentially given by the matrix element of a product of n-particle density operators.

For a variety of applications which have gained some interest in recent years the conditions under which (6) holds are too restrictive.

For instance, the investigation of equal-time commutation relations can completely reduce to the consideration of expressions of the form[7,8]

$$\langle \Psi | [j^0(g_\epsilon, h^{(i,j,r-i-j)}), j^\nu(0)] | \Phi \rangle, \quad (8)$$

with

$$h^{(i,j,r)}(\mathbf{x}) = (x^1)^i(x^2)^j(x^3)^r,$$

$$i,j,r \geq 0 \quad \text{finite}.$$

In order to connect the matrix elements above to (off-mass shell) scattering amplitudes, one always has to use a generalized Gauss theorem in one or the other form.[9-13]

This means one rewrites the total charges, i.e., their moments, as four-dimensional integrals over the corresponding divergences of the currents, which in turn are interpreted as interpolating meson fields via the principle of Haag[14] and Zimmermann[15]. In a rigorous treatment of this step, one has to calculate by means of expansions of the type (6) the contributions of infinite timelike surfaces.[9,13]

Speculative applications of (6) to cases not covered by the proof in Ref. 4 lead to the resolution of curiosities like frame dependences of sum rules.[9,13]

Therefore, it is worthwhile to generalize the asymptotic series (6) in two directions: (i) To extend (6) to the case of more general "smearing functions" in the space variable $\mathbf{x}$ than S_3; for instance to $0'_{c3}$ distributions and at least polynomials or 0_{M3} functions.[16] (ii) To consider instead of one single field operator $A(f^{x^0})$, a product of fields $\Pi_i A_i(f_i^{x_i^0})$ (which do not necessarily destroy the vacuum state).

Both these generalizations will be the main topic of the present paper. A further interesting problem is the origin of the leading terms in the asymptotic series like (6). By this we mean the interpretation of these terms as contributions from certain intermediate particle states. We will show that they all originate from *disconnected* intermediate particle states.

All our results follow from the general postulates of Wightman fields with the exception of strict locality.[1,17,18] Instead of strict locality we assume only almost locality.[3] In detail we require for the fields: (i) The fields $A_i(f)$ smeared with test functions $f(x)$ from S_4 are operators with a dense domain in a Hilbert space $\mathfrak{h}$. (ii) Translational invariance. (iii) Spectrum condition, i.e., all states in $\mathfrak{h}$ have real masses and positive energies. (iv) Almost locality[3]; roughly this means that the commutator of two fields vanishes in spacelike directions faster than any inverse polynomial.

According to Borchers[5] these assumptions imply the smearing of the fields over the space variables to be superfluous. After smearing in time, the fields $A_i(g;\mathbf{x})$ are operator-valued bounded C^∞-functions[6] in $\mathbf{x}$. Moreover their n-fold truncated vacuum expectation values are from $S_{3(n-1)}$ in the $n-1$ difference variables $\xi_i = \mathbf{x}_i - \mathbf{x}_{i+1} (i = 1, \ldots, n-1)$. This property, called (B) in the following, plays the central part in our proofs. For simplicity we will give the proofs only for scalar particles. The generalizations to other particles are trivial.

II. TIMELIKE ASYMPTOTIC BOUNDS

In this section we are going to prove some timelike asymptotic falloff properties for truncated vacuum expectation values of field operators.

Let $B_i(x)$ be an exact one-particle excitation operator[1] built from an almost local field $A_i(x)$:

$$\begin{aligned} B_i(x) &=: \int d^4y f_{M_i}(y-x) A_i(y) \\ &= \int d^4p e^{ipx} \tilde{f}_{M_i}(p) \tilde{A}_i(p), \end{aligned} \tag{9}$$

with $f_{M_i}(p) \in S_4$ and support f_{M_i} concentrated in a hose around the discrete one-particle mass shell $p^2 = M_i^2$ in the forward cone. It has the properties

$$\begin{aligned} &B_i(x)|0\rangle = 0, \\ &B_i(x)^*|0\rangle = \text{one-particle state}. \end{aligned} \tag{10}$$

Furthermore let $f^{(+)}_{M_i a_i}(x)$ be a smooth positive-frequency solution of the Klein–Gordon equation:

$$\tilde{f}^{(+)}_{M_i a_i}(x) =: \frac{1}{(2\pi)^{3/2}} \int d^4p\, \delta_+(p^2 - M_i^2) \tilde{f}_i^{(+)}(\mathbf{p}) e^{-ipx} \tag{11}$$

with

$$\tilde{f}_i^{(+)}(\mathbf{p}) = \Psi_{M_i a_i}(\mathbf{p}) \langle M_i, \mathbf{p}|B_i(0)^*|0\rangle^{-1} \tag{12}$$

and $\Psi_{M_i a_i}$ specified in the introduction. From B_i and $f^{(+)}_{M_i a_i}$ we form the *LSZ* operator:

$$\begin{aligned} B_i(f_i^{(+)};t) &=: i \int_{t=x^0} d^3x B_i(x)^* \overleftrightarrow{\frac{\partial}{\partial x^0}} f^{(+)}_{M_i a_i}(x) \\ &= \int d^4p \tilde{B}_i(p)^* \frac{\omega_i(\mathbf{p}) + p^0}{2\omega_i(\mathbf{p})} \tilde{f}_i^{(+)}(\mathbf{p}) e^{-i[\omega_i(\mathbf{p}) - p^0]t}, \end{aligned} \tag{13}$$

which when applied to the vacuum creates a time-independent one-particle state

$$\frac{d}{dt} B_i(f_i^{(+)} t)^*|0\rangle = 0. \tag{14}$$

Finally we introduce the following notations for various spaces of functions [16]: (i) 0_{Mn} denotes the space of all polynomial bounded C^∞-functions of n variables. (ii) The Fourier space of 0_{Mn}, that is, the space of all strongly decreasing distributions, is denoted by $0'_{cn}$. (iii) By 0^p_{Mn} we denote the subclass of 0_{Mn} functions which are fourier transforms of $0'_{cn}$ distributions with point support. (0^p_{Mn} contains for instance all polynomials of n variables and the trigonometric functions.)

With these preparations we can now formulate our first theorem

Theorem 1: Let $\{\tilde{f}^{(+)}_{M_i a_i}(\mathbf{P}_i); i = 1, \ldots, r\} \in S_3$ and $\{f^{(+)}_{M_j b_j}(\mathbf{P}_j); j = r+1, \ldots, s+r\} \in S_3$ be two sets of wavefunctions with nonoverlapping support in velocity space. Let $\{A_i(x), i = 1, \ldots, n\}$ be a set of almost local operators and $D(t, \tau_1, \tau_2, x_1^0, \ldots, x_n^0)$ a differential monomial of arbitrary degree. Then we have for all r, s with $r \geq 2$ or $s \geq 2$ and for all $g_i(x^0) \in S_1; h_j(\mathbf{x}_j) \in 0'_{c3} \cup 0^p_{M_3}$:

$$\begin{aligned} &|D\langle 0|\prod_{i=1}^{r} B_i(f_i^{(+)}, t + \tau_1) \\ &\times \prod_{\alpha=1}^{n} A_\alpha(g_\alpha^{t+x_\alpha^0}, h_\alpha) \prod_{j=1}^{s} B_{r+j}(f_{r+j}^{(+)}; t + \tau_2)^* \Big| 0\Big\rangle^T| \end{aligned}$$

$$\leq (1+t^2)^{-N}(1+\tau_1^2)^{L_0}(1+\tau_2^2)^{\bar{L}_0}$$
$$\times \prod_{j=1}^{n}[1+(x_j^0)^2]^{L_j} C(t,\tau_1,\tau_2;x^0\cdots|g\cdots,h\cdots), \tag{15}$$

with N any arbitrary positive integer, $\{\bar{L}_0, L_j\}$ some finite positive integers. $C(\cdots/\cdots)$ is a bounded continuous function in t when τ_i, x_j^0 are kept fixed and vice versa.

Proof: Without loss of generality we assume $s \geq 2$ and $\{f^{(+)}_{M_i b_i}\}$ nonoverlapping. Multiplying out all the differential operators, we find that the left-hand side of (15) is bounded by a finite sum of terms of the form

$$I(t,\tau_1,\tau_2;x_1^0,\ldots,x_n^0) = : \Big|\int \prod_{i=1}^{r} d^3y_i d^4z_i f_i^{(+)}(t+\tau_1,\mathbf{y}_i)^{*()} f_{M_i}(z_i^0 - t - \tau_1, \mathbf{z}_i - \mathbf{y}_i)$$
$$\times \int \prod_{j=1}^{s} d^3y_{r+j} d^4z_{r+j} f_{r+j}^{(+)}(t+\tau_2,\mathbf{y}_{r+j})^{()} f_{M_{r+j}}(z_{r+j} - t - \tau_2, \mathbf{z}_{r+j} - \mathbf{y}_{r+j})^{*()} \int \prod_{\alpha=1}^{n} d^4u_\alpha g_\alpha(u_\alpha^0 - t - x_\alpha^0)^{()} h_\alpha(\mathbf{u}_\alpha)$$
$$\times \langle 0|\prod_{i=1}^{r} A_i(z_i) \prod_{l=1}^{n} A_l(u_l) \prod_{j=1}^{s} A_{r+j}(z_{r+j})^*|0\rangle^T|. \tag{16}$$

Here $()$ indicates possible time derivatives. By translational invariance, the Fourier transform of the truncated vacuum expectation value has the form

$$\Big\langle 0\Big|\prod_{i=1}^{r}\tilde{A}_i(P_i)\prod_{l=1}^{n}\tilde{A}_l(k_l)\prod_{j=1}^{s}A_{r+j}(-P_{r+j})^*\Big|0\Big\rangle^T$$
$$= \delta\Big(\sum_{i=1}^{r}P_i + \sum_{l=1}^{n}k_l - \sum_{j=1}^{s}P_{r+j}\Big)\hat{W}_{r+n+s}(P_1,\ldots,P_r,k_1,\ldots,k_n,P_{r+1},P_{r+1},\ldots,P_{r+s-1})^T, \tag{17}$$

where according to property (B) $\hat{W}_{r+n+s}(\cdots)$ is from $S_{3(r+n+s-1)}$ after smearing in the energy variables with functions from $S_{(r+n+s-1)}$. Introducing everywhere in (16) the Fourier transforms, we obtain

$$I(t,\tau_1,\tau_2;x_1^0,\ldots,x_n^0) = \Big|\int \prod_{j=1}^{n} d^3k_j \tilde{h}_j(\mathbf{k}_j) \int \prod_{i=1}^{r+s-1} d^3p_i$$
$$\times \chi(\tau_1,\tau_2;x_1^0,\ldots,x_n^0|\mathbf{p}_1,\ldots,\mathbf{p}_r,\mathbf{k}_1,\ldots,\mathbf{k}_n,\mathbf{p}_{r+1},\ldots,\mathbf{p}_{r+s-1})$$
$$\times \exp\{-i\Omega(\mathbf{p}_1,\ldots,\mathbf{p}_{r+s-1},\mathbf{k}_1,\ldots,\mathbf{k}_n)t\}\Big|, \tag{18}$$

where we have used the abbreviations

$$\Omega(\mathbf{p}_1,\ldots,\mathbf{p}_{r+s-1},\mathbf{k}_1,\ldots,\mathbf{k}_n) = : \sum_{i=1}^{s-1}\omega_{r+i}(\mathbf{p}_{r+i}) - \sum_{j=1}^{r}\omega_j(\mathbf{p}_j) + \omega_{r+s}\Big(\sum_{i=1}^{r}\mathbf{p}_i + \sum_{l=1}^{n}\mathbf{k}_l - \sum_{j=1}^{s-1}\mathbf{p}_{r+j}\Big), \tag{19}$$

$$\bar{\Omega}(\mathbf{p}_1,\ldots,\mathbf{p}_{r+s-1},\mathbf{k}_1,\ldots,\mathbf{k}_n) = : \sum_{i=1}^{s-1}\omega_{r+i}(\mathbf{p}_{r+i}) + \omega_{r+s}\Big(\sum_{i=1}^{r}\mathbf{p}_i + \sum_{l=1}^{n}\mathbf{k}_l - \sum_{j=1}^{s-1}\mathbf{p}_{r+j}\Big), \tag{20}$$

$$\chi(\tau_1,\tau_2,x_1^0,\ldots,x_n^0|\mathbf{p}_1,\ldots,\mathbf{p}_r,\mathbf{k}_1,\ldots,\mathbf{k}_n,\mathbf{p}_{r+1},\ldots,\mathbf{p}_{r+s-1}) = : \Big(\frac{1}{2\omega_{r+s}}f_{r+s}^{(+)}\Big)^{*()}\Big(\sum_{j=1}^{r}\mathbf{p}_j + \sum_{l=1}^{n}\mathbf{k}_l - \sum_{i=1}^{s-1}\mathbf{p}_{r+i}\Big).$$
$$\times \int \prod_{j=1}^{n} dk_j^0 \tilde{g}(k_j^0)^{()} \int \prod_{i=1}^{r} \frac{dp_i^0}{2\omega_i(\mathbf{p}_i)} \tilde{f}_i^{(+)}(\mathbf{p}_i)^{()} \tilde{f}_{M_i}(p_i)^{()} * \int \prod_{l=1}^{s-1} \frac{dp_{r+l}^0}{2\omega_{r+l}(\mathbf{p}_{r+l})} \tilde{f}_{r+l}^{(+)}(\mathbf{p}_{r+l})^{*()} f_{M_{r+l}}(p_{r+l})^{()}$$
$$\times f_{M_{r+l}}\Big(\sum_{j=1}^{r}p_j + \sum_{l=1}^{n}k_l - \sum_{i=1}^{s-1}p_{r+i}\Big)\hat{W}_{r+n+s}(p_1,\ldots,p_r,k_1,\ldots,k_n,p_{r+1},\ldots,p_{r+s-1})$$
$$\times \exp\Big[i\Big(\sum_{j=1}^{r}(\omega_j(\mathbf{p}_j) - p_j^0)\tau_1 - \sum_{\alpha=1}^{n}k_\alpha^0 x_\alpha^0\Big)\Big] \exp\Big[-i\Big(\bar{\Omega}(\mathbf{p}_1,\ldots,\mathbf{p}_{r+s-1},\mathbf{k}_1,\ldots,\mathbf{k}_n) - \sum_{j=1}^{r}p_j^0 - \sum_{\alpha=1}^{n}k_\alpha^0\Big)\tau_2\Big]. \tag{21}$$

$\chi(\tau_1,\tau_2,x_1^0,\ldots,x_n^0|\mathbf{p}_1,\ldots,\mathbf{p}_r,\mathbf{k}_1,\ldots,\mathbf{k}_n,\mathbf{p}_{r+1},\ldots,\mathbf{p}_{r+s-1})$ may be considered as a continuous linear functional from $S_{r+s+n-1}$ depending on the $3(r+s-1)+4n+2$ parameters τ_i, x_j^0, p_l, k_j. As a function of these parameters it has the following three important properties.

1. It is from $S_{3(r+n+s-1)}$ in all momentum variables and has pairwise nonoverlapping support in the velocity space corresponding to the set $\{\mathbf{p}_{r+1},\ldots,\mathbf{p}_{r+s-1}\}$.

2. It has also pairwise nonoverlapping support in the velocity space corresponding to the set of variables

$$\Big\{\mathbf{p}_{r+1},\ldots,\mathbf{p}_{r+s-1},\sum_{j=1}^{r}\mathbf{p}_j + \sum_{i=1}^{n}\mathbf{k}_i - \sum_{l=1}^{s-1}\mathbf{p}_{r+l}\Big\}.$$

3. It is a polynomial bounded C^∞-function (0_{Mn} function) in the time variables $\{x_\alpha^0, \tau_1, \tau_2 | \alpha = 1,\ldots,n\}$.

Properties 1 and 2 follow from the corresponding property of the wavefunctions $f^{(+)}\cdots(P)$ and the truncated Wightman function $\hat{W}_{r+s+n}$ which due to (B) is from $S_{3(n+r+s-1)}$ after smearing in the energy variables.

The truncated Wightman function $W_{r+n+s}(P_1,\ldots,P_r,k,\ldots,k_n,P_{r+1},\ldots,P_{r+s-1})^T$ is a tempered distri-

bution in the energy variables $P_i^0; k_j^0$. Multiplication by functions $g_j(k_j^0) f \cdots (P_i)\, S_1 \times S_4$ results in a strongly decreasing distribution (from $0'_{c(r+n+s-1)}$) in $\{k_j^0, P_i^0\}$. However, the Fourier transform of a $0'_{c(r+n+s-1)}$ distribution is a $0_{M(r+n+s-1)}$ function. This proves property 3.

If $h_j(\mathbf{x}) \in 0'_{c3}$ then $\tilde{h}_j(\mathbf{P})$ is from 0_{M3}. This means the whole integrand in (18) apart from $\exp\{-i\Omega(\cdots)t\}$ has again the properties 1–3. If $h_j(\mathbf{x}) \in 0^p_{M3}$, then the Fourier $\tilde{h}_j(\mathbf{k})$ is of the form

$$\tilde{h}_j(\mathbf{k}) = \sum_{i=1}^{R} P_i\left(\frac{\partial}{\partial k^l}\right) \delta(\mathbf{k} - \mathbf{a}_{ij}), \tag{22}$$

where $P_i(\partial/\partial k_l)$ are polynomials in the differential operators $\partial/\partial k_l; l = 1, 2, 3$. In this case we can perform the $\mathbf{k}$ integrations in (18) by means of the δ functions. After this operation, the integrand of the remaining $\mathbf{p}$ integrals has again the properties 1–3, with $\mathbf{k}$ replaced by a_{ij}.

In both cases we can follow the classical argument of Hepp[1,19] to show that the integrals in (18) are functions from S_1 in t. Combining this with property 3, we get

$$\begin{aligned} I(t, \tau_1, \tau_2, x_1^0, \ldots, x_n^0) &\leq (1 + t^2)^{-N} \\ &\times (1 + \tau_1^2)^{L_0} (1 + \tau_2^0)^{L_0} \prod_{j=1}^{n} [1 + (x_j^0)^2]^{L_j} \\ &\times \overline{C}(t, \tau_1, \tau_2, x_1^0, \ldots, x_n^0 | g \ldots, k \ldots). \end{aligned} \tag{23}$$

Here N is an arbitrary positive integer and $\{\overline{L}_0, L_i\}$ are some finite positive integers. $C(\cdots | g \cdots, h \cdots)$ is a positive bounded continuous function in t for fixed $\{\tau_i, x_j^0\}$ and vice versa. This proves our theorem.

For $h_i(\mathbf{x}) \in 0'_{c3}$ and $D(t, \tau_1, \tau_2, x_1^0, \ldots, x_n^0)$ of the form

$$D = \overline{D}\, \frac{\partial^2}{\partial \tau_1 \partial \tau_2},$$

we can prove the following generalization of Theorem 1.

Corollary 1: If the assumptions of Theorem 1 hold and if moreover $h_i(\mathbf{x})$ is restricted to $0'_{c3}$, then for all $n, r, s \geq 0$, we have

$$\begin{aligned} &\left| \overline{D}\, \frac{\partial}{\partial \tau_k} \left\langle 0 \middle| \prod_{i=1}^{r} B_i(f_i^{(+)}; \tau_1 + t) \right. \right. \\ &\left. \left. \times \prod_{l=1}^{n} A_l(g_l^{x_l^0 + t}, h_l) \prod_{j=1}^{s} B_{r+j}(f_{r+j}^{(+)}; \tau_2 + t)^* \middle| 0 \right\rangle^{(T)} \right| \\ &\leq G_k(\tau_k + t) C_k(x_1^0, \ldots, x_2^0 | g \cdots, h \cdots) \\ &\text{for } k = 1, 2, \end{aligned} \tag{23'a}$$

$$\begin{aligned} &\left| \overline{D}\, \frac{\partial^2}{\partial \tau_1 \partial \tau_2} \left\langle 0 \middle| \prod_{i=1}^{r} B_i(f_i^{(+)}, \tau_1 + t) \right. \right. \\ &\left. \left. \times \prod_{l=1}^{n} A_l(g_l^{x_l^0 + t}; h_l) \prod_{j=1}^{s} B_{r+j}(f_{r+j}^{(+)}, \tau_2 + t)^* \middle| 0 \right\rangle^{(T)} \right| \\ &\leq G_1(\tau_1 + t) G_2(\tau_2 + t) C_0(x_1^0, \ldots, x_2^0 | g \cdots, h \cdots), \end{aligned} \tag{23'b}$$

with $G_i(t)$ being positive function from S_1, and $c_i(x_1^0, \ldots, x_n^0 | g \cdots, h \cdots)$ are polynomial bounded continuous functions.

Proof: We restrict ourselves to the proof of the last relation. First, we consider the untruncated matrix elements. Applying Schwartz inequality twice, the left-hand side of (23'b) is bounded by

$$\begin{aligned} &\Pi(t + \tau_1, t + \tau_2 | x_1^0 \ldots, x_n^0) \\ &= \left\| \frac{d}{d\tau_1} \prod_{i=1}^{r} B_i(f_i^{(+)}; \tau_1 + t)^{*()} \middle| 0 \right\rangle \right\| \\ &\cdot \left\| \frac{d}{d\tau_2} \prod_{j=1}^{s} B_{r+j}(f_{r+j}^{(+)}; \tau_2 + t)^{*()} \middle| 0 \right\rangle \right\| \\ &\times \left\| \prod_{\alpha=n}^{1} A_\alpha(g_\alpha^{x_\alpha^0 + t}, h_\alpha)^{*()} \prod_{\beta=1}^{n} A_\beta(g_\beta^{x_\beta^0 + t}, h_\beta)^{()} \right. \\ &\left. \times \frac{d}{d\tau_k} \prod_l B_l(f_l^{(+)}; \tau_k + t)^{*()} \middle| 0 \right\rangle \right\|, \\ &k = 1 \text{ or } 2. \end{aligned} \tag{24}$$

According to Hepp[1] and Eq. (14) the first two factors are from S_1. The last factor may be developed into truncated matrix elements Then all terms which contain only factors with at most one $B_i(\cdots)$ and or one $B_j(\cdots)^*$ vanish by means of (14). In the remaining terms we can apply Theorem 1. This proves the corollary for the untruncated matrix elements and all $n, r, s \geq 0$. From the recursive definition of the truncated matrix elements,[1,18] corollary 1 then follows also for the latter ones.

The following important property is an obvious consequence of property (B).

Corollary 2: If the assumptions of Corollary 1 hold and if moreover some or all $h_j(\mathbf{x})$ are of the form

$$h_{\mathbf{y}_j}(\mathbf{x}) = \frac{\partial^{(n)}}{(\partial x^1)^{n_1} (\partial x^2)^{n_2} (\partial x^3)^{n_3}} \delta(\mathbf{x} - \mathbf{y}_j),$$

then the functions $C_i(x_1^0, \ldots, x_n^0 / g \cdots, h_{\mathbf{y}_j})$ occurring on the right-hand sides of (23'a) and (23'b) are bounded in all arguments $\mathbf{y}_j$ by a constant. Theorem 1 and Corollary 1 are already sufficient to derive the asymptotic series for $h_j(\mathbf{x}) \in 0'_{c3}$. Unfortunately Corollary 1 does not hold if at least one $h_j(\mathbf{x})$ is from 0^p_{M3}, since then the third factor on the right-hand side of (24) is infinite. However, the following theorem provides the necessary bounds to cover also this case.

Theorem 2: Let $\{f_i^{(+)}(\mathbf{P}_i); i = 1, \ldots, r\} \in S_3$ and $\{\tilde{f}_{r+j}^{(+)}(\mathbf{P}_{r+j}); j = 1, \ldots, s\} \in S_3$ be two sets of wavefunctions with nonoverlapping support in velocity space and $\{A_i(x_i); i = 1, \ldots, n\}$ a set of almost local operators. Then we have for all integers $r, s, n \geq 0$, all $g_i(x^0) \in S_1, h_j(\mathbf{x}_j) \in 0'_{c3} \cup 0^p_{M3}$ and for $t > 0$:

$$\begin{aligned} &\left| \int_0^\infty d\tau_k \frac{\partial}{\partial \tau_k} \int \prod_{j=1}^{n} d^3 x_j h_j(\mathbf{x}_j) \right. \\ &\times \left\langle 0 \middle| \prod_{i=1}^{r} B_i(f_i^{(+)}; \tau_1 + t) \prod_{l=1}^{n} A_l(g_l^{x_l^0 + t}; \mathbf{x}_l) \right. \\ &\left. \left. \times \prod_{j=1}^{s} B_{r+j}(f_{r+j}^{(+)}; \tau_2 + t)^* \middle| 0 \right\rangle^T \right| \\ &\leq t^{-N} \overline{C}_k(x_1^0, \ldots, x_n^0 | g \cdots, h \cdots), \quad k = 1, 2, \end{aligned} \tag{25a}$$

and

$$\Big| \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 \int \prod_{j=1}^n d^3x_j h_j(\mathbf{x}_j) \frac{\partial^2}{\partial\tau_1\partial\tau_2} \times \Big\langle 0 \Big| \prod_{i=1}^r B_i(f_i^{(+)}; t+\tau_1) \prod_{l=1}^n A_l(g_l^{x_l^0+t}, \mathbf{x}_l) \times \prod_{j=1}^s B_{r+j}(f_{r+j}^{(+)}; \tau_2 + t)^* \Big| 0 \Big\rangle^T \Big| \leq t^{-N} \bar{C}(x_1^0, \ldots, x_n^0 | g \cdots, h \cdots). \quad (25b)$$

Here N is any arbitrary positive integer and $\bar{c}_{(i)}(x_1^0, \ldots, x_n^0 | g \cdots, h \cdots)$ are for fixed finite x_j^0 some finite constants. The bounds (25a) and (25b) remain true for $t < 0$ if $\int_0^\infty d\tau_k$ is replaced by $\int_{-\infty}^0 d\tau_k$.

Proof: For all $h_j(\mathbf{x}) \in 0'_{c3}$, Theorem 2 is an obvious consequence of Corollary 1. For simplicity we restrict ourselves to prove the second inequality for all $h_j(\mathbf{x}_j)$ being from 0^p_{M3}.

Dropping all variables which are not essential for the following arguments we introduce the shorthand notation:

$$F(\tau_1, \tau_2; t, \mathbf{x}_1, \ldots, \mathbf{x}_n) = : \frac{\partial^2}{\partial\tau_1\partial\tau_2} \times \Big\langle 0 \Big| \prod_{i=1}^r B_i(f_i^{(+)}; t+\tau_1) \prod_{l=1}^n A_l(g_l^{x_l^0+t}; \mathbf{x}_l) \times \prod_{j=1}^s B_{r+j}(f_{r+j}^{(+)}; t+\tau_2)^* \Big| 0 \Big\rangle^T. \quad (26)$$

Similar to the proof of Theorem 1, the function $F(\cdots)$ may be represented by [take in (18) $\tilde{h}_j(\mathbf{k}) = \exp\{i\mathbf{k}\mathbf{x}_j\}$]

$$F(\tau_1, \tau_2; t, \mathbf{x}_1, \ldots, \mathbf{x}_n) = \int \prod_{j=1}^n d^3k_j \int \prod_{i=1}^{r+s-1} d^3p_i \times \hat{\chi}(\tau_1, \tau_2; x_1^0, \ldots, x_n^0 | \mathbf{p}_1, \ldots, \mathbf{p}_r, \mathbf{k}_1, \ldots, \mathbf{k}_n, \mathbf{p}_{r+1}, \ldots, \mathbf{p}_{r+s-1}) \exp\{i \sum_\alpha \mathbf{k}_\alpha \mathbf{x}_\alpha\} \times \exp\{-i\Omega(\mathbf{p}_1, \ldots, \mathbf{p}_{r+s-1}; \mathbf{k}_1, \ldots, \mathbf{k}_n)t\}. \quad (27)$$

$\{\Omega, \bar{\Omega}\}$ are defined in (19) and (20). $\hat{\chi}$ is a sum of expressions of the form (21) with the Wightman functions $\hat{W}_{r+n+s}(\cdots)^T$ replaced by

$$\Big[\bar{\Omega}(\mathbf{p}_1, \ldots, \mathbf{p}_{r+s-1}, \mathbf{k}_1, \ldots, \mathbf{k}_n) - \sum_{j=1}^r p_j^0 - \sum_{l=1}^n k_\alpha^0\Big] \Big[\sum_{i=1}^r (\omega_i(\mathbf{p}_i) - p_i^0)\Big] \hat{W}_{r+n+s}(\cdots).^T$$

Especially $\hat{\chi}$ has the properties 1–3 stated in the proof of Theorem 1. Therefore we can again apply the trick of Hepp[1] and deduce from the representation (27) that

$$\frac{\partial^{|m|}}{\partial\tau_1^{m_1}\partial\tau_2^{m_2}} F(\tau_1, \tau_2; t, \mathbf{x}_1, \ldots, \mathbf{x}_n) \in S_{3n+1}$$

for every fixed $\tau_i (i = 1, 2)$ and all $m_1 m_2$. Furthermore we obtain from Corollaries 1 and 2:

(i) $D(t, \mathbf{x}_1, \ldots, \mathbf{x}_n) F(\tau_1, \tau_2; t, \mathbf{x}_1, \ldots, \mathbf{x}_n) \in S_2$

for every fixed $(3n+1)$-tuple $\{t, \mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and arbitrary differential monomials D.

(ii) $|F(\tau_1, \tau_2; t, \mathbf{x}_1, \ldots, \mathbf{x}_n)| \leq G_1(t+\tau_1) G_2(t+\tau_2)$,

with $G_i \in S_1$ *independent of* $\mathbf{x}_i (i = 1, \ldots, n)$. From these three properties it follows immediately that

$$\sup_t \Big| t^N \prod_{j=1}^n h_j(\mathbf{x}_j) F(\tau_1, \tau_2; t, \mathbf{x}_1, \ldots \mathbf{x}_n) \Big| \leq \bar{C}_N(\tau_1, \tau_2) \Big[1 + \sum_{i=1}^n \mathbf{x}_i^2\Big] \quad (28a)$$

and

$$\sup_{t>0} |t^N F \tau_1, \tau_2; t, \mathbf{x}_1, \ldots, \mathbf{x}_n)| \leq C_N / [(1+\tau_1^2)(1+\tau_2^2)]^M \quad (28b)$$

for $\tau_i \geq 0, h_j(\mathbf{x}) \in 0^p_{M3}$ and all integers $M, N \geq 0$.

Now let $E_i(\tau_1, \tau_2, \mathbf{x}_1, \ldots, \mathbf{x}_n), i = 1, 2$ be two C^∞-functions with the properties

$$\sum_{i=1}^2 E_i(\cdots) \equiv 1,$$

$$E_1(\tau_1, \tau_2, \mathbf{x}_1, \ldots, \mathbf{x}_n) = \begin{cases} 1 \text{ for } \Big(\tau_1^2 + \tau_2^2 + \sum_{i=1}^n \mathbf{x}_i^2\Big)^{1/2} \leq R \\ 0 \text{ for } \Big(\tau_1^2 + \tau_2^2 + \sum_{i=1}^n \mathbf{x}_i^2\Big)^{1/2} \geq R + \Delta R. \end{cases} \quad (29)$$

Consider the integrals

$$I_i^N = : \sup_{t>0} t^N \Big| \int_0^\infty d\tau_1 \int_0^\infty d\tau_2 \int \prod_{j=1}^n d^3x_j h_j(\mathbf{x}_j) \times E_i(\tau_1, \tau_2, \mathbf{x}_1, \ldots, \mathbf{x}_n) F(\tau_1, \tau_2, t, \mathbf{x}_1, \ldots, \mathbf{x}_n) \Big|.$$

Due to the conditions (28a), (28b), and (29) I_1^N is obviously a finite number for every integer $N \geq 0$.

Furthermore the two bounds (28a) and (28b) imply that for arbitrary positive integers N, M, and $\tau_i \geq 0$,

$$\sup_{t>0} \Big| t^N \prod_{j=1}^n h_j(\mathbf{x}_j) F(\tau_1, \tau_2; t, \mathbf{x}_1, \ldots, \mathbf{x}_n) \Big| \sim \mathcal{O}\Big[\Big(\tau_1^2 + \tau_2^2 + \sum_{i=1}^n \mathbf{x}_i^2\Big)^{-M}\Big] \quad (30)$$

holds along every straight line through the origin $\{\tau_1, \tau_2, \mathbf{x}_1, \ldots, \mathbf{x}_n\} = 0$ in the direct product of the half-space $H_{2+} = \{\tau_1, \tau_2; \tau_i \geq 0\}$ and the $3n$-dimensional Euclidean space E_{3n}.

However, (30) in turn guarantees that also I_2^N is finite for every integer $N \geq 0$. This may be seen by introducing polar coordinates in $H_{2+} \otimes E_{3n}$. This proves Theorem 2.

By means of these timelike asymptotic bounds, it is now very easy to derive the asymptotic series of the field operators.

III. ASYMPTOTIC SERIES

In this section we proceed in two steps. We first demonstrate the essential arguments of the derivation for the simple case of a product to two almost local fields. After this the generalization to an arbitrary number of fields will be obvious. Furthermore we assume for simplicity that the

vacuum expectation value of a single field A_i vanishes:

$$\langle 0|A_i(f)|0\rangle = 0. \tag{31}$$

In the opposite case we have only to add further terms to the following series in an obvious way.

A. Products of Two-Field Operators

We define the connected part[20] of many particle matrix elements by the following formula, in which the summations run over all possible one-particle states[2] of the theory:

$$\begin{aligned}
\Big\langle \Psi_{ex} \prod_{j=1}^{2} A_j(f_j) \Big| \Phi_{ex} \Big\rangle^C =: & \Big\langle \Psi_{ex} \Big| \prod_{j=1}^{2} A_j(f_j) \Big| \Phi_{ex} \Big\rangle \\
& - \sum_i \Big(\Big\langle 0 \Big| \prod_{j=1}^{2} A_j(f_j) \Big| i \Big\rangle \langle \Psi_{ex} | a_{ex}(i) | \Phi_{ex} \rangle + \Big\langle i \Big| \prod_{j=1}^{2} A_j(f_j) \Big| 0 \Big\rangle \langle \Psi_{ex} | a_{ex}(i)^* | \Phi_{ex} \rangle \Big) \\
& - \sum_{i,j} \Big(\Big\langle i \Big| \prod_{r=1}^{2} A_r(f_r) \Big| j \Big\rangle \langle \Psi_{ex} | a_{ex}(i) a_{ex}(j) | \Phi_{ex} \rangle \\
& + \langle 0 | A_1(f_1) | i \rangle \langle 0 | A_2(f_2) | j \rangle \langle \Psi_{ex} | a_{ex}(i) a_{ex}(j) | \Phi_{ex} \rangle \\
& + \langle i | A_1(f_1) | 0 \rangle \langle j | A_2(f_2) | 0 \rangle \langle \Psi_{ex} | a_{ex}(i)^* a_{ex}(j)^* | \Phi_{ex} \rangle \Big) \\
& - \sum_{i,j,k} \{ [\langle 0 | A_1(f_1) | i \rangle \langle j | A_2(f_2) | k \rangle + (1 \leftrightarrow 2)] \langle \Psi_{ex} | a_{ex}(j)^* a_{ex}(i) a_{ex}(k) | \Phi_{ex} \rangle \\
& + [\langle j | A_1(f_1) | 0 \rangle \langle k | A_2(f_2) i \rangle + (1 \leftrightarrow 2)] \langle \Psi_{ex} | a_{ex}(j)^* a_{ex}(k)^* a_{ex}(i) | \Phi_{ex} \rangle \} \\
& - \sum_{i,j,k,l} \langle i | A_1(f_1) | j \rangle \langle k | A_2(f_2) | l \rangle \langle \Psi_{ex} | a_{ex}(i)^* a_{ex}(k)^* a_{ex}(j) a_{ex}(l) | \Phi_{ex} \rangle .
\end{aligned} \tag{32}$$

Now the connected part defined by this formula has some strong falloff properties if one or both time variables move to infinity.

Theorem 3: Let Ψ_{ex} and Φ_{ex} be asymptotic many-particle states from D_{ex} with nonoverlapping support in velocity space. Let furthermore at most one $f_i(x)$ be from $S_1 \otimes 0^p_{M3}$ and the others from $S_1 \otimes 0'_{c3}$. Then for $t > 0$ if ex = out or $t < 0$ if ex = in we have

$$\begin{aligned}
& |\langle \Psi_{ex} | A_1(f_1^{t+y_1^0}) A_2(f_2^{t+y_2^0}) | \Phi_{ex} \rangle^C| \\
& \qquad \leq |t|^{-N} C(f_1, f_2 | y_1^0, y_2^0),
\end{aligned} \tag{33}$$

with N any arbitrary positive integer: L_i some finite positive integer and $C(f_1, f_2 | y_1^0, y_2^0)$ some finite constant.

Proof: It is sufficient to prove the theorem for outgoing states containing n_1-, respectively n_2, particles with nonoverlapping support in velocity space. By means of the strong asymptotic condition

$$s\text{-}\lim_{t \to +\infty} \Phi^{n_i}_{t} = \Phi^{n_i}_{out} \tag{34}$$

for

$$\Phi^{n_i}_t =: \prod_{r=1}^{n_i} B_r(f_r^{(+)}; t)^* | 0 \rangle, \tag{35}$$

we replace the outgoing states $\Phi^{n_2}_{out}$, $\Psi^{n_2}_{out}$ by states of the form (35):

$$\begin{aligned}
& \Big\langle \Psi^{n_1}_{out} \Big| \prod_{j=1}^{2} A_j(f_j^{t+y_j^0}) \Big| \Phi^{n_2}_{out} \Big\rangle \\
& = \Big\langle \Phi^{n_1}_t \Big| \prod_{j=1}^{2} A_j(f_j^{t+y_j^0}) \Big| \Phi^{n_2}_t \Big\rangle \\
& \quad + \int_t^\infty ds \, \frac{\partial}{\partial s} \Big\langle \Phi^{n_1}_t \Big| \prod_{j=1}^{2} A_j(f_j^{t+y_j^0}) \Big| \Phi^{n_2}_s \Big\rangle \\
& \quad + \int_t^\infty ds \, \frac{\partial}{\partial s} \Big\langle \Phi^{n_1}_s \Big| \prod_{j=1}^{2} A_j(f_j^{t+y_j^0}) \Big| \Phi^{n_2}_t \Big\rangle \\
& \quad + \int_t^\infty ds \int_t^\infty du \, \frac{\partial^2}{\partial s \partial u} \\
& \quad \times \Big\langle \Phi^{n_1}_s \Big| \prod_{j=1}^{2} A_j(f_j^{t+y_j^0}) \Big| \Phi^{n_2}_u \Big\rangle .
\end{aligned} \tag{36}$$

We show first that the last three terms of (36) are bounded by the right-hand side of (33) separately.

Case 1: All $f_i(x) \in S_1 \otimes 0'_{c3}$: Consider the second term

$$\begin{aligned}
& \Pi_t(y_1^0, y_2^0) \\
& \quad =: \int_0^\infty ds \, \frac{\partial}{\partial s} \Big\langle \Phi^{n_1}_t \Big| \prod_{j=1}^{2} A_j(f_j^{t+y_j^0}) \Big| \Phi^{n_2}_{s+t} \Big\rangle \\
& \quad = \int_0^\infty ds \, \frac{\partial}{\partial s} \Big\langle 0 \Big| \prod_{j=n_1}^{1} B_j(f_j^{(+)}; t) \prod_{k=1}^{2} A_k(f_k^{t+y_k^0}) \\
& \qquad \times \prod_{i=1}^{n_2} B_{n_1+i}(f_{n_1+i}^{(+)}; s+t)^* \Big| 0 \Big\rangle .
\end{aligned} \tag{37}$$

Since a one-particle excitation operator $B_l(f^{(+)}, t)^*$ applied to the vacuum state results in a time-independent state (14), the time derivative d/ds of all vacuum expectation values of the type occurring in the integrand of (37) with at most one $B_j(f^{(+)}; s)^*$ vanish. The same is then also true for the corresponding truncated expectation values.

Developing the integrand of (37) into truncated expectation values, the differential operator kills all terms except those with at least one factor containing at least two one-particle excitation operators $B_l(f^{(+)}; S)^*$. From Corollary 1 we obtain for the surviving ones the bound

$$|\Pi_t(y_1^0, y_2^0)| \leq (1 + t^2)^{-N} C(f_1, f_2, y_1^0, y_2^0). \tag{38}$$

This proves the desired bound for the second term of (36). The third and fourth term can be treated in exactly the same way by means of Theorems 1 and 2, respectively, together with (14).

Case 2: One $f_i(x) \in S_1 \otimes 0^p_{M3}$: The only difference from Case 1 has its origin in the occurence of a vacuum expectation value

$$\langle 0|A_1(f_1^{t+y_1^0})A_2(f_2^{t+y_2^0})|0\rangle \tag{39}$$

multiplied by truncated vacuum expectation values of the operators $B_r(f^{(+)};t)^{(+)}$ in the development of the integrand of (37) into truncated functions. The latter factors guarantee the desired bound in t as above. However, the term (39) does exist if and only if at most one $f_i(x)$ is from $S_1 \otimes 0^p_{M3}$. This follows from property (B) stated in Sec. I.[21,22]

There remains the first term on the right-hand side of (36):

$$\begin{aligned} I_t(y_1^0, y_2^0) &=: \Big\langle \Phi_t^{n_1} \Big| \prod_{r=1}^{2} A_r(f_r^{t+y_r^0}) \Big| \Phi_t^{n_2} \Big\rangle \\ &= \Big\langle 0 \Big| \prod_{j=n_1}^{1} B_j(f_j^{(+)};t) \prod_{r=1}^{2} A_r(f_r^{t+y_r^0}) \\ &\quad \times \prod_{i=1}^{n_2} B_{n_1+i}(f_{n_1+i}^{(+)};t)^* \Big| 0 \Big\rangle . \end{aligned} \tag{40}$$

In the decomposition into truncated vacuum expectation values, all terms with at least one factor containing at least two field operators $B_j(f^{(+)};t)$ or two $B_l(f^{(+)},t)^*$ satisfy the bound (33) due to Theorem 1. Therefore we have to pick out only those terms which contain exclusively factors with at most one $B_i(f^{(+)};t)$ and/or at most one $B_r(f^{(+)};t)$. Besides the two-point functions

$$\langle 0|B_i(f_i^{(+)};t)B_j(f_j^{(+)};t)|0\rangle , \tag{41}$$

such possible factors are

$$\begin{aligned} &\Big\langle 0 \Big| \prod_{r=1}^{2} A_r(f_r^{t+y_r^0}) \Big| 0 \Big\rangle , \\ &\Big\langle 0 \Big| \prod_{r=1}^{2} A_r(f_r^{t+y_r^0}) B_j(f_j^{(+)};t)^* \Big| 0 \Big\rangle^T \\ &\quad = \Big\langle 0 \Big| \prod_{r=1}^{2} A_r(f_r^{t+y_r^0}) \Big| j \Big\rangle , \\ &\Big\langle 0 \Big| B_i(f_i^{(+)};t) \prod_{r=1}^{2} A_r(f_r^{t+y_r^0}) \Big| 0 \Big\rangle^T \\ &\quad = \Big\langle i \Big| \prod_{r=1}^{2} A_r(f_r^{t+y_r^0}) \Big| 0 \Big\rangle , \\ &\Big\langle 0 \Big| B_i(f_i^{(+)};t) \prod_{r=1}^{2} A_r(f_r^{t+y_r^0}) B_j(f_j^{(+)};t)^* \Big| 0 \Big\rangle^T \\ &\quad = \Big\langle i \Big| \prod_{r=1}^{2} A_r(f_r^{t+y_r^0}) \Big| j \Big\rangle^T , \end{aligned} \tag{42}$$

and are all possible products ($j = 1, 2$) of the following terms:

$$\begin{aligned} &\langle 0|A_j(f_j^{t+y_j^0})B_r(f_r^{(+)};t)^*|0\rangle^T = \langle 0|A_j(f_j^{t+y_j^0})|r\rangle , \\ &\langle 0|B_r(f_r^{(+)};t)A_j(f_j^{t+y_j^0})|0\rangle^T = \langle r|A_j(f_j^{t+y_j^0})|0\rangle , \\ &\langle 0|B_i(f_i^{(+)};t)A_j(f_j^{t+y_j^0})B_r(f_r^{(+)};t)^*|0\rangle^T \\ &\quad = \langle i|A_j(f_j^{t+y_j^0})|r\rangle . \end{aligned} \tag{43}$$

Each of the matrix elements (42) and each of the products of (43) is multiplied by a product of the time-independent two-point functions (41) build from all the remaining one-particle excitation operators. These products of two-point functions define according to the Haag-Ruelle scattering theory[1,18] the scalar products of asymptotic particle states.

Applying the following relation which follows from the canonical commutation relations of the asymptotic fields:

$$\begin{aligned} &\sum_{r_1,\dots,r_{m-n}=1,\{r_i\neq r_j\}}^{m} \\ &\prod_{s=1}^{m-n} \delta_{\alpha_s\beta_{r_s}} \Big\langle \Psi \Big| \prod_{\{j=r_1,\dots,r_{m-n}\}}^{m} a_{\mathrm{ex}}(\beta_j)^* \Big| 0 \Big\rangle \\ &\quad = \Big\langle \Psi \Big| \prod_{s=1}^{m-n} a_{\mathrm{ex}}(\alpha_s) \prod_{j=1}^{m} a_{\mathrm{ex}}(\beta_j)^* \Big| 0 \Big\rangle , \end{aligned} \tag{44}$$

and going back from the truncated expressions (42) to the untruncated ones, these terms which do not satisfy the bound (33), deliver exactly the sums in the definition of the connected matrix elements (32). This proves our theorem.

Remark: Theorem 3 can be extended to the case that both $f_i(x) \in S_1 \otimes 0^p_{M3}$ if one considers

$$\Big\langle \Psi_{\mathrm{ex}} \Big| \prod_{i=1}^{2} A_i(f_i) \Big| \Phi_{\mathrm{ex}} \Big\rangle^C - \Big\langle 0 \Big| \prod_{i=1}^{2} A_i(f_i) \Big| 0 \Big\rangle \langle \Psi_{\mathrm{ex}} \| \Phi_{\mathrm{ex}} \rangle$$

instead of the connected part alone.

B. The General Case

We start again with the construction of connected many-particle matrix elements analogous to (32). As it is obvious from the proof of Theorem 3 we have to collect from the decomposition of the matrix element

$$\Big\langle 0 \Big| \prod_{j=1}^{n} B_j(f_j^{(+)};t) \prod_{\alpha=1}^{n} A_\alpha(f_\alpha^{t+y_\alpha^0}) \prod_{i=1}^{n_2} B_{n_1+i}(f_{n_1+i}^{(+)};t)^* \Big| 0 \Big\rangle \tag{45}$$

into truncated vacuum expectation values all those terms, which consist exclusively of factors containing at most one $B_j(f_j^{(+)};t)$ and/or at most one $B_i(f_i^{(+)};t)^*$, and to subtract them from the original matrix element. These possible factors are, besides the two-point functions of the one-particle excitation operators, the following ones:

$$\begin{aligned} &\langle 0|A_{i_1}\dots A_{i_r}|0\rangle^T , \\ &\Big\langle 0 \Big| B_j(f_j^{(+)};t) \prod_{l=1}^{r} A_{i_l} B_s(f_s^{(+)};t)^* \Big| 0 \Big\rangle^T \\ &\quad = \Big\langle j \Big| \prod_{l=1}^{r} A_{i_l} \Big| s \Big\rangle^T , \\ &\Big\langle 0 \Big| B_j(f_j^{(+)};t) \prod_{l=1}^{r} A_{i_l} \Big| 0 \Big\rangle^T = \Big\langle j \Big| \prod_{l=1}^{r} A_{i_l} \Big| 0 \Big\rangle^T , \\ &\Big\langle 0 \Big| \prod_{l=1}^{r} A_{i_l} B_s(f_s^{(+)};t)^* \Big| 0 \Big\rangle^T = \Big\langle 0 \Big| \prod_{l=1}^{r} A_{i_l} \Big| s \Big\rangle^T , \\ &r = 1, 2, \dots, n. \end{aligned} \tag{46}$$

Here $\{i_1, \dots, i_r\}$ is any subset of $\{1, \dots, n\}$ with $i_1 < i_2 < \cdots < i_r$.

In the first instance we collect from all possible products of the expressions (46) those with a fixed number $s \leq n$ of one-particle excitation operators $B_j(f^{(+)};t)^*$ and a fixed number $r \leq n$ of operators $B_j(f^{(+)};t)$. These are given by

$$\begin{aligned}\Big(\prod_{\alpha=1}^{r\leq n} j_\alpha \Big\| A_1 \cdots A_n \Big\| \prod_{\beta=1}^{s\leq n} j_{r+\beta}\Big) =: \sum_{m=0}^{\min\{r,s\}} \\ \times \sum_{\substack{\text{part}\\ \{1,\ldots,n\}}} \prod_{\alpha=1}^{m} \langle j_\alpha | A_{i_{k_\alpha}} \cdots A_{i_{k_\alpha+t_\alpha}} | j_{r+\alpha}\rangle \\ \times \prod_{\beta=m+1}^{r} \langle j_\beta | A_{i_{k_\beta}} \cdots A_{i_{k_\beta+t_\beta}} |0\rangle^T \qquad (47) \\ \times \prod_{\delta=m+1}^{s} \langle 0 | A_{i_{k_{r+\delta}}} \cdots A_{i_{k_{r+\delta}+t_{r+\delta}}} | j_{r+\delta}\rangle^T \\ \times \langle 0 | A_{i_{\zeta_1}} \cdots A_{i_{\zeta_1+\epsilon_1}} |0\rangle^T \cdots \langle 0 | A_{i_{\zeta_l}} \cdots A_{i_n} |0\rangle^T.\end{aligned}$$

Here $\sum_{\text{part}}$ extends over all partitions $\{i_{k_1} \ldots i_{k_1+t_1}\}, \{i_{k_2} \ldots i_{k_2+t_2}\} \cdots \{i_{\zeta_l} \cdots i_n\}$ of $\{1,\ldots,n\}$ into disjoint subsets such that in every truncated matrix element on the right-hand side of (47) the order of the field operators is the same as in $(\cdots \| A_1,\ldots,A_n \| \cdots)$ on the left-hand side. Furthermore in order to avoid double counting of factors, we impose the following ordering condition: Within each of the four different groups of truncated matrix elements the factors are ordered according to increasing index of the first field A_i; that means

$$\begin{aligned} &i_{k_1} < i_{k_2} < \cdots < i_{k_m}, \\ &i_{k_{m+1}} < i_{k_{m+2}} < \cdots < i_{k_{m+r}}, \\ &i_{k_{m+r+1}} < i_{k_{m+r+2}} < \cdots < i_{k_{r+s}}, \\ &i_{\zeta_1} < i_{\zeta_2} \cdots < i_{\zeta_l}. \end{aligned}$$

Finally if two or more Fermi fields occur among the A_i and B_j, every term on the right-hand side of (47) is to be multiplied by a factor $(-1)^{N_F}$. N_F is the number of interchanges of Fermi fields which is met in going from the natural order $\{j_1,\ldots,j_r, 1,\ldots,n, j_{r+1},\ldots,j_{r+s}\}$ to the order of the particular term in question.

Each expression of the type (47) with $0 \leq r,s \leq n$ is multiplied by products of two-point functions built from the remaining $(n_1 - r)$ fields $B_j(f^{(+)}_{M_j b_j};t)$ and $(n_2 - s)$ fields $B_{r+j}(f^{(+)}_{M_{r+j}b_{r+j}},t)^*$. By means of

$$\begin{aligned} &B_j(f_j^{(+)};t)|0\rangle = 0, \\ &\frac{d}{dt}B_j(f_j^{(+)};t)^*|0\rangle = 0, \end{aligned} \qquad (48)$$

and Eq. (44), we obtain for the latter products:

$$\begin{aligned}\Big\langle 0 \Big| \prod_{i=1}^{n_1} a_{\text{ex}}(i) \prod_{t=n_1+1}^{n_2} a_{\text{ex}}(t)^* \Big| 0 \Big\rangle \\ \{i\neq j_1,\ldots,j_r\} \qquad \{t\neq j_{r+1},\ldots,j_{r+s}\} \qquad (49) \\ = \Big\langle \Psi_{\text{ex}}^{n_1} \Big| \prod_{\alpha=1}^{r} a_{\text{ex}}(j_\alpha)^* \prod_{\beta=1}^{s} a_{\text{ex}}(j_{r+\beta}) \Big| \Phi_{\text{ex}}^{n_2} \Big\rangle,\end{aligned}$$

where $\Psi_{\text{ex}}^{n_1}, \Phi_{\text{ex}}^{n_2}$ are the strong asymptotic limits of the states

$$\Phi_t^{n_i} = \prod_{l=1}^{n_i} B_l(f_l^{(+)};t)^*|0\rangle.$$

Multiplying (47) and (49) together and summing over all r,s with $0 \leq r;\ s \leq n$ and over all one-particle quantum numbers[2] j, we obtain all possible terms from the decomposition of the matrix element (45) into truncated matrix elements which contain exclusively factors with at most one $B_i(f^{(+)};t)^*$ and/or at most one $B_j(f^{(+)};t)$. If we subtract all these sums from the original matrix element, we obtain the so called *connected* many-particle matrix element:

$$\begin{aligned}\langle \Psi_{\text{ex}} | A_1(f_1) \ldots A_2(f_2) | \Phi_{\text{ex}}\rangle^c \\ =: \langle \Psi_{\text{ex}} | A_1(f_1) \ldots A_n(f_n) | \Phi_{\text{ex}}\rangle \\ - \sum_{r,s=0}^{n} \sum_{j_1,\ldots,j_{r+s}} \Big\{ \sum_{m=0}^{\min\{r,s\}} \sum_{\substack{\text{part}\\ \{1,\ldots,n\}}} \\ \times \prod_{\alpha=1}^{m} \langle j_\alpha | A_{i_{k_\alpha}} \cdots A_{i_{k_\alpha+t_\alpha}} | j_{r+\alpha}\rangle^T \\ \times \prod_{\beta=m+1}^{r} \langle j_\beta | A_{i_{k_\beta}} \cdots A_{i_{k_\beta+t_\beta}} |0\rangle^T \qquad (50) \\ \times \prod_{\delta=m+1}^{s} \langle 0 | A_{i_{k_{r+\delta}}} \cdots A_{i_{k_{r+\delta}+t_{r+\delta}}} | j_{r+\delta}\rangle^T \\ \times \langle 0 | A_{i_{\zeta_1}} \cdots A_{i_{\zeta_1+\epsilon_1}} |0\rangle^T \cdots \langle 0 | A_{i_{\zeta_l}} \cdots A_{i_n} |0\rangle^T \Big\} \\ \times \Big\langle \Psi_{\text{ex}} \Big| \prod_{\alpha=1}^{r} a_{\text{ex}}(j_\alpha)^* \prod_{\beta=1}^{s} a_{\text{ex}}(j_{r+\beta}) \Big| \Phi_{\text{ex}} \Big\rangle.\end{aligned}$$

The second sum in (50) runs over all possible one-particle states[1] j of the theory; the last sum is explained after Eq. (47). With this construction of the connected matrix elements the generalization of Theorem 3 to an arbitrary number of fields obviously reads as follows.

Theorem 4: Let Ψ_{ex} and Φ_{ex} be asymptotic many-particle states from D_{ex} with nonoverlapping support in velocity space. If at most one $f_i(x)$ is from $S_1 \otimes 0^p_{M3}$ and all others from $S_1 \otimes 0'_{c3}$, then we have for $t > 0$ if ex = out or $t < 0$ if ex = in:

$$\begin{aligned}\Big| \Big\langle \Psi_{\text{ex}} \Big| \prod_{s=1}^{n} A_s(f_s^{t+y_s^0}) \Big| \Phi_{\text{ex}} \Big\rangle^c \Big| \\ \leq |t|^{-N} C(f_1,\ldots,f_n | y_1^0,\ldots,y_n^0), \qquad (51)\end{aligned}$$

with N any arbitrary positive integer and $C(f_1, \cdots, y_n^0)$ a finite positive constant.

The proof of this theorem follows in exactly the same way as that of Theorem 3. With the considerations above which led us to the definition of the connected matrix elements (50) this proof is a repetition of the arguments from that of Theorem 3.

By the two theorems of this section we have extended the formula of Araki and Haag to an arbitrary number of field operators and a class of "smearing functions" in the space variables large enough to cover most applications of current interest.

IV. THE ORIGIN OF THE LEADING TERMS

It is quite interesting to see which parts of the original matrix elements (certain intermediate states or parts of them) contribute to the leading terms in the asymptotic series of the last section.

We restrict ourselves to the discussion of a product of two-field operators and Bose particles. The generalization to arbitrary products and Fermi particles is then again a pure matter of combinatorics and introduction of Fermi factors $(-1)^{N_F}$.

Definition 1: A many-particle matrix element of a field is called *n-fold disconnected*, which we indicate by

$$\langle \underbrace{\Psi^{N_1}_{\mathrm{ex}\{\alpha_i\}}}_{n} | A(f) | \Phi^{N_2}_{\mathrm{ex}\{\beta_j\}} \rangle \equiv \Big\langle 0 \Big| \prod_{i=1}^{N_1} a_{\mathrm{ex}}(\alpha_i) \underbrace{A(f) \prod_{j=1}^{N_2} a_{\mathrm{ex}}(\beta_j)^*}_{n} \Big| 0 \Big\rangle$$

if n particles of the ket (bra) run unaffected by the field $A(f)$ to the left (right) and destroy n particles of the bra (ket):

$$\langle \underbrace{\Psi^{N_1}_{\mathrm{ex}\{\alpha_i\}}}_{n} | A(f) | \Phi^{N_2}_{\mathrm{ex}\{\beta_j\}} \rangle$$

$$=: \sum_{r_1,\ldots,r_{N_2-n}=1,\,\{r_i\neq r_j\}}^{N_2} \frac{1}{(N_2-n)!} \times \Pi(\beta_{r_1},\ldots,\beta_{r_{N_2-n}})^{-1} \times \Big\langle 0 \Big| \prod_{i=1}^{N_1} a_{\mathrm{ex}}(\alpha_i) \prod_{j=1,\,\{j\neq r_1,\ldots,r_{N_2-n}\}}^{N_2} \times a_{\mathrm{ex}}(\beta_j)^* A(f) \prod_{t=1}^{N_2-n} a_{\mathrm{ex}}(\beta_{r_t})^* \Big| 0 \Big\rangle. \tag{52}$$

Here $\Pi(\beta_1,\ldots,\beta_m)$ is defined by

$$\Pi(\beta_1,\ldots,\beta_m) =: Z_1!\,Z_2!\cdots Z_k! \tag{53}$$

if among the quantum numbers β_i $Z_1; Z_2; \cdots; Z_k$ are equal.

In order to calculate the right-hand side of (52) we introduce a complete orthonormal set of many-particle states

$$|\epsilon_1,\ldots,\epsilon_m\rangle_{\mathrm{ex}} = \Pi(\epsilon_1,\ldots,\epsilon_m)^{-1/2} \prod_{i=1}^{m} a_{\mathrm{ex}}(\epsilon_i)^* |0\rangle,$$
$$\sum_{m=0}^{\infty} \frac{1}{m!} \sum_{\{\epsilon_i\}} |\epsilon_1,\ldots,\epsilon_m\rangle\langle\epsilon_1,\ldots,\epsilon_m| = 1 \tag{54}$$

between $a_{\mathrm{ex}}(\beta_{N_2})^*$ and $A(f)$. Then only the terms with $m = N_1 - n$ particles can contribute.

$$\langle \underbrace{\Psi^{N_1}_{\mathrm{ex}\{\alpha_i\}}}_{n} | A(f) | \Phi^{N_2}_{\mathrm{ex}\{\beta_j\}} \rangle = \frac{1}{(N_1-n)!(N_2-n)!} \sum_{r_1,\ldots,r_{N_2-n}=1,\,\{r_i\neq r_j\}}^{N_2} \sum_{\epsilon_1,\ldots,\epsilon_{N_1-n}} \Pi(\beta_{r_1},\ldots,\beta_{r_{N_2-n}})^{-1}$$
$$\times \Pi(\epsilon_1,\ldots,\epsilon_{N_1-n})^{-1} \Big\langle 0 \Big| \prod_{s=1}^{N_1-n} a_{\mathrm{ex}}(\epsilon_s) A(f) \prod_{t=1}^{N_2-n} a_{\mathrm{ex}}(\beta_{r_t})^* \Big| 0 \Big\rangle \Big\langle 0 \Big| \prod_{i=1}^{N_1} a_{\mathrm{ex}}(\alpha_i) \prod_{l=1}^{N_1-n} a_{\mathrm{ex}}(\epsilon_l)^*$$
$$\times \Big| \prod_{j=1,\,\{j\neq r_1,\ldots,r_{N_2-n}\}}^{N_2} a_{\mathrm{ex}}(\beta_j)^* \Big| 0 \Big\rangle = \frac{1}{(N_1-n)!(N_2-n)!} \sum_{\epsilon_1,\ldots,\epsilon_{N_1-n}} \sum_{\gamma_1,\ldots,\gamma_{N_2-n}}$$
$$\times \Pi(\epsilon_1,\ldots,\epsilon_{N_1-n})^{-1} \Pi(\gamma_1,\ldots,\gamma_{N_2-n})^{-1} \Big\langle 0 \Big| \prod_{l=1}^{N_1-n} a_{\mathrm{ex}}(\epsilon_l) A(f) \prod_{k=1}^{N_2-n} a_{\mathrm{ex}}(\gamma_k)^* \Big| 0 \Big\rangle$$
$$\times \sum_{r_1,\ldots,r_{N_2-n}=1,\,\{r_i\neq r_j\}}^{N_2} \prod_{s=1}^{N_2-1} \delta_{\gamma_s \beta_{r_s}} \Big\langle 0 \Big| \prod_{i=1}^{N_1} a_{\mathrm{ex}}(\alpha_i) \prod_{l=1}^{N_1-n} a_{\mathrm{ex}}(\epsilon_l)^*$$
$$\times \prod_{j=1,\,\{j\neq r_1,\ldots r_{N_2-n}\}}^{N_2} a_{\mathrm{ex}}(\beta_j)^* \Big| 0 \Big\rangle.$$

By means of Eq. (44) we finally obtain for our n-fold disconnected matrix element:

$$\langle \underbrace{\Psi^{N_1}_{\mathrm{ex}\{\alpha_i\}}}_{n} | A(f) | \Phi^{N_2}_{\mathrm{ex}\{\beta_j\}} \rangle = \frac{1}{(N_1-n)!(N_2-n)!} \sum_{\epsilon_1,\ldots,\epsilon_{N_1-n}} \sum_{\gamma_1,\ldots,\gamma_{N_2-n}} \Pi(\epsilon_1,\ldots,\epsilon_{N_1-n})^{-1}\, \Pi(\gamma_1,\ldots,\gamma_{N_2-n})^{-1}$$
$$\times \Big\langle 0 \Big| \prod_{l=1}^{N_1-n} a_{\mathrm{ex}}(\epsilon_l) A(f) \prod_{k=1}^{N_2-n} a_{\mathrm{ex}}(\gamma_k)^* \Big| 0 \Big\rangle \Big\langle \Psi^{N_1}_{\mathrm{ex}\{\alpha_i\}} \Big| \prod_{s=1}^{N_1-n} a_{\mathrm{ex}}(\epsilon_s)^* \prod_{t=1}^{N_2-n} a_{\mathrm{ex}}(\gamma_t) \Big| \Phi^{N_2}_{\mathrm{ex}\{\beta_j\}} \Big\rangle. \tag{55}$$

Comparing this formula with Eq. (32) we see that the second, third, and fourth terms on the right-hand side are given by

$$\sum_i \langle 0 | A_1(f_1) A_2(f_2) | i \rangle \langle \Psi^{N_1}_{\mathrm{ex}} | a_{\mathrm{ex}}(i) | \Phi^{N_2}_{\mathrm{ex}} \rangle = \delta_{N_1,N_2-1} \langle \underbrace{\Psi^{N_1}_{\mathrm{ex}}}_{N_1} | A_1(f_1) A_2(f_2) | \Phi^{N_2}_{\mathrm{ex}} \rangle, \tag{56}$$

$$\sum_i \langle i | A_1(f_1) A_2(f_2) | 0 \rangle \langle \Psi^{N_1}_{\mathrm{ex}} | a_{\mathrm{ex}}(i)^* | \Phi^{N_2}_{\mathrm{ex}} \rangle = \delta_{N_1,N_2+1} \langle \underbrace{\Psi^{N_1}_{\mathrm{ex}}}_{N_1-1} | A_1(f_1) A_2(f_2) | \Phi^{N_2}_{\mathrm{ex}} \rangle, \tag{57}$$

$$\sum_{i,j}\langle i|A_1(f_1)A_2(f_2)|j\rangle\langle\Psi_{ex}^{N_1}|a_{ex}(i)^*a_{ex}(j)|\Phi_{ex}^{N_2}\rangle = \delta_{N_1,N_2}\langle\underbrace{\Psi_{ex}^{N_1}|A_1(f_1)A_2(f_2)}_{N_1-1}|\Phi_{ex}^{N_2}\rangle. \tag{58}$$

The remaining terms of the asymptotic sereis (32) originate from disconnected intermediate states between the two-field operators A_i.

Definition 2: We call a normalized intermediate N-particle state

$$|\varphi_{ex\{\gamma_t\}}^{N}\rangle = \Pi(\gamma_1,\ldots,\gamma_N)^{-1/2}\prod_{t=1}^{N}a_{ex}(\gamma_t)^*|0\rangle$$

(n_1,n_2)-fold disconnected and denote this by

$$\frac{1}{N!}\sum_{\gamma_1,\ldots,\gamma_N}\langle\underbrace{\Psi_{ex\{\alpha_i\}}^{N_1}|A_1(f_1)}_{n_1}|\varphi_{ex\{\gamma_t\}}^{N}\rangle \times\langle\underbrace{\varphi_{ex\{\gamma_t\}}^{N}|A_2(f_2)}_{n_2}|\Phi_{ex\{\beta_j\}}^{N_2}\rangle, \tag{59}$$

if n_1 particles of the intermediate ket run unaffected by the field A_1 to the left and destroy n_1 particles of the outer bra and if n_2 particles from the intermediate bra run unaffected by the field A_2 to the right and destroy n_2 particles of the outer ket.

Each factor in the expression (59) is of course the n_i-fold disconnected matrix element introduced in Definition 1 and explicitly given by (55). Inserting (55) into (59) it is straightforward to indentify the remaining terms of (32) with disconnected matrix elements.

The result is

$$\sum_{i,j}\langle 0|A_1(f_1)|i\rangle\langle 0|A_2(f_2)|j\rangle\langle\Psi_{ex}^{N_1}|a_{ex}(i)a_{ex}(j)|\Phi_{ex}^{N_2}\rangle = \delta_{N_1,N_2-2}\sum_{\{\gamma_s\}}\langle\underbrace{\Psi_{ex}^{N_1}|A_1(f_1)|\varphi_{ex\{\gamma_s\}}^{N_1+1}}_{N_1}\rangle\langle\underbrace{\varphi_{ex\{\gamma_s\}}^{N_1+1}|A_2(f_2)|\Phi_{ex}^{N_2}}_{N_1+1}\rangle\frac{1}{(N_1+1)!}, \tag{60}$$

$$\sum_{i,j}\langle i|A_1(f_1)|0\rangle\langle j|A_2(f_2)|0\rangle\langle\Psi_{ex}^{N_1}|a_{ex}(i)^*a_{ex}(j)^*|\Phi_{ex}^{N_2}\rangle = \frac{\delta_{N_1,N_2+1}}{(N_1-1)!}\sum_{\{\gamma_s\}}\langle\underbrace{\Psi_{ex}^{N_1}|A_1(f_1)|\varphi_{ex\{\gamma_s\}}^{N_1-1}}_{N_1-1}\rangle\langle\underbrace{\varphi_{ex\{\gamma_s\}}^{N_1-1}|A_2(f_2)|\Phi_{ex}^{N_2}}_{N_1-2}\rangle, \tag{61}$$

$$\sum_{i,j,k}\{\langle 0|A_1(f_1)|i\rangle\langle j|A_2(f_2)|k\rangle + \langle j|A_1(f_1)|i\rangle\langle 0|A_2(f_2)|k\rangle\}\langle\Psi_{ex}^{N_1}|a_{ex}(j)^*a_{ex}(i)a_{ex}(k)|\Phi_{ex}^{N_2}\rangle$$
$$= \frac{\delta_{N_1,N_2-1}}{(N_1+1)!}\sum_{\{\gamma_s\}}\{\langle\underbrace{\Psi_{ex}^{N_1}|A_1(f_1)|\varphi_{ex\{\gamma_s\}}^{N_1+1}}_{N_1}\rangle\langle\underbrace{\varphi_{ex\{\gamma_s\}}^{N_1+1}|A_2(f_2)|\Phi_{ex}^{N_2}}_{N_1}\rangle + \langle\underbrace{\Psi_{ex}^{N_1}|A_1(f_1)|\varphi_{ex\{\gamma_s\}}^{N_1}}_{N_1-1}\rangle$$
$$\times\langle\varphi_{ex\{\gamma_s\}}^{N_1}|A_2(f_2)|\Phi_{ex}^{N_2}\rangle\} - \sum_{i,j}\langle 0|A_1(f_1)i\rangle\langle i|A_2(f_2)|j\rangle\langle\Psi_{ex}^{N_1}|a_{ex}(j)|\Phi_{ex}^{N_2}\rangle, \tag{62}$$

$$\sum_{i,j,k}\{\langle i|A_1(f_1)|0\rangle\langle j|A_2(f_2)|k\rangle + \langle i|A_1(f_1)|k\rangle\langle j|A_2(f_2)|0\rangle\}\langle\Psi_{ex}^{N_1}|a_{ex}(i)^*a_{ex}(j)^*a_{ex}(k)|\Phi_{ex}^{N_2}\rangle$$
$$= \frac{\delta_{N_1,N_2+1}}{(N_1-1)!}\sum_{\{\gamma_s\}}\{\langle\underbrace{\Psi_{ex}^{N_1}|A_1(f_1)|\varphi_{ex\{\gamma_s\}}^{N_1-1}}_{N_1-1}\rangle\langle\underbrace{\varphi_{ex\{\gamma_s\}}^{N_1-1}|A_2(f_2)|\Phi_{ex}^{N_2}}_{N_1-2}\rangle + \langle\underbrace{\Psi_{ex}^{N_1}|A_1(f_1)|\varphi_{ex\{\gamma_s\}}^{N_1}}_{N_1-1}\rangle$$
$$\times\langle\underbrace{\varphi_{ex\{\gamma_s\}}^{N_1}|A_2(f_2)|\Phi_{ex}^{N_2}}_{N_1-1}\rangle\} - \sum_{i,j}\langle i|A_1(f_1)|j\rangle\langle j|A_2(f_2)|0\rangle\langle\Psi_{ex}^{N_1}|a_{ex}(i)^*|\Phi_{ex}^{N_2}\rangle, \tag{63}$$

$$\sum_{i,j,k,l}\langle i|A_1(f_1)|j\rangle\langle k|A_2(f_2)|l\rangle\langle\Psi_{ex}^{N_1}|a_{ex}(i)^*a_{ex}(k)^*a_{ex}(j)a_{ex}(l)|\Phi_{ex}^{N_2}\rangle = \frac{\delta_{N_1,N_2}}{N_1!}\sum_{\{\gamma_s\}}\langle\Psi_{ex}^{N_1}|A_1(f_1)|\varphi_{ex\{\gamma_s\}}^{N_1}\rangle$$
$$\times\langle\varphi_{ex\{\gamma_s\}}^{N_1}|A_2(f_2)|\Phi_{ex}^{N_2}\rangle - \sum_{i,j,k}\langle i|A_1(f_1)|k\rangle\langle k|A_2(f_2)|j\rangle\langle\Psi_{ex}^{N_1}|a_{ex}(i)^*a_{ex}(j)|\Phi_{ex}^{N_2}\rangle. \tag{64}$$

There remain the last terms on the right-hand sides of Eq. (62)–(64) to be identified with disconnected matrix elements.

Obviously these terms are equal to the one-particle intermediate state contributions to the n-fold disconnected matrix elements (56)–(58). Explicitly they read, respectively,

$$\delta_{N_1,N_2-1}\sum_\gamma\langle\underbrace{\Psi_{ex}^{N_1}|A_1(f_1)|\varphi_{ex;\gamma}^{1}}_{N_1}\rangle\langle\varphi_{ex;\gamma}^{1}|A_2(f_2)|\Phi_{ex}^{N_2}\rangle,$$
$$\delta_{N_1,N_2+1}\sum_\gamma\langle\underbrace{\Psi_{ex}^{N_1}|A_1(f_1)|\varphi_{ex;\gamma}^{1}}_{N_1-1}\rangle\langle\varphi_{ex;\gamma}^{1}|A_2(f_2)|\Phi_{ex}^{N_2}\rangle,$$
$$\delta_{N_1,N_2}\sum_\gamma\langle\underbrace{\Psi_{ex}^{N_1}|A_1(f_1)|\varphi_{ex;\gamma}^{1}}_{N_1-1}\rangle\langle\varphi_{ex;\gamma}^{1}|A_2(f_2)|\Phi_{ex}^{N_2}\rangle. \tag{65}$$

With these identifications, we may rewrite our asymptotic series (32) in terms of disconnected matrix elements:

$$\langle \Psi_{ex}^{N_1} | A_1(f_1^{X_1^0}) A_2(f_2^{X_1^0}) | \Phi_{ex}^{N_2} \rangle$$

$$= \sum_{N>0} \sum_{n>0} \sum_{(m_1,m_2)>0} \sum_{\{\gamma_1,\ldots,\gamma_N\}} \frac{1}{N!} \langle \Psi_{ex}^{N_1} | A_1(f_1^{X_1^0}) | \varphi_{ex\{\gamma_1,\ldots,\gamma_N\}}^{N} \rangle \langle \varphi_{ex\{\gamma_1,\ldots,\gamma_N\}}^{N} | A_2(f_2^{X_2^0}) | \Phi_{ex}^{N_2} \rangle$$

$$\times \{(1-\delta_{N,1})\delta_{m_1,0}\delta_{m_2,0}(\delta_{N_1,N_2-1}\delta_{n,N_1} + \delta_{N_1,N_2+1}\delta_{n,N_1-1} + \delta_{N_1,N_2}\delta_{n,N_1-1})$$

$$+ \delta_{n,0}[\delta_{N_1,N_2-1}\delta_{N,N_1+1}\delta_{m_1,N_1}\delta_{m_2,N_1+1} + \delta_{N_1,N_2+2}\delta_{N,N_1+1}\delta_{m_1,N_1}\delta_{m_2,N_1+1}\delta_{N_1,N_2}\delta_{N,N_1}\delta_{m_1,N_1-1}\delta_{m_2,N_1-1}$$

$$+ \delta_{N_1,N_2-1}\delta_{m_2,N_1}(\delta_{N,N_1+1}\delta_{m_1,N_1} + \delta_{N,N_1}\delta_{m_1,N_1-1})$$

$$+ \delta_{N_1,N_2+1}\delta_{m_1,N_1-1}(\delta_{N,N_1-1}\delta_{m_2,N_1-2} + \delta_{N,N_1}\delta_{m_2,N_1-1})]\} + \langle \Psi_{ex}^{N_1} | A_1(f_1^{X_1^0}) A_2(f_2^{X_2^0}) | \Phi_{ex}^{N_2} \rangle^C, \tag{66}$$

where the connected part (last term) has the properties described in Theorem 3.

V. CONCLUDING REMARKS

In the foregoing sections we have reduced the timelike asymptotic behavior of many-particle matrix elements of almost local fields to that of one-particle and vacuum matrix elements plus a remainder which decreases faster than any inverse polynomial. There remain two problems: (i) What is the asymptotic behavior of these latter matrix elements? (ii) How does the asymptotic behavior in time depend on the smearing functions in the space variables? The complete solution of these two problems goes beyond the scope of the present paper. It will be published in a separate paper. Here we will only present a discussion of the two simplest cases in order to show what one has to expect.

A. Vacuum–One-Particle Matrix Elements

Since for $f(x) \in S_1 \otimes 0'_{c3}$ the matrix element

$$\langle Ma\gamma | A(f^X) | 0 \rangle = (2\pi)^{3/2} \int d^4p \Theta(p^0) \delta(p^2 - M^2) \times \exp\{ipx\} \tilde{f}(-p) \Psi_{Ma}(\mathbf{p})^* \langle Ma\gamma | A(0) | 0 \rangle \tag{67}$$

is a smooth solution of the Klein–Gordon equation, it is bounded by $(x^0)^{-3/2}$ [1,18]

$$\langle Ma\gamma | A(f^{X_0}) | 0 \rangle \underset{X^0 \to \pm\infty}{\sim} C(x^0)^{-3/2}, \qquad |C| < \infty. \tag{68}$$

However, this decrease changes rapidly if $f(x)$ is from $S_1 \otimes 0^p_{M3}$. Take for instance:

$$f_{(r,s,k)}(x) = \varphi(x^0)(x^1)^r(x^2)^s(x^3)^k, \qquad \varphi(x^0) \in S'_1,$$

$$\tilde{f}_{(r,s,k)}(p) = (2\pi)^{3/2}(i)^{r+s+k} \times \tilde{\varphi}(p^0) \frac{\partial^{r+s+k}}{(\partial p^1)^r(\partial p^2)^s(\partial p^3)^k} \delta(\mathbf{p}). \tag{69}$$

Introducing (69) into (67) we can integrate and obtain:

$$\langle Ma\gamma | A(f^t_{(r,s,n-r-s)}) | 0 \rangle = i^n(2\pi)^3 \frac{\partial^n}{(\partial p^1)^r(\partial p^2)^s(\partial p^3)^{n-r-s}} \times \left\{ \frac{1}{\omega_M(\mathbf{p})} \tilde{\varphi}(-\omega_M(\mathbf{p})) \Psi_{Ma}(\mathbf{p})^* \langle M\mathbf{p}\gamma | A(0) | 0 \rangle \times \exp\{i\omega_M(\mathbf{p})t\} \right\}_{p=0}. \tag{70}$$

In other words, in general, we obtain a polynomial of degree $(n-1)$ in t multiplied by an oscillating function, if the wavefunction $\psi_{Ma}(\mathbf{p})$ does not happen to vanish together with sufficiently many derivatives for $\mathbf{p} = 0$.

B. One Particle–One-Particle Matrix Elements

It can be shown by means of the stationary phase method[4] that the matrix element

$$\langle Ma\gamma | A(f^t) | mb\epsilon \rangle = (2\pi)^{3/2} \int \frac{d^3p}{2\omega_M(\mathbf{p})} \int \frac{d^3k}{2\omega_m(\mathbf{k})} \times \tilde{f}(\omega_m(\mathbf{k}) - \omega_M(\mathbf{p})) \psi_{Ma}(\mathbf{p})^* \psi_{mb}(\mathbf{k}) \times \langle M\mathbf{p}\gamma | A(0) | m\mathbf{k}\epsilon \rangle \exp\{i[\omega_M(\mathbf{p}) - \omega_m(\mathbf{k})]t\} \tag{71}$$

vanishes for $t \to \pm\infty$ at least like t^{-3} if $f(x) \in S_1 \otimes 0'_{c3}$ (that means $\tilde{f}(\mathbf{p}) \in S_1 \otimes 0_{M3}$).

However, if we take for $f(x)$ the function (69) from $S_1 \otimes 0^p_{M3}$, we can again perform one integration. Performing furthermore the differentiations with respect to P we get

$$\langle Ma\gamma | A(f^t_{(r,s,n-r-s)}) | mb\epsilon \rangle = \sum_{m=0}^{n} t^m \int d^3\mathbf{p} a_m^{r,s}(\mathbf{p}) \exp\{i[\omega_M(\mathbf{p}) - \omega_m(\mathbf{p})]t\}, \tag{72}$$

with $a_m^{r,s}(\mathbf{p})$ being from S_3. For $M \neq m$, the integral in (72) is from S_1 in t as it easily follows by a change of variables.

However, if $M = m$, we again obtain a polynomial increase of degree n for $t \to \pm\infty$. The coefficients of this polynomial can be expressed by physical form factors and their derivatives with respect to the invariant momentum transfer.[13]

As we have seen the asymptotic behavior in the time variable of the leading terms depends very sensitively on the choice of the smearing functions in the space variables $\mathbf{x}$. In contrast to the wide spread belief it depends critically on the *asymptotic behavior* of the smearing functions in space-

like directions. In both cases discussed here, it changes from a decrease to a polynomial increase if the smearing over the space variables is changed from decreasing to increasing.

ACKNOWLEDGMENT

I want to thank Dr. M. Rinke for pointing out an error in the original preprint of this article.

[1] K. Hepp, in *Particle Symmetries and Axiomatic Field Theory* (Gordon and Breach, New York, 1965), Vol. 1.
[2] In the following we will abbreviate the set of numbers $\{M_i, a_i, \gamma_i\}$ by their index i.
[3] For the exact definition of almost locality, see[1] Axiom IV′ on p. 182. Local fields are also almost local.
[4] H. Araki and R. Haag, Commun. Math. Phys. **4**, 77 (1967).
[5] H. J. Borchers, Nuovo Cimento **33** 1600 (1964).
[6] In Ref. 5 the boundedness was proved only for local fields. However one can easily extend the proof for almost local fields.
[7] A. H. Völkel, Phys. Rev. D **1**, 3377 (1970).
[8] A. H. Völkel, Phys. Rev. (to be published).
[9] B. Schroer, and P. Stichel, Commun. Math. Phys. **3**, 258 (1966).
[10] S. Fubini, G. Furlan, and C. Rossetti, Nuovo Cimento **40A**, 1171 (1965).
[11] S. Okubo, Nuovo Cimento **41A**, 586 (1966).
[12] S. Fubini and G. Furlan, Ann. Phys. (N.Y.) **48**, 322 (1968).
[13] Uta Völkel, and A. H. Völkel, Nuovo Cimento **63A**, 203 (1969).
[14] R. Haag, Phys. Rev. **112**, 669 (1958).
[15] W. Zimmermann, Nuovo Cimento **10**, 597 (1958).
[16] L. Schwartz, *Théorie des Distributions* (Hermann, Paris, 1959), Vol. II.
[17] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
[18] R. Jost, *The General Theory of Quantized Fields* (Am. Math. Soc., Providence R.I., 1965).
[19] K. Hepp, Commun. Math. Phys. **1**, 95 (1965).
[20] The sense of this terminology will become clear in Sec. IV.
[21] H. Reeh, F. Physik **16**, 687 (1968).
[22] P. Stichel, Springer Tracts in Mod. Physics **50**, 110 (1969).

The Existence and Completeness of Generalized Wave Operators for Spherically Symmetric, Asymptotically Coulomb Potentials

Michael O'Carroll
Departamento de Matematica, Ponlificia Universidade Catolica, Rua Marques de Sao Vicente, 209, Rio de Janeiro, GB-ZC-20, Brazil
(Received 22 April 1971)

By applying a theorem of Kuroda, we prove the existence and continuous completeness of the generalized wave operators $W_\pm(H_2, H_1)$, $W_\pm(H_1, H_2)$, where $W_\pm(H_j, H_i) = s - \lim e^{iH_j t} e^{-iH_i t} \times P_i$, in the Hilbert space $L^2(R^3)$. P_i is the projection operator on the subspace of absolute continuity of H_i. H_1 is the self-adjoint Hamiltonian for a particle in a pure Coulomb potential $V_c = ze^2/|x|$, and H_2 is the self-adjoint Hamiltonian for a system described by the potential function $V_c + V$, where V is a real-valued, measurable function of $x \in R^3$, spherically symmetric $[V(x) = V(r = |x|)]$, satisfies the condition

$$\int_0^R r^2 |V(r)|^2 dr + \int_0^\infty (1 + r)^\delta |V(r)|^i dr < \infty$$

for some $R(0 \leq R < \infty)$, some $0 < \delta < 1$ with $i = 1, 2$, and is continuous except at $r = 0$. In conjunction with our result, we obtain a bound for the radial Coulomb Green's function. Dropping the continuity assumption on V, we have absolutely continuous completeness of the wave operators.

This is a brief paper to show the existence and continuous completeness of the generalized wave operators in $L^2(R^3)$ for a system described by a potential function $V_c + V$. The function $V_c = ze^2/|x|, x \in R^3$, is the Coulomb potential and V is a spherically symmetric, real, measurable potential function which satisfies, with $r = |x|, x \epsilon R^3$,

$$\int_0^R r^2 |V(r)|^2 dr + \int_R^\infty (1 + r)^\delta |V(r)|^i dr < \infty, \quad i = 1, 2, \qquad (1)$$

for some $R(0 < R \leq \infty)$ and some $0 < \delta < 1$. As the potential is spherically symmetric, the self-adjoint Coulomb Hamiltonian H_1 in $L^2(R^3)$ is defined as the direct sum of Hamiltonians $H_1 = \sum_{l=0} H_1^l$, H_1^l acting in $\mathcal{H}^l = L^2(0, \infty) \times L_l^2(\Omega)$. $L_l^2(\Omega)$ is the $2l + 1$ dimensional Hilbert space spanned by the spherical harmonics and H_1^l acts as the identity on this factor. In $L^2(0, \infty)$, H_1^l is taken as the self-adjoint operator obtained from the theory of eigenfunction expansions by suitably restricting the domain of the differential operator

$$L_1^l = -\frac{\hbar^2}{2m}\frac{d^2}{dr^2} + \frac{\hbar^2}{2m}\frac{l(l+1)}{r^2} + \frac{ze^2}{r}. \qquad (2)$$

For the characterization of the domain $D(H_1^l) \subset L^2(0, \infty)$, see Kodaira[1] or Stone.[2] For $l = 0$, the additional condition $u(0) = 0$ is imposed on $u \in D(H_1^0)$ to obtain a self-adjoint H_1^0, as the point $r = 0$ is in the limit circle case. From the theory of eigenfunction expansions and a knowledge of the explicit solutions, the operator $(H_1^l - \lambda)^{-1}$ can be explicitly constructed and is an integral operator of the Carleman type for $\lambda \notin \Sigma(H_1^l)$. The spectrum of H_1^l, $\Sigma(H_1^l)$, consists of a countably infinite discrete spectrum in $[-z^2m^2e^4/2\hbar^2(l+1)^2, 0)$ which accumulates at 0 and is given by $\lambda_n = -z^2m^2e^4/2\hbar^2(l + n + 1)^2 (n = 0, 1, \ldots)$ for $Z < 0$ (for $Z > 0$ there is no point spectrum) and $\lambda \in [0, \infty)$ belongs to the absolutely continuous spectrum. The point $\lambda = 0$ is not an element of the point spectrum. We denote by P_1^l the projection operator on the subspace of absolute continuity of H_1^l. $H_2^l = H_1^l + V$ with $D(H_2^l) = D(H_1^l) \subset D(V)$ when V obeys the conditions of (1). We will need the following theorem of Kuroda[3].

Theorem 1: Let K_1 be a self-adjoint operator in a separable Hilbert space X and let V be a closed symmetric operator in X, such that V is relatively bounded with respect to K_1 with bound less than 1. Then $K_2 = K_1 + V$ is self-adjoint with $D(K_2) = D(K_1)$.

$$|V|^{1/2}(K_1 - \zeta)^{-1}$$

like directions. In both cases discussed here, it changes from a decrease to a polynomial increase if the smearing over the space variables is changed from decreasing to increasing.

ACKNOWLEDGMENT

I want to thank Dr. M. Rinke for pointing out an error in the original preprint of this article.

[1] K. Hepp, in *Particle Symmetries and Axiomatic Field Theory* (Gordon and Breach, New York, 1965), Vol. 1.
[2] In the following we will abbreviate the set of numbers $\{M_i, a_i, \gamma_i\}$ by their index i.
[3] For the exact definition of almost locality, see[1] Axiom IV′ on p. 182. Local fields are also almost local.
[4] H. Araki and R. Haag, Commun. Math. Phys. **4**, 77 (1967).
[5] H. J. Borchers, Nuovo Cimento **33** 1600 (1964).
[6] In Ref. 5 the boundedness was proved only for local fields. However one can easily extend the proof for almost local fields.
[7] A. H. Völkel, Phys. Rev. D **1**, 3377 (1970).
[8] A. H. Völkel, Phys. Rev. (to be published).
[9] B. Schroer, and P. Stichel, Commun. Math. Phys. **3**, 258 (1966).
[10] S. Fubini, G. Furlan, and C. Rossetti, Nuovo Cimento **40A**, 1171 (1965).
[11] S. Okubo, Nuovo Cimento **41A**, 586 (1966).
[12] S. Fubini and G. Furlan, Ann. Phys. (N.Y.) **48**, 322 (1968).
[13] Uta Völkel, and A. H. Völkel, Nuovo Cimento **63A**, 203 (1969).
[14] R. Haag, Phys. Rev. **112**, 669 (1958).
[15] W. Zimmermann, Nuovo Cimento **10**, 597 (1958).
[16] L. Schwartz, *Théorie des Distributions* (Hermann, Paris, 1959), Vol. II.
[17] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
[18] R. Jost, *The General Theory of Quantized Fields* (Am. Math. Soc., Providence R.I., 1965).
[19] K. Hepp, Commun. Math. Phys. **1**, 95 (1965).
[20] The sense of this terminology will become clear in Sec. IV.
[21] H. Reeh, F. Physik **16**, 687 (1968).
[22] P. Stichel, Springer Tracts in Mod. Physics **50**, 110 (1969).

The Existence and Completeness of Generalized Wave Operators for Spherically Symmetric, Asymptotically Coulomb Potentials

Michael O'Carroll
Departamento de Matematica, Ponlificia Universidade Catolica, Rua Marques de Sao Vicente, 209, Rio de Janeiro, GB-ZC-20, Brazil
(Received 22 April 1971)

By applying a theorem of Kuroda, we prove the existence and continuous completeness of the generalized wave operators $W_\pm(H_2, H_1)$, $W_\pm(H_1, H_2)$, where $W_\pm(H_j, H_i) = s - \lim e^{iH_j t} e^{-iH_i t} \times P_i$, in the Hilbert space $L^2(R^3)$. P_i is the projection operator on the subspace of absolute continuity of H_i. H_1 is the self-adjoint Hamiltonian for a particle in a pure Coulomb potential $V_c = ze^2/|x|$, and H_2 is the self-adjoint Hamiltonian for a system described by the potential function $V_c + V$, where V is a real-valued, measurable function of $x \in R^3$, spherically symmetric $[V(x) = V(r = |x|)]$, satisfies the condition

$$\int_0^R r^2 |V(r)|^2 dr + \int_0^\infty (1 + r)^\delta |V(r)|^i dr < \infty$$

for some $R(0 \leq R < \infty)$, some $0 < \delta < 1$ with $i = 1, 2$, and is continuous except at $r = 0$. In conjunction with our result, we obtain a bound for the radial Coulomb Green's function. Dropping the continuity assumption on V, we have absolutely continuous completeness of the wave operators.

This is a brief paper to show the existence and continuous completeness of the generalized wave operators in $L^2(R^3)$ for a system described by a potential function $V_c + V$. The function $V_c = ze^2/|x|, x \in R^3$, is the Coulomb potential and V is a spherically symmetric, real, measurable potential function which satisfies, with $r = |x|, x \epsilon R^3$,

$$\int_0^R r^2 |V(r)|^2 dr + \int_R^\infty (1 + r)^\delta |V(r)|^i dr < \infty, \quad i = 1, 2, \qquad (1)$$

for some $R(0 < R \leq \infty)$ and some $0 < \delta < 1$. As the potential is spherically symmetric, the self-adjoint Coulomb Hamiltonian H_1 in $L^2(R^3)$ is defined as the direct sum of Hamiltonians $H_1 = \sum_{l=0} H_1^l$, H_1^l acting in $\mathcal{H}^l = L^2(0, \infty) \times L_l^2(\Omega)$. $L_l^2(\Omega)$ is the $2l + 1$ dimensional Hilbert space spanned by the spherical harmonics and H_1^l acts as the identity on this factor. In $L^2(0, \infty)$, H_1^l is taken as the self-adjoint operator obtained from the theory of eigenfunction expansions by suitably restricting the domain of the differential operator

$$L_1^l = -\frac{\hbar^2}{2m}\frac{d^2}{dr^2} + \frac{\hbar^2}{2m}\frac{l(l+1)}{r^2} + \frac{ze^2}{r}. \qquad (2)$$

For the characterization of the domain $D(H_1^l) \subset L^2(0, \infty)$, see Kodaira[1] or Stone.[2] For $l = 0$, the additional condition $u(0) = 0$ is imposed on $u \in D(H_1^0)$ to obtain a self-adjoint H_1^0, as the point $r = 0$ is in the limit circle case. From the theory of eigenfunction expansions and a knowledge of the explicit solutions, the operator $(H_1^l - \lambda)^{-1}$ can be explicitly constructed and is an integral operator of the Carleman type for $\lambda \notin \Sigma(H_1^l)$. The spectrum of H_1^l, $\Sigma(H_1^l)$, consists of a countably infinite discrete spectrum in $[-z^2m^2e^4/2\hbar^2(l+1)^2, 0)$ which accumulates at 0 and is given by $\lambda_n = -z^2m^2e^4/2\hbar^2(l + n + 1)^2 (n = 0, 1, \ldots)$ for $Z < 0$ (for $Z > 0$ there is no point spectrum) and $\lambda \in [0, \infty)$ belongs to the absolutely continuous spectrum. The point $\lambda = 0$ is not an element of the point spectrum. We denote by P_1^l the projection operator on the subspace of absolute continuity of H_1^l. $H_2^l = H_1^l + V$ with $D(H_2^l) = D(H_1^l) \subset D(V)$ when V obeys the conditions of (1). We will need the following theorem of Kuroda[3].

Theorem 1: Let K_1 be a self-adjoint operator in a separable Hilbert space X and let V be a closed symmetric operator in X, such that V is relatively bounded with respect to K_1 with bound less than 1. Then $K_2 = K_1 + V$ is self-adjoint with $D(K_2) = D(K_1)$.

$$|V|^{1/2}(K_1 - \zeta)^{-1}$$

is a Hilbert-Schmidt operator for some $\zeta \notin \Sigma(K_1)$ (the spectrum of K_1); then $W_\pm(K_2, K_1)$ and $W_\pm(K_1, K_2)$ exist and are absolutely continuous complete.

$$W_\pm(K_j, K_i) = s - \lim_{t\to\pm\infty} e^{iK_j t} e^{-iK_i t} P_i,$$

where P_i is the projection operator on the subspace of absolute continuity of K_i. Also if K_1 is lower-bounded then so is K_2.

We apply Theorem 1 by setting $X = L^2(0,\infty)$, $K_1 = H_1^l$, $P_1 = P_1^l$. V is taken as the maximum multiplication operator by $V(r)$ on $L^2(0,\infty)$ and $|V|^{1/2} = |V^*|^{1/2} = |V(r)|^{1/2}$. Denoting K_2 by H_2^l we have the following theorem.

Theorem 2: For each l with V satisfying the conditions of (1), the generalized wave operators $W_\pm(H_2^l, H_1^l)$, $W_\pm(H_1^l, H_2^l)$ exist and are absolutely continuous complete. H_2^l is lower-bounded.

Proof: Using the integral representation of Wichman and Woo[4] for the Carleman integral operator $(H_1^l - \lambda)^{-1}$, i.e., the radial Coulomb Green's function, we show that the Carleman integral operators $V(H_1^l - \lambda)^{-1}$ and $|V|^{1/2}(H_1^l - \lambda)^{-1}$ are Hilbert–Schmidt for sufficiently large negative λ. This calculation is carried out in the Appendix. Since every $v \in D(H_1^l - \gamma) = D(H_1^l)$ can be written as $v = (H_1 - \gamma)^{-1}u$ for some $u \in L^2(0,\infty)$ and $\gamma \notin \Sigma(H_1^l)$ and $V(H_1^l - \gamma)^{-1}$ is Hilbert–Schmidt, $D(V) \supset DH_1^l)$. Furthermore $||V(H_1^l - \gamma)^{-1}|| \to 0$ as $\gamma \to -\infty$ from (A30) so that V has H_1^l bound zero. Thus the conditions of Theorem 1 are satisfied.

Turning to the three-dimensional problem, we define a self-adjoint H_i to be the direct sum $\sum_{l=0}^{\infty} \oplus H_i^l$ acting in the direct sum of Hilbert spaces $\mathcal{H} = \sum_{l=0}^{\infty} \oplus \mathcal{H}^l$. $\mathcal{H}$ is isomorphic to $L^2(R^3)$ as shown by Green and Lanford.[5] Defining $P_i = \sum_{l=0}^{\infty} P_i^l$, then P_i is the projection operator on the subspace of absolute continuity of H_i. From the fact that the wave operators

$$W_\pm(H_j, H_i) = s - \lim_{t\to\pm\infty} e^{iH_j t} e^{-iH_i t} P_i$$

exist iff the $W_\pm(H_j^l, H_i^l)$ exist for each l[6] and are equal to the direct sum of the $W_\pm(H_j^l, H_i^l)$, we then have (noting Theorems 1, 2) the following:

Theorem 3: The generalized wave operators $W_\pm(H_2, H_1)$, $W_\pm(H_1, H_2)$ exist and are continuous complete on $L^2(R^3)$, when V satisfies the condition of Eq. 1 and is continuous except at $r = 0$.

Proof: The continuous completeness rather than absolutely continuous completeness follows from the fact that the continuous spectrum consists of an absolutely continuous spectrum only as shown by Kodaira.[1]

If H_1 is the free particle Hamiltonian rather than the Coulomb Hamiltonian, then letting $\delta = 0$ in (1), Theorem 3 still holds as seen from the estimate of (A31). However this case is covered by a more general theorem of Kuroda[6] using the theory of forms. We tried to apply his method to the Coulomb problem but we were not able to obtain the necessary estimates. It would be interesting to work out the relation between the time-dependent theory and the time-independent theory for the Coulomb problem as done by Ikebe,[7] Green and Lanford[5] for the case of shorter-range potentials. In doing this the definition of the scattering amplitude for systems which are asymptotically Coulomb may be clarified.

APPENDIX

1. Estimates for the Kernel of the Operator $(H_1 - \lambda)^{-1}$ (the Coulomb Radial Green's Function)

The Coulomb radial Green's function for each l is an integral operator whose real and imaginary parts are Carleman kernels.[2] We denote it by $G_l(x, y:\lambda)$, $x, y \in R$. It is the inverse of the operator $H_1^{(l)} - \lambda$ ($\lambda \in C$ not an element of the spectrum of $H_1^{(l)}$), where $H_1^{(l)}$ is the self-adjoint operator determined by the formal differential operator

$$L_l = -\frac{\hbar^2}{2m}\frac{d^2}{dr^2} + \frac{\hbar^2}{2m}\frac{l(l+1)}{r^2} + \frac{ze^2}{r^2}$$

acting in $L^2(0,\infty)$ as defined by Stone[2] and Kodaira.[1] The operator $(H_1^{(l)} - \lambda)^{-1}$ is bounded and defined on all $L^2(0,\infty)$ for $\lambda \notin \Sigma(H_1^l)$. In the case of $l = 0$, we impose the boundary condition $u(0) = 0$ for elements $u \in D(H_1^0)$.

For $\lambda \notin \left[-\frac{\hbar^2\gamma^2}{2}, \infty\right)$ with

$$\epsilon = \frac{\gamma}{p}, \quad p = \left(\frac{2m\lambda}{\hbar^2}\right)^{1/2}, \quad \mathrm{Imp} > 0, \quad \gamma = \frac{ze^2m}{\hbar^2}, \tag{A1}$$

we use the integral representation of Wichmann and Woo[4] for the kernel $G^{(l)}(x, y; \lambda)$. We have with $x, y \in (0,\infty)$, $\infty > y > x > 0$, $x = r'$, $y = r''$,

$$G_l(r', r'', \lambda) = \frac{-ipr'r''}{\Gamma(1+i\epsilon)\Gamma(1-i\epsilon)} \int_0^1 dt \int_1^\infty ds\, [s(1-t)]^{i\epsilon}\,[t(s-1)]^{-i\epsilon} \times \frac{\partial^2}{\partial t\,\partial s}\,[ste^{ip[r'(1-t)-r''(1-s)]} j_l(pr't) h_l^{(1)}(pr''s)], \tag{A2}$$

where $j_l(z)$, $h_l^{(1)}(z)$ are the spherical Bessel and Hankel functions. Also $G^l(x, y, \lambda)$ for $\infty > x > y > 0$ is $G^l(y, x, \lambda)$. Setting $\epsilon = 0$, we obtain the free particle Green's function $G_l^f(x, y, \lambda)$:

$$G_l^f(x', y'', \lambda) = -ipr'r'' j_l(pr') h_l^{(1)}(pr''). \tag{A3}$$

We use the representations

$$j_l(z) = \frac{z^l}{2^{l+1}l!} \int_{-1}^{1} e^{izs}(1-s^2)^l ds, \tag{A4}$$

$$h_l^{(1)}(z) = \left(\frac{c_{l+1}}{z^{l+1}} + \frac{c_l}{z^l} + \cdots + \frac{c_1}{z^1}\right) e^{iz}. \tag{A5}$$

Note that for $z = i\rho, \rho > 0$, we have

$$h_l^{(1)}(i\rho) \propto \left(\frac{a_1}{\rho} + \frac{a_2}{\rho} + \cdots \frac{a_{l+1}}{\rho^{l+1}}\right) e^{-\rho}, a_i > 0. \quad \text{(A6)}$$

Also for $\lambda < -\gamma^2 h/2m$, we have $\epsilon = \gamma/ik$, where $p = ik$ and $|\epsilon| < 1$. In making the estimates for (A2) we consider the t integral with the associated r' terms and the s integral with the associated r'' terms separately.

t integral:

$$M_1(k, r', l) = r' e^{-kr''} \int_0^1 dt I_1(t, \epsilon) \frac{d}{dt} J_1(t, k, r') \quad \text{(A7)}$$

with

$$I_1(t, \epsilon) = (1-t)^{i\epsilon} t^{-i\epsilon}, \qquad i\epsilon = \gamma/k,$$

$$J_1(t, k, r') = te^{kr't} \frac{(kr')^l}{2^{l+1} l!} \int_{-1}^1 e^{-kr's}(1-s^2)^l ds.$$

J_1 is a positive increasing function of t and so is dJ_1/dt for $0 \le t \le 1$. We majorize M_1 of (A7) by

$$|M_1| \le r' e^{-kr'} \left(\int_0^1 dt |I_1|^\alpha\right)^{1/\alpha} \left(\int_0^1 dt \left|\frac{d}{dt} J_1\right|^q\right)^{1/q}, \quad \text{(A8)}$$

where $1/\alpha + 1/q = 1$, $\alpha, q > 1$. For $k > 2|\gamma|\alpha$, we have

$$|M_1| \le r' e^{-kr'} \left(\int_0^1 dt \left|\frac{d}{dt} J_1\right|^q\right)^{1/q}. \quad \text{(A9)}$$

Writing $q = 1 + n$, we have

$$|M_1| \le r' e^{-kr'} \left\{\left|\frac{d}{dt} J_1\right|_{t=1}^n\right\}^{1/q} \left\{\int_0^1 dt \left|\frac{d}{dt} J_1\right|\right\}^{1/q}$$

$$\le r' e^{-kr'} \left|\frac{d}{dt} J_1\right|_{t=1}^{n/q} \{[J_1]_{t=1}\}^{1/q}. \quad \text{(A10)}$$

Performing the integrals for $\frac{d}{dt} J_1|_{t=1}, J_1|_{t=1}$, we have

$$|M_1| \le r' e^{-kr'} \left\{e^{2kr'} \frac{(kr')^l}{(1+kr')^l}^{n/q} \frac{(kr')^l}{(1+kr')^{l+1}} e^{2kr'}\right\}^{1/q} \le \frac{C_1}{k} \frac{(kr')^{l+1}}{(1+kr')^{l+1/q}} e^{kr'}, \quad k > 2|\gamma|\alpha, \quad 1/\alpha + 1/q = 1. \quad \text{(A11)}$$

In (A7)–(A10), we have omitted numerical constants which do not depend on k, r'. In (A11), C_1 depends only on l being independent of k, r' for $k > 2|\gamma|\alpha$. From (A11), we obtain the free particle estimate on setting $q = 1$.

s-integral:

$$M_2(k, r'', l) = r'' \int_1^\infty ds I_2(s, \epsilon) \frac{d}{ds} J_2(s, k, r'') \quad \text{(A12)}$$

with

$$I_2(s, \epsilon) = s^{i\epsilon}(s-1)^{-i\epsilon},$$

$$J_2(s, k, r'') = se^{kr''(1-s)} h_l^{(1)}(ikr''s).$$

$d/ds\, J_2(s, k, r'')$ has the form

$$\frac{d}{ds} J_2(s, k, r'') \alpha e^{kr''} \left\{\left(\frac{a_2}{\rho^2} + \cdots \frac{la_{l+1}}{\rho^{l+1}}\right) e^{-2\rho} + 2\left(a_1 + \frac{a_2}{\rho} + \cdots \frac{a_{l+1}}{\rho^l}\right) e^{-2\rho}\right\} \quad \text{(A13)}$$

with $\rho = kr''s$. The expression in (A13) is a monotone decreasing function of s along with all its derivatives. However, we have not been able to exploit this fact. M_2 is then a sum of terms $T_n (1 \le n \le l+1)$, the typical one being majorized by

$$T_n \alpha \frac{r'' e^{kr''}}{(kr)^n} \int_1^\infty \left|\frac{s^{i\epsilon}(s-1)^{-i\epsilon}}{s^n}\right| e^{-2kr''s}. \quad \text{(A14)}$$

For $n \ge 2$, we have

$$T_n < \frac{1}{k} \frac{e^{-kr''}}{(kr'')^{n-1}}, \quad k > 2|\gamma|, \quad \text{(A15)}$$

by setting $e^{-kr''s}$ equal to $e^{-2kr''}$ in (A14). For T_0, we have

$$T_0 \le \frac{1}{k} (kr'')^{\gamma/k} e^{-kr''}, \quad k > 2|\gamma|, \quad \text{(A16)}$$

by changing variables to $s' = s - 1$ and then to $u = kr''s'$. Similarly,

$$T_1 \le \frac{1}{k} \frac{(kr'')^{\gamma/k}}{kr''} e^{-kr''}, \quad k > 2|\gamma|. \quad \text{(A17)}$$

Thus

$$|M_2(k, r', l) \le \frac{C_1}{k} \frac{(1+kr'')^{l+|\gamma|/k}}{(kr'')^l} e^{-kr''}, \quad k > 2|\gamma|. \quad \text{(A18)}$$

In (A18) C_1 depends on l only for $k > 2|\gamma|$ and we obtain the free particle estimate on setting $|\gamma| = 0$. Finally using (A11) and (A18) in (A12), we have for $p = ik$, $0 \le x < y < \infty$,

$$|G_l(x, y; \lambda)| \le Ck \left(\frac{(kx)^{l+1}}{k(1+kx)^{l+1/q}} e^k\right) \left(\frac{(1+ky)^{l+|\gamma|/k}}{k(ky)^l} e^{-ky}\right), \quad \text{(A19)}$$

for $k > 2|\gamma|\alpha, \alpha^{-1} + q^{-1} = 1$.

2. Estimates for the Operator $F(H_1^l - \lambda)^{-1}$

We now consider the operator $FG_l = F(H_1^l - \lambda)^{-1}$, where F is a real-valued, measurable function of $x \in [0, \infty)$ considered as a maximum multiplication operator on $L^2(0, \infty)$. We want to obtain conditions on F such that $D(F) \supset D(H_1^l)$ as this is part of the relatively bounded condition. To satisfy the condition that the H_1^l bound of F is less than 1, we also want $\|F(H_1^l - \lambda)^{-1}\|$ to go to zero as $\lambda \to -\infty$. To this end note that for $k > 2|\gamma|\alpha$, the range of G_l is the domain $D(H_1^l - \lambda)$ as the domain of G_l ranges through all $L^2(0, \infty)$. Thus if FG_l has finite Hilbert–Schmidt norm, we have

$$\|FG_l u\| \le \|FG_l\| \, \|u\| \le \|FG_l\|_{H\text{-}S} \|u\| \forall u \in L^2(0, \infty) \quad \text{(A20)}$$

so that $D(F) \supset D(H_1^l)$. Using the estimates of (A19) for G_l, we obtain sufficient conditions on F [see Eq. (A30)] so that $\|FG_l\|_{H\text{-}S}$ is finite. We have

$$\|FG_l\|_{H\text{-}S}^2 = \int_0^\infty \int_0^\infty |F(x)|^2 |G_l(x, y, \lambda)|^2 dxdy$$

$$= \int_{y=0}^{\infty} \int_{x=0}^{y} [\quad] dx dy + \int_{x=0}^{\infty} \int_{y=0}^{x} [\quad] dy dx$$
$$\equiv R_1 + R_2. \tag{A21}$$

We consider the first term of (A21) which satisfies

$$R_1 \leq \frac{C}{k^2} \int_{y=0}^{\infty} dy \frac{(1+ky)^{2(l+|\gamma|/k)}}{(1+ky)^{2l}} e^{-2ky} \times \int_0^y \frac{(kx)^{2(l+1)}}{(1+kx)^{2(l+1/q)}} e^{2kx} |F(x)|^2 dx = R_1' + R_1''. \tag{A22}$$

We break (A22) further into integrals R_1' and R_1'', where the R_1' integral is over the triangle $0 \leq x \leq y$ and $0 \leq y \leq R$ for some $0 \leq R \leq \infty$ and the R_1'' integral extends over the rest of the triangle $0 \leq x \leq y$ and $0 \leq y \leq \infty$. Thus

$$R_1' \leq \frac{C}{k^2} \int_0^R \left(\frac{(1+ky)^{2(l+|\gamma|/k)}}{(ky)^{2l}} e^{-2ky} \right) \times \int_0^y \frac{(kx)^{2(l+1)}}{(1+kx)^{2(l+1/q)}} e^{2kx} |F(x)|^2 dx$$
$$\leq \frac{C}{k^2} \int_0^R dy e^{-2ky} (1+ky)^{2|\gamma|/k} \int_0^y (kx)^2 e^{2kx} |F(x)|^2 dx$$
$$\leq C(1+kR)^{2|\gamma|/k} \int_0^R dy e^{-2ky} \int_0^y x^2 e^{2kx} |F(x)|^2 dx, \tag{A23}$$

noting that $s^l/(1+s)^l$ is monotonic increasing with maximum value 1 for $s \in [0, \infty)$. Changing variables to $x' = x, z = y - x$, we have

$$R_1' \leq C(1+kR)^{2|\gamma|/k} \int_0^R dz e^{-2kz} \int_0^R x'^2 |F(x')|^2 dx'$$
$$\leq \frac{C(1+kR)^{2|\gamma|/k}}{k} \int_0^R x'^2 |F(x')|^2 dx'. \tag{A24}$$

Similarly,

$$R_1'' \leq \frac{C}{k^2} \int_R^{\infty} dy e^{-2ky} \frac{(1+ky)^{2(l+|\gamma|/k)}}{(ky)^{2l}} \times \left[\int_0^R + \int_R^y \left(\frac{(kx)^{2(l+1)}}{(1+kx)^{2(l+1/q)}} \right) \times e^{2kx} |F(x)|^2 dx \right]$$
$$\leq C \int_R^{\infty} dy e^{-2ky} e^{+2kR} (1+ky)^{2|\gamma|/k} \int_0^R x^2 |F(x)|^2 dx + \frac{C}{k^2} \int_R^{\infty} e^{-2ky} (ky)^{2(1-q^{-1}+|\gamma|/k)} dy \int_R^y e^{2kx} |F(x)|^2 dx \tag{A25}$$

$$R_1'' \leq \frac{C}{k} (1+kR)^{2|\gamma|/k} \int_0^R x'^2 |F(x')|^2 dx' + \frac{C}{k^{3-2\beta}} \left(\int_R^{\infty} |F(x')|^2 dx' + \int_R^{\infty} x'^{2\beta} |F(x')|^2 dx' \right). \tag{A26}$$

Note that we have

$$k > 2|\gamma|\alpha, \quad 1/\alpha > 2|\gamma|/k, \quad \alpha^{-1} + q^{-1} = 1,$$

so $1 > 1 - q^{-1} = \alpha^{-1} > 2|\gamma|/k$ and

$$1 + |\gamma|/k > 1 - q^{-1} + |\gamma|/k \equiv \beta = \alpha^{-1} + |\gamma|/k > 3|\gamma|/k,$$

and we may choose β arbitrarily close to zero by taking k, α large. Thus combining (A24) and (A26) we have

$$R_1 \leq \frac{C}{k} (1+kR)^{2|\gamma|/k} \int_0^R x^2 |F(x)|^2 dx + \frac{C}{k^{3-2\beta}} \int_R^{\infty} (1+x)^{2\beta} |F(x)|^2 dx. \tag{A27}$$

Letting $|\gamma|, \beta \to 0$ in (A27), we obtain the free particle estimate.

We now obtain estimates for R_2:

$$R_2 \leq \frac{C}{k^2} \int_{x=0}^{\infty} \int_{y=0}^{x} |F(x)|^2 \frac{(1+kx)^{2(l+|\gamma|/k)}}{(kx)^{2l}} e^{-2kx} \times \frac{(ky)^{2(l+1)}}{(1+ky)^{2(l+1/q)}} e^{2ky} dy dx. \tag{A28}$$

Replace

$$\frac{(ky)^{2(l+1)}}{(1+ky)^{2(l+1/q)}}$$

by its maximum value in $0 \leq y \leq x$,

$$\frac{(kx)^{2(l+1)}}{(1+kx)^{2(l+1/q)}},$$

and perform the y integral to obtain

$$R_2 \leq \frac{C}{k^3} \int_{x=0}^{\infty} |F(x)|^2 \frac{(kx)^2}{(1+kx)^{2(q^{-1}-|\gamma|k^{-1})}} dx$$
$$\leq \frac{C}{k} \int_{x=0}^R x^2 |F(x)|^2 dx + \frac{C}{k^3} \int_R^{\infty} F(x)|^2 \frac{(kx)^{2(q^{-1}-|\gamma|k^{-1}+\beta)}}{(1+kx)^{2(q^{-1}-|\gamma|k^{-1})}} dx$$
$$\leq \frac{C}{k} \int_{x=0}^R x^2 |F(x)|^2 dx + \frac{C}{k^{3-2\beta}} \int_R^{\infty} |F(x)|^2 (1+x)^{2\beta} dx. \tag{A29}$$

Combining (A27) and (A29), we obtain the desired estimate for (A21)—namely,

$$\| FG_l \|_{H-S}^2 \leq \frac{C}{k} (1+kR)^{2|\gamma|/k} \int_0^R x^2 |F(x)|^2 dx + \frac{C}{k^{3-2\beta}} \int_R^{\infty} (1+x)^{2\beta} |F(x)|^2 dx. \tag{A30}$$

With γ and β set equal to zero, we obtain estimates for the case where G_l is the free particle Green's function G_l^f, i.e.,

$$\| FG_l^f \|_{H-S}^2 \leq \frac{C}{k} \int_0^R x^2 |F(x)|^2 dx + \frac{C}{k^3} \int_R^{\infty} |F(x)|^2 dx. \tag{A31}$$

---

1 K. Kodaira, Am. J. Math. 71, 921 (1949).
2 M. H. Stone, *Linear Transf. in Hilbert Space* (A.M.S., New York, 1932). See Chap. X, Vol. XV.
3 S. T. Kuroda, J. Math. Soc. (Japan) 11, 247 (1959).
4 E. H. Wichmann and C. H. Woo, J. Math. Phys. 2, 178 (1961).
5 T. A. Green and O. E. Lanford III, J. Math. Phys. 1, 139 (1960).
6 S. T. Kuroda, J. Math. Phys. 3, 933 (1962).
7 T. Ikebe, Arch. Ratl. Mech. Anal. 5, 1 (1960).

Addendum: Linear Random Operator Equations in Mathematical Physics, III

[J. Math. Phys. 12, 1948 (1971)]

G. Adomian
Department of Mathematics, University of Georgia, Athens, Georgia 30601

(Received 19 October 1971)

Part of footnote 20 on page 1955 was omitted. The addition should read as follows:

Since L^{-1} involves space as well as time variables, $G(t,\tau)$ is a simplified notation for $G(t,\tau,\bar{r})$ and the last term is actually:

$$\int_{-\infty}^{\infty}\int_{v} G(t,\tau,\bar{r})(\partial^2/\partial r^2)\alpha(\bar{r},\tau,\omega)y(\bar{r},\tau,\omega)dvd\tau.$$

For notational simplicity we have not shown specific dependence on ω or even $\bar{r}$; it is quite common to suppress ω. The integral is interpreted as being over the appropriate space. $\int d\tau$ means a double integral over a volume v and over all values of τ. In the expressions which follow for $y_1, y_2, \cdots$ and $K_2, K_3, \cdots$, the double integrals are actually quadruple integrals, etc.

Errata: Relativistic Fields Due to a Particle in a Grounded Cylindrical Box

[J. Math. Phys. 11, 1295 (1970)]

Richard L. Liboff*
Faculté des Sciences, Université Libre de Bruxelles, Brussels, Belgium

(Received 14 October 1971)

The last equation in the right column of p. 1301 should read

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix}_h \neq 0, \qquad z \leq ct$$

and the first clause in the sentence following should be deleted. These homogeneous contributions (given in quadrature form in text) must be added, with appropriate sign, to Eqs. (9a) and (9b) to obtain the correct fields for the case of the semi-infinite pipe. They vanish in the limit $t \to \infty$ so that Eq. (10) is correct as stated for the case of the completely infinite pipe. Other corrections follow.

page	*Location*	*Change*	*to*
1296	bottom right	$2\pi q\delta(r)$	$4\pi q\delta(r)$
1297	right column in $E_z^{(+)\prime}$ (two places)	$k_j J_1^2$	J_1^2
1297	bottom right	$\mathrm{sgn}(z-vt)$	1
1298	right column in E_z	$\sinh K_j z/L$ $\cosh(K_j-\Omega_j t)$	$\cosh K_j z/L$ $\sinh(K_j-\Omega_j t)$
1300	left column in E_z	$\sinh K_j z/L$ $\cosh(K_j-\Omega_j t)$	$\cosh K_j z/L$ $\sinh(K_j-\Omega_j t)$
1302	Eq. 10	4	2
1302	Eq. 9b in E_z	$\sinh\gamma kz$	$\cosh\gamma kz$
1302	middle of right column	$\cosh K$	$\cosh K(1-z/L)$

* Permanent address: College of Engineering, Cornell University, Ithaca, New York 14850.